2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

Robust Nonlinear Adaptive Backstepping Controller Design for Unmanned Autonomous Vehicles

T. K. Roy*[1], M. F. Pervej[1], F. K. Tumpa[1], L. C. Paul[2], and M. I. Sarkar[1]

[1]Dept. of Electronics & Telecommunication Engineering
[1]Rajshahi University of Engineering & Technology, Rajshahi-6204, Bangladesh
[2]Dept. of Electronic and Telecommunication Engineering
[2] Pabna University of Science & Technology, Pabna-6600, Bangladesh
*Email: roy.kanti03@gmail.com

***Abstract*—In this paper, a nonlinear robust adaptive controller design for unmanned autonomous vehicles (UAVs) for controlling the hovering flight is proposed. The proposed controller is designed recursively based on the control Lyapunov theory where the mass of the UAV within the model is assumed as an unknown parameter. To show the robustness property of the proposed controller, the effects of external disturbances are also considered within the UAV system model during the design of the controller. The unknown parameter (mass) is estimated using the adaptation law and incorporated in the final control law and then the overall stability of the whole system is confirmed through the negative semi-definiteness of control Lyapunov functions (CLFs). Note that the proposed controller is designed in such a way that it is adaptive to the unknown parameter of the UAV and robust to external disturbances. Finally, the performance of the designed controller is tested using a MATLAB simulation model under hovering flight condition and the performance is also compared with an existing nonlinear robust adaptive backstepping controller. Simulation results illustrate the robustness of the designed controller over the existing controller in terms of rejecting external wind gusts.**

Keywords—Control Lyapunov function; external uncertainty; unmanned autonomous vehicle; parametric uncertainty; robust adaptive backstepping controller

I. INTRODUCTION

During the last two decades, UAVs especially unmanned autonomous helicopters (UAHs) have been attracted a great deal of interest over the fixed-wing aircraft due to their several advantages such as compact size, ability to vertical take-off/landing, hovering for a long period of time etc. Among these abilities, hovering and vertically take-off/landing property of an UAH is very essential. Though an UAH has several merits over the fixed-wing aircraft, the control of an UAH is more challenging as compared to its full-scale counterpart to achieve full flight wrapping due to its inherent nonlinear unstable characteristic. In addition, another main difficulty of an UAH is the strong nonlinearities which mainly arise from the cross-couplings among the tail rotor, main rotor, and aerodynamic uncertainties [1]. Furthermore, as it is an under-actuated mechanical system with six degrees of freedom (6-DOF), the control in the presence of any kind disturbances such as external or internal is extremely difficult task. Under this situation, to provide a better performance of an UAH a nonlinear robust adaptive controller design is very essential.

Traditionally, owing to the effortlessness and ease execution, linear controllers are widely used to control of a small-scale helicopter which are mainly designed based on the linear approximation model around a fixed operating point [2], [3]. However, these linear controllers are not able to track the desired flight trajectory when the desired flight path is far away from the original point for which the controller is initially designed, i.e., under a wide variation of operating points these linear controllers are not suitable. To overcome these limitations of linear controllers, recently, the design of several advance nonlinear control schemes have been gained popularity to control the UAH under different flight operating mode in the presence of large disturbances [4]-[7].

A most commonly used nonlinear control technique is the feedback linearization (FBL) controller which has been widely used to obtain the control input for different nonlinear systems by transforming the nonlinear system to linear system [8], [9]. Based on the FBL control scheme, a nonlinear robust controller was proposed in [10] for rejecting the external wind gust effect, in which the external wind gust, is considered to be the sum of a preset number of sinusoids with unidentified amplitudes, frequencies, and phases. But, the main drawback of this control approach is that to meet the desired control objective, the system parameters and model should be known accurately and unerringly. However, in practice, it is quite difficult to know the exact values of the UAH system parameters. Therefore, to overcome this limitation of FBL controller, a nonlinear controller should be designed in such a way that which is insensitive to the parameters variations and external disturbances.

The aforementioned problems of FBLC can be resolved using another nonlinear sliding mode controller (SMC) which

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

is less sensitive to parameters variations and external disturbances. To control the nonlinear flight of an UAH, a nonlinear SMC based on the time-varying sliding surface is proposed in [11], [12]. Although SMC is less sensitive to parameter variations and external disturbances and can be operated under a wide range of operating point, the selection of the time-varying sliding surface is extremely a difficult task and makes its complicated. Under this situation, to overcome the parameter sensitivity problem of FBL controllers and selection of time-varying sliding surface of SMCs the nonlinear adaptive backstepping controller is an alternative control schemes as discussed in [13]-[16], [25]-[29] for different applications. In this method, stabilizing controls Lyapunov function is constructed in every step during the design of the controller and it can dynamically also estimate the unknown parametric or external uncertainties by using the online parameter adaptation laws. The authors of this paper would like to mentioned here that in [13]-[16], the controller is designed to control the horizontal position of an UAH under consideration of both parametric uncertainties and external disturbances. In order to achieve better control performance, a nonlinear robust backstepping controller is proposed in [4], [17] to control the longitudinal and lateral dynamics of an UAH in the presence of external disturbances. Another robust H_∞ controller to control the longitudinal and lateral dynamics of a BELL 205 helicopter is proposed in [18] where the model uncertainty is considered as an unknown parameter. A similar approach is also proposed in [19] to control the horizontal position of an UAV in the presence of horizontal wind gusts. Though the horizontal position of an UAH is successfully controlled in the aforementioned literature, the control of vertical dynamics is not clear in these papers. To address this limitation of these papers, an H_∞ control approach is proposed in [20] to control the altitude along with attitude of an UAV for a hovering flight in the presence of vertical wind gusts. To control the vertical motion of an UAH during the landing condition a nonlinear controller is proposed in [21]. However, in this paper, the main prominence was on the parametric uncertainties rather than exterior disturbances. Moreover, how to control the longitudinal and lateral position of an UAH is not also clear in this paper. To overcome this limitation, a robust backstepping controller is proposed [22] to control the horizontal position along with the vertical height of an UAH in the presence of external wind gusts. However, the controller is designed based on the linearized model of the UAH which is only applicable for the hovering flight condition. Meanwhile, the hovering flight control of an UAH especially in a gusty environment is always a tough task to the control engineer due to the intrinsic nonlinearity, dynamical unsteadiness and trembling operating conditions of an UAH. To control the hovering flight of an UAH under aforementioned conditions, a neural network based adaptive control scheme is proposed in [23]. However, the main drawback of this control scheme is that the controller is designed based the one-DOF flight condition of an UAH which is really tough to operate in a real system.

It is well known that the gross weight of an UAH could be changed during the operation of flight for several reasons such as, fuel consumption, external airdrop etc. which have significant effect on the stability of an UAH system. Since the main rotor thrust of an UAH is used to overcome the UAH's gravity, so the mass variation during the flight operation condition is an important perturbation for the force equations. To grip this situation, a nonlinear adaptive backstepping controller is designed in [24]. However, external wind gust effect is taken into account during the design of the controller In addition; no one has yet designed a robust adaptive backstepping controller by considering the effect of mass variation during the operation of flight along with the external wind gusts.

The aim of this paper is to design a robust adaptive bacstepping controller to control the outer loop of an UAH under the consideration of parametric uncertainties along with external disturbances. Note that backstepping is a recursive nonlinear control design method, which provides an alternative to the FBL control scheme. The benefit of this control scheme is that it can gain from the nonlinear stabilizing terms rather than completely eliminate them as like the FBL controller. To prove the stability of the whole system of an UAH, CLFs are formulated at different stages during the design process of the controller. Finally, a MATLAB/ SIMULINK software based simulation model is used to evaluate the effectiveness of the designed controller for hovering flight mode. To show the superiority of the proposed controller the performance is also compared with an existing controller as proposed in [22].

The rest of the paper is organized in the following approach. The mathematical model of an UAH and control problem of this paper is discussed in Section (II). The Section (III) presents the design procedure of the proposed controller based on the mathematical as discussed in Section (II). The gust model that used in the simulation is briefly discussed in Section (IV). Section (V) discusses the simulation results. Finally, the conclusion of the work is presented in Section (VI).

II. Dynamical Modeling and Problem Formulation

This section is intended to describe the dynamical model of an UAH with the single main rotor and tail rotor. The nonlinear rigid body dynamics differential equations of an UAH in terms of its translational and rotational dynamics can be expressed as follows:

$$\dot{\varsigma} = V \tag{1}$$

$$m\dot{V} = Rf_b + mg\varepsilon_3 \tag{2}$$

$$\dot{\eta} = \pi\omega \tag{3}$$

$$I\dot{\omega} = \omega \times I\omega + M \tag{4}$$

where m is the total mass of an UAH, I is the inertia matrix, $\varsigma = [x, y, z]^T$ is the position vector, $V = [v_x, v_y, v_z]^T$ is the linear velocity vector, and $\omega = [p, q\ r]^T$ represents the angular velocity rate vector. The gravitational force ($mg\varepsilon_3$) is unambiguously included in where ε_3 is a unit vector with one in the third place. In this paper, the Euler angle vector $\eta = [\phi, \theta, \psi]^T$ is defined in the roll, pitch and yaw sequence. Based on the Euler angle vector definition, the rotation matrix R from the body frame to inertia frame of an UAH in terms of the Euler angles can be represented as follows:

$$R=\begin{bmatrix} c_\theta c_\psi & s_\phi s_\theta c_\psi - c_\phi s_\psi & c_\phi s_\theta c_\psi + s_\phi s_\psi \\ c_\theta s_\psi & s_\phi s_\theta s_\psi + c_\phi c_\psi & c_\phi s_\theta s_\psi - s_\phi c_\psi \\ -s_\theta & s_\phi c_\theta & c_\phi c_\theta \end{bmatrix} \quad (5)$$

where the compressed notation *c* denotes for *cos (.)* and *s* for *sin(.)*. A conventional single main rotor helicopter has four independent control inputs which can be defined as follows:

$$u_c = [T_m, T_t, a_1, b_1]^T$$

there the components T_m and T_t represent the magnitude of the main rotor and tail rotor thrusts on an UAH, respectively. It is well known that the leading force of an UAH is produced due to the main and tail rotor thrusts. However, for the simplification, most of the researchers design the controllers by considering only the main and tail rotor dynamics while ignoring the horizontal stabilizer, vertical stabilizer, and fuselage stabilizer dynamics. Based this statement, the force generated by the main rotor (f_b) in the (x_b, y_b, z_b) direction in the body frame of an UAH can be written as follows:

$$f_b = [X_m, Y_m, Z_m]^T \quad (6)$$

However, in order to include the effects of the rotor flapping on the fuselage rigid body dynamics of an UAH, the forces and moments need to be expressed in terms of rotor flapping. Thus, the force components generated by the main rotor thrust can be written as follows:

$$X_m = -T_m sin(a_1) \quad (7)$$

$$Y_m = T_m sin(b_1) \quad (8)$$

$$Z_m = -T_m \quad (9)$$

By the way, the tail rotor generates force only in the y_b direction. Therefore, the force generated by the tail rotor can be represented by the following equation:

$$[X_t, Y_t, Z_t]^T = [0, -T_t, 0]^T \quad (10)$$

Now, after including the UAH's weight, the complete force vector equation can be written as follows:

$$f_b = \begin{bmatrix} -T_m \sin(a_1) \\ T_m \sin(b_1) - T_t \\ -T_m \end{bmatrix} + R^T \begin{bmatrix} 0 \\ 0 \\ mg \end{bmatrix} \quad (11)$$

Similarly, the net moments which acting on an UAH can be written as follows:

$$M = \begin{bmatrix} L \\ R \\ N \end{bmatrix} = \begin{bmatrix} L_b b_1 + Y_m M_z + T_t T_z \\ M_a a_1 + X_m M_z \\ M_Q + Y_m M_x + T_t T_x \end{bmatrix} \quad (12)$$

where L_b and M_a are the effective flap-stiffness constants, M_x and M_z are the main rotor lengths, T_x and T_z are the tail rotor lengths with respect to the centre of the body's mass of an UAH. There exist a simple relationship between the drag vector (M_Q) and main rotor thrust (T_m) which can be written as follows:

$$M_Q = C_m T_m^{1.5} + D_m \quad (13)$$

where $C_m = \frac{1}{\Omega\sqrt{2\rho A}}$ is the coefficient expressing the relationship between the main rotor thrust and drag; D_m is the initial drag of the main rotor when the blade pitch is zero. The aerodynamic forces and moments owing to the empennage and fuselage are negligible during the hovering and low-velocity flight so these forces are not taken into account during the design of the controller. One of the important factor of mass changes which mainly occurs owing to the airdrop or fuel consume during the operation of flight which has a significant impact on the main rotor thrust [23] to maintain the dynamical stability of an UAH. Moreover, the stability of the UAH system will be disturbed due to the external wind gusts. Therefore, a controller should be designed in such way that which can handle these problems and stabilize the system. Based on the above assumptions, the dynamical model of the outer loop of an UAH can be written as follows:

$$\dot{\varsigma} = V + \varepsilon_1 \quad (14)$$

$$\dot{V} = R f_b \theta + g\sigma_3 + \varepsilon_2 \quad (15)$$

where $\theta = \frac{1}{m}$ is the parametric uncertainty, ε_1, and ε_1 are the bounded external disturbances within the outer loop model of an UAH. When the UAH system is subjected to the variations in parameters and external disturbances; the proposed controller should have the capability to maintain the UAH system stability. The design of a robust nonlinear adaptive outer loop controller based on (14) and (15) is shown in the following section in order to accomplish the desired control objectives.

III. Robust Adpative Controller Design

The aim of this section is to design an outer loop controller, using a robust adaptive backstepping control approach to control the numerous corporeal quantities (e.g. position, linear velocity) during hovering flight mode condition of an UAH. Thus, the proposed controller should be designed in such a way that all tracking errors can converge to zero for any initial condition of an UAH. The controller design procedure is separated into two distinct steps which are ornately discussed as follows.

Design step 1: According to the design purpose, the position tracking error variable can be defined as follows:

$$e_1 = \varsigma - \varsigma_d \quad (16)$$

where $\varsigma_d$is the desired position of an UAH which is constant. Using (14), the derivative of e_1 can be written as:

$$\dot{e}_1 = V + \varepsilon_1 \quad (17)$$

Here, V is a virtual control variable and its desired value V_d is a stabilizing function for (17), for that another error variable need to be defined which can be represented as follows:

$$e_2 = V - V_d \quad (18)$$

By substituting (18) into (17), yields

$$\dot{e}_1 = V_d + e_2 + \varepsilon_1 \quad (19)$$

In order to stabilize the position error dynamics of an UAH as represented by (19), the CLF can be written as:

$$W_1 = \frac{1}{2} e_1^T e_1 \quad (20)$$

whose time derivative is

$$\dot{W}_1 = e_1^T \dot{e}_1 \quad (21)$$

By substituting the value of $\dot{e}_1$, (21) can be written as:

$$\dot{W}_1 = e_1^T(V_d + e_2 + \varepsilon_1) \quad (22)$$

Now, the objective of this step is to design the stabilizing function V_d in such a way that which makes $\dot{W}_1 \leq 0$. Therefore, the stabilizing function V_d is chosen as:

$$V_d = -\alpha e_1 \quad (23)$$

where α is a positive constant which can be used to tune the output response. Then, (22) can be written as:

$$\dot{W}_1 = -\alpha|e_1|^2 + e_1^T(e_2 + \varepsilon_1) \quad (24)$$

where $|.|$ denotes the Euclidean norm throughout this paper. It can be seen that (24) will be negative semi-definite if $e_1 = 0$ and this will happen if $e_2 = 0$. Hence, the next control objective is to drive e_2 to zero which is discussed in the following step.

Design step 2: The time derivative of V_d as represented by (23) can be written as follows:

$$\dot{V}_d = -\alpha(V + \varepsilon_1) \quad (25)$$

As $\dot{e}_1 = V + \varepsilon_1$. The velocity error dynamics along with (15) and (25) can be written as:

$$\dot{e}_2 = Rf_b\theta + g\sigma_3 + \varepsilon_2 + \alpha(V + \varepsilon_1) \quad (26)$$

Since θ is an unknown parameter; it is possible to rewrite (26) as follows by assuming $\hat{\theta}$ is an estimated value of the unknown parameter.

$$\dot{e}_2 = Rf_b(\tilde{\theta} + \hat{\theta}) + g\sigma_3 + \varepsilon_2 + \alpha(V + \varepsilon_1) \quad (27)$$

where $\tilde{\theta} = \theta - \hat{\theta}$ is the estimation error of the unknown parameter θ. In order to handle the unknown parameter, the final CLF can be written as:

$$W_2 = W_1 + \frac{1}{2}e_2^T e_2 + \frac{1}{2\gamma}\tilde{\theta}^T\tilde{\theta} \quad (28)$$

where γ is the adaptation gain which is selected in such a way that the estimation error converges to zero. Now the time derivative of Lyapunov function W_2 is

$$\dot{W}_2 = \dot{W}_1 + e_2^T\dot{e}_2 - \frac{1}{\gamma}\tilde{\theta}^T\dot{\hat{\theta}} \quad (29)$$

By inserting (24) and (27) into (29), it can be written as:

$$\dot{W}_2 = -\alpha e_1^T e_1 + e_1^T e_2 + e_2^T\left(Rf_b\hat{\theta} + g\sigma_3 + \varepsilon_2 + \alpha V\right) - \frac{1}{\gamma}\tilde{\theta}^T\left(\dot{\hat{\theta}} - \gamma e_2^T Rf_b\right) + e_1^T\Delta_1 \quad (30)$$

where $\Delta_1 = \varepsilon_1(e_1^T + \alpha)$. Now

$$Rf_b = -\frac{1}{\hat{\theta}}\left(e_1 + g\sigma_3 + \alpha V + \beta e_2 + F_2 sgn(e_2)\right) - e_1^T F_1 sgn(e_1) \quad (31)$$

where sgn is a signum function which can be written as follows:

$$sgn(e) = \begin{cases} +1 & if \quad e > 0 \\ 0 & if \quad e = 0 \\ -1 & if \quad e < 0 \end{cases} \quad (32)$$

The external disturbances Δ_1 and ε_2 are assumed to be bounded by known constants F_1 and F_2 that is $|\Delta_1| \leq F_1$ and $|\varepsilon_2| \leq F_2$. In such cases, (30) can be rewritten as follows:

$$\dot{W}_2 = -\alpha e_1^T e_1 - \beta e_2^T e_2 - \frac{1}{\gamma}\tilde{\theta}^T\left(\dot{\hat{\theta}} - \gamma e_2^T Rf_b\right) + e_2^T\left(\varepsilon_2 - F_2 sgn(e_2)\right) + e_1^T\left(\Delta_1 - F_1 sgn(e_1)\right) \quad (33)$$

Since $\tilde{\theta}^T$ is not available, the best way to cancel this term from $\dot{W}_2$ by selecting the following adaptation law

$$\dot{\hat{\theta}} = \gamma e_2^T Rf_b \quad (34)$$

Finally, the resulting derivative of W_2 can be written as follows:

$$\dot{W}_2 = -\alpha e_1^T e_1 - \beta e_2^T e_2 + e_2^T\left(\varepsilon_2 - F_2 sgn(e_2)\right) + e_1^T\left(\Delta_1 - F_1 sgn(e_1)\right) \quad (35)$$

Now using the Schwartz inequality, we have

$$\dot{W}_2 \leq -\alpha|e_1|^2 - \beta|e_2|^2 + |e_2|(|\varepsilon_2| - F_2|e_2|) + |e_1|(|\Delta_1| - F_1|e_1|) \quad (36)$$

which assures that the time derivative of Lyapunov function is negative definite. Thus, from (36), it is obvious that the error dynamics of the system is asymptotically stable. Therefore, the derived robust adaptive backstepping control law stabilizes the whole system. The gust model which is used for the simulation to test the robustness property of the designed controller is discussed in the following section.

IV. GUSTS MODEL

To analysis the robustness of the designed controller in a gusty environment, wind gusts can be treated as spectral turbulence or discrete. For spectral turbulence gusts, the Von Karman and Dryden turbulence models are normally used as the external disturbances. However, the Von Karman model has been widely used between these two models. Anyway, due to the computational complexity of the Von Karman model, we will not use this model in this paper. There are several sources for wind gust models based upon the empirical data which consists of band limited white noise which is passing through appropriate forming filters. On the other hand, the Dryden turbulence model is scaled with respect to altitude, velocity and wing span of an UAH. Therefore, the filters that are used to generate the Dryden wind gust model for atmospheric disturbance can be written as follows:

$$H_u(s) = \sigma_u\sqrt{\frac{2L_u}{\pi U}}\frac{1}{1+\frac{L_u}{U}s} \quad (37)$$

$$H_v(s) = \sigma_v\sqrt{\frac{2L_v}{\pi U}}\frac{1+\frac{\sqrt{3}L_v}{U}s}{(1+\frac{L_v}{U}s)^2} \quad (38)$$

$$H_w(s) = \sigma_w\sqrt{\frac{2L_w}{\pi U}}\frac{1+\frac{\sqrt{3}L_w}{U}s}{(1+\frac{L_w}{U}s)^2} \quad (39)$$

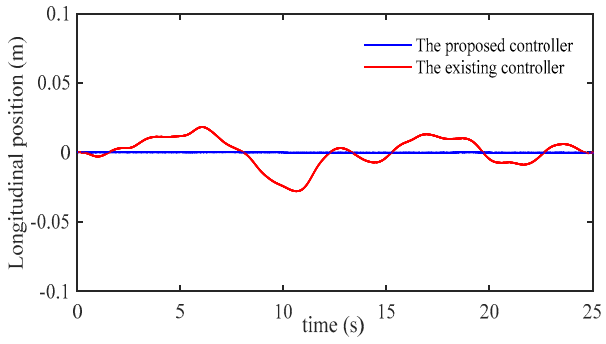
where the symbols have their usual meaning which can be found in [20]. Simulation studies are conducted in the following section to show the effectiveness of the designed controller.

V. CONTROLLER PERFORMANCE EVALUATION

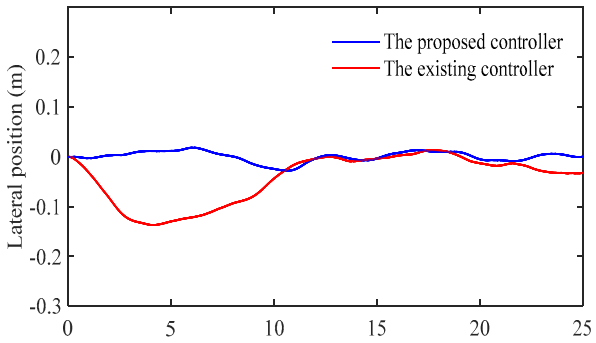
In this section, the performance of the designed robust adaptive controller is evaluated on a VARIO XLC helicopter platform using the MATLAB/SIMULINK software simulation model by considering the parameters those are engaged in real applications for an UAH. The simulation is completed for the hovering flight stabilization of a VARIO XLC helicopter under the consideration of external wind gusts effect on the system. The initial and desired positions of a VARIO XLC helicopter are set to x=0 m, y=0 m and z=-2 m, respectively. Furthermore, in order to verify the robustness of the designed controller, the bound on the external uncertainties are selected as: F_1= 0.2 m/s and F_2= 0.4 m/s^2, respectively based on the variation of linear velocity and acceleration due to the external wind gusts. Since the performance of the designed controller

is affected by the choice of control parameters, so the tuning parameters for the controller are considered as: k_1 =2 and k_2=5, and the adaptation gain parameter is chosen as: γ_1 =10. Atmospheric external wind gust is also included with the simulation model to show the real-world environment effect on the designed controller. In order to show the superiority of the designed controller, the performance of the designed controller is also compared with an existing controller as proposed in [22].

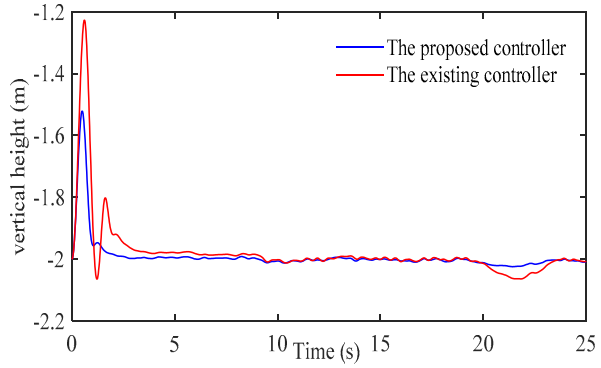
The position responses of the VARIO XLC helicopter are shown in Fig. 1, from where it can be seen that the position responses are settled down to their desired equilibrium position in a much faster way with designed controller (solid blue line) as compared to the existing controller (solid red line).



(a) Longitudinal position



(b) Lateral position

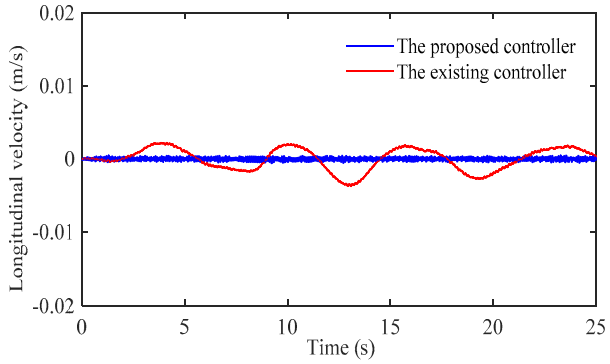


(c) Vertical height

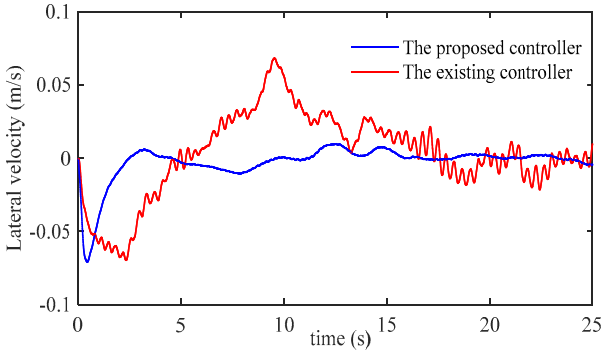
Fig. 1. Position responses of a VARIO XLC helicopter

From Fig. 1, it is also obvious that the designed controller can damp out the frequency of oscillation and also can eliminate the steady-state error in a better way as compared to the existing controller.

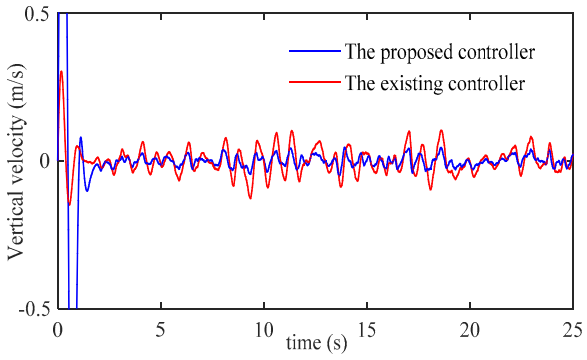
In order to ensure the stable hovering flight operation of a VARIO XLC helicopter, the linear velocity should be zero. However, the linear velocity will be disturbed when the external disturbances will be added to the system. Under this situation, the linear velocity responses of a VARIO XLC helicopter with both controllers are shown in Fig. 2. From Fig. 2, it can be seen that speed of convergence to the desired point of the designed controller (solid blue line) is much faster than of the existing controller (red solid line). The corresponding external wind gusts response is shown in Fig. 3, which is used in simulation to test the controller performance.



(a) Longitudinal velocity



(b) Lateral velocity



(c)Vertical velocity

Fig. 2. Linear velocity responses of a VARIO XLC helicopter

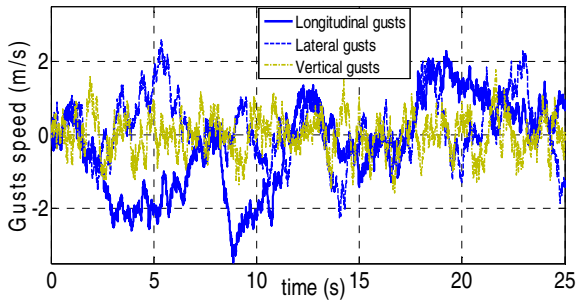


Fig. 3. Test wind gusts response

From simulation results, it is obvious that the designed controller is more effective and able to provide better system performances than the existing controller in terms of settling the responses to their desired steady-state operating points.

VI. CONCLUSION

In this paper, a recursive method is proposed to design a nonlinear robust adaptive backstepping controller for an UAH in order to improve the hovering flight performance in the presence of both parametric uncertainties and external disturbances. Based on this approach, a robust adaptive control law for uncertain parameters is designed in this paper and the external disturbances are also bounded to avoid the deterioration of the designed controller performance. From the simulation results, it is evident that the designed controller is able to maintain desired steady-state hovering flight condition in a better way as compared to the existing robust backstepping controller. Future work will deal with the justification of the designed control scheme in a real system e.g., flight test under the consideration of both parametric and external uncertainties to prove the practicability of the designed controller in the true life.

REFERENCES

[1] T. K. Roy, M. Garrat, H. R. Pota, and H. Teimoori, "Hover flight control of a small helicopter using robust backstepping and PID," in Proc. of the IEEE 10th WCICA, 6-8 July 2012.

[2] H. Shim, T. J. Koo, and F. Hoffmann, "A comprehensive study of control design for an autonomous helicopter," in Proc. of the 37th IEEE International Conf. on Decision and Control, pp. 3653-3658, 1998.

[3] X. H. Xia, and Y. J. Ge, "Finite-horizon optimal linear control for autonomous soft landing of small-scale helicopter," in Proc. of the IEEE Int. Conf. on Information and Automation, pp. 1160-1164, 2010.

[4] T. K. Roy, M. Garrat, H. R. Pota, and H. Teimoori, "Robust control for longitudinal and lateral dynamics of small scale helicopter, in Proceedings of the 31st IEEE CCC, 25-27 July, 2012.

[5] L. Guo, C. Melhuish, and Q. Zhu, "Towards neural adaptive hovering control of helicopters," in Proceedings of the IEEE International Conf. Control Applications, pp. 54-58, 2002.

[6] T. K. Roy and M. Garrat, "Altitude control of an unmanned autonomous helicopter via robust backstepping controller under horizontal wind gusts," in Proc. of the IEEE 7th Int. Conf. on Electrical and Computer Engineering, pp. 771-774, 20-22 December 2012.

[7] T. K. Roy, M. Garrat, H. R. Pota, and H. Teimoori, "Robust altitude control of an unmanned autonomous helicopter using backstepping," in Proc. of the 10th WCICA, pp. 1650-1654, 6-8 July 2012.

[8] N. Kim, A. Calise, N. Hovakimyan, J. V. R. Prasad, and J. E. Corban, "Adaptive output feedback for high bandwidth flight control," Journal Guide, Control Dynamic, vol. 25, no.6, pp. 993-1002, 2002.

[9] T. J. Koo, and S. Sastry, "Output tracking control design of a helicopter model based on approximation linearization," in Proceedings of the 37th International Conf. on Decision and Control, pp. 3635-3640, 1998.

[10] K. A. Danapalasingam, J. J. Leth, A. la C. Harbo, and M. Bisgaard "Robust helicopter stabilization in the face of wind disturbance" in Proc. of the 49th IEEE Conf. on Decision and Control, Dec., 2010.

[11] A. Salihbegovic and M. Hebibovic, "Attitude tracking of the small-scale helicopter system using disturbance observer based sliding mode control," in Proc. of the 22nd Med. Conf. of, pp. 1578-1583, 2014.

[12] Y. Wang, X. Xu and Y. Dai, "A location tracking control for unmanned small scale helicopter using sliding mode controller," in Proc. of the 6th Int. Conf. on IHMSC, pp. 152-155, 2014.

[13] T. K. Roy, "Robust adaptive control for longitudinal and lateral dynamics of a small scale helicopter," IJCCNet, vol. 1, no. 3, pp. 22 – 34, 2012.

[14] T. K. Roy, "Horizontal position control of a small scale helicopter using adaptive backstepping controller," IJCCNet, vol. 2, no. 2, pp. 1 – 14, May 2013.

[15] M. Krstic, I. Kanellakopoulos, and P. Kokotovic, "Nonlinear and adaptive control design" Wiley, 1995.

[16] T. K. Roy, "Longitudinal and lateral dynamics control of a small scale helicopter using adaptive backstepping controller under horizontal wind gusts," Asian Transaction on Engineering, vol. 3, no. 1, March 2013.

[17] T. K. Roy, M. Garrat, H. R. Pota, and H. Teimoori, "Robust backstepping control for longitudinal and lateral dynamics of small scale helicopter," Journal of University of Science and Technology of China, vol. 42, no. 7, July 2012.

[18] D. J. Wallcer, M.C. Turner, A. J. Smerlas, ME. Strange, and A.W. Gubbels, "Robust control of the longitudinal and lateral dynamics of the BELL 205 helicopter," in Proc. of the ACC, June 1999.

[19] X. Yang, M. Garratt, and H. R. Pota, "A nonlinear position controller operations of rotary-wing UAVs," 18th IFAC World Congress, August 28- September 2, 2011.

[20] S. Suresh, P. Kashyab, and M. Nabi, "Automatic take-off control system for helicopter an H∞ approach," in Proc. of the 11th Conf. Control, Automation, Robotics and Vision, 7-10 December, 2010.

[21] A. Isidori, L. Marconi, and A. Serrani, "Robust nonlinear motion control of a helicopter," IEEE Transactions on Automatic Control, vol. 48, no. 3, March 2003.

[22] T. K. Roy, "Position control of a small helicopter using robust backstepping," in Proc. of the 7th IEEE Int. Conference on Electrical and Computer Engineering, pp. 787-790, 20-22 December 2012.

[23] L. Guo, C. Melhuish, and Q. Zhu, "Towards neural adaptive hovering control of helicopters," in Proc. IEEE International Conf. Control Applications, pp. 54-58, September 2002.

[24] C.-T. Lee, and C.-C. Tsai, "Adaptive backstepping integral control of a small-scale helicopter for airdrop missions," Asian Journal of Control, vol. 12, no. 4, pp. 531-541, July 2010.

[25] T. K. Roy and A. A. Suman, "Adaptive backstepping controller for altitude control of a small-scale helicopter by considering the ground effect compensation," in Proc. of the IEEE 2nd Int. Conference on Informatics, Electronics and Vision (ICIEV), pp. 1-5, May 2013.

[26] T. K. Roy, "Robust backstepping control for small helicopter," Masters of Engineering thesis, UNSW, Canberra, Australia, November 2012.

[27] T. K. Roy, M. Morshed, F. K. Tumpa, M. F. Pervej, "Robust adaptive backstepping speed controller design for a series DC motor," in Proc. of IEEE Int. Women in Eng. Conf. on Elect. and Computer Eng., December 2015.

[28] T. K. Roy, M. F. Pervej, and F. K. Tumpa, "Adaptive controller Design for grid current regulation of a CSI based PV system (accepted)," in Proc. of the 2nd IEEE ICECTE, December 2016.

[29] T. K. Roy, M. F. Pervej, F. K. Tumpa, and L. C. Paul, "Nonlinear adaptive controller design for velocity control of a DC motor driven by a DC-DC buck converter using backstepping approach (accepted)," in Proc. of the 2nd IEEE ICECTE, December 2016.

Cloud-Based Solution for Improvement of Response Time of MySQL RDBMS

Sayed Md. Fahim Fahad
Department of Computer Science and Engineering
Jahangirnagar University
Dhaka, Bangladesh
fahimfahad92@gmail.com

Mohammad Shorif Uddin
Department of Computer Science and Engineering
Jahangirnagar University
Dhaka, Bangladesh
shorifuddin@juniv.edu

***Abstract* - MySQL is an open-source very popular RDBMS to store and manage data. But it faces difficulties of prolonged response time to handle big data. On the other hand, cloud is a perfect solution for managing and querying big data but it is not suitable to process small data due to delayed response time. In this paper, we propose a cloud-based solution that will improve response time and performance of the system to handle big as well small data. We combine MySQL and cloud to store and manage all the data. We also focus on migration of data between MySQL and cloud. It is found that the proposed solution performs better than the only MySQL solution.**

Index Terms – MySQL; Big data; Hive; Migration; Response time

I. INTRODUCTION

Performance is always an important issue for any kind of system. To function, a database server needs four fundamental resources: CPU, memory, disk, and network. If any one of these resources is weak or overloaded, then the database server is very likely to perform poorly [1]. The effective maximum table size for MySQL databases is usually determined by operating system constraint on file sizes, not by MySQL internal limits [2]. On Windows 32-bit platforms, it is not possible by default to use more than 2GB of RAM within a single process, including MySQL. This is because the physical address limit on Windows 32-bit is 4GB and the default setting within Windows is to split the virtual address space between kernel (2GB) and user/applications (2GB). Windows systems have about 4,000 ports available for client connections, and after a connection on a port closes, it takes two to four minutes before the port can be reused. In situations where clients connect and disconnect to and from the server at a high rate, it is possible for all available ports to be used up before closed ports become available again. If this happens, the MySQL server appears to be unresponsive even though it is running. Ports may be used by other applications running on the machine as well, in such case the number of ports available to MySQL is lower [3]. On the other hand, the amount of data produced by the users in the last few years is enormous. The systems that are running on MySQL are not really suitable for handling big data. Usually, the response time increases radially [4], [5].

It is important to find a better solution for these types of systems to improve response time. Cloud computing can be a very good solution for handling huge amount of data. It brings remarkable changes in the IT industry. Cloud based system provides lot of advantages like fast automated recovery from failures, better performance, scalability, elasticity, etc. [6],[7].

Previously, Bezavada and Cheppalli [8] in their paper drew an analogy for data management between the traditional relational database systems and the big data techniques. They tried to find the business insights of current user records. Ionuț [9] presented a set of best practices for moving PHP web applications from a traditional hosting to a cloud based one. Ruchi Nanda [10] focused on the study of different existing DBMS in the cloud computing domain. Rakesh et al. [11] showed the comparison between SQL and HiveQL. HiveQL is the query language for Hive. Dhurba [12] showed how hive plays an important role to manage petabyte level data in Facebook. Pushpalatha and sudhee [13] showed how Hive interface can help to process huge data easily.

In this paper, we tried to improve the response time of the system by combining MySQL and cloud. Necessary conditions for combining MySQL and cloud are also discussed.

The rest of this paper is organized as follows. Section II provides the methodology of the proposed system. Section III provides the experimentation, and finally, the conclusions are drawn in Section IV.

II. METHOD

A. Data migration reasoning

Migration from MySQL to cloud requires huge reengineering. If we migrate the database from MYSQL to cloud there will be a development cost. The development cost may increase depending on the system complexity. More importantly not every system is suitable for migration. The systems that only use big data should use cloud only. Combination will not be beneficial if the systems do not use big data and metadata at the same time. Here the term “meta data” means the information we get by analysing big data. The necessary conditions for combination will discuss later. We can reduce reengineering cost by combining MySQL and cloud. The system will perform all the regular functionalities to collect data from clients and store them in MySQL. Data validation, concurrency control, constraints definition, security management issues will not be changed. The only thing needs to be noted that the transfer of data from MySQL to cloud after a specific time period. This period will be defined by the system admin. Admin will determine what type

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

of data will be transferred to cloud. As the data is transferred to cloud, query related to those data will be submitted to cloud.

B. Necessary conditions

As we said earlier, not every system is suitable for this solution. The systems that use metadata and huge client data are suitable for combining MySQL and cloud. Client data are the big data here. So these will be transferred to cloud. As cloud performs better on big data, metadata should remain in MySQL. Some systems use client's data for further analysis and produce business reports or statistical outputs. These types of data also should remain in MySQL.

C. Technology selection

There are few cloud technologies that we can use to store migrated data. They are Hive, Pig Latin, and HBase. The selection should be done considering the kind of system we are using.

Hive provides us data warehousing facilities on top of an existing Hadoop cluster. It supports parallel processing. Query Language popularly known as HiveQL which is similar to SQL. Schema definition is a must. User can create tables and store data in Hive just like MySQL. It provides us with various tools for easy extraction, transformation and loading of data. Hive is good for analysis. So systems that require analyzing data and generating reports, can use Hive [11], [12], [13].

Pig Latin is a dataflow language. Programmers write scripts to process data. Data is processed using parallel processing technique. The schema definition is optional. Data is loaded to cloud before it is processed using pig. It is suitable for those who would like to use cloud as per demand. Data will be loaded to cloud using pig scripts [14], [15], [16], [17], [18].

Hbase is a database that is used to store and retrieve data. HBase enables fast read and write functions with scalability. So HBase should be used when system needs random, real time read/write access to big data. It can be used for online processing system [19], [20]. For our example system, we use MySQL and Hive to prove the theory.

D. Migration Proposition

The data migration must be done carefully and efficiently, as it is the most important step. Errors must be checked. The migration flow chart is given in Fig. 1. In the flow chart, we used a term "upload_time" which is determined by system admin. Upload time is the time when MySQL table data will be migrated to hive.

During the configuration time, system admin must set the value of upload_time. This will also determine the frequency of data migration from MySQL to cloud. When the system starts, it will check the upload_time and if the current_time matches the upload_time then data will be migrated. We used CSV files to hold data temporarily during migration. Hive can easily read data from CSV files. After the successful migration data will be deleted from MySQL and CSV files.

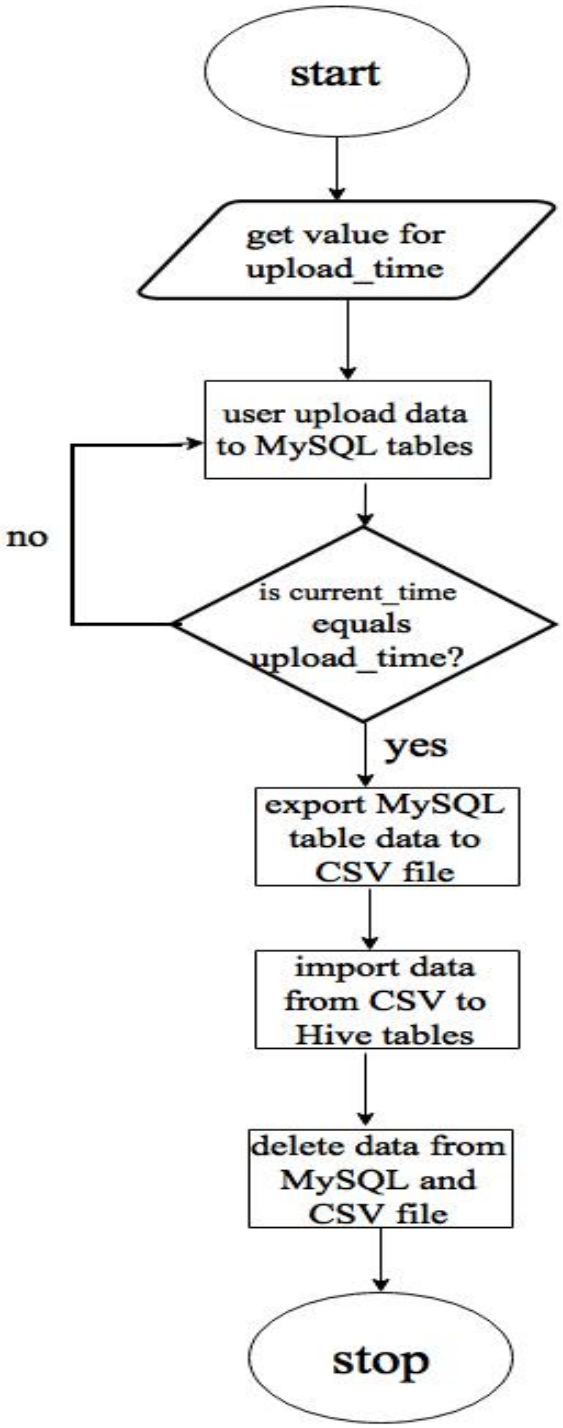


Fig 1: Data handling flow chart of the proposed migration model

III. EXPERIMENTATION

To test the theory, we used a demo project "Patient health history management" that stores and manages patient health records and simple metadata related to patients [total number

of cancer affected patients in the previous year**].** We did experimentation ranging data records from 1200 to 309600 using the following three different types of query. The database is designed and developed by us and is available upon request.

Database schema: Table name (fields) (approximate size of one record)

user (id, firstname, lastname, address, nid, father_nid, mother_nid, DOB, blood_grp, active, telephone, gender) (One record requires around 154B)
doctor (id, user_id, educational qualification, hospital, experience) (One record requires around 3.52KB)
health_history (id, user_id, doctor_id, date, type_of_disease, description) (One record requires around 1.15MB)
medical_reports (id, health_history_id, report, summary) (One record requires around 8.025MByte)
user_summary (id**,** type_of_user, no_of_active_user) (One record requires around 5.33KB)
health_record_summary (id, type_of_disease, year, no_of_patient) (One record requires around 5.33KB)

Machine configuration:
For MySQL: Application server (Apache Tomcat 7), Processor 2.20GHz, Memory 4GB
For Hive: 2 Laptops (2.20GHz, 8GB RAM)

Fig. 2 shows the result of the first query. First query is a complex query which requires three join operations. It enquires for the patient detail health history and all the medical reports related to patient. Surely this type of query will have to handle the big data situation at some point of time. We tested the query response time for both MySQL and Hive. MySQL performs better for smaller table size. But when the size grows with time, Hive starts to perform better. When the data size is greater than 208800 records, Hive response time is better than MySQL.

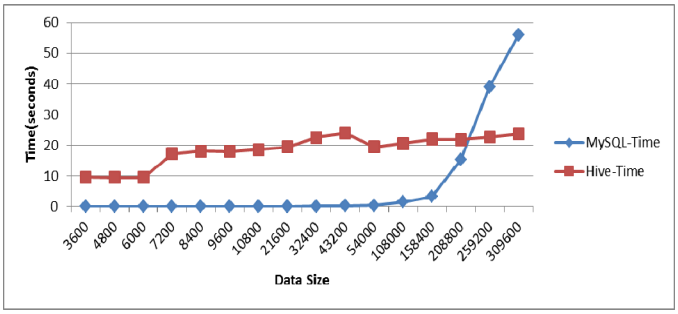


Fig. 2: Big data query for a single patient detail with health history and medical reports.

Fig. 3 shows the result of the second query. Second query used two join operations. This will list the patients treated by a specific doctor. It is much simpler query comparing to the first one. But it will also face the big data situation. When the data size is nearly 259200, Hive performs better than MySQL. The greater the data size, the better the Hive's performance.

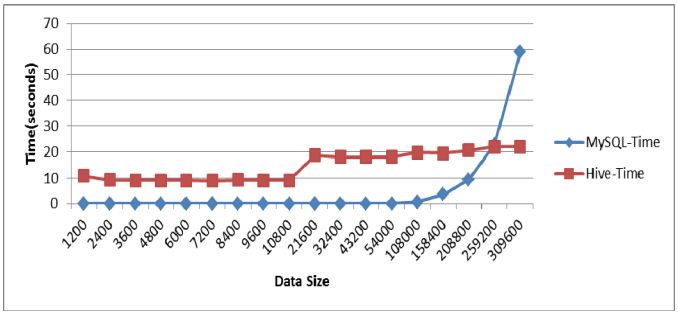


Fig. 3: Big data query for a single doctor with list of patients treated.

Fig. 4 shows the result of the third query. Third query is on metadata and the record size is 309600. We analysed patient data and stored these in another table. This is a very small table. From these analysed data we get the statistical information. The query is the simplest one - not a big data related query. It will get the number of cancer affected patients. The query will not require any join operation. As the table size is not big, MySQL needs only fraction of a second but Hive needs more than 5 seconds. So, MySQL outperformed Hive on metadata queries.

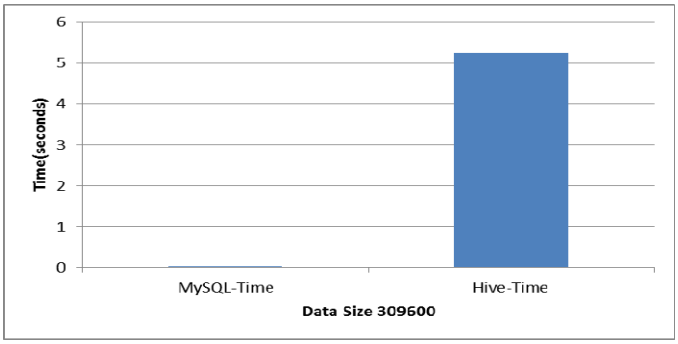


Fig 04: Meta data query for total number of cancer affected patients.

A methodology may revive.

IV. Conclusion

In this paper, we combined MySQL and cloud to handle the systems that use big data and metadata at the same time. Results showed that the response time gets better when we combine MySQL and cloud. From the above, we can see that response time gets better for Hive with the increased data size and complex queries. MySQL performs better for small data and simpler queries. Systems those require complex query processing and huge data management process may use this solution with an ease. We have not tested our solution under various conditions (for example, big data sample, big cluster

size, powerful MySQL server) yet. In the future, we will test it for different situations and much bigger data sample.

REFERENCES

[1] Baron Schwartz, “10 essential performance tips for MySQL”. Article. 2012. [Online]. Available at “http://www.infoworld.com/article/2616674/database-administration/10-essential-performance-tips-for-mysql.html”. [Accessed on: December 29,2015]

[2] “Limits on Table Size”. MySQL Documentation. [Online]. Available at “https://dev.mysql.com/doc/mysql-reslimits-excerpt/5.6/en/table-size-limit.html”. [Accessed on: December 28,2015]

[3] “Windows Platform Limitations”, MySQL documentation. [Online]. Available at “https://dev.mysql.com/doc/mysql-reslimits-excerpt/5.1/en/limits-windows.html”. [Accessed on: December 15,2015]

[4] Peter Zaitsev, "Why MySQL could be slow with large tables?”. Article. 2006. [Online]. Available at https://www.percona.com/blog/2006/06/09/why-mysql-could-be-slow-with-large-tables/. [Accessed on: December 28, 2015]

[5] MySQL limitations. 2016. [Online]. Available at http://searchit channel.techtarget.com/feature/MySQL-limitations. [Accessed on: December 29, 2015]

[6] M. A. Hadi, "Overview of Cloud Computing Towards to Future Networks", International Journal of Computer Science and Innovation, vol 2015, issue 2, pp. 68-78, 2015. [Accessed on: December 15, 2015]

[7] Richi Jennings, “5 Financial Benefits of Moving to the Cloud”. Article. 2015. [Online]. Available at “http://www.webroot.com/ie/en/business/resources/articles/cloud-computing/five-financial-benefits-of-moving-to-the-cloud”. [Accessed on: December 16, 2015]

[8] Prasadu Bezavada and Sulakashan Cheppalli, "Data Modeling Considerations in Telecommunications using Hadoop", International Journal of Scientific Engineering and Technology Research, vol 3, issue 30, pp. 5935-5938, 2014. [Accessed on: December 15, 2015]

[9] Ionut VODÄ,, "Migrating Existing PHP Web Applications to the Cloud", Informatica EconomicÄ, vol 18, no 4, pp. 62-72, 2014. [Accessed on: December 16, 2015]

[10] Ruchi Nanda, "Review of Query Processing Techniques of Cloud Databases", Suresh Gyan Vihar University Journal of Engineering & Technology, vol 1, issue 2, pp. 12-16, 2015. [Accessed on: December 15, 2015]

[11] Rakesh Kumar, Neha Gupta, Shilpi Charu, Somya Bansal and Kusum Yadav, "Comparison of SQL with HiveQL", International Journal for Research in Technological Studies, vol 1, issue 9, pp. 28-30 August. 2014. [Accessed on: December 16, 2015]

[12] Dhruba Borthakur, "Petabyte Scale Data at Facebook". Presented at XLDB Conference at Stanford University. 2012. Available at “https://www.conf.slac.stanford.edu/xldb2012/talks/xldb2012_wed_1105_Dhruba Borthakur.pdf”. [Accessed on: December 28, 2015]

[13] N. Pushpalatha and P. sudhee, "Data processing in big data by using Hive interface", International Journal of advance research in computer science and management studies, vol 3, issue 4, pp. 272-277, April. 2015. [Accessed on: December 15, 2015]

[14] Pig (programming tool). 2015. [Online]. Available at https://en.wikipedia.org/wiki/Pig_(programming_tool). [Accessed on: December 15, 2015]

[15] Jeff Barr, "Pig Latin-High Level Data Processing with Elastic MapReduce", AWS Blog. 2009. [Online]. Available at “https://aws.amazon.com/blogs/aws/pig-latin-high-level-data-processing-with-elastic-mapreduce/”. [Accessed on: December 15, 2015]

[16] Data Management in the Cloud, PIG LATIN AND HIVE. 2014. [Online]. Available at “http://datalab.cs.pdx.edu/education/clouddbmswin2014/notes/CloudDB-2014-Lect10-full.pdf”. [Accessed on: December 15, 2015]

[17] White paper. MapR, Hive, and Pig on Google Compute Engine. 2014. [Online]. Available at “https://www.mapr.com/sites/default/files/mapr_hive_pig_ on_gce_4.pdf”. [Accessed on: December 28, 2015]

[18] Difference between Pig and Hive-The Two Key Components of Hadoop Ecosystem. 2014. [Online]. Available at “https://www.dezyre.com/article/difference-between-pig-and-hive-the-two-key-components-of-hadoop-ecosystem/79”. [Accessed on: December 29, 2015]

[19] Apache Hbase. 2011. [Online]. Available at “http://hortonworks.com/hadoop/hbase/”. [Accessed on: December 15, 2015]

[20] Steven Haines, "Introduction to HBase, the NoSQL Database for Hadoop”. Article. 2014. [Online]. Available at http://www.informit.com/articles/article.aspx?p=2255108. [Accessed on: December 15, 2015].

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

An Android Based Security Alert System for Female

Sanjida Sharmin, Md. Khaliluzzaman*, Sayeda Fauzia Khatun and Shajeda Khanam
Dept. of Computer Science and Engineering
International Islamic University Chittagong (IIUC)
Chittagong- 4203, Bangladesh
ssharmin114@gmail.com, *khalil_021@yahoo.co.in, fauziarafa@gmail.com, jahinshamu0308@gmail.com

***Abstract* — In continuously upgrading world, women belief in their self-worth. They have the same participation like men in almost every sector of life. But lives of women have become so vulnerable these days. Security of their lives is one of the burning questions. Everywhere they have to face unwanted incidents. Considering all the incidents towards female, this idea of Bluetooth security device aligned with an android mobile application came in consideration. During crisis, women just to press button from that security device and an automatic message of the victim's location information will be sent to selected numbers through the application. An automated message will also be sent to the nearest police station. Moreover a voice call will be sent to the first number of contact list.**

Keywords—Security; Bluetooth; location; button; application

I. INTRODUCTION

In modern era believes in women empowerment. Women are contributing in every sector of the country. But today female's security at night and even at day time is big concern. Most of the time, they are in the victims and the criminal stays free. Rape, acid violence, dowry related violence, sexual harassment, mug etc. are very common indecent in our country, which is a very shameful fact. Every day these situations are found in the daily newspaper whose are full of violence against women.

The result of VAW Survey 2011 identified that as many as 87% of currently married women have ever experienced any type of violence by current husband and 77% reported any type of violence faced during the past 12 months from the survey time. The higher percentage of any type of violence is predominantly contributed by psychological violence. Almost 90% of those who have ever violated by current husband have the past 12 month's experience of violence which implies the persistence nature of violence by the spouse.

Besides, a survey is conducted to measure the Gender-based violence in [1]. According to this survey, it is seen that in Bangladesh 66% of married women are experienced violence by current as well as previous husbands, while 98% have ever been violated by either current or previous husbands. So, overall survey depicts that the women security has become very essential.

In this section provides a descriptive summary of some technique that have been implemented and tested for women security. The Dhrubajyoti Gogoi et al. [2] proposes an Android Based Emergency Alert Button system which describes about an SOS application developed in android platform. The uniqueness of this application apart from other SOS application available is that the user need not spent time navigating inside the phone menu; unlock the screen, to trigger the service.

Jorge Zaldivar et al. [3] propose a method providing Accident Detection in Vehicular Networks through OBD-II Devices and Android-based Smartphone. Alexandre Bartel et al. [4] introduced automatically securing permission based software by reducing the attack surface: an application to android was proposed.

Sharma et al. [5] describes an android application which interprets the message a mobile device receives on possible intrusion and subsequently a reply (Short Message Service) SMS which triggers an alarm/buzzer in the remote house making others aware of the possible intrusion.

After analyzing all the ideas, an idea is emerged which is very helpful for women for helping themselves. This work consists of an android mobile application and a Bluetooth security device. The individuality of this application apart from other application available is that the user does not need to spent time to unlock the screen, open the application to search inside the contact list to make voice call and write message. At the emergency situations, user can directly press the button of the ring and thereby, a signal from the device will be received by the android mobile. After receiving signal an automatic message of their location information will be sent to selected numbers through the application and also a voice call will be sent to one of the predefined numbers. Many applications and device available which sends a custom message to the predefined number but not the location of the user .This feature of the application not only helps to find the exact location of the person in problem but also will help the nearest police to trace the location of incident immediately.

The paper is organized as follows. In Section II proposed work is described. In the next section performance result is explained. The paper is concluded in Section IV.

II. SYSTEM IMPLEMENTATION

In this section, described the system work flow in details. The flow diagram shows the main target of this work. The working flow is divided into five major parts that are:

1) connecting the security device with the mobile android application, 2) security device signal process by the mobile application, 3) retrieve the GPS location, 4) retrieve the

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

emergency contacts, and 5) finally, message and voice call is sent to the selected contact numbers with the victim's present location.6) also, sending message to the nearest police station. The workflow of the proposed work is shown in Fig. 1.

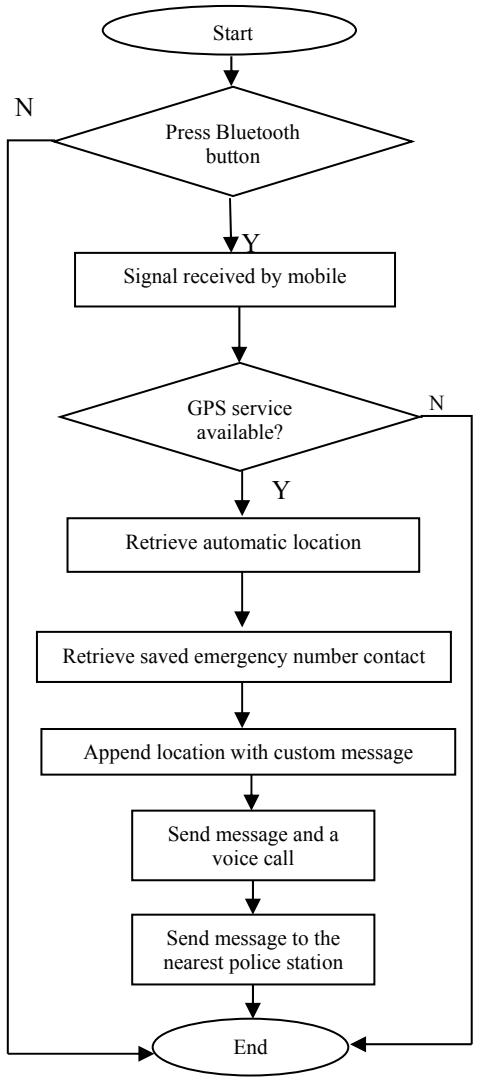


Fig 1. Workflow of the proposed system.

A. *Connecting security device with the mobile application*

In the initial stage, the security device has to be turned on. Then android mobile need to be discovered the device. And complete the pair to the android cell phone. It requires security code to be paired to make the connection secured. Fig. 2 shows the Bluetooth device which sends the signal. The Bluetooth device is in the finger ring. It will be activated when the finger ring is pressed.

B. *Security device signal process by the mobile application*

This section describes the procedure of the mobile application that process the signals come from the security device. The security device has a button for sending signals to the mobile application. At the emergency situation, the user will press the security button and it will send a signal to the mobile application. The mobile will receive the signal. After that, the mobile application performs the operation to send the user information to the predefined contact lists. For that, the mobile application has to be retrieved the GPS location of the user android smart phone.

C. *Retrieve the GPS location*

In this section describs the process of retrieving the GPS lication from the user's current locaiton through the android mobiel applcation. For this step, user needs to keep smart phone's GPS on. As soon as, the application will get the signal from the security device which is in the user's finger ring, it will start its main work on it. The application will retrieve the GPS location of the user's present location. Fig. 3 shows result of retriving user's GPS location.

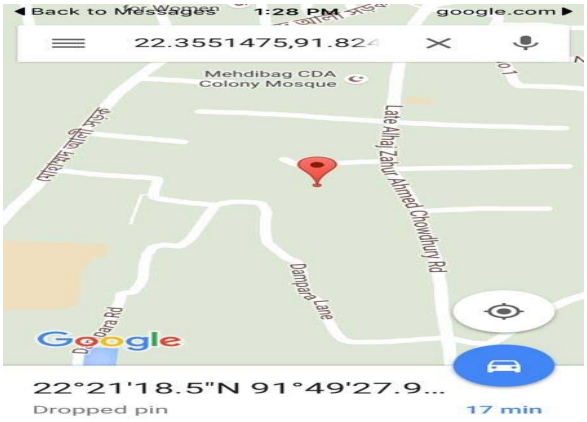


Fig 2. Retrieve the GPS location of the user's current location.

D. *Retrieve the emergency contacts*

This section described the process of retrieving the specific number of emergency contacts from pre define contact list. After getting the GPS of user's current location, the mobile application will retrieve previously defined emergency contact numbers. User can add as much number as the user want and modify it later on. Fig. 3 shows the customized contact list. When user press the security device, the current location of the user will be send to those contacts that are in the top of the contact list. The number of the contact to which the message will be send is defined by the user. However, voice call will only send to the top of the contact number. That means, voice call is passed only one contact number and message will be send a list of the contact number that is defined by the user.

Fig 3. The customized contact list.

E. Passing message and voice call to the pre-difine contacts

This section defines the procedure of sending the message to the pre selected contact list. For that, current location of the GPS is retrieved and the contacts are selected in the previous section. Now, mobile application appends the retrieval user current location GPS link to the predefined alert message. Then the message will be sent to the emergency contacts that are selected previously. User can modify the message also as per user's personal situations.

To access the message with user GPS location, the receiver also needs to keep his mobiles GPS on to get the location and direction towards it. Fig. 4 shows the message received by the receiver of specific contact list. After receiving the user message, the receiver just accesses the link to see the user's current location through the GPS system. From this GPS system the receiver can take necessary action to save the victim.

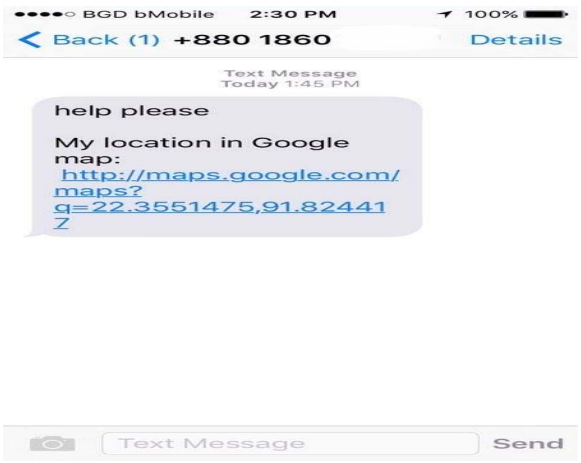


Fig 4. Sending message to the pre define contact numbers.

Fig. 5 shows the direction towards the victim's current location. According to the figure, the receiver retrieves the current location of the victim after opening the getting message which shows direction of victims via Google map.

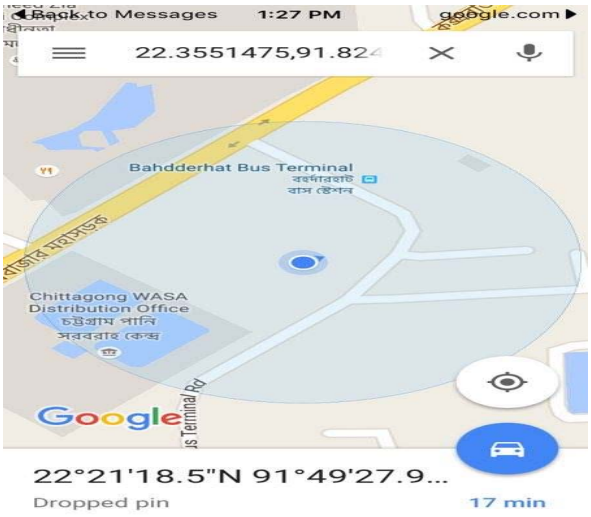


Fig 5. Direction towards victims' location.

After sending the message to the pre-defined contact list, the application sends a voice call to the top contact number shown in Fig. 6. It is an automatic process of this application. When application completes the message sending task, it is automatically forward a voice call to the specific number. From this voice call the receiver can understand the current situation of the victim. The receiver also gets the message with user's GPS location, so the receiver can take action to help the user according to the situations.

Fig 6. Sending voice call to the top contact number.

F. Sending message to the nearest police station

After sending the voice call the location of the nearest police station is traced. The direction of the police station is shown in Fig. 7 and the customize message is shown in Fig. 8. A customized message is also sent to the nearest police station.

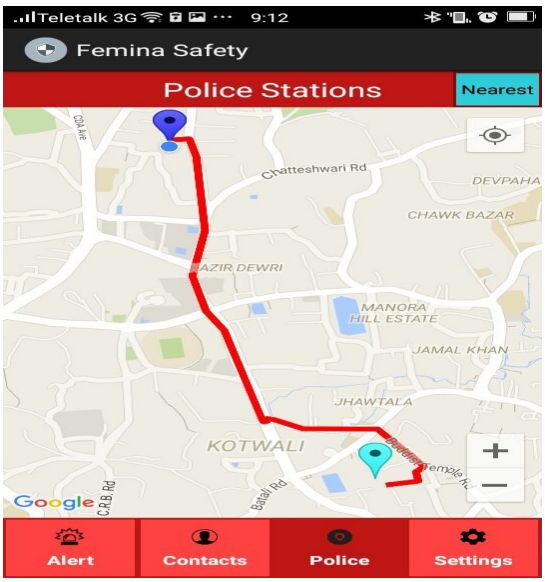


Fig 7. Location of the nearest police station

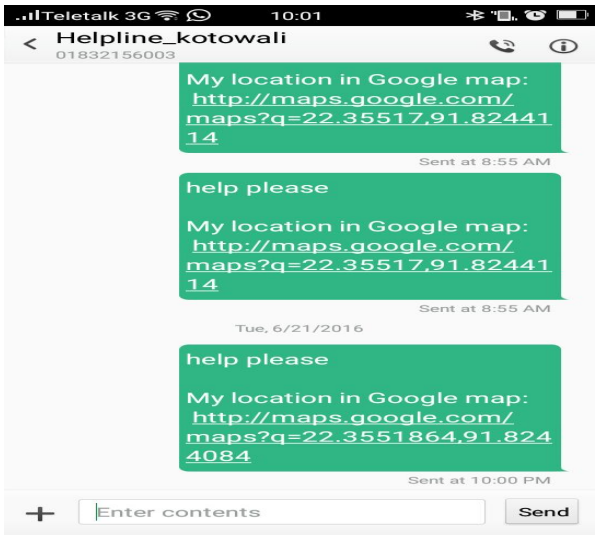


Fig 8. Sending customized message to the nearest police station.

III. Performace Analysis

This section explains the performance analysis of the proposed system. The criterion that is considered in this section is the distance between the security device and android phone. This criterion is selected because of during the accidental situation the phone in which the application is running would be thrown away from the security device.

In the experiment it is tested whether it's getting the signal from different distances and sending the message and voice call to the correct contact numbers. The experimental result is shown in Table I.

TABLE I. Performance Analysis Considering Distance

Distance (m)	Different Parameters		
	Signal	Location	Message
2	Yes	Found	Sent
5	Yes	Found	Sent
8	Yes	Found	Sent
10	Yes	Found	Sent
12	No	Not found	Not sent
14	No	Not found	Not sent

From the experimental result it is seen that the mobile android application recognized the security device signal efficiently with in the distance of 10 m. After that it is not recognized the signal from the security device. So, it is ensure that the system working perfectly within 10 m.

IV. Conclusion

In this paper a very generic approach is presented to reduce the attack on women and ensure the maximum security if this security device is used in extensively, it can be possible to reduce the violence towards the women and provide them a safe and secured life. For this system, user's need a security device made of Bluetooth device and an android mobile application. The proposed system works properly with the distance between the security device and the mobile phone is 10 m.

Yet, the system has some limitations. Both sender and receiver need to keep the GPS on. The system not working properly with the distance between the security device and the mobile phone is more than 10 m.

References

[1] Md. Alamgir Hossen "Measuring Gender-based violence: Results of the Violence Against Women (VAW) Survey in Bangladesh," In 5th global forum on gender stistics.

[2] Dhrubajyoti Gogoi, Rupam Kumar Sharma, "Android Based Emergency Alert Button," In International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, vol.2, Issue-4, March 2013.

[3] Alexandre Bartel, Jacques Klein, Yves Le Traon,Martin Monperrus, "Automatically Securing Permission-Based Software by Reducing the Attack Surface: An Application to Android," In arXiv:1206.5829v2 [cs.CR] 20 Mar 2013.

[4] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[5] R. K. Sharma, A. Mohammad, H. Kalita and D. Kalita, "Android interface based GSM home security system," In 2014 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), Ghaziabad, 2014, pp. 196-201.

[6] Li Ma, Lei GU, Jin Wang "Research and Development of Mobile Application for Android Platform" In International Journal of Multimedia and Ubiquitous Engineering Vol.9, No.4 (2014), pp.187-198

[7] Zhang Hong-Han1 , Wang Ru-Jun1 "Research on Application based on Android System" In 2nd International Conference on Computer and Applications.

[8] Jorge Zaldivar, Carlos T. Calafate, Juan Carlos Cano, Pietro Manzoni "Providing accident detection in vehicular networks through OBD-II devices and Androidbased smartphones" In 5th IEEE workshop on user mobility and vehicular networks.

[9] Xianhua Shu, Zhenjun Du, Rong Chen "Research on Mobile Location Service Design Based on Android"

[10] Zeshan A Rajput, Samuel Mbugua, David Amadi, Viola Chepng'eno,Jason Saleem,Yaw Anokwa,Carl Hartung,Gaetano Borriello " Evaluation of an Android-based mHealth system for population surveillance in developing countries"

[11] Suhas Holla, Mahima M Katti "Android Based Miobile Application Development and its Security" In International Journal of Computer Trends and Technology- vol.3,Issue-3- 2012

[12] Akanksha Singh, Arijit Pal, Bijay Rai, "GSM Based Home Automation, Safety andSecurity System Using Android Mobile Phone" In International Journal of Engineering Research & Technology (IJERT) Vol. 4 Issue 05, May-2015

[13] H. Bing, "Analysis and Research of System Security Based on Android," In 2012 Fifth International Conference on Intelligent Computation Technology and Automation (ICICTA), Zhangjiajie, Hunan, 2012, pp. 581-584.

[14] Moon-Hee Park, Jin-Hyuk Hong and Sung-Bae Cho, "Location-Based Recommendation SytemUsing Bayesian User's Preference Model in Mobile Devices," in Proceedings of Ubiquitous Intelligence and Computing (UIC), pp. 1130 - 1139, 2007.

[15] R. Piyare and M. Tazil, "Bluetooth based home automation system usingcell phone," in IEEE 15th International Symposium on Consumer Electronics, Singapore, 2011, pp. 192 - 195.

Analyzing MRI Segmentation Based on Wavelet and BEMD using Fuzzy C-Means Clustering

Md. Khaliluzzaman and Lamia Iqbal Dolon
Dept. Of Computer Science and Engineering
International Islamic University Chittagong (IIUC)
Chittagong - 4203, Bangladesh
khalil_021@yahoo.co.in, lamiaiq04@yahoo.com

Kaushik Deb*
Dept. Of Computer Science and Engineering
Chittagong University of Engineering & Technology (CUET)
Chittagong - 4349, Bangladesh
*debkaushik99@cuet.ac.bd

***Abstract* - Segmentation of images means a great matter for the medical field treatment purpose. For the extraction of brain polyps, magnetic resonance image (MRI) processing contributes in a wide range. Usually it works in two ways: white matter and gray matter. The extraction of any type of issues helps in submissions of image segmentation like in medical report analysis, in preparation of radiotherapy, in formation of medical treatment etc. The main purpose of this paper is the Fuzzy C-Means (FCM) clustering exploitation by the help of Wavelet and Bi-dimensional Empirical Mode Decomposition (BEMD), as for the aim of improving the eminence of MR noisy images. To gain the best image segmentation method, in this paper the signal to noise ratio (SNR) rates were calculated by the data set of FCM clustering. As in the medical term of MRI segmentation, the experiment has done with synthetic WEB Images of brain that has verified the robustness and proved with efficiency with the applicable approach.**

Keywords - Image segmentation; Magnetic Resonance Imaging (MRI); Fuzzy C-means; SNR; Wavelet; BEMD

I. INTRODUCTION

In the area of image processing, joint space-spatial frequency illustrations have gained spatial consideration and it also happens in vision and pattern recognition sector. For medical image analysis purpose, Magnetic Resonance Imaging (MRI) segmentation is a very stimulating and substantial phase. This step helps in various medical treatment purposes. MRI is a nonintrusive process as it is functioned for identifying internal muscles and organs.MR images have very different nature and it doesn't have any linear features. Partial volume effect (PVE) is an issue which defines that a pixel can contain more than one tissue. MR images performance can be affected by PVE and this can occur as of blurred frontier between relic passions of the similar tissue as over the spatial domain it is not constant of image as well as geometric distortion [5].

With the aim of developing the requirements for the MRI segmentation, researchers and scientist have been developed many segmentation methods over past few years. Identification of tissues and organs from the MR images is the main problem in segmentation. To solve those problems some spontaneous works have been done and those are described in the literature review segment. The main objective of MR image segmentation is to define the measurements of lesions, tissues, as well as organs existent in an image obtained from the patient. The segmentation measurements and the changing of those results with respect to time may help in the phase of patient treatment. It also helps to analysis and diagnosis forecasting of patients over exploration.

In a precise image segmentation process also helps to identify region of concern. The foremost approach of segmentation is to make image supplementary simple and expressive.

There remain many classification techniques in medical image segmentation but there is not any standard classification technique. On the basis of commonly used technique, the classification is done into various attitudes like as region based segmentation techniques which are sight for region recognition from the particular properties [6], another basis of physical evidence for irregular as well as regular tissues [2]. The Edge base segmentation techniques that are identified the edges among the sections of the image with the various features [7]. The knowledge based methods i.e., artificial neural networks. Other methods are based on data fusion method [1][9], based on Hybrid Method [4] and Random Markov method [10].

In some times the FCM is being used for the image segmentation. This is a multi-resolution process employed by the discrete wavelet transform (DWT). It is one of the good methods used for image segmentation in real time. By this process firstly the features are extracted from image, and then perform the selection, and finally segmentation is performed using the K-means clustering algorithm.

Fuzzy C-Means (FCM) based on image segmentation method had been applied in [11]. There proposed an algorithm which defined the size of window dynamically as for mining apposite inclusive spatial data from images. Moreover, without adding any penalty term there used the standard objective function of FCM. For describing one of each pixel clearly there used an n-dimensional feature. That method showed satisfactory results on real images but it failed in classification of five classes (i.e. background, cortical bone, malleable jawbone, adipose flesh and muscle) in the MRI of thigh.

Empirical Mode Decomposition (EMD) image function is operated in [12]. Bi-dimensional Empirical Mode Decomposition (BEMD) usually used in image segmentation of MRI and CT scans. By using the simple fusion rules, the fusion of the generated Bi-dimensional Intrinsic Mode Functions (BIMFs) are done. They had done comparison of the BEMD based fusion results by two efficient synthesis

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

techniques: wavelet fusion and curve let fusion. BEMD performance had been calculated by using three well known synthesis image quality matrices; i.e., Peak Signal-to-Noise Ratio (PSNR), Mutual Information Parameter (MIP) and Structure Similarity Index Metric (SSIM).

In [13], worked on the image processing technique that is Wavelet for MR and CT scan image decomposition. That method involved de-noising of the images. For that, there used wiener filter that removed salt and pepper noise. Their main aim of image de-noising was noise reduction and feature preservation. In there, wavelet transformation was done by wavelet filters. They used different wavelet transformations i.e. Daubechies, Symlets, and Coiflets for gaining optimal results from the input images. But for noisy MR images this method didn't worked so well as like for CT scan images.

Work on brain tumor segmentation by using improved FCM has been done in [14]. In the paper, for decreasing noise ratio and to increase the contrast level there had preprocessed the test input images. Then the work was done on the segmentation of brain images by improved Fuzzy C-Means (FCM) algorithm to display the abnormal activity like growth of tumor. After that, texture features (GLCM) had removed from the images. In that extraction phase, for various distances and directions the calculation of statistical measurements had been done by gray level co-occurrence matrix. But there main disadvantage occurred in the edge detection of tumor as they can't sharpen every single pixel at a high contrast.

In [15], proposed a method for MR image segmentation by using the Wavelets and FCMs. The work is based on the image segmentation obtained by MRI is a challenging task as for the inherent signal noises and in homogeneity. Discrete Wavelet Transform has been applied to MRI image to excerpt great details and after some dispensation on high pass image. Their proposed method includes the FCM dissection algorithm functional to the Kirch's edge detection mask wavelet transformed image that used to increase the edge feature in the image. In the paper, authors are worked with hybrid method with help of wavelets and Fuzzy C-means. They extracted the high filtered image with the help of wavelets and Kirch's edge detection mask that had been applied to intensify extra brink specifics that afford satisfactory edge facts. Usually MR images are dispersed; incorrect brink collection can cause absconding polyp portion or tab a lot in decent physical forms as polyps. These types of methods are not that much dependable. For human body, it provides improved imagining of soft muscles. But only along the Discrete Wavelet transformation (DWT) is not much effective because the discrete value can miss some important value which may be useful for the survey.

For the segmentation of MRI, clustering is the most functioning or exploitable technique. Clustering used to classify pixels into curriculums and it does not have any idea of previous evidence or exercise. It categorizes pixels interested in the same class with maximum possibility. It may discover unsystematic pixels which do not have any class probability. The working out of clustering techniques has done by means of pixel structures with possessions of every session [16].

In this paper proposed a method to determine the best technique from the two decomposition techniques i.e., Wavelet and BEMD. These techniques are used before applying the image clustering technique i.e. FCM on MR images. For that, firstly apply the decomposition technique on the input MRI image. After that, FCM clustering technique is used on the outcomes of the decomposition techniques. Finally, calculate value of SNR to select the best segmentation technique from the segmentation approaches.

The paper is arranged as follows. The proposed framework is explained step by step in Section II. In Section III, the experimental results are described. Finally, the conclusion is given in Section IV.

II. PROPOSED METHOD

In this section describes the proposed method which is about the segmentation of MRI brain images with signal noise ratio. In the proposed methodology two important phases have involved i.e., extraction of features and clustering techniques. Extraction of feature process has done by Wavelet decomposition (2D) and BEMD. The working criterion of wavelet decomposition has divided into two segments i.e. outputs in low pass and outputs in high pass. The low pass values are the estimated elements and the high pass values are the exhaustive elements. In images Dubechies-1 (DAUB1) Wavelet is employed to acquire the wavelet features. The extraction of the feature has done in both Wavelet as well as BEMD decomposition. The results from the decomposition techniques are given into FCM which is functioned as a vector of the feature. That result was gained from the previous clustering step. The outputs of the MR images are divided into white and gray matter segments. After that, the value of the SNR is calculated for segmented output. Finally, determines the best segmentation technique based on the SNR value of the MR image. Fig. 1 shows the proposed method.

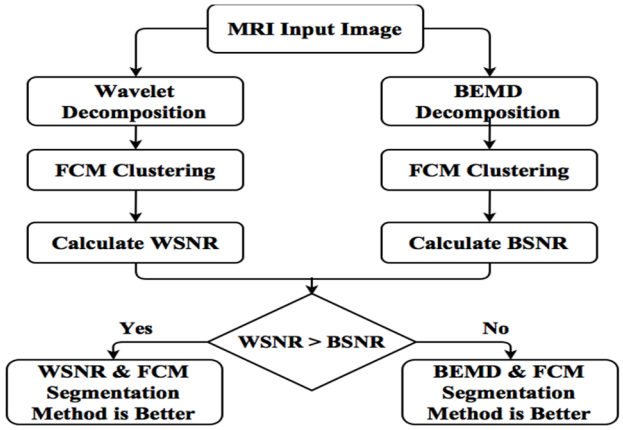


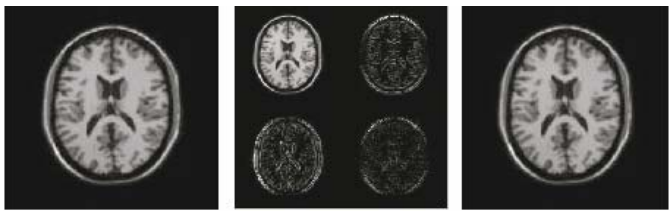
Fig. 1. The proposed method.

A. *Wavelet Decomposition*

The wavelet transform which is based on the Fourier transform is used in the image processing. This technique is becomes even more easy to transfer and analyze and also easier to compress the MR images. Wavelet Transformation (WT) characteristics share properties which is the basic functions of the family, mostly limited provision for the

frequency domains and unique domains as well as scalability. In this family, there provides Haar, Coiflets, Biorthogonal, Daubechies, Reverse Biorthogonal, and Symlets. In this research paper, the family of Daubechies is used functionally. In the signal and image processing applications like segmentation, classification, de-noising and compression, Daubechies family is widely used.

The input MR image, wavelet decomposition of input image, and wavelet reconstruction to input image from decomposition image are shown in Fig. 2(a), 2(b), and 2(c) respectively.



(a) (b) (c)

Fig. 2. Processing experiment of wavelet decomposition and construction: a) Input MR image, b) Wavelet decomposition of MR image, and c) wavelet reconstruction of MR image.

B. Bi-dimensional Empirical Mode Decomposition(BEMD)

The EMD method has been established in [17], it is an algorithm of signal processing which has been recognized to be capable of perfectly exploring adaptive, non-stationary, and non-linear data by gaining regional structures as well as time-frequency distribution (TFD). For that, firstly decomposition of data/signal has done perfectly into its intrinsic mode function (IMF), in the next step finds the TFD of the signal/data from each intrinsic mode function with the help of utilization of the instantaneous frequency and Hilbert transform. The Hilbert-Huang transform (HHT) [17] is known by that process of above. The Bi-dimensional EMD (BEMD) technique is the process of two-dimensional (2D) data/images in a decomposition method.

The BEMD method is used in this paper based on IMFs to de-noise and decompose the MRI with the weighted threshold value of IMFs. After applying this method in the MR image the image noises are distributed in the different frequencies mainly high as well as intermediate frequency. The shifting procedure is used to attain the frequency elements in the image.

A sampled signal decomposes the selecting method by the EMD. This process is created by two obligations. The IMF has equal quantity of zero crossings as well as extrema in first process. In second process, respect to the local mean, IMF is symmetric. Also, it accepts those have two extrema at least. In the final stage, de-noising is achieved by reconstruct the source image. It is also mentioned that the segmentation quality is influenced by smoothness of the image data. Fig. 3 shows the experimental example of BEMD on MR images. Here, 48 iterations are performed for IMF to achieve the required response.

C. Fuzzy-C Means (FCM)

In image segmentation, FCM procedure is an unsubstantiated technique of clustering. FCM idea depends on two or more clustering data classes only by recognized amount of courses.

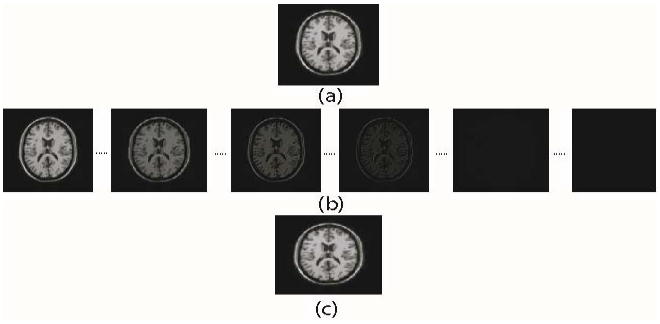


Fig. 3. Processing experiment of FCM: a) Input MR image, b) IMF (1,...,6,…,12,…,20,…,48),…, residue image, and c) reconstructed MR Image.

FCM algorithm is based on the minimization objective function shown in bellow.

$$J(U, c_1, c_2, \dots, c_c) = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^{m} \, d_{ij}^{2} \quad (1)$$

here, the value of u is within 0 and 1. The value of the cancroids cluster I is define as C_i , the distance of the Euclidean between i^{th} and j^{th} cancroids data point is d_{ij}. The weighting function is m[1,∞]. The identified data sample of Fuzzy portioning is approved by optimization objection function which is the iterative process shown in bellow.

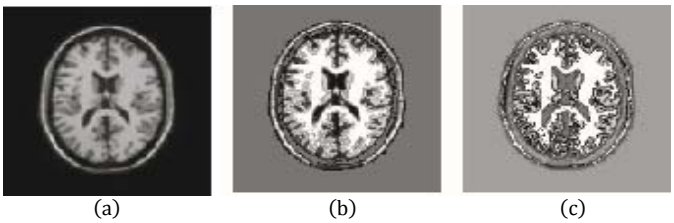
$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{kj}} \right)^{2/(m-1)}} \quad (2)$$

$$c_{ij} = \frac{\sum_{j=1}^{n} u_{ij}^{m} x_j}{\sum_{j=1}^{n} u_{ij}^{m}} \quad (3)$$

The stopping moment of this iteration is

$$max_{ij} \left\{ |u_{ij}^{(k+1)} - u_{ij}^{(k)}| \right\} < \epsilon \quad (4)$$

where, the value of the termination criterion within 0 and 1, the iteration step is k. The local minimum is J_m.



(a) (b) (c)

Fig. 4. Processing experiment of segmentation: a) Input MR image, b) segmented MR image using Wavelet & FCM, and c) segmented MR image using BEMD & FCM.

III. EXPERIMENTAL RESULTS

The experimental results of some stair images are explained in this section. The experiments were instigated on MRI web images in MATLAB environment. The MRI generated by various percent of noise that are 0%, 2%, 4%, 6%, 8%, 10% and with different parity of INU that are 0%, 30% and 50%. All the experimental MR images are taken from the Brain Web Database.

The Database is at the McConnell Brain Imaging Centre of the Montreal Neurological Institute, McGill University. Fig. 5(a), Fig. 5(b), and Fig. 5(c) are shown the example the brain MR images.

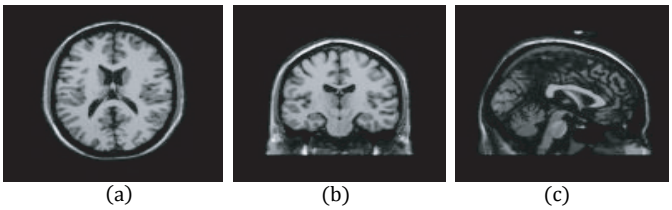


Fig. 5. T1 simulated brain web images: (a) MR sample 1 image, (b) MR sample 2 image and (c) MR sample 3 image.

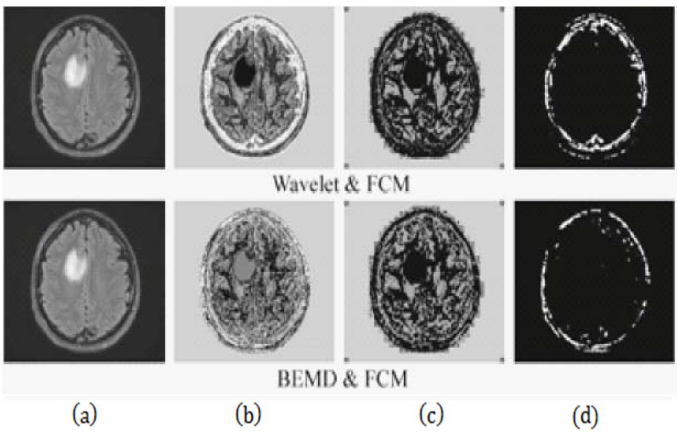


Fig. 6. Processing experiment of smooth MR image for sample 1: a) original MR image, b) segmented MR image, c) Gray Matter, and d) White Matter.

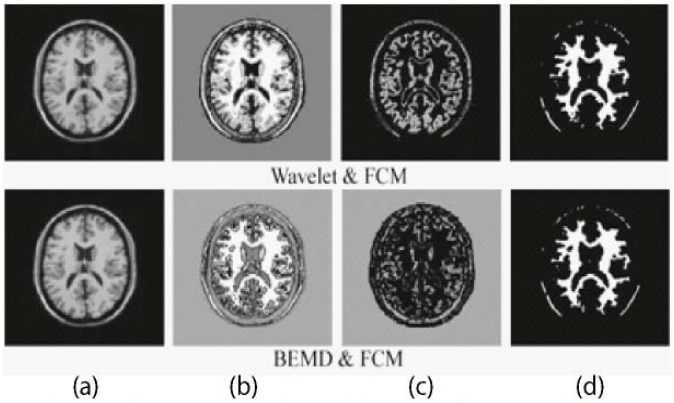


Fig. 7. Processing experiment of smooth MR image for sample 2: a) original MR image, b) segmented MR image, c) Gray Matter, and d) White Matter.

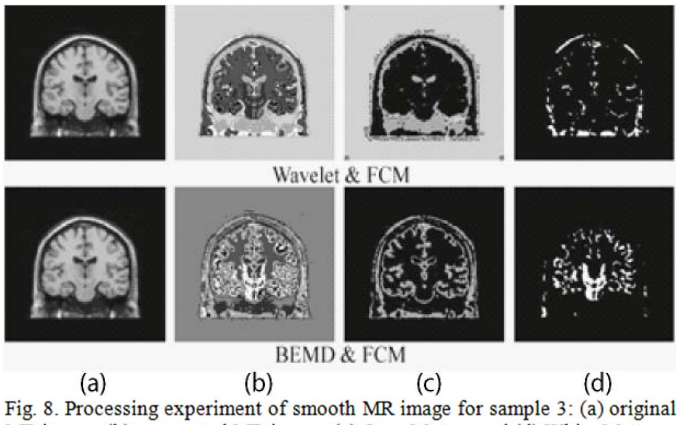


Fig. 8. Processing experiment of smooth MR image for sample 3: (a) original MR image, (b) segmented MR image, (c) Gray Matter, and (d) White Matter.

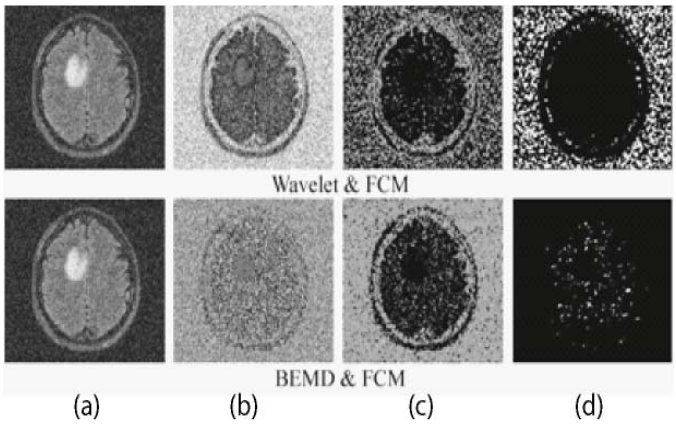


Fig. 9. Processing experiment of noisy MRI image for sample 1: (a) original MR image, (b) segmented MR image, (c) Gray Matter, and (d) White Matter.

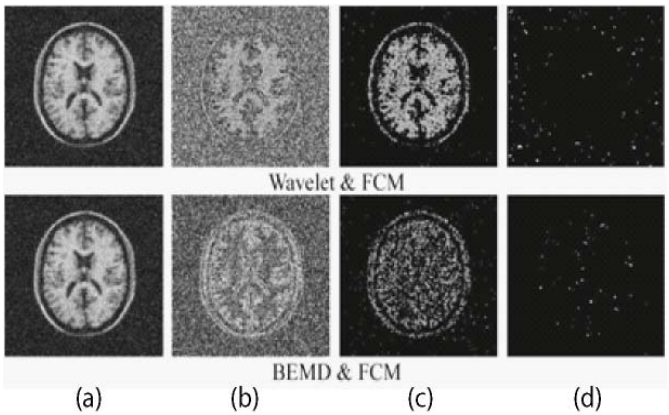


Fig. 10. Processing experiment of noisy MR image for sample 2: (a) original MR image, (b) segmented MR image, (c) Gray Matter, and (d) White Matter.

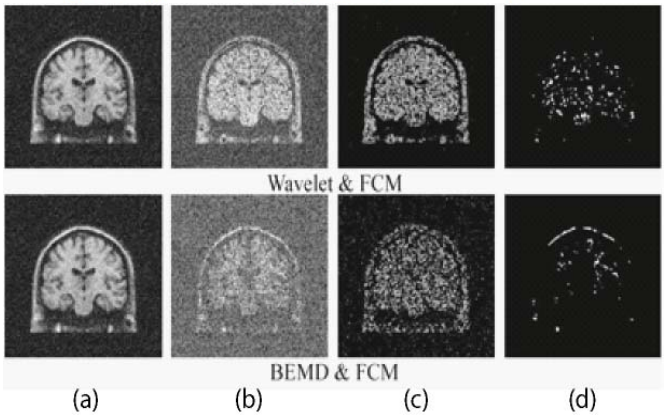


Fig. 11. Processing experiment of noisy MRI image for sample 3: (a) original MR image, (b) segmented MR image, (c) Gray Matter, and (d) White Matter.

The processing experimental example of smooth MRI sample image 1, sample image 2 and sample image 3 are shown in Fig. 6, Fig. 7, and Fig. 8 respectively. Another processing experimental example of Gaussian noisy MRI sample image 1, sample image 2, and sample image 3 are shown in Fig. 9, Fig. 10, and Fig. 11 respectively. In the all processing experiments apply the wavelet and BEMD with the FCM clustering.

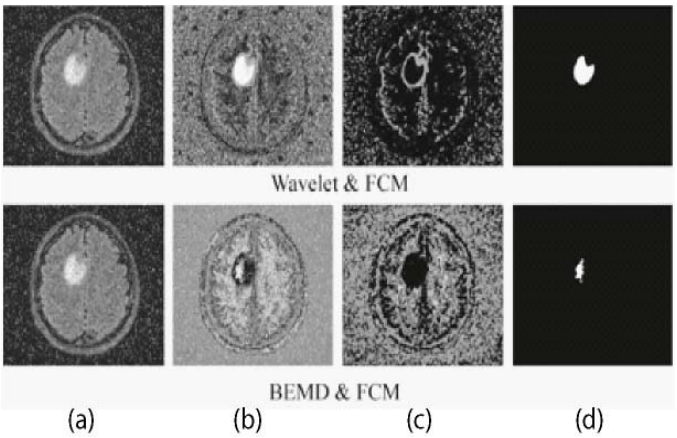


Fig. 12. Processing experiment of 10% Salt & Pepper noisy MR image for sample 1: a) original MR image, b) segmented MR image, c) Gray Matter, and d) White Matter.

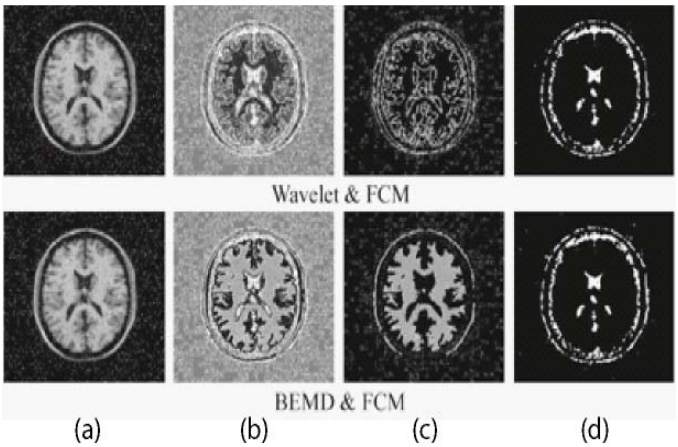


Fig. 13. Processing experiment of 10% Salt & Pepper noisy MR image for sample 2: a) original MR image, b) segmented MR image, c) Gray Matter, and d) White Matter.

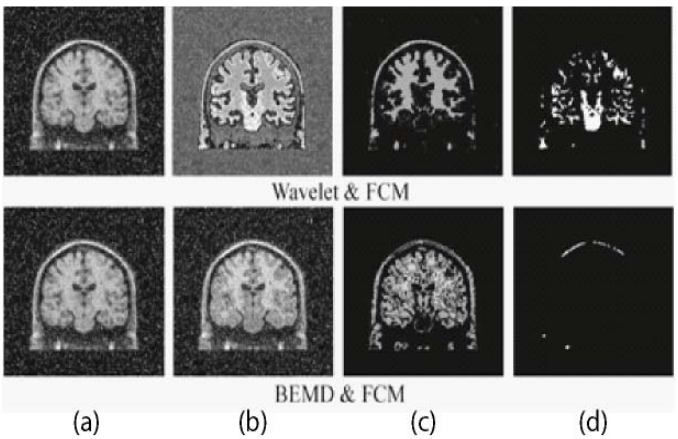


Fig. 14. Processing experiment of 10% Salt & Pepper noisy MR image for sample 3: a) original MR image, b) segmented MR image, c) Gray Matter, and d) White Matter.

The processing experimental example of 10% Salt & Pepper noisy MRI sample image 1, sample image 2, and sample image 3 are shown in the Fig. 11, Fig. 12, and Fig. 13 respectively. In these experiments apply the wavelet and BEMD with FCM clustering.

The segmentation presentation of evaluation in this paper is dignified with the Eq. (5).

$$SNR = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N} f(x,y)^2}{\sum_{x=1}^{M}\sum_{y=1}^{N} S(x,y)^2} \quad (5)$$

Where, the input image is defined as $f(x,y)$ and the segmentation image is defined as $s(x,y)$.

In this paper the best segmentation method is computed through the value of the SNR which is calculated by the Eq. (5). The value is calculated both for the wavelet and FCM as well as BEMD and FMC. The SNR value is calculated for the MRI smooth image as well as for the MRI image by adding 10% salt & Pepper noise and Gaussian noise. The value of the SNR for the sample MR image 1, sample MR image 2, and sample MR image 3 is shown in Table I.

TABLE I: SNR OF THE INPUT IMAGES

Sample MRI image	Input MRI with Smooth, Adding Gaussian and 10% Salt & Pepper noise	SNR	
		Wavelet & FCM	BEMD & FCM
1	Smooth	1.0175	1.0977
	mean= 0 and variance =0.025	0.7749	0.7789
	10% Salt & Pepper noise	0.8692	0.8731
2	Smooth	0.4249	0.4259
	mean=0 and variance =0.025	0.6588	0.6654
	10% Salt & Pepper noise	0.5042	0.5241
3	Smooth	0.3992	0.3997
	mean= 0 and variance =0.025	0.6452	0.6472
	10% Salt & Pepper noise	0.4955	1.0011

Fig. 15 shows the value of SNR for smooth MRI sample image 1, MRI sample image 2, and MRI sample image 3. Fig. 16 and Fig. 17 shows the SNR value for these MRI images which are added by the noise of Gaussian and 10% Salt & Pepper respectively.

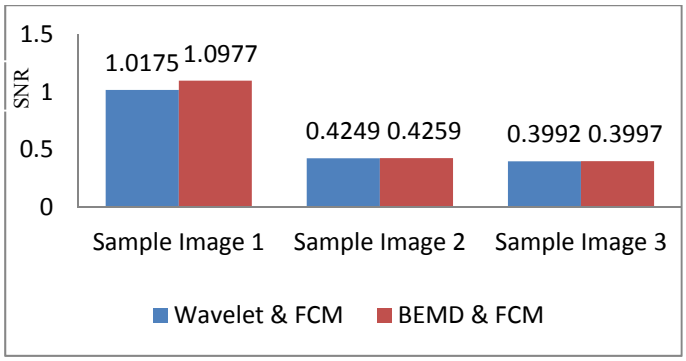


Fig. 15. SNR values of Wavelet & FCM and BEMD & FCM for smooth images.

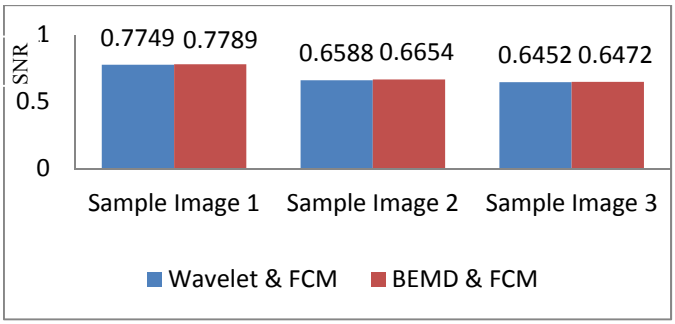


Fig. 16. SNR values of Wavelet & FCM and BEMD & FCM for noisy images (Gaussian noise).

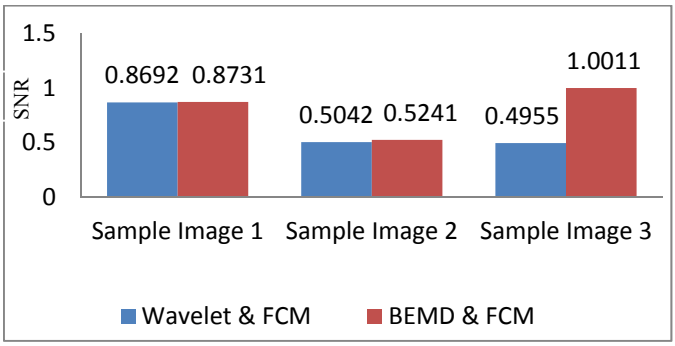


Fig. 17. SNR values of Wavelet & FCM and BEMD & FCM for noisy images (10% Salt & Pepper noise).

From the Fig. 15, Fig. 16, and Fig. 17 it has been seen that the SNR value for the BEMD and FCM i.e. BSNR shows the superior value for smooth and noisy i.e., Gaussian, 10% Salt & Pepper noise MRI image. According to this SNR value it can be mentioned that BEMD decomposition method is better with respect to wavelet method.

IV. CONCLUSIONS

The paper represents an efficient approach and strong evidences for MR image segmentation. For that, firstly decomposition method i.e. BEMD as well as wavelet are employed on the MR images for extracting the MR image features. After that, decomposition image is clustered by using the FCM algorithm. By this process the MR image is converted to segment image. Then, the SNR value is calculated from the segmented image both for BEMD and FCM as well as wavelet and FCM. By this SNR value better approach is detected. The best method has founded robust beside the various MR images. The processing examples with unreal Brain Web image which validated effectiveness as well as forcefulness of the superior methodology for the noisy and smooth MR images. In future, the research work will be extended through enhancing the segmentation performance by different version of segmentation technique for various types of noisy (white noise) MR images.

References

[1] A. S. S. A. S. Ouarda, "MR Brain Real Images Segmentation Based Modalities Fusion and Estimation Et Maximization Approach." International Journal of Advanced Computer Science & Applications, Vol. 7, No. 1, pp. 267-273, 2016.

[2] R. Sampath, and A. Saradha, "Alzheimer's Disease Image Segmentation with Self-Organizing Map Network," Journal of software, Vol. 10, No. 6, pp. 670-680, 2015.

[3] A. Sindhuja and V. Sadasivam, "Wavelet based segmentation using optimal statistical features on breast images," ICTACT Journal on Image & Video Processing, Vol. 4, No. 4, 2014.

[4] B. S. Anami and P. H. Unki, "A combined fuzzy and level sets based approach for brain MRI image segmentation," Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), 2013 Fourth National Conference on Digital Object Identifier, 10.1109/NCVPRIPG.2013.6776216. pp 1-4, 2013.

[5] H. Shamsi, H. Seyedarabi, and S. Erfani, "MRI image segmentation based on new fuzzy c-means algorithm," International Journal on Computer Science and Engineering, Vol. 3, No. 8, 2011.

[6] A. Kouhi, H. Seyedarabi, and A. Aghagolzadeh, "A Modified FCM Algorithm for MRI Brain Image Segmentation," In 2011 7th Iranian Conference on Machine Vision and Image Processing, pp. 1-5, IEEE, 2011.

[7] G. B. Aboutanos, J. Nikanne, N. Watkins and B. Dawant, "Model Creation and Deformation for the Automatic Segmentation of the Brainin MR Images," IEEE Transactions on Biomedical Engineering, Vol. 46, No. 11, 1999.

[8] D. A. Karras, and B. G. Mertzios, "On Edge Detection in Mri Using the Wavelet Transform and Unsupervised Neural Networks", EC-VIP-MC 2003. 4th EURASIP Conference focused on Video I Image Processing and Multimedia Communications, Zagreb, Croatia, 2-5 July 2003.

[9] L. Gui, R. Lisowski, T. Faundez, P. S. Huppi, F. Lazeyras, and M. Kocher, "Automatic Segmentation of Newborn Brain Mri Using Mathematical morphology," In 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 2026-2030, IEEE, 2011.

[10] Tohka, J. ,Dinov, I. D., Shattuck, D.W. , Toga, A.W. ,"Brain MRI tissue classification based on local Markov random fields," Magnetic Resonance Imaging, Volume 28, Issue 4, pp. 557-573, May 2010.

[11] P. Antonio, "FCM-Based Method for MRI Segmentation of Anatomical Structure," In International Conference on Bioinformatics and Biomedical Engineering (pp. 175-183). Springer International Publishing, 2016.

[12] A. T. Alshawi, F. E. A. El-Samie and S. A. Alshebeili, "Magnetic resonance and computed tomography image fusion using bidimensional empirical mode decomposition," In 2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP) , pp: 413 – 417, 2015.

[13] S. Dixit, M. R. Kini, M. D. Patil and S. Chaithra, "Brain Tumor Detection from Clinical Ct and Mri Using Wavelet Based Image Fusion Technique," In PARIPEX-Indian Journal of Research, Vol. 5, No. 6, 2016.

[14] A.M. Nichat, and S. A. Ladhake, "Brain Tumor Segmentation and Classification Using Modified FCM and SVM Classifier," Brain, Vol. 5, No. 4, 2016.

[15] N. Noreen, K. Hayat and S. A. Madani, "MRI segmentation through wavelets and fuzzy C-means," In World Applied Sciences Journal, 13, pp.34-39, 2011.

[16] S. Luo, "Automated Medical Image Segmentation Using a New Deformable Surface Model," In IJCSNSInternational Journal of Computer Science and Network Security, Vol. 6, No. 5A, pp. 109-115, 2006.

[17] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu and H.H. Shih, Q. Zheng and H. H. Liu, " The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," In Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, Vol. 454, No. 1971, pp. 903-995,. The Royal Society, 1998.

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

RFI and SQLi based Local File Inclusion Vulnerabilities in Web Applications of Bangladesh

Afsana Begum, Md. Maruf Hassan, Touhid Bhuiyan, Md. Hasan Sharif
Software Engineering Department
Daffodil International University, Dhaka, Bangladesh
afsana.swe@diu.edu.bd, maruf.swe@diu.edu.bd, t.bhuiyan@daffodilvarsity.edu.bd, hasan543@diu.edu.bd

***ABSTRACT* - People nowadays cannot think of even a single moment without the internet. Doubtlessly, web applications are currently the key to all change in the world. The features and facilities of web applications make our lives easier and also demonstrate different channels of communication and exchange of data. The number of internet user is increasing day by day in Bangladesh. Web applications are not only encouraging internet users to easily receive diversified services, but also creating new opportunities in every business sector. With the help of web applications, businesses can easily enhance the quality of their services to consumers with minimal effort. However, at the same time, many threats have arisen as a result of the misuse of this technology. Cyber attacks could emerge as a major threat to the digital transformation of Bangladesh, given the vulnerability in web applications caused by careless coding during the development of such applications. During our review, we observed that the existence of Local File Inclusion (LFI) vulnerability in the web applications of Bangladesh was very critical. This paper explores in detail the harmful web application vulnerability attack, Local File Inclusion (LFI) based on Remote File Inclusion (RFI) as well as Structured Query Language Injection (SQLi), and their impact on the applications of Bangladesh.**

Keywords- Web Application Vulnerability; RFI and SQLi based Local File Inclusion (LFI); Remote File Inclusion (RFI)

I. INTRODUCTION

In the past four years, web applications have changed communications and information sharing systems of Bangladesh in unimaginable ways. Now approximately all sectors are trying to provide their services through web applications apart from their traditional channels in order to meet the demands of their consumers. More than 63 million people are using the internet in their daily lives for personal or organizational purposes. [14]

Nowadays, the people of Bangladesh have gradually come to depend on the web for all kinds of online services. They are using web applications in various sectors for different purposes, such as online monetary transactions, payments of gas bills and electric bills, file sharing, online learning, online shopping, etc. Since web applications have become a mandatory daily need of users, there is a high risk that cyber criminals can exploit the weaknesses of such web applications. It has been observed in this study that most of the web applications have been developed in PHP language because of its prevalent and easy-to-develop characteristics.

The web applications of Bangladesh are found to be insecure because of the careless use of functions, especially in the PHP language. Problems in coding result in different types of vulnerability inside the web applications. It has become a huge challenge to ensure the security of millions of existing vulnerable web applications. Proper security skills, however, would help the developer design a secure web application with minimal errors.

Some common types of vulnerability occur in web applications like SQLi, LFI, RFI, XSS and so on. In this paper, we will focus on LFI exploitation based on RFI and SQLi. If a web application has some codes that will dynamically refer to an external script, it represents the existence of RFI vulnerability in the web application. In this case, the main target of the attacker is to exploit the referencing function so that they can upload malware (e.g. backdoor shells). This exploitation can occur from a remote domain. On the other hand, SQL Injection is the process that allows a user to insert data into the input field without proper filtering. In this case, intentional SQL commands are inserted into the server, and the output of the SQL command can be shown. Again, LFI is the exploitation of a web application by injecting a file that already exists in the web application. In this paper, we will discuss the above vulnerabilities and the techniques of its exploitation.

This paper is organized into five sections: Literature Review is in Section II while LFI and its exploitation types, along with its working process, are respectively explained in Section III and IV. After performing data analysis, results and statistics are presented in Section V. The Conclusion is provided in Section VI.

II. LITERATURE REVIEW

A number of researches has been conducted on different web vulnerabilities, such as SQL Injection, XSS, CSRF, buffer overflow, broken authentication, local file inclusion (LFI), etc. [1], [2], [7], [8]. We have found some review on Cross-site scripting, State Integrity, Insecure Direct Object References, Failure to Restrict URL Access, Invalidated Redirects and Forwards, Insufficient Transport Layer Protection, Security Mis-configuration and Insecure Cryptographic Storage vulnerabilities [11] [12]. Moreover, some analyses or surveys on the percentage of random vulnerable web applications were found [13]. A few studies on SQL Injection, XSS and CSRF have been conducted in the Bangladesh region [3], [4], [5], [6], [7], [9]. Although we did find some studies about LFI with a certain degree of perspective, we did not get a complete study about LFI vulnerability in Bangladesh. Our paper we will discuss the impact of LFI, based on SQLi and RFI vulnerability, on web applications in Bangladesh.

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

III. LFI, RFI AND SQLI VULNERABILITY

A. Local File Inclusion (LFI):

Local File Inclusion (LFI) is a web application's vulnerability wherein it allows a user to include different files located in the web application on the server machine. This vulnerability can be oppressed through the exploitation of 'dynamic file inclusion' mechanisms implemented in the targeted web application. The vulnerability occurs due to improper use of programming built-in functions/methods and also because the user's selected parameter is not validated and sanitized.

```
Line 01: <html>
Line 02: <body>
Line 03: <?php
Line 04: "include/".include($_GET['filename']."php");   // use of include() function
Line 05: ?>
Line 06: </body>
Line 07: </html>
```

Code 01: Example of LFI vulnerable code in PHP language

Code 01 is an example of an LFI vulnerable code, which is implemented in PHP language. In Line 04, the statement allows the user to include any type of '.php' file located in the web server machine. Using this, the user of that application can access several sensitive files that are supposed to be restricted for the general user.

B. Remote File Inclusion (RFI):

Remote File Inclusion (RFI) is one of the weaknesses in a web application by which it remotely accepts any type of user input e.g. code, command, etc. RFI vulnerability is almost the same as LFI except that RFI can only be directory writeable. This vulnerability occurs when a webpage of the application receives a path of a certain file to be included as input without ensuing proper validation and sanitized process during development. In our study, it was observed that this type of vulnerability usually exists in PHP/JSP and ASP scripts.

C. Structured Query Language Injection (SQLi):

SQLi is a code injection technique used to attack data-driven applications, in which tricky SQL syntax is inserted into a user field for execution. When a user's input is incorrectly filtered for string, literal escape characters embedded in SQL statements get unexpectedly executed on the system. So, SQL injection allows an attacker to spoof identity, tamper with existing data, and/or modify or destroy the whole database system.

```
Line01: UserName = getRequestString("UserName");
Line02: UserPass = getRequestString("UserPass");
Line03: SQL = 'SELECT * FROM USERS WHERE Name ='" + UserName + "' AND Pass ='" + UserPass + "'"
Line04: database.execute(SQL)
```

Code 02: Example of SQLi vulnerable code

Lines 01 and 02 are predefined variables for storing username and password of a user in Code 02. In Line 03 of the above code is the SQL statement for authenticating the user. Line 04 of the code is used to execute the inserted query or statement. If a user inserts ' 'pass' or '1'='1--+- ' in the password field, the code will execute in the program, as below:

SELECT id FROM USERS WHERE Name= 'username' AND Pass ='password' or 1=1'

Since 1=1 is always truth, the server will execute the above inserted query against the database and SQL injection will occur.

IV. TYPES OF LFI EXPLOITATION

There are mainly two types of LFI exploitation techniques used in practice. However, LFI exploitation can be conducted in other two bases i.e. RFI and SQLi bases. In this paper, we will not only discuss the types of general LFI exploiting technique but also the types of RFI- and SQLi-based exploitation techniques. The details are explained as follows:

A. General LFI Exploitation:

When attackers find the existence of the LFI vulnerability, they will figure out the list of the sensitive files e.g. /etc/passwd, /etc/user, /etc/shadow etc. with the intention to misuse. The LFI exploitation technique is categorized into the following two types:

i) $GET Method Exploitations:

For $GET exploitations, the infected URL may look like this: http://www.website.com/downloads.php? file=contact.php. To identify the LFI vulnerability of the above URL, the attacker will modify the parameter/ value of file as /etc/passwd like e.g. http://www.website.com/downloads.php? file=/etc/passwd. If the execution of the given modification replies the following code, it indicates that the site is LFI vulnerable.

```
root:x:0:0:root:/root:/bin/bash
bin:x:1:1:bin:/bin:/sbin/nologin
daemon:x:2:2:daemon:/sbin:/sbin/nologin
adm:x:3:4:adm:/var/adm:/sbin/nologin
lp:x:4:7:lp:/var/spool/lpd:/sbin/nologin
sync:x:5:0:sync:/sbin:/bin/sync
shutdown:x:6:0:shutdown:/sbin:/sbin/shutdown
User:/var/ftp:/sbin/nologin
nobody:x:99:99:Nobody:/:/sbin/nologin
dbus:x:81:81:System message bus:/:/sbin/nologin
rpc:x:32:32:Rpcbind Daemon:/var/cache/rpcbind:/sbin/nologin
usbmuxd:x:113:113:usbmuxd user:/:/sbin/nologin
vcsa:x:69:69:virtual console memory owner:/dev:/sbin/nologin
rtkit:x:499:497:RealtimeKit:/proc:/sbin/nologin avahi-
autoipd:x:170:170:Avahi
SSH:/var/empty/sshd:/sbin/nologin
tcpdump:x:72:72::/:/sbin/nologin oprofile:x:16:16:Special user
account to be used by OProfile:/home/oprofile:/sbin/nologin
sebl:x:500:500:sebl:/home/sebl:/bin/bash mysql:x:27:27:MySQL
```

Figure 01: Example output of /etc/passwd file for LFI vulnerable site in case of parameter modification

Figure 01 shows the example of the output of /etc/passwd file for LFI vulnerable site in case of parameter modification. In the reply, system discloses sensitive information like root user, password, SSH login credentials, etc.

ii) $POST Method Exploitations:

Developers usually design the data processing techniques through the HTTP POST methods to impose security features. Therefore, the user will not be able to easily view the transformation data. However, the intruder can view the data from session cookies and can perform LFI exploitation even in the

$POST method protection. The exploitation can be performed because of the improper use of different functions/ methods. In the cookies, this hidden filename information exists with other data in the format ' ?file=/etc/passwd'.

Developers also store different types of files in different directories to ensure security. As a result, the attacker may not find the file that they are looking for. To bypass this type of security, the attacker will perform directory transversal query execution for getting the file they seek. The following simple command (../) is used to change the recent directory.

? file=../../../../../../../../etc/passwd.namefile=nt

B. RFI-based LFI Exploitation:

RFI and LFI vulnerability are nearly similar except that RFI can access "/proc/self/environment" directory. Therefore, RFI exploitation will result in more harm for the web applications as compared to general LFI vulnerability. If the directory is writeable in the web applications URL, the attacker can modify the parameter as below.

http://www.website.com/downloads.php?
file=/proc/self/environment

The site replies with the following code after entering the given modification in the URL of the web application.

C. SQLi-based LFI Exploitation:

Structure query language injection is a process to retrieve data from a database directly by inserting SQL queries from injectable perimeters and the data displayed on the web browser. For SQLi-based LFI, exploitations are divided in two categories, as discussed below:

i) Union Base Exploitation:

The malicious syntax for exploiting LFI vulnerability is replaced by the vulnerable columns number that is shown in the browser during injection.

http://www.website.com/downloads.php? id=10' and false union select 1, load_file ('/etc/passwd'), 3,4--+-

After executing the above manual modification, the output will be in the web browser in a format like Figure 01, where the system discloses sensitive information like root user, password, SSH login credentials, etc.

ii) Error Base Exploitation:

Error base injections take advantages of poor error handling in the web application.[] For error base exploitation, it is not necessary to count all columns of the database of the web application. Some built-in functions like floor(), rand() forces the database to retrieve data from database with some limitations. The following is the example of error base SQL injection.

http://www.website.com/downloads.php? id=10' and OR 1
GROUP BY CONCAT_WS(0x3a,
'/etc/passwd',FLOOR(RAND(0)*2)) HAVING MIN(0) OR 1--+-

Here, concat_ws() function is use to concatenate multiple functions, floor() that returns a random floating point where 0x3a is the hex value of ' : '.

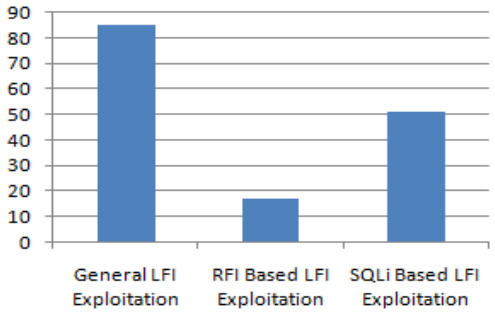
These query returns all the root passwords, the same as Figure 01. The attacker will find the username and password from the database for SSH login. The password that the intruder receives from the output of query is usually encrypted in MD5 format. If the intruder can access '/proc/self/environment', it will enable them to update/modify or insert malicious codes directly to the specific web server instead of the root user credentials.

V. RESULTS AND STATISTICS

In this study, we have formulated the sample size mechanism using a universal calculator, provided by G*Power 3.1.9.2. In this research, we examined 153 LFI vulnerable websites of Bangladesh. We found three types of LFI vulnerability in 153 websites in the sample. We had chosen manual black box testing approach to collect data for this study. We analysed this dataset, based on LFI exploitation type, level of access in host system, and level of risk created by LFI vulnerability. The analysis is discussed below:

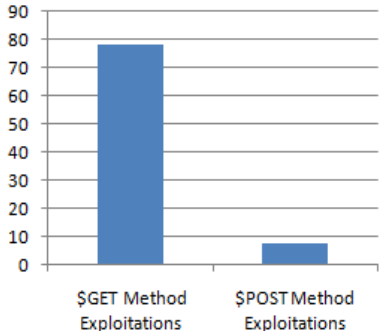
A. Analysis of LFI exploitation type:

Graph 01 shows that among 153 LFI-vulnerable web-site/applications, 85 sites can be exploited by using the general type of LFI exploitation. RFI-based LFI and SQLi-based types of LFI exploitation were found in 17 and 51 websites respectively.



Graph 01: LFI Exploitation Types

We also found 78 websites could be exploited by $GET method whereas 7 sites can be exploited by $POST method in general LFI exploitations.

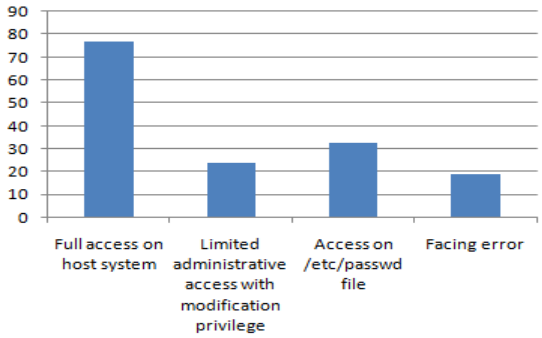


Graph 02: General LFI Exploitation Type

After analysing the above statistics, it was observed that a good number of Bangladeshi websites has been built without adequate protection against LFI exploitation during the development of the application.

B. Analysis based on level of access in host system:

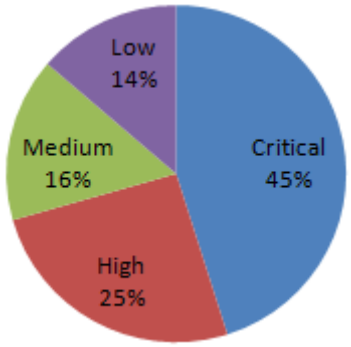
In this review, we have segregated the level of access into four categories that an intruder can achieve after exploiting the LFI vulnerability. The statistics are presented in Graph 03. Among 153 LFI vulnerable websites, 77 websites can compromise their full access on the host system after the exploitation. Limited administrative access with modification privilege can be achieved in 24 websites. Access on /etc/passwd file could be gained in 33 sites. Only 19 websites faced error messages.



Graph 03: Level of Access in Host System

c. Analysis based on the level of risk in creating LFI vulnerability:

The risk level is identified by the impact of compromises in critical information that could be retrieved from the websites through LFI exploitation. We have categorized the level of risk in four different criteria i.e. Critical, High, Medium, and Low. The statistics of the given risk level are shown in Graph 04. In this dataset, 45% risk is considered Critical, since the intruder can obtain full access and control over the host server and in the web application's admin panel. Here, 25% of websites are marked as High-risk, which can be exploited easily by the outsider ad will cause serious harm to the host server. In this review, 16% and 14% sites are considered Medium and Low respectively in terms of risk involved after the LFI vulnerability is exploited.



Graph 04: Level of Risk Caused by LFI Vulnerability

VI. CONCLUSION

In this paper, we evaluated 153 LFI vulnerable websites of Bangladesh. We explored RFI and SQLi-based LFI vulnerability and techniques for its exploitation in this study. We also figured out the impact in the case of exploitation of this vulnerability in the web applications of Bangladesh. This type of vulnerability is one of the most common weaknesses in the web applications, in which the attacker can easily misuse it by altering system resources/by running a shell script to obtain access to any confidential information. Our research in this area is a continuous process. We will conduct our investigation on different sectors of Bangladesh and we will also work on different solutions of this problem. Our observations about the existence of this vulnerability in web applications indicate the developer's carelessness e.g. reducing parameter population, secure use of php functions like require(), include(), etc. The effects of the exploitation of this vulnerability can be alarming in Bangladesh if proper measures are not taken on time. Careful coding practices and regular monitoring during development of the web applications would minimize the risk of those silly mistakes that can create serious vulnerability in the applications.

VII. REFERENCES

[1] O. B. Al-Khurafi and M. A. Al-Ahmad, "Survey of Web Application Vulnerability Attacks," 4th International Conference on Advanced Computer Science Applications and Technologies (ACSAT), Kuala Lumpur. pp. 154–158, 2015. *(references)*

[2] M. S. Tajbakhsh and J. Bagherzadeh, "A sound framework for dynamic prevention of Local File Inclusion," Information and Knowledge Technology (IKT), 7th Conference on, Urmia. pp. 1–61, 2015.

[3] T. Farah, M. Shojol, M. Hassan and D. Alam, "Assessment of vulnerabilities of web applications of Bangladesh: A case study of XSS & CSRF," Sixth International Conference on Digital Information and Communication Technology and its Applications (DICTAP), Konya. pp. 74–78, 2016.

[4] D. Alam, M. A. Kabir, T. Bhuiyan and T. Farah, "A Case Study of SQL Injection Vulnerabilities Assessment of .bd Domain Web Applications." Fourth International Conference on Cyber Security, Cyber Warfare, and Digital Forensic (CyberSec), Jakarta. pp. 73–77, 2015.

[5] T. Farah, D. Alam, M. N. B. Ali and M. A. Kabir, "Investigation of Bangladesh region based web applications: A case study of 64 based, local, and global SQLi vulnerability," IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE), Dhaka. pp. 177–180, 2015.

[6] D. Alam, T. Bhuiyan, M. A. Kabir and T. Farah, "SQLi vulnerability in education sector websites of Bangladesh", 2015 Second International Conference on Information Security and Cyber Forensics (InfoSec), Cape Town. pp. 152–157, 2015.

[7] B. Rexha, A. Halili, K. Rrmoku and D. Imeraj, "Impact of secure programming on web application vulnerabilities," IEEE International Conference on Computer Graphics, Vision and Information Security (CGVIS), Bhubaneswar. pp. 61–66, 2015. [Accessed: 29 September 2015]

[8] N. M. Vithanage and N. Jeyamohan, "WebGuardia—an integrated penetration testing system to detect web application vulnerabilities," International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai. pp. 221–227, 2016.

[9] T. Bhuiyan, D. Alam and T. Farah, "Evaluating the Readiness of Cyber Resilient Bangladesh. Journal of Internet Technology and Secured Transactions (JITST)," Vol. 4, No. 1, ISSN 2046-3723,2016.

[10] R. Johari and P. Sharma. "A Survey On Web Application Vulnerabilities (SQLIA,XSS) Exploitation and Security Engine for

SQL Injection," International Conference on Communication Systems and Network Technologies, 2012.

[11] N. E. E. Moussaid and A. Toumanari. "Web Application Attacks Detection: A Survey and Classification," International Journal of Computer Applications (0975–8887) , Volume 103–No. 12, October 2014.

[12] P. V. Ami and S.C. Malav, "Top Five Dangerous Security Risks Over Web Application," International Journal of Emerging Trends & Technology in Computer Science, 2013.

[13] M.S. Tajbakhshand and J. Bagherzadeh, "A Sound Framework for Dynamic Prevention of Local File Inclusion," 7th International Conference Information and Knowledge Technology, 2015.

[14] Internet-subscribers-bangladesh-august-2016. 2016. [Online]. Available: www.btrc.gov.bd/content/Internet-subscribers-bangladesh-august-2016 [Accessed: 25 September 2016].

Design and Optimization of An E-shaped Wearable Antenna Working in ISM Band

Md.Saheb Ali, Khaleda Ali
Electronics and Telecommunication Engineering
University of Liberal Arts Bangladesh
Road 4/a ,Dhanmondi, Dhaka ,Bangladesh.
sahebalirs@gmail.com
Khaleda.ali@ulab.edu.bd

Abstract-**A flexible planar textile antenna suitable for wearable applications has been designed at 2.4GHz. The effects of the dielectric substrates of the micro strip antenna has been investigated. Such materials are: fleece, denim and Velcro. Both edge mounted and vertical mounted excitation configurations are considered. It is observed that for edge mount, micro strip line feeding, antenna with fleece provides highest gain (7.59dBi). However the impedance bandwidth with the same feeding is higher for the antennas with denim substrate (25%). Moreover the impedance bandwidth can be enhanced (36%) by using vertical feed. The numerical study has been performed using CST microwave studio.**

Keywords-Textile antenna, wearable electronics, radiation pattern.s

I. INTRODUCTION

In recent years, the area of wireless communications has exhibited a significant progress. Body centric wireless communications (BCWC) is one of the key contributors towards such expansion. Both the academic and industrial community have shown tremendous interest in this field. From health care to sports applications and even in the defence sectors BCWC flaunts its suitability. A scalable and robust body area network requires smaller and flexible antennas that can be placed near human body. Textile antennas provide logical solutions to such problems. These antennas are mainly composed of a conductive patch as radiating element attached to textile materials as dielectric substrates. This structure can be easily integrated into the clothing [1]. Such combination of microstrip with high impedance fabric surface minimizes the back radiation and make it more appropriate for on-body communications. Moreover they offer several advantages including low budget, short profile, and minimal weight.

Over last few years different research works have been accomplished on textile antennas [2]-[8]. In [2] Massey has investigated a quarter wavelength patch antenna using copper plated rip-stop nylon as the radiating structure and foam as the substrate. In [3] and [4], fabric antennas are fabricated from knitted copper and from woven conductive fabrics respectively. In [5], [6] aracon and flectron are considered as textile substrates. However such antennas due to their knitted structure and expensive textile substrates do not provide much flexibility and high gain.

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

Therefore, the aim of this paper is to design and investigate a simple textile antenna with easily available and cheap fabric materials incorporated with a copper tape. To perform such operation, three fabrics are considered: fleece, denim and velcro. E-shaped copper tape can be integrated on to the substrate. Both the probe feeding and microstrip line feeding are incorporated in the study.

II. SIMULATION SETUP

The numerical study has been performed using CST microwave studio .Three sets of simulations are performed using three different textile materials as dielectric substrates: fleece, denim and velcro [9]. .A copper (annealed) conductor is used as the E-shaped patch. The graphical representation of the antenna with two feeding techniques are given in Figure 1 and Figure 2.

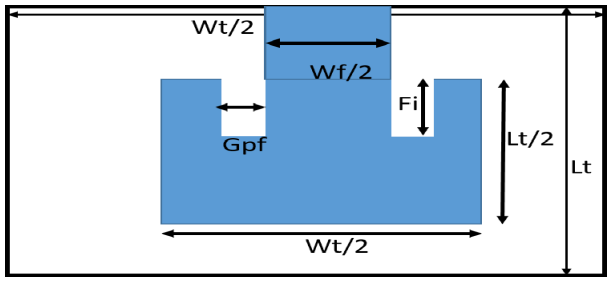


Figure 1. E-shaped textile antenna fed by microstrip line

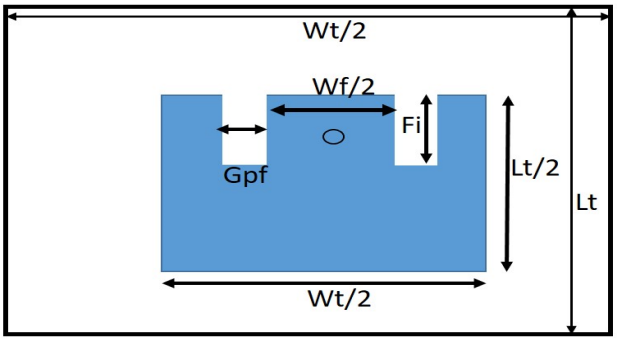


Figure 2. E-shaped textile antenna feed by coaxial probe

The dimension of the substrate and patch are obtained from the equations (1-5) found from literature [10], [11].

$$Wt = \frac{c}{2fs\sqrt{\frac{\epsilon r+1}{2}}} \quad (1)$$

$$\epsilon\, eff = \frac{\epsilon r+1}{2} + \frac{\epsilon r-1}{2}\left[1 + 12\frac{h}{Wt}\right]^{-\frac{1}{2}} \quad (2)$$

$$Leff = \frac{c}{2f\sqrt{\epsilon eff}} \quad (3)$$

$$\Delta L = 0.412h\frac{(\epsilon eff+0.3)(\frac{W}{h}+0.264)}{(\epsilon eff-0.258)(\frac{W}{h}+0.8)} \quad (4)$$

$$Lt = Leff - 2\Delta L \quad (5)$$

Here, fs is the Resonance Frequency; Wt is the Width of the Patch; ϵ_r is the relative Permittivity of the dielectric substrate; h is the thickness of the substrate; c is the Speed of light; εeff is the effective relative Permittivity of the dielectric substrate; Leff is the effective Length; ΔL is the extension of the length and Lt is the length of the patch.

Based on these formulae, the optimized values of the antenna dimensions are given in Table 1. Since decreasing the dielectric constant values of the substrate may enhance the resonant frequency [9] therefore the antenna dimensions for different substrates are kept different.

Parameter	Fleece	Denim	Velcro
Wt (mm)	121.9	109.6	115.5
Lt (mm)	109.1	93.9	10
Gpf	1	3	1
Fi (mm)	16	16	16
Wf (mm)	4.065	4.065	4.065

Table 1.Dimensions of the proposed textile antenna

The dielectric constant and loss tangent of the substrates (given in Table 2) range from 1-1.6.

Textile materials	Fleece	Velcro	Denim
Dielectric constant	1	1.34	1.6
Loss tangent	0	0.006	0.07

Table 2. Dielectric properties of the textile materials

To consider the effect of the feeding technique two different types of feeds are used: edge mounted feeding and vertical probe feeding. The dimensions (width×length) of microstrip line and of the coaxial probe (RG174) are provided in Table 3.

Feeding technique					
Microstrip line			*Coaxial Probe(RG174)*		
Dimension	Width (mm)	Length (mm)	Parameter	Inner conductor radius (mm)	outer conductor radius (mm)
value	4.065	45.5	value	0.24	0.99

Table 3: Dimension of the microstrip line and Coaxial probe

In edge mounted excitation a microstrip line is added to the E-shaped patch where the coaxial transmission is assumed to be attached at the edge of the substrate.

In vertical probe feed, inner conductor of the coaxial line is considered to be passed through the textile material intersecting the microstrip patch. A RG174 SMA coaxial line is considered to be connected in this case.

The location of the probe feed is selected at a point where the impedance matches.

III. SIMULATION RESULTS

Different parameters of the antenna are investigated in the concerned study.

(A)*S-parameter*: The s_{11} values of the textile antenna are provided in Figure 3. Three different curves are obtained for three different dielectric substrates. It can be seen that the bandwidth are different for three different materials. The numerical values of the bandwidth and minimum values of return loss are presented in table 4. It is evident from the Table 4 that denim causes the highest bandwidth. However the reflection coefficient is lowest for Fleece (-22.64 dB).Such difference occur due to the variation of dielectric properties of the textile substrates. Fleece having the lowest dielectric constant and zero value of loss tangent represent itself as the best candidate for textile antenna.

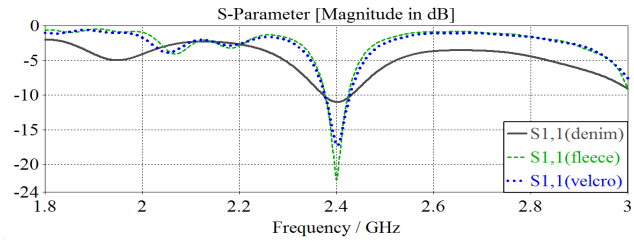


Figure 3. Return loss of the three textile materials

Textile materials Used in antenna	Bandwidth (MHz)	Return loss (dB)
Fleece	45.6	-22.64
Denim	58.6	-11
Velcro	52.1	-17.75

Table 4: Bandwidth and return loss values of the antennas with different textile materials

(B)*VSWR*: VSWR being a function of return loss indicates the effect of antenna mismatch .It is observed from the Figure 4 that for all the textile substrates VSWR≤2 at 2.4 GHz. This indicates that only 10% of power is reflected back.

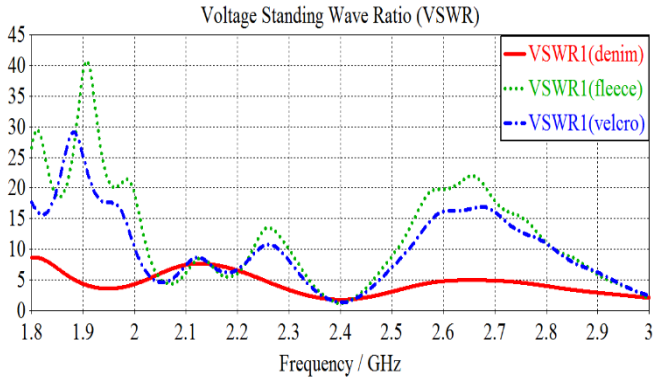


Figure 4. VSWR of the of the antennas with different textile materials

The VSWR of the antennas are given in the following table 5

Textile materials	VSWR
Fleece	1.16763
Denim	1.78345
Velcro	1.31168

Table 3. Minimum VSWR at 2.4 GHz

Moreover, it is observed from Table 5 that minimum value of VSWR is achieved for Fleece and it goes maximum when a denim substrate is used.

(C)*Radiation Pattern*: The radiation pattern of the antenna is plotted in Figure (5), (6) and (7).

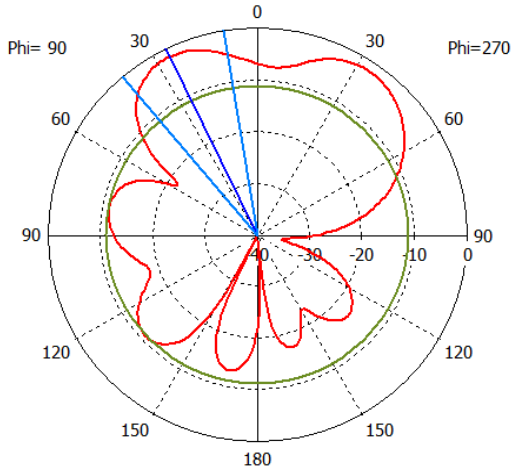


Figure 5. Polar radiation pattern due to the fleece fabrics

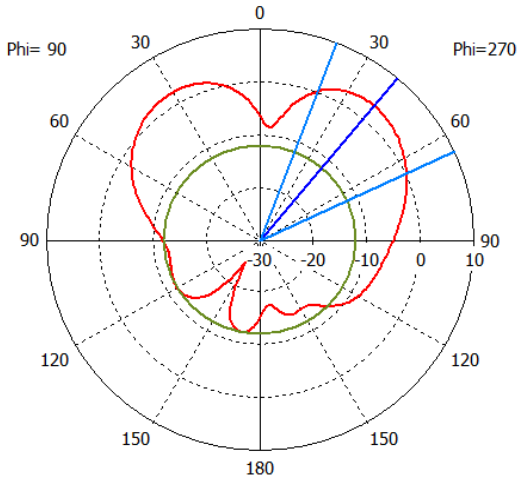


Figure 6. Polar radiation pattern due to denim fabrics

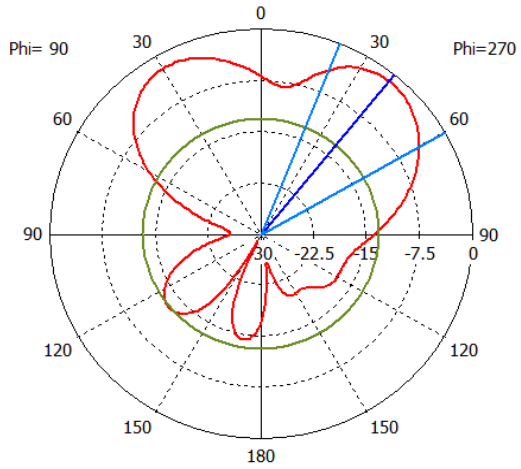


Figure 7. Polar radiation pattern due to velcro fabrics

It is observed that the beamwidth is highest for denim. But the gain values are greater for velcro and fleece.
The directivity, radiation efficiency, gain values and impedance bandwidths are provided in Table 6.

Textile materials	Band width (MHz)	Directivity (dBi)	Radiation efficiency (dB)	Gain (dBi)	Impedence bandwidth (%)
Fleece	45.6	8.4	-0.811	7.59	19
Denim	58.6	7.44	-5.74	1.66	25
Velcro	52.1	7.95	-1.39	6.56	22
Fleece (Coaxial probe feed)	84.64	9.12	-0.1375	8.98	36

Table 6: Comparison of Antenna parameters

Here textile antenna with fleece (fed at the edge) provides the highest gain of 8.59 dBi. The gain value is lowest for denim. However denim shows a high impedance bandwidth of 25%.
The effects of feeding techniques are provided in Figure (8) and (9). A clear difference is visible for Microstrip line and coaxial feed. For microstrip line, due to fleece, the gain and impedance bandwidth are.7.59dB and 19% whereas for probe feed 9dB gain and 36% impedance bandwidth is achieved.

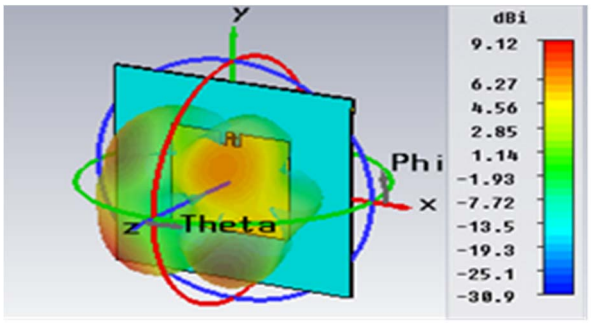


Figure 8. 3-D radiation pattern for coaxial probe feed

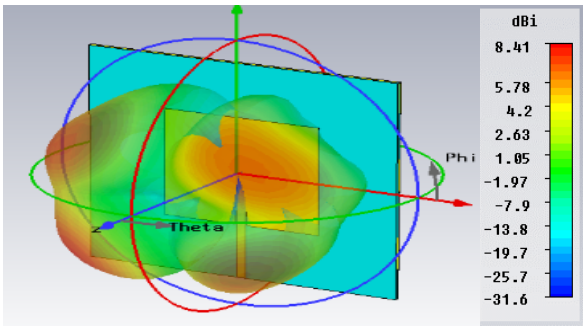


Figure 9. 3-D radiation pattern for microstrip line feed

IV. Conclusions

In this paper a flexible textile antenna has been proposed working at 2.4 GHz. Three different textile materials are considered as dielectric substrates. The maximum dimensions of the antenna with fleece as a dielectric constant is 12.19X10.91cm^2.Such antenna provides the highest gain of 7.59dBi for microstrip line feed with an impedance bandwidth of 19%. On the other hand the impedance bandwidth of the antenna with denim has higher value with a very low gain of 1.66dB.Among three cases, the antenna with velcro has an optimum performance with considerably good gain (6.56dB) and impedance bandwidth of 22%.The consequences of two feeding techniques are also taken into account. It is observed that vertical probe feed can enhance the gain to almost 9dB and impedance bandwidth can be doubled to 36%. Such flexible antennas would be easy and cheap to fabricate which makes them highly suitable for wearable applications.

References

[1] C. Hertleer, H. Rogier, L. Vallozzi, and L. V. Langenhove, "A textile antenna for off-body communication integrated into protective clothing for firefighters," *IEEE Transactions on Antennas and Propagation*, Vol. 57, No. 4, 919–925, Apr. 2009.,

[2]P. J. Massey, "Mobile Phone Antennas Integrated Within Clothing," *IEE 2001 Eleventh International. Conference on Antennas and Propagation*, Vol. 1, Publ.No. 480, Manchester, U. K., 2001, pp. 344–347.

[3] P. Salonen, and H. Hurme, "A Novel Fabric WLAN Antenna for Wearable Applications,"*2003 IEEE Antennas and Propagation Society Int. Symp.*, Columbus, OH, Vol. 2, 2003, pp. 700–703.

[4] M.Tanaka, and J.-H. Jang, "Wearable Microstrip Antenna," *2003 IEEE Antennas and Propagation Society Int. Symp.*, Columbus, OH, Vol. 2, 2003, pp. 704–707.

[5]P. Salonen., et al., "Effect of Conductive Material on Wearable Antenna Performance: A Case Study of WLAN Antennas," *2004 IEEE Antennas and Propagation Society International Symp.*, Vol. 1, Monterey, CA, 2004, pp. 455–458.

[6]M. Rizwan, Y. R.-Samii and L. Ukkonen " Circularly polarized textile antenna for 2.45GHz" *2015 IEEE MTT-S 2015 International Microwave Workshop Series on* ,pp. 21-22.

[7] S. Khan, V. K Singh, B**.** Naresh , " *Textile Antenna Using Jeans Substrate for Wireless Communication Application*," *International Journal of Engineering Technology Science and Researc*h,Volume 2, Issue 11 November 2015

[8]S. Subramaniam, S .Dhar, K. Patra and B. Gupta "Miniaturization of Wearable Electro-textile Antennas using *Minkowski Fractal Geometry*," *2014 IEEE Antennas and Propagation Society International Symposium (APSURSI),*pp. 309 - 310

[9]R. Salvado, C. Loss, R. Gonçalves, and P. Pinho, "*Textile materials for the design of wearable antennas: A survey," J. Sensors*, vol. 12, pp. 15841–15857, 2012

[10] C. A. Balanis, "*Antenna Theory: Analysis and Design*, 3rd Edition," John Wiley & Sons, 2005.

[11]I. Locher, M. Klemm, T. Kirstein, and G. Troster, "Design and Characterization of Purely Textile Patch Antennas," *IEEE Trans. Adv. Packag.*, vol. 29, no. 4, pp. 777–788, 2006

[12]P. S. Hall and Y. Hao, *Antennas and Propagation* for Body-Centric Wireless Communications. Artech House, 2006.

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

Lip Contour Extraction Using Elliptical Model

Shahnewaz Ali
Omniabit S.r.l
Milan, Italy
shahnewaz_ali@yahoo.com

***Abstract*— Lip features and contour detection is an important aspect of computer vision, having many application domains. This research conveys various aspects of the related topics and presents an elliptical mathematical approach followed by image processing methods such as morphological operation to extract lip contour. Lip contour detection starts from mouth area which is segmented from face image area and then it fits an elliptical contour on lip area by evaluating image features such as edges and corner points thus identify lips and its properties. Later morphological approach is introduced to eliminate and compensate contour points. This application is suitable to run on embedded electronic devices specially considered Advanced RISC Machine, well known as ARM. Face detection itself a complicated task thus core application perform several image processing steps to define face area properly. In the domain of Human Computer Interface System human skin color is determined by using color map. This is also a necessary component for face area detection. This research conveys necessary information regarding this issue such as color space and classification which have pivotal role to define Lip Contour from a detected face area.**

Keywords— lip contour; ARM processor; elliptical model; skin color map; morphological processing; computer vision; human computer interfaces

I. INTRODUCTION

Many computer vision applications require Lip Contour detection mechanism or Lip Features extraction capability to serve their purposes. Human expression recognition, intelligent video conference system etc. can be considered potential sectors where lip contour detection and changes of the contour properties therefore lips movement play a significant role to formulate a solution. In these mentioned Human Computer Interface Systems, human presence in a video frame or in an image plays a pivotal role. If a system becomes incapable to detect faces, it will fail to perform other steps as well. Skin color map, a color map to distinguish skin, has a significant importance here and it is a complex domain itself due to skin color varies from man to man. There are significant research has been taken in this area to identify and unify skin color model globally [8][9]. Moreover, embedded electronic system and devices have achieved an enormous development. Currently multi core embedded devices are available at reasonable price. The reference model considered Reduced Instruction Set Computing (RISC) architecture based processor which is well known as an Advanced RISC Machine (ARM).

This paper proposed an elliptical mathematical approach to identify lip contour and run over ARM embedded processor. To obtain image at run time, system utilize video frame. System captures video frame through the attached Universal Serial BUS (USB) camera and recognize human face in the image, a run time captured video frame. To justify face area and before proceeding other steps, detected face area is evaluated with skin color model as well. After successful face detection, for each face, mouth region is cropped. Edge detection mechanism then takes place and later edges are evaluated to find corners point, a point where neighbor pixels have different rotational angle. Using these corner points, mathematical approach has been obtain which also validate each edge features corresponds to lip feature. Finally morphological operation takes place to eliminate and compensate contour points. Model system has been tested with embedded ARM Linux Board which is equipped with a USB camera.

This paper is organized as section II. illustrates literature overview of related technologies and theories, section III. illustrates Working methodology, system implementation and achievements, Section IV. Results and Discussions and finally section V. contains conclusion which describes possible uses and achievements summary.

II. LITERATURE

A. Embedded System Board

An embedded system board basically consists of computational unit known as central processing unit, peripherals and internal component such as register and memory. Modern embedded system board support various size of memory from some Mega Bytes to Giga Bytes. Every processor's architecture follows a defined set of Instructions, known as Instruction Set Architecture (ISA).

ARM processor supports Reduced Instruction Set Computing (RISC) architecture [1]. This research tested on ARM 32-bit RISC processor. Apart from the processor architecture, embedded system board usually consist Operating System also known as kernel. Presently, there are many Linux based Operating System exists, more common are Debian based operating system, OpenSUSE, ARCH Linux. Operating System provides a way to use embedded board and processor as a whole. The research concern about the peripheral and networking stack of Operating System due to image capturing devices can be accessible through Networking Protocol e.g. Hyper Text Transfer Protocol (HTTP) or direct access via device peripherals such as Universal Serial Bus (USB).

B. Color Domain

Color Space contains color values by defining different color components. Usually a color is often referred to RGB

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

color space, where R is referred to Red color component, G stands for Green color and finally B components represents Blue color components [8]. Apart from this color space several color space such as HSV(Hue, Saturation, and Value), HSI (Hue, Saturation, and Intensity) or YCbCr color space are considered and have a pivotal role in image processing scope.

Opting a color space is an important step for skin color discrimination. Research shows YCbCr have better effect to detect skin pixel or nonskin pixel [2][8] where Y stands for Luminance, Cb stands to Chromatic blue and Cr corresponds to Chromatic red [8]. It is possible to convert RGB color space image to YCbCr color space image. Considering Open Computer Vision Library (OpenCV), it has a built in function to convert captured image to YCbCr image.

III. WORKING METHODOLOGY

In Human Computer Interaction System lip contour takes a pivotal role. To solve this problem some researcher proposed template based methodology [4]. This research work presents elliptical model by estimating ellipse parameters and then extract lip contour area form frontal face images. The basic building blocks are described in Fig. 1. Further the working process can be break down into the following sub sections.

A. Capture Image

Image can be obtained via USB camera module or using Networking protocol. OpenCV provides API to capture image directly if the camera module is physically attached to the device. In addition, system can also read still images. The main objective of this subtask is to provide frontal images to the consequent task block. At this stage the raw image is represented into RGB color space.

B. Region of Interest Detection

Before entering into lip contour detection method, it is required to identify region of interest. This step intended to detect presence of a face. This phase consists the method described by Viola-Jones [5]. It is widely used to detect face in real time[5]. System converts input RGB images to Gray scale by taking the average value of RGB componets.

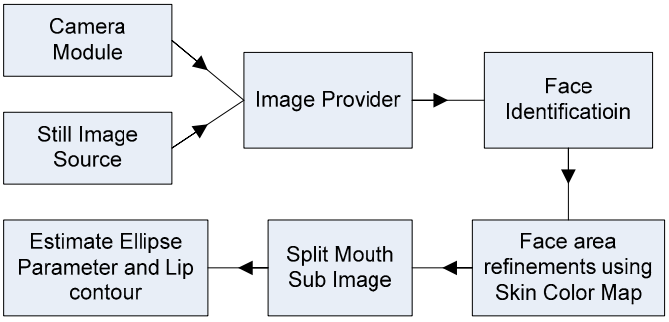


Fig. 1. Building Blocks of Lip Contour extraction Process

Face detection and classification can be achieved by Haar feature-based cascade classifiers[7], OpenCV provides a trained classifier to serve this functionality. According to this classifier the algorithm needs to be trained by a lot of positive images and negative images so that it can classify face features. But in some cases it detect face region where no face exist [6]. Some researcher evaluates this by using increased face area window [6], however, this work evaluates chromatic information to refine the detected area using YCbCr color space [8]. The principle of this refinement is based on skin color pixel or nonskin color pixel. If the detected image area contains any nonskin pixel according to the color space the pixel area will be rejected.

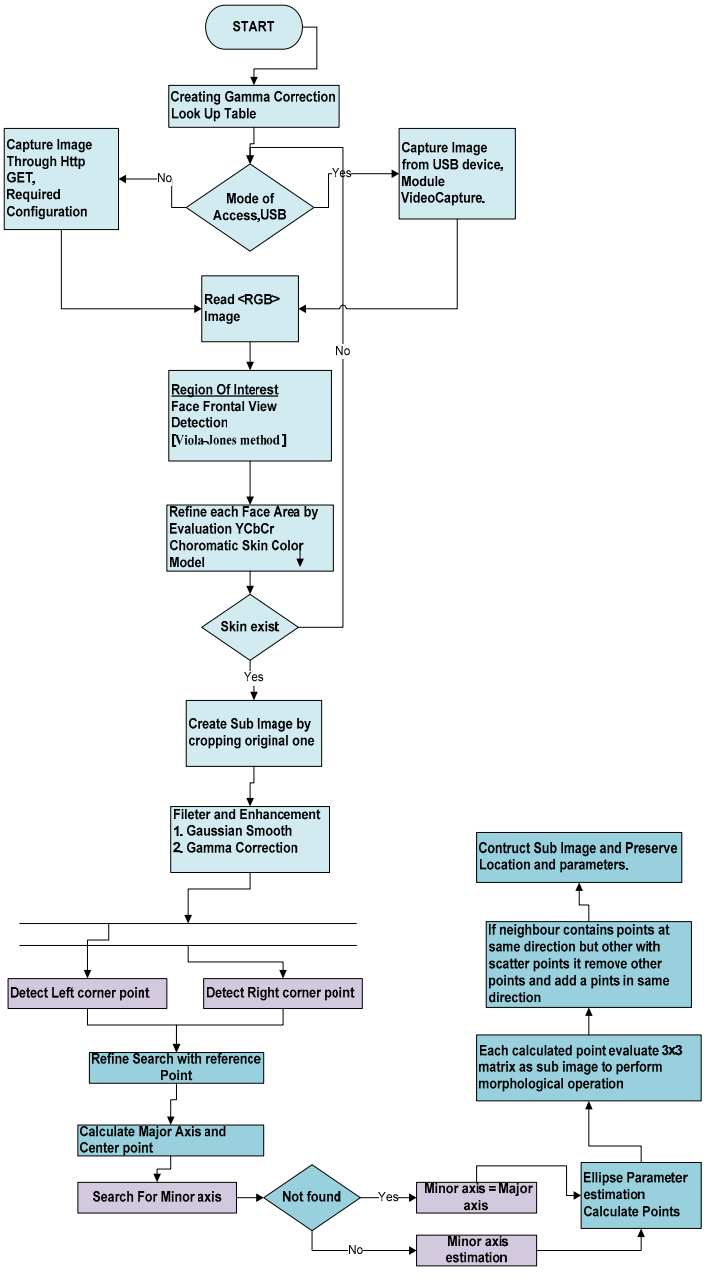


Fig. 2. Application Working Process

C. Image cropping

Image frame that contains face will be cropped by two third heights, the reason behind that lip area remains in this region over the total face area [6][10]. This cropped image will be

passed through the filter phase before performing contour detection.

D. Image Enhancement and Filter

Target image frame is converted into Gray scale image. Converted image pass through the Gaussian filter to reduce noises and to have a smooth image. Gaussian smoothing filter or Gaussian kernel widely used to remove spatial noises from the image. Core of this filter is Gaussian kernel which can be defined in 2D space as,

$$\text{Kernel } G(x,y)= \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (1)$$

Greater standard deviation (σ) performs more image smoothing, using OpenCv it is possible to achieve the value of standard deviation dynamically from the kernel size. Kernel size is the sub block of the pixels where Gaussian convolution will take place for each pixel. This experiment use kernel size 5x5 to have a moderate image smoothing.

On the other side, Gamma correction is an image enhancement method which increase the image overall quality. The benefit of Gamma correction method over gray scale images are contrast ratio improvement and beneficent for features extractions process [12] such for edges. It is a nonlinear pixel wise operation thus it has a computational cost. The reference image is cropped into the region where lips exist therefore this sub image has smaller dimension than original image which makes this operation almost negligible. Gamma operation can be defined as,

$$X_0 = AX_i^{\gamma} \quad (2)$$

Where X_0denotes output pixel value, A stands for Gain and often it is considered equal to 1. X_i refers to the input pixel value. γ value is considered 0.5, research finds these value is suitable for image enhancing [11]. Moreover, the sub image is converted into gray scale therefore it is possible to create a Look Up Table (LUT) to eliminate computation cost at run time. Look Up Table contains gray scale pixel value from 0 to 255 and compute its Gamma correction value that is rounded to maximum value of 255. Finally this value is written to the LUT, at run time module looks in to the LUT to find corresponding Gamma Correction value.

E. Edge Detection

Image frame that contains face will be cropped as mentioned above. This cropped image will be passed through the filter phase before performing contour detection. This research considers edges are the features of the lip contour detection, however integral projection and edge angle detection [10] or Harris corner point are the alternative choices to defined lip contour features.

Fig. 3. Detected Edges using Canny Edge detector

To define lip contour, Canny Edge detector is considered to detect edges from the sub image. Canny Edge detector can detect edges in low error rate, edges are accurately localized and finally noises have low effect on edge detection.

F. Determine contour

Binary image contains only the edge and nonedge pixels. Here the elliptical model starts to recognize lip contour by estimating ellipse parameters. This paper consider the presence of edges on the corresponding pixel position, considering the 3x3 matrix stated as below,

TABLE I. 3X3 MATRIX ANGLE POSITIONS

x1,y1		x1,y3
	X,Y	
x3,y1		x3,y3

The change of angle can be localized by evaluating the pixel tuples (x1, y1), (x1, y3), (x3,y1), (x3,y3) at point X,Y. Slope is defined by,

$$\theta = \arctan\left(\frac{\Delta y}{\Delta x}\right) \quad (3)$$

This information is useful to define a pixel as a peak point having upward and downward changes. For each edges these information preserve tuple contains (location, $\theta_{up}, \theta_{down}$). To compensate edges detection or some mechanical noise during image procurement 5x5 matrix is evaluate.

TABLE II. 5X5 MATRIX ANGLE POSITIONS

x0,y0				x0,y4
	x1,y1		x1,y3	
		X,Y		
	x3,y1		x3,y3	
x4,y0				x4,y4

By evaluating this matrix, if the sub matrix (1;3) or (3;4) exists this indicates a curvature point as illustrated in the Fig. 4. The sub matrix tuple expresses a portion of matrix as (start column; end column) respect to the original matrix.

Fig. 4. Corner Point characteristics

Initial state of the algorithm finds the horizontal vertices points of ellipse. There will be two vertex points on horizontal axis, each vertex point preserve the above properties. Therefore left vertex point have minimum X-coordinate value and at least one point exist on coordinate I and coordinate IV.

If the point consist any directional changes but not both but it contains points where value of angle resides, $\theta = 90$ or 270, algorithms consider this is a new point and continue to search missing direction until boundary condition has been reached. If found then these two points will be coalesced into first point otherwise both of them will be rejected.

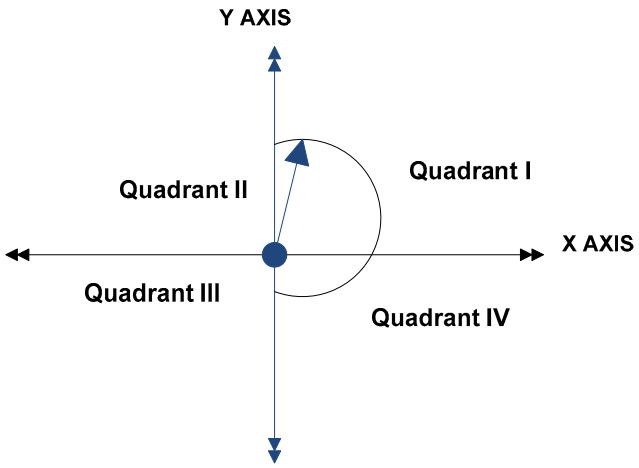


Fig. 5. Geometric Quadrant use to search angle points

Similarly, for right vertex, the algorithm considers maximum X- coordinate value where two directional changes confined. Algorithm proceeds for the left corner point to right by incrementing X coordinate each time.

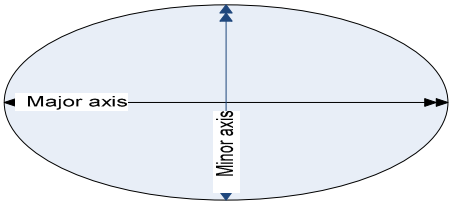


Fig. 6. Ellipse Model

After finding, two vertices, algorithm draws the major axis of the ellipse. Middle point of the major axis is the intersection point of major axis and minor axis. According to the Coordinate Geometry this can calculated by using the formula stated below,

$$X_{mid} = \left(\frac{x_{left}+x_{right}}{2}\right) and\ Y_{mid} = \left(\frac{y_{left}+y_{right}}{2}\right) \quad (4)$$

From the center point it proceeds searching vertical vertices along the axis Y. If not found, it will interpolate that it exist on the point where the distance between the center point and the vertex is equal to the distance from the center point to the any of the horizontal vertex therefore the X intercept and Y intercept become equal. In mathematics, ellipse, which is not centered at origin (0,0) but at a point having coordinate value (h,k), defined as,

$$\frac{(x-h)^2}{a^2} + \frac{(y-k)^2}{b^2} = 1 \quad (5)$$

Solving (5) all the points of the ellipse, whose parameters are obtained from lip contour structure, can be computed. Morphological operation will take place to refine points using sub image of 3x3 matrix. During this operation erosion and dilation are performed which consider 9 neighbor original image points to the point which obtained by the ellipse. Fig. 7, depicts resulting contour obtained from the above method and superposition of the both images, original and processed elliptical shape depicted in Fig. 8.

Fig. 7. Resulting lip contour obtained from the method

Fig. 8. Detected contour is super positioned into the original edge detected image

IV. Results and Discussion

System has been tested using different still images and video frames. System assumes, all the images are captured from frontal view and having straight in focused. However, test has been performed with Frontal Face Images considering different condition and it used still images and partial set of frontal images data base known as MUCT database category ‘a’. It is observed that with different lighting condition, shadow, camera focus angle and posture of frontal face edge detection outcome changes thus affects detection approach.

This method is able to extract lip contour from a complex edge space as depicted Fig. 9. In various conditions such as light reflection, focused angled, open and close mouth condition and different lips structure, this method is able to extract lip contour and performance error index observed started from 1.4% but depends on edges. This method relies on detected edges. It is observed that, when lip width ratio varies between upper and lower lips the error also increased and same occurrences also observed when corner points are missing or having many corner points. In addition, if the face angle increases the error rate may increase if it encountered lack of sufficient edge points. Moreover, in some case, it is observed that face skin and lip color are too similar, and the corner point is not observed through edge detection.

On the other side, in some cases it is mostly appreciated to have another mathematical model to construct Lip-Content Sub Image along with Anthropometric theory. The system can be improved by using K-Means clustering theory to coalesce couple of scattered point into one point, which is left for the future task. During the experiments, system can able to define lip contour on both situation open mouth position and close mouth, it preserves the minor axis changes. In the field of Human Computer Interaction, this information is available to recognized person activity such as smile, silent, talking etc.

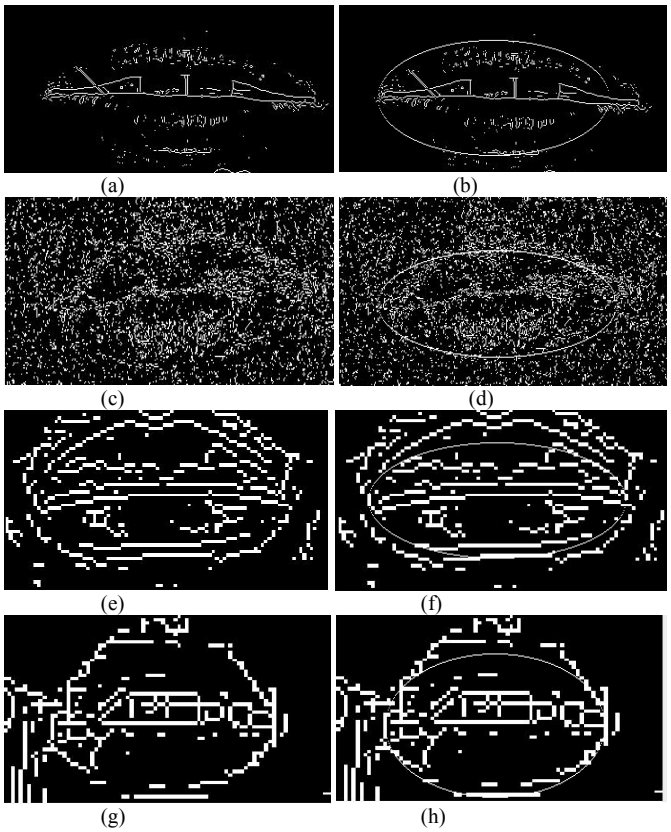


Fig. 9. Model Test Simulation at different conditions, errors are correspond to the edged points (a) edges of angled focused frontal face image having moderate contrast, two corner points are detected properly (b) Ellipse model, considering edges it encountered approximately 1.4% deviation (c) edges of frontal face image having angled focused and greater lighting reflection condition (d) Ellipse model, considering edges it encountered approximately 3.12% deviation (e) Edges of lips having different Lip width ratio (f) Ellipse model, faced more error, approximately 20% (g) inadequate Edges, open mouth condition contains noises on left side and inner mouth edges, corner points are evaluated through recurrent point search (h) Ellipse model faces error on upper lips approximately 6%.

V. CONCLUSION

The rudimentary goal is to extract lip contour has been achieved and tested. Lip contour has been parameterized by ellipse model. Human Computer Interaction (HCI) System such as human gestures and expression detection, intelligent video conference system currently is a growing sector where lip contour detection and mouth movement [13] takes a pivotal role to define whether human is talking or to identify his/her facial expression and events. By evaluating contour parameter changes, it can be possible to identify person activity. Therefore, this research focused on the lip contour detection where precondition is assumed that contour detection phase image is taken from frontal view. This calibration phase, person remains mouth in close position. Considering open mouth, inner contour is not observed. Instead of that it focused to the outer contour. Change of minor axis, from mouth close position to open and vice versa can be recognized a human state. On the other side changes on major axis as well as minor axis can be clustered into different phases such smile or other expression.

Implementation of this research has been tested with Matlab R2008b version as well as with OpenCv 3.1. Host computing is used during test are Intel (R) Core i3 and ARM 32 bit embedded Linux system.

ACKNOWLEDGMENT

Evaluating edges, a lip contour extraction approach using elliptical model has been presented in this paper. Author would like to thanks to the reviewers for their precious reviews that helped author to improve the paper overall quality.

REFERENCES

[1] Y. Yuan, W. Wu, L. Hou, S. Geng, Z. Zhou and Q. Liu, "The Study of SoC Architecture Design Based on 32-bit Embedded RISC Processor," 2012 Second International Conference on Intelligent System Design and Engineering Application, Sanya, Hainan, 2012, pp. 1339-1342.

[2] B. D. Zarit, B. J. Super and F. K. H. Quek, "Comparison of five color models in skin pixel classification," Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, 1999. Proceedings. International Workshop on, Corfu, 1999, pp. 58-63.

[3] S. L. Phung, A. Bouzerdoum and D. Chai, "Skin segmentation using color and edge information," Signal Processing and Its Applications, 2003. Proceedings. Seventh International Symposium on, 2003, pp. 525-528 vol.1.

[4] N. S. Nayak, R. Velmurugan, P. C. Pandey and S. Saha, "Estimation of lip opening for scaling of vocal tract area function for speech training aids," Communications (NCC), 2012 National Conference on, Kharagpur, 2012, pp. 1-5.

[5] P. Viola and M. Jones, "Robust real-time face detection," Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on, 2001, pp. 747-747.

[6] S. Jain, P. C. Pandey and R. Velmurugan, "Lip contour detection for estimation of mouth opening area," 2015 Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), Patna, 2015, pp. 1-4.

[7] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, 2001, pp. I-511-I-518 vol.1.

[8] A. Kumar and S. Malhotra, "Real-time human skin color detection algorithm using skin color map," Computing for Sustainable Global Development (INDIACom), 2015 2nd International Conference on, New Delhi, 2015, pp. 2002-2006.

[9] M. J. Jones and J. M. Rehg, "Statistical color models with application to skin detection," Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on., Fort Collins, CO, 1999, pp. 280 Vol. 1.

[10] S. Wasista, Setiawardhana and F. Z. Rochim, "Lips feature detection using camera and ARM 11," Electronics Symposium (IES), 2015 International, Surabaya, 2015, pp. 123-127.

[11] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 2005, pp. 886-893 vol. 1.

[12] X. Guan, S. Jian, P. Hongda, Z. Zhiguo and G. Haibin, "An Image Enhancement Method Based on Gamma Correction," Computational Intelligence and Design, 2009. ISCID '09. Second International Symposium on, Changsha, 2009, pp. 60-63.

[13] Y. Xiong, B. Fang and F. Quek, "Detection of Mouth Movements and its Applications to Cross-Modal Analysis of Planning Meetings," 2009 International Conference on Multimedia Information Networking and Security, Hubei, 2009, pp. 225-229.

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

Multiclass Motor Imagery Classification for BCI Application

Md. Mamun or Rashid
Department of Biomedical Engineering,
Khulna University of Engineering & Technology
Khulna-9203, Bangladesh
mamunbme@yahoo.com

Mohiuddin Ahmad
Department of Electrical and Electronic Engineering,
Khulna University of Engineering & Technology
Khulna-9203, Bangladesh
mohiuddin.ahmad@gmail.com

***Abstract*— Motor Imagery (MI) is now highly adapted to control machine or computer by interfacing with brain or mind. This paper proposes a method to differentiate left, right and feet motor imagery movement according to two and three class classification using statistical features of the EEG signal of the subjects. For this purpose the collected EEG signals of three subjects are segmented and feed to discrete wavelet transform (DWT). DWT decomposes each EEG segmented signal to collect the relevant features. ANN classifies the three left, right and feet movement class trials data. Classification accuracy varies with respect to subject. This straightforward method can be used to design a well-organized BCI system with better accuracy.**

Keywords—motor imagery; discrete wavelet transform; neural network; brain machine interface; statistical feature.

I. INTRODUCTION

In broad aspects of MI, MI is specially supervised and efficient procedure for the patients who suffer from Motor neuron diseases (MNDs). MNDs are neurological disorders that affect mostly motor neurons; that command over the voluntary muscles of the body. MI is also consulted for the paralyzed patients or if the motor neurons are dead because of disease or accident.

Among many accomplished research studies, authors [1] developed a method to decompose the EEG signal using short time Fourier transform with CSP and SVM for better classification accuracy. Results verified that single channel EEG signals taken from both sensorimotor and forehead areas can classify four class motor imagery movement. Authors in [2] used CSP for feature extraction and FDA for dimension reduction of the feature vectors. Finally SVM classifies the four class motor imagery movement individually from the rest position. Authors in [3] classify left hand, right hand, feet and tongue, a multi-class imagery movement based on ERD and sample entropy using SVM classifier. Researchers in [4] introduced a method to distinguish between simple and compound limb motor imagery with help of CSP and SVM. They adopted ERSP, PSE and spatial distribution coefficient method to prove the comparison. In [5] researchers used SVM, LDA and KNN classifiers to multiclass motor imagery using two different datasets. In [6] authors proposed FBCSP algorithm to separate multiclass motor imagery and evaluates performance in terms of kappa value. Researchers in [7] introduced a new technique to collect MRICs automatically and differentiate three class LH, RH and foot imagery data. In [8], authors used 22 electrodes over the scalp and multi-class discrimination accomplished using ICA and CSP algorithm. In [9] authors proposed Biomimetic Pattern Recognition (BPR) for 2 class motor imagery classification and compares performance between SVM and LDA. In [10], authors proposed wavelet-CSP algorithm with ICA to improve the classification rate of SVM in terms of kappa value.

In this work, we propose an easy method of classifying multiclass EEG MI signal and with acceptable accuracy rate.

The rest of this paper is arranged as follows: Section II describes materials and proposed method. Section III illustrates the results and discussion is in section IV. Finally section V concludes the paper.

II. MATERIALS & PROPOSED METHODOLOGY

A. EEG Data Collection & Dataset Description

This research work exploits a publicly available dataset [11]. Many subjects perform two class and three class motor imagery functions during data acquisition. In our works three class data for three subjects (A, B, C) are used. Fourteen electrodes have been placed over the subject's sensorimotor area for data acquisition as shown in Fig. 1. All three subjects imagine left hand (LH), right hand (RH) and both feet (FT) movement on different days for several sessions. Subjects are requested to sit in front of a computer desk and to imagine as the cue indicated. Motor imagery task starts after the trigger and continuous to 3-10 seconds and ends each trial followed by 2 seconds of short break.

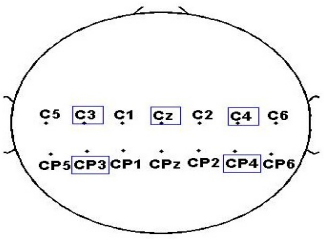


Fig. 1. Fourteen electrode placement for data acquisition.

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

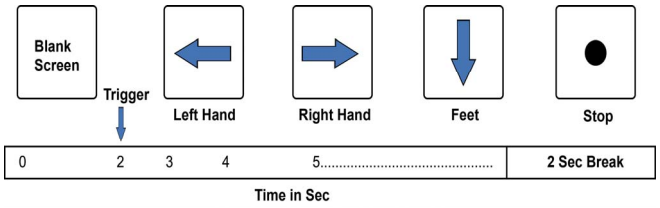


Fig. 2. Block diagram of the experimental paradigm for data acquisition.

The whole experimental procedure with corresponding time is shown in Fig. 2 above as given in [11].

B. Experimental Flowchart and Channel Selection

The complete experimental working flow is shown in Fig. 3 below. First and foremost is to record the motor imagery data from subjects and all subjects are required to perform imagery task well for better results. Recorded signal then segmented according trials and decomposed using DWT to collect statistical characteristics for each signal. These feature vectors afterwards provided to NN to evaluate the classification accuracy of the proposed method.

EEG signals from the subjects are recorded using two devices of g.tec and Neuroscan devices. Among the 14 electrodes 5 electrodes data are recorded for 3 class motor imagery movement. Those 5 electrodes are indicated in figure-1 in square box. But further for our research purpose most significant channels, channels C3 and C4 are taken from both hemispheres.

C. Feature Extraction using DWT

DWT is a prominent tool for EEG signal decomposition to the required level depending on the dominant frequency. In our work db4 wavelet is used to perform level 5 decomposition of each EEG segment. Wavelet toolbox provides 12 statistical characteristics, 5 are chosen as features. In Fig. 4, approximate coefficient is A5 and detail coefficients D1, D2, D3, D4 and D5 are shown, which are outcome by consecutive low pass and high pass filtering to level 5. Each EEG segment DWT decomposes into increasingly finer detail based on two sets of basic functions [12] as follows:

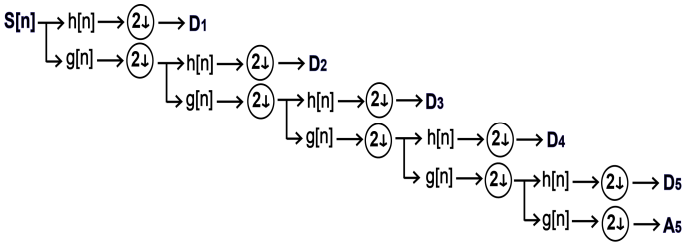


Fig. 4. Motor imagery EEG signal decomposition using db 4 wavelet.

$$S(t)=\sum_{k}2^{j_0/2}a_{j_0}(k)\phi(2^{j_0}t-k)+\sum_{j=j_0}^{\infty}\sum_{k}2^{j/2}d_j(k)\psi(2^j t-k) \quad (1)$$

Where $\varphi(t)$ and $\psi(t)$ are the basic scaling and mother wavelet function respectively.

$$a_j(k)=\int_{-\infty}^{\infty}2^{j/2}x(t)\phi(2^j t-k)dt \quad (2)$$

$$d_j(k)=\int_{-\infty}^{\infty}2^{j/2}x(t)\psi(2^j t-k)dt \quad (3)$$

Here $a_j(k)$ and $d_j(k)$ are the approximation and detail coefficients of wavelet respectively. In equation (1), the first summation is an approximation of $S(t)$ and second term sums more detail component [12].

D. NN Design

Depending on the input feature vectors and on the number of class of the classifier ANN is designed. So 5 input and 3 output for the corresponding input and output class respectively is used in the NN. While using 2 classes, 2 target vectors are used for the corresponding 2 class classification. In the Fig. 5 designed NN is shown with input, hidden and output layer. 5-45 neurons in the hidden layer are tried, but better results using 10-15 neurons. Trials number varies with respect to the subject; subject A completes 270 trials, 90 trials for each class. Subject B performs 174 trials, 58 for each class and subject C imagines 180 trials, 60 trials for each class.

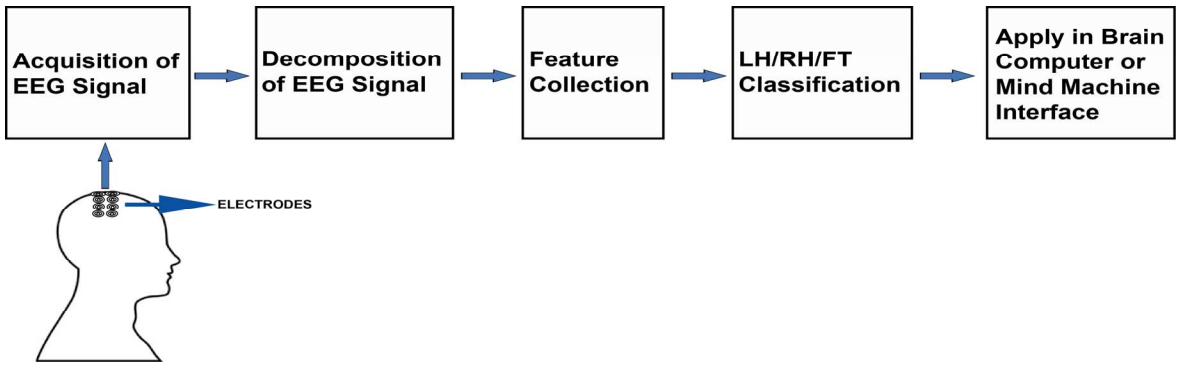


Fig. 3. Block diagram of the multiclass motor imagery classification for BCI application.

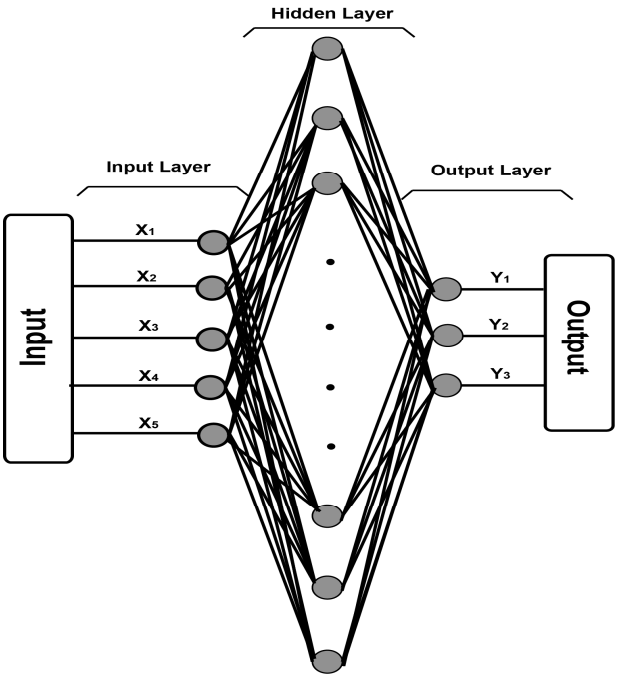


Fig. 5. NN design for 3 class MI classification.

III. Result

A. DWT Decomposition Result

All trials from all subjects are decomposed through DWT and 5 statiscal parameters of these EEG segments are collected to create feature vectors for NN. Five statistical parameters are mean, median, standard deviation, median absolute deviation and mean absolute deviation of the each EEG signal or trial. Fig. 6, Fig. 7, and Fig. 8. depicts the decomposition of LH, RH and FT trials of subject A respectively.

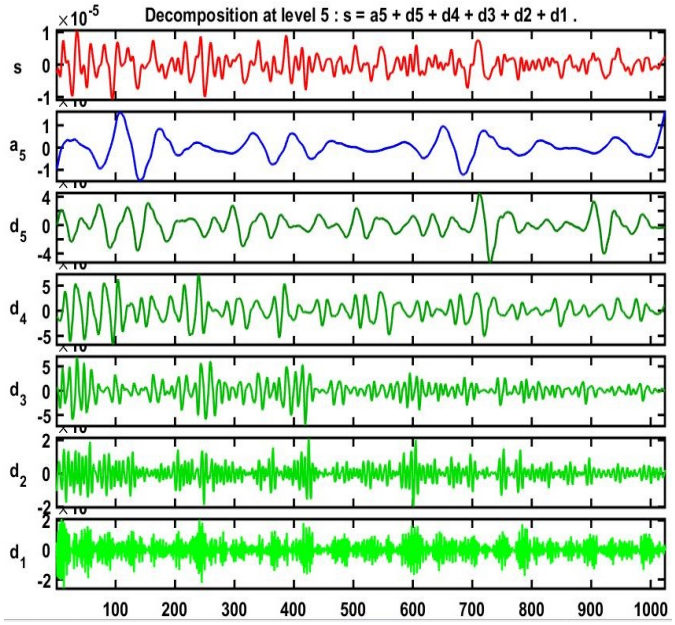


Fig. 6. Decomposition of LH imagery signal of subject A.

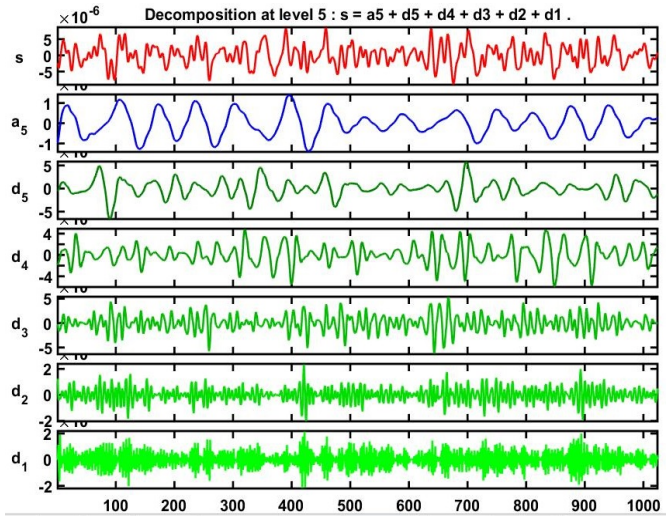


Fig. 7. Decomposition of RH imagery signal of subject A.

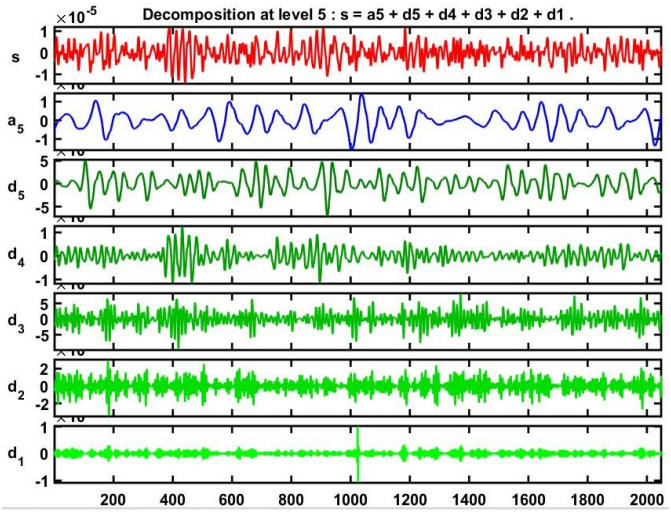


Fig. 8. Decomposition of FT imagery signal of subject A.

B. NN Classification Result

Feature vectors are given to NN subject wise and 70% of them are counted for training, 15% for testing and 15% are for validation. Resulting, such as for subject A; 188, 41 and 41 trials are for 3 class and 126, 27, 27 trials for 2 class are randomly distributed for training, testing and validation respectively. Confusion matrixes are listed in Table I and Table II, and classification accuracies are shown in Table III.

TABLE I. Confusion matrix for LH and RH classification

Subject A		Subject B		Subject C	
LH	**RH**	**LH**	**RH**	**LH**	**RH**
61 33.9%	20 11.1%	39 33.6%	21 18.1%	60 50.0%	0 0.0%
29 16.1%	70 38.9%	19 16.4%	37 31.9%	0 0.0%	60 50.0%

TABLE II. CONFUSION MATRIX FOR LH, RH AND FT CLASSIFICATION

Subject A			Subject B			Subject C		
LH	**RH**	**FT**	**LH**	**RH**	**FT**	**LH**	**RH**	**FT**
57 21.1%	18 6.7%	0 0.0%	37 21.3%	21 12.1%	21 12.1%	60 33.3%	0 0.0%	2 1.1%
32 11.9%	55 20.4%	2 0.7%	15 8.6%	29 16.7%	2 1.1%	0 0.0%	57 31.7%	2 1.1%
1 0.4%	17 6.3%	88 32.6%	6 3.4%	8 4.6%	35 20.1%	0 0.0%	3 1.7%	56 31.1%

TABLE III. CLASSIFICATION RESULTS FOR LH-RH AND LH-RH-FT CLASS

Subject	**LH-RH Classification**		**LH-RH-FT Classification**	
	Accurately classified	**Misclassified**	**Accurately classified**	**Misclassified**
A	72.8%	27.2%	74.1%	25.9%
B	65.5%	34.5%	58.0%	42.0%
C	100.0%	0.0%	96.1%	3.9%

Table I and Table II represents the confusion matrix for two and three class respectively for all subjects. In 2 classes classification trials are divided as class wise to 50% and for 3 classes to 33.3%. So in Table I for subject A, 33.9% (61) trials are correctly classified from 50% (90) trials for LH and 38.9% (70) trials are accurately discriminate from 50% (90) trials for RH. Best results come for subject C with 50% (60) trials out of 50% (60). In Table II for subject A, from 33.3% (90) trials, correctly separated 21.1% (57), 20.4% (55) and 32.6% (88) trials for LH, RH and FT respectively.

C. *Regression Plot*

Regression plot illustrates the connection between outputs and targets. In Fig. 9, Fig. 10 and Fig. 11, regression plot for individual subjects are depicted with their corresponding R values. R value indicates that how closely related outputs and targets. If the value is 1 proves they are closely related or if 0 then randomly related. In the figure inside the square box, the dash line represents the best fit position and solid line as in Fig. 9, the blow line shows the original fits of the data. Circles portray the distribution of the data. Subject C indicates the best regression value for all R=0.94, nearly 1 and subject B shows regression value for all R=0.45, close to 0.50. So subject C's data best fits and finally results maximum classification accuracy. Here regression plots are drawn for LF, RH and FT (3 class data) feature vectors for subject A, B and C.

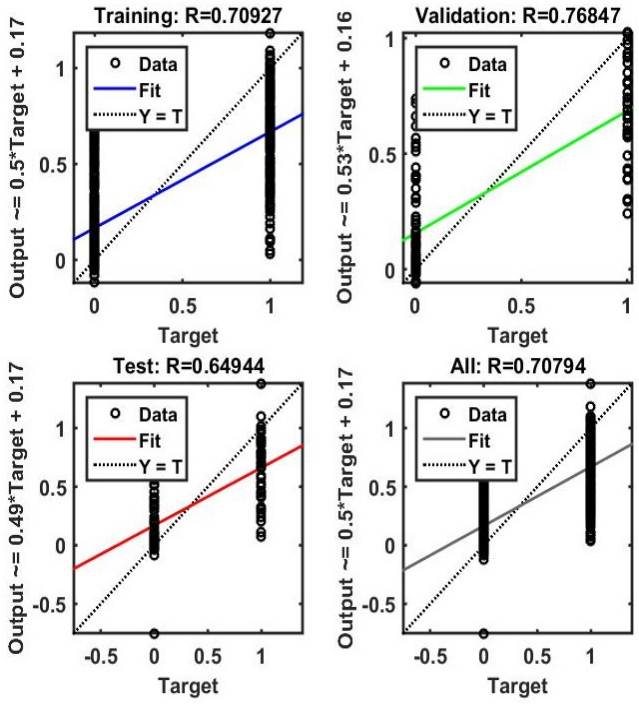


Fig. 9. Regression curve of a random trial of subject A.

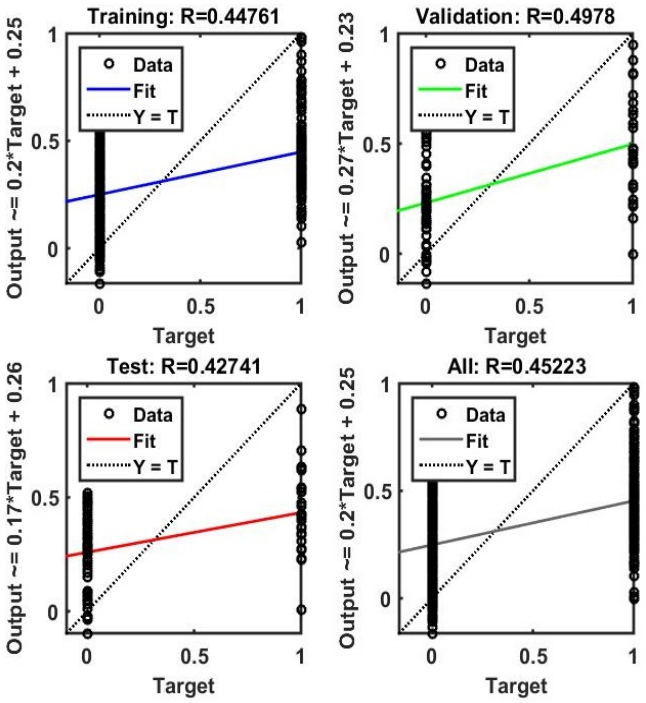


Fig. 10. Regression curve of a random trial of subject B.

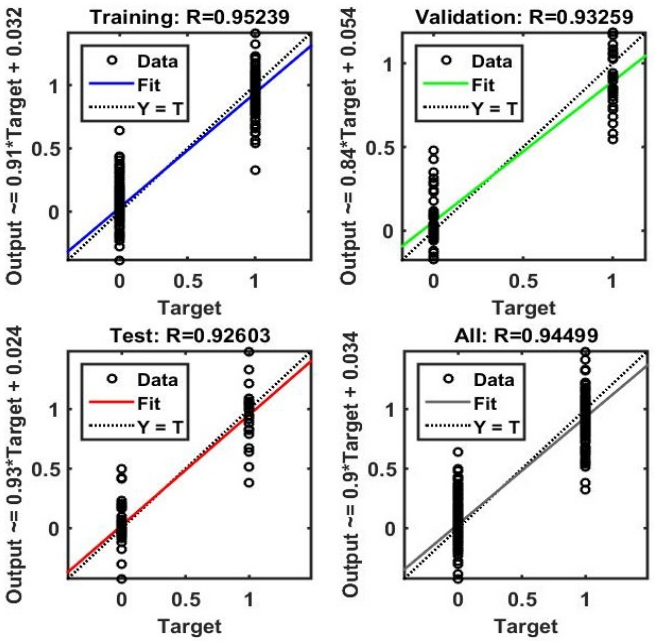


Fig. 11. Regression curve of a random trial of subject C.

IV. Discussion

Classification results prove that using DWT MATLAB toolbox for feature extraction and NN classifier it is possible to classify 3 class motor imagery movement (LH/RH/FT) utilizing statistical features of the EEG signal. Among three subjects classification results, dataset for subject C shows best seperation accuracy, for 2 class (LH/RH) and 3 class accuracy is 100% and 96.% respectively. But dataset of subject B shows average accuracy results for both 2 and 3 classes. Now after analyzing and comparing the regression plot and R value of the subject B's and C's dataset, plots shows data unfited for B's dataset and well fited for subject C's dataset. R values for B's is 0.50 (approximately) and for subject C's is above 0.90. So if the datasets of subject's B and A fits well as subject's C then classification results would be better for all subjects. However our proposed method can be used to classify 3 class motor imagery movement with precise accuracy as subject C for various BCI application. Accuracy rate varies ± 4% for all class except subject C's 2 class accuracy rate which shows 100 % always.

In Table IV below, we have presented a comparison table where we juxtapose our research work with other relevant studies and research works with related important information.

V. Conclusion

Left hand, right hand and feet classification is the mostly used in the BCI application to execute primary command using neural activity. To perform elementary task with perfectness it is very necessary to discriminate between those orders from cerebral cortex. This research work proves that our proposed method is able to LH/RH/FT motor imagery signal with acceptable accuracy. So this procedure can be applied to design an efficient BCI system to differentiate multiclass or three class neural activities.

TABLE IV. Comparision with relevent studies

Authors	Data Set	Electrodes	Class	Subjects	Methods	Performance
Sheng Ge [1]	Dataset IIIa, BCI competition	C3, Cz, C4, Fp1, Fpz, Fp2,	4	K3, K6 and L1	CSP and STFT, SVM	Fp2 - 73.4, 78.3, 75.2 C4 - 71.3, 88.1, 71.2
Aarathi Kumar [2]	BCI competition IV	---	4		CSP and SVM	Overall accuracy 91.11%.
Deng Wang [3]	Dataset IIa BCI competition IV	C3, Cz , C4	4	9 subjects	ICA and SVM	S1-71.43, S2-67.50, S3-64.29, S4-57.50, S5-87.86, S6-58.93, S7-85.71, S8-79.29, S9-73.93.
Weibo Yi [4]	Author's	C3, Cz, C4	Multiclass	10 subjects	CSP and SVM	Highest accuracy 84% Mean accuracy 70%.
Ridha Djemal [5]	IIa, BCI competition IV IVa, BCI competition III	C3, Cz, C4	3	9 subjects 4 subjects	FFT, AR and SVM, LDA and KNN Classifier	For IIa: 86.06%; For IVa: 93.3%
Zheng Yang Chin [6]	Dataset IIa BCI Competition IV	---	4	9 subjects	FBCSP	Mean Kappa values ranges 0.31-0.57
Bangyan Zhou [7]	Dataset IIa BCI Competition IV	---	3	9 subjects	ICA and SVM	Highest accuracy 86.96% and average accuracy 67.64%.
M Naeem [8]	---	C3, C4, Cz, CP1, CP2, CPz	4	8 subjects	ICA and CSP	Classification accuracies between 33% and 84%
Yan Wu [9]	Five datasets, BCI Competition and Author's	C3, Cz, C4	2	---	CSP and Biomimetic Pattern Recognition, LIBSVM	SVM – average 82.33 % LDA – average 80.43 % BPR – average 85.56 %
BAI Xiaoping [10]	Data sets from BCI Competition IV	C3, Cz, C4	4	9 subjects	Wavelet-CSP with ICA and SVM	Average kappa coefficient of 0.68
Yijun Wang [13]	---	C3, C4	3	5 subjects	CSP and LDA Classifier	Average online 79.48% and offline 85.00%
Proposed Method	Data sets provided by the Dr. Cichocki's Lab [11]	C3, C4	3	3 subjects	DWT and NN	2 class: Highest 100% and average: 79.43%. 3 class: Highest 96.1% and average: 76.06%.

Acknowledgment

This work was supported by Higher Education Quality Enhancement Project (HEQEP), UGC, Bangladesh; under Subproject "Postgraduate Research in BME", CP#3472, KUET, Bangladesh.

References

[1] S. Ge, R. Wang, and D. Yu, "Classification of Four-Class Motor Imagery Employing Single-Channel Electroencephalography," *PLoS One*. 9(6), Jun 20 2014.

[2] A. Kumar and N. P. V, "4-Class Motor Imagery Classification for Post Stroke rehabilitation using Brain-Computer Interface," *International Journal of Engineering Sciences & Research Technology,* 5(8), 649–654. August 15, 2016.

[3] D. Wang, D. Miao, and G. Blohm, "Multi-class motor imagery EEG decoding for brain-computer interfaces," *Frontiers in neuroscience*, 6:151, Oct 9 2012.

[4] W. Yi, S. Qiu, H. Qi, L. Zhang, B. Wan, and D. Ming, "EEG feature comparison and classification of simple and compound limb motor imagery," *Journal of NeuroEngineering and Rehabilitation*, 10:106, 12 October 2013.

[5] R. Djemal, A. G. Bazyed, K. Belwafi, S. Gannouni, and W. Kaaniche, "Three-Class EEG-Based Motor Imagery Classification Using Phase-Space Reconstruction Technique," *Brain Sciences*, 6(3), 23 August 2016.

[6] Z. Y. Chin, K.. K. Ang, C. Wang, C. Guan, and H. Zhang, "Multi-class Filter Bank Common Spatial Pattern for Four-Class Motor Imagery BCI," Proceeding of *31st Annual International Conference of the IEEE EMBS*, pp. 571-574, 2-6 September, 2009.

[7] B. Zhou, X. Wu, L. Zhang, Z. Lv, and X. Guo, "Robust Spatial Filters on Three-Class Motor Imagery EEG Data Using Independent Component Analysis," *Journal of Biosciences and Medicines*, 2, 43-49. April 2014.

[8] M. Naeem, C. Brunner, R. Leeb, B. Graimann, and G. Pfurtscheller, "Seperability of four-class motor imagery data using independent components analysis," *Journal of Neural Engineerin,*3(3):208-16, 27 June 2006.

[9] Y. Wu and Y. Ge, "A novel method for motor imagery EEG adaptive classification based biomimetic pattern recognition," *Neurocomputing*, vol. 116, pp. 280–290, 20 September 2013.

[10] X. Bai, X. Wang, S. Zheng, and M. Yu, "The Offline Feature Extraction of Four-class Motor Imagery EEG Based on ICA and Wavelet-CSP," Proceedings of the *33rd Chinese Control Conference*, 28-30 July 2014.

[11] http://www.bsp.brain.riken.jp/~qibin/homepage/Datasets_files/EEG_BCI_MI_AllSub.zip

[12] D. Gajic, Z. Djurovic, S. Di Gennaro, and F. Gustafsson, "Classification of EEG signals for detection of epileptic seizures based on wavelets and statistical pattern recognition," *Biomedical Engineering: Applications, Basis and Communications*, vol. 26, iss. 02, 21 February 2014.

[13] Y. Wang, B. Hong, X. Gao, and S. Gao, "Implementation of a Brain-Computer Interface Based on Three States of Motor Imagery," Proceedings of the *29th Annual International Conference of the IEEE EMBS*, 23-26 August, 2007.

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

Multiple Sensors Based Fire Extinguisher Robot Based on DTMF, Bluetooth and GSM Technology with Multiple Mode of Operation

Humayun Rashid, Iftekhar Uddin Ahmed, Aasim Ullah, Fahim Newaz,
Mohammad Sijanur Rahaman Robin, S M Taslim Reza

Electrical and Electronic Engineering (EEE), International Islamic University Chittagong (IIUC), Chittagong-4314, Bangladesh
Email: raahat.rashid09@gmail.com

***Abstract*—In this paper, design and development of a multiple sensors based fire extinguisher robot is proposed and implementation is demonstrated with a brief discussion of construction and operation. The developed fire extinguisher robot can be operated in multiple modes using the DTMF and Bluetooth remote control as well as GSM and GPS technology. Basically, three different sensors of flame sensor, temperature sensor, and smoke sensor have been used to ensure proper detection of fire. The robot can be controlled using both DTMF remote control and Android smartphone and can be operated in three different modes. The first mode allows full autonomous operation of the robot which can be activated by the user or by the robot itself based on the situation. The second mode is a line following mode where robot follows a black drawn line to detect fire and the third mode is complete manual operation using remote control.**

Keywords—Fire Extinguisher Robot, Arduino Mega, DTMF, GSM, Bluetooth

I. INTRODUCTION

Detection of fire alongside with extinguishment is a detrimental work that risks the health as well as the existence of a flame extinguisher person in the hazard but through utilizing a robot to execute fire detection and extinguishing in a fire-prone area, loss of lives and undesired incidents can be avoided in a considerable number [1]. The day by day progress of advanced technology has made it feasible to develop different types of household and industrial robot and automation. The definition of the robot states that a system with the capability of executing human tasks or behaving in a human-like manner is regarded as robot [2]. Continuous research and developments are going on for obtaining a reliable and effective method which can be enforced to develop a firefighting robot to detect and extinguish the fire to lessen the risk of injury to victims.

A Firefighting robot is an independent ground vehicle [3] which should have two main functions, ability to detect fire and the ability to extinguish the fire. A small fire extinguisher system along with various sensors are attached to a fire extinguisher robot for proper performance. The appropriate use of the robot will make sure that the fire combating as well as recovery exercises might be maintained without having place flame fighters life at danger through utilizing making use of automation technological innovation as an alternate choice of human [2].

The design approach and implementation of a fire extinguisher robot is presented in this paper where three types of sensors of flame sensor, smoke sensor, and temperature sensor have been used for fire identification to make the detection process more reliable. Multiple control system has been implemented to make the robot more efficient to extinguish the fire. After detecting fire using three sensors, the robot is programmed to send a text message with location coordinates to its user's mobile using GSM technology or notification to its Bluetooth paired android phone to get command for a specific operation mode. The user can set any of the three modes of operation using DTMF remote control or Android Smartphone. After confirmation of mode selection, the robot will start extinguishing the fire using water from water tank utilizing a DC pump motor as well as a servo motor. The pipe of the water tank is flexible and attached to a servo motor which allows the robot to spray the water from 10^o to 170^o.The proposed model is divided into two major section; design and implementation as well as result analysis and discussion which have been discussed briefly.

II. LITERATURE REVIEW AND PREVIOUS WORK

A brief study of previous works has been conducted for understanding the limitations of the previously developed robot to implement new features to the proposed model. A paper on developing of a firefighter robot has been presented in [4] where the robot would search for fire by following a navigation system of line tracker and a camera was used for getting the exact position of the target place. A custom arm was designed using servo motors to implement the extinguishing device of the robot. Another similar automatic fire extinguisher robot has been designed in [5] using fuzzy logic along with thermostat sensor to detect fire as well as Wi-Fi enabled camera to locate fire accident. The robot was designed wirelessly in order to operate it from a remote location. Another work on firefighting robot has been presented in [6] using Arduino and flame sensor to detect the fire along with a navigator to help the robot to go to the location autonomously. This robot was designed mainly as a path guidance where it also worked as a fire extinguisher in the emergency situation. Another work is presented in [7] where the smoke detector was used to locate the source of the

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

fire employing two PIC16F877 microcontrollers; one of which was used for delivering a message to the owner and to alert the robot about a fire as well as another one was used to navigate the robot to the fire source.

III. Overview Of Proposed System

The functional block diagram of the proposed robot is illustrated in figure 1 and figure 2 illustrates the mechanical structure. The robot will utilize three sensors to detect temperature, smoke, and flame to implement the fire detection unit whereas one or maximum two sensors were used in the previous developed of fire extinguisher robot. When fire accidents occur, the temperature rises suddenly with smoke and flames. These three sensors will continuously monitor all-around of its range to detect any kind sudden rise of temperature, smoke or flame. If any of the sensors of gas or temperature can detect any abnormal parameters, data will be delivered instantly to the microcontroller.

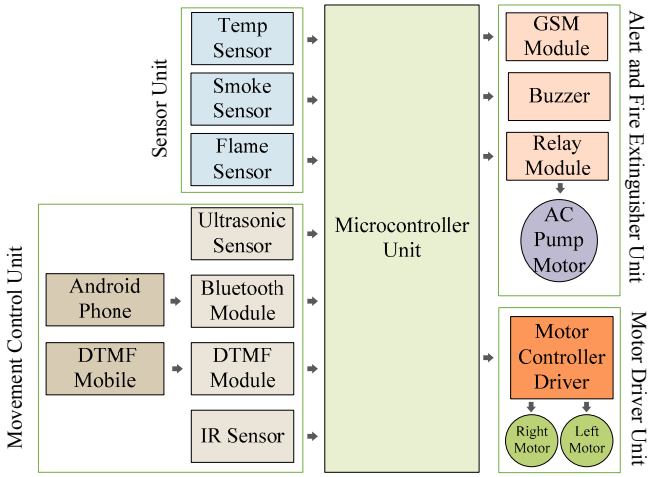


Fig. 1. Functional Block diagram of the proposed fire extinguisher robot.

Whenever the microcontroller can detect that a fire accident is happening it will instantly contact with its user. There will be two controlling methods, the first method is using an android phone and the second method is to use a DTMF remote control which allows controlling the robot from anywhere in the world using GSM technology. If the smartphone is paired for controlling the robot, the system will notify the user by sending a notification to the paired Bluetooth android phone. An app will be developed to receive notification and control the robot. If a DTMF system is utilized, the robot will send a message to the user's cell phone using GSM and GPS module to notify about fire accident with the coordinates of the accident's location and wait for user's action. The robot will also be able to send a text message to local fire service or security center with coordinates of the exact location. The system will be designed in a way that if the flame sensor can detect any kind of flame, the fire extinguishing unit will be activated instantly. But if temperature sensor and gas sensor can detect a sudden change of parameters, the user will be able to set three modes of operations to control the robot using both Android apps and DTMF mobile phone. The modes are: a) Autonomous mode b) Line following Mode c) Manual Control.

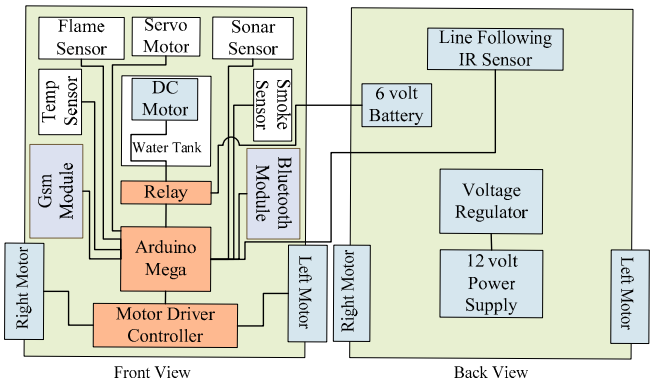


Fig. 2. Mechanical structure of the proposed fire extinguisher robot.

If there is no response from the user, an alert notification will be sent again to the robot from the microcontroller so that user can understand that no step is taken in the accident place. After sending the final notification the robot will automatically start its operation by selecting autonomous mode which will be set as the default mode of operation.

If the user gives instruction to activate any specific mode, the robot will follow the command by executing user's instruction. For line following mode, it will follow a black line to search the flame. The manual mode will be operated by a user and the robot will move according to user's instruction. The manual mode can be controlled both using DTMF mobile phone and android app. In all three modes, the fire extinguishing system will be automatically activate after the flame detection. The robot will stop its movement and start the DC motor pump with a flexible water pipe attached to a servo with the ability to rotate 10^{o} to 170^{o} to spray water to the target place efficiently. A water tank will be attached to the extinguishing unit to store water and after the competition of extinguishing operation, the robot will stop DC motor and go back to its previous position.

IV. Design And Implementation

A. Microcontroller Unit and Motor Driver Controller

The project is implemented using Atmega microcontroller platform known as Arduino Mega. It consists of 54 digital as well as 16 analog input/output pins with 4 UARTs and16 MHz crystal oscillator [10]. The Mega 2560 board is specially chosen because of its compatibility with most sensors and modules. The employed motor driver module circuit for firefighting robot is known as L293D motor driver controller consists of 4 inputs and 4 outputs to control two DC motors. One of the main facility of L293D is to provide up to 600mA current at voltages variation from 4.5 V to 36V [9]. The direction of rotation of DC motor in forward and reverse can be controlled through combining different input logics. Another strong reason to choose this specific IC for movement control is the ability to control the speed of two motors using PWM from microcontroller which is required to control the robot for different modes. The basic circuit arrangement of L293D is shown in figure 3. The circuit diagram in figure 5 shows that the four inputs are connected to the digital pin of D9, D10, D11 and D12.

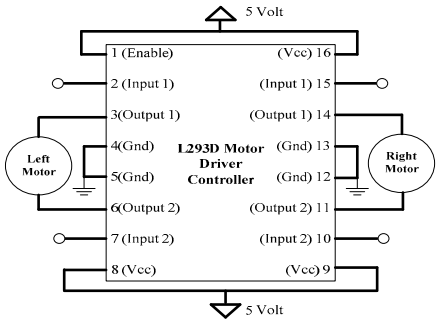


Fig. 3. Circuit diagram of L293D based motor driver controller module.

B. Sensors

1) Temperature Detection: LM35 is a highly accurate inherent calibrated temperature sensor to detect the abrupt rise of temperature with attributes of low linear output impedance as well as less heating issue and capable of operating from a voltage range of 4 volt to 30 Volt with the drain current lower than 60 µA [11]. It is comprised of three pins of Vcc, Out & Ground. Out has been linked with the analog pin A0 of Arduino Mega as shown in the circuit diagram of figure 5.

2) Smoke Detection: A smoke sensor made of a micro aluminum oxide ceramic tube known as MQ9 is utilized to detect smoke. The sensor combined with a sensitive layer of Tin Dioxide with high sensitiveness to carbon monoxide for detecting smoke or gas in presence of carbon monoxide (CO) [12]. MQ9 includes of measuring electrode and necessary work conditions of sensitive components are provided by a heater [12]. Four pins have been used to get signals and other two pins have been utilized to provide heating current. The analog output pin of the gas sensor has been connected to the analog pin A1 of Arduino Mega.

3) Flame Detection: The chosen flame sensor has properties of sensitiveness to radiation and flame spectrum along with the capability of detection of origin of the ordinary light source. The flame sensor has been attached to the analog pin of A2 of Arduino Mega.

C. Development of DTMF , Bluetooth ,GSM and GPS

1) DTMF Remote Control: MT8870 is the best choice for developing DTMF remote control. A mobile has been attached to the system to receives a signal via DTMF receiver. Figure 4 is showing the circuit diagram of the MT8870 DTMF module.

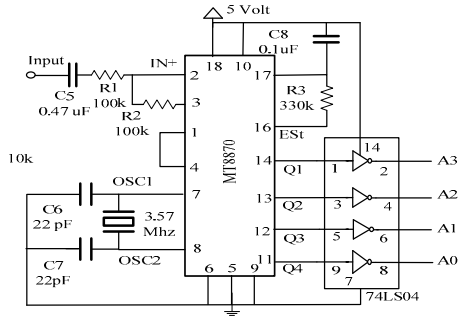


Fig. 4. Circuit diagram of MT8870 based DTMF Module.

After making a phone call to the DTMF receiver connected phone, symbol pulse is provided to the system when any button is pressed. The outputs of DTMF module are Q1, Q2, Q3 and Q4 which has been interfaced with the digital pin of Arduino Mega D2, D3, D4, and D5.

2) Bluetooth Remote Control: The Bluetooth module HC05 has been employed with the system with digital pin D0 and D1 of Arduino Mega using main function pins of RX and TX for receiving and transmitting data. Radio waves are utilized by Bluetooth technology to interconnect with other peripheral devices for data exchanging.

3) GSM and GPS Module: The 808 GSM GPS module has been connected with D14 and D15 of Arduino Mega and provides location coordinates using GPS. The microcontroller is able to send SMS using GSM to user and fire extinguisher unit. The sim808 module incorporates Quad-Band network as well as brings together GPS technology for satellite routing [14]. It has ultra-low power absorption as well as 22 tracking and 66 receiver channels [14].Indoor localization by A-GPS as well as control capability through AT instruction by using UART is additionally obtainable.

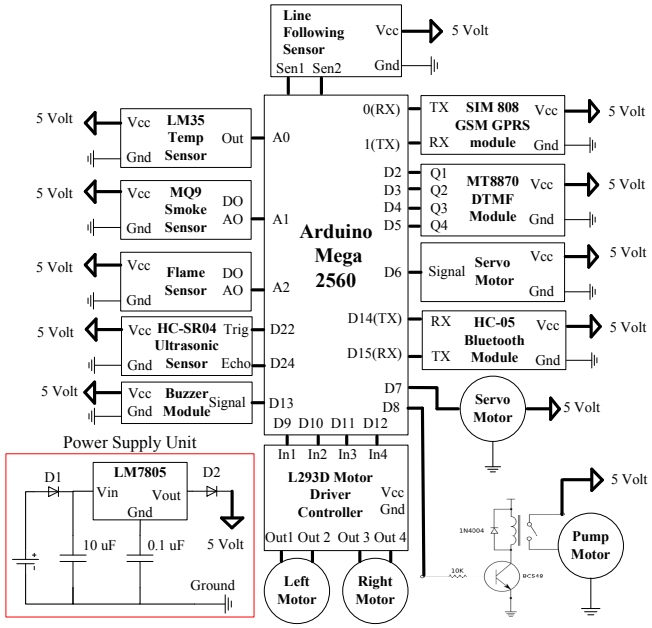


Fig. 5. Circuit diagram of proposed fire extinguisher robot.

D. Modes of operations

1) Development of Line Following method: Three modes of control system have been developed using line following, autonomous and manual control. The line following sensor has been implemented using two sets of IR LED with photodiode arrangement for each motor. To identify a black line from a white surface, the reflected IR light from the white surface falls on the photodiode that can be detected by microcontroller due to voltage variation. A Comparator known as LM358 is used for comparing the output voltage with a reference voltage for ensuring proper operation and adjusting the sensitivity.

2) Development of Autonomous Control: An ultrasonic sensor has been interfaced with Arduino Mega for autonomous

controlling of the robot. The ultrasonic sensor is noted as HC-SR04 which comprises of 4 pins of Vcc, ground, Trig as well as Echo and 2cm to 400cm non-contact distance can be measured to locate obstacle in front of the robot using the sensor [13]. Trig and echo have been attached to the digital pin of D22 and D24 of Arduino mega. Signal is sent by Trig pin and the detection of the signal is determined by Echo.

3) Development of Manual Control: Manual control requires no extra hardware interfacing with the microcontroller. It can be fully controlled using the DTMF phone or Android app after activating the manual mode by user.

E. Fire Extinguisher System with Servo Motor and pump

The fire extinguisher system has been designed using a relay module along with a water tank, DC pump motor and a servo motor. Relay module has been used with a transistor BC 547 which has been linked to the digital pin of D8 of the microcontroller to run DC pump motor. When fire extinguisher system will be activated the relay will switch on the DC pump motor to pump the water out of the tank and spray on the fire through a pipe which will be rotating from 10º -170º and in reverse direction continuously to ensure proper fire extinguishing. The overall hardware arrangement of the fire extinguisher robot is shown in figure 6.

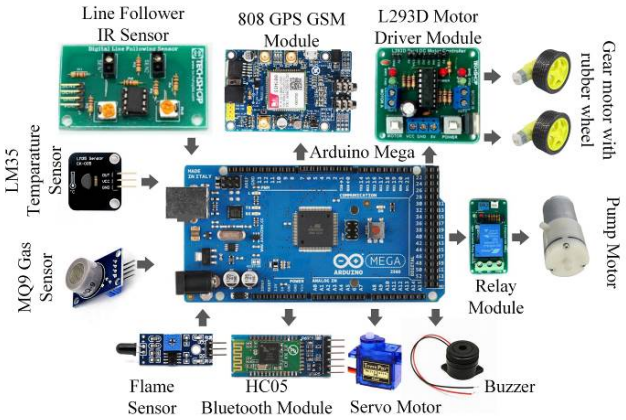


Fig. 6. Hardware arrangement of the fire extinguisher robot.

F. Programming

The programming for the microcontroller of the robot in presented through a flowchart in figure 7. After installing the robot the sensor will initiate hardware at first and check for the selected control mode of either DTMF or Android. An android app will be developed for controlling the robot which will utilize Bluetooth to pair with the robot. Figure 8 is showing the proposed user interface of the android app. If a fire incident happens in any place, the sensors will detect and provide notification to the microcontroller. The sensors will provide analog values to the microcontroller, so after getting the value microcontroller will convert the values into digital format using A/D converter. Then converted value will be compared to the predetermined values of the sensors to detect fire. Once fire detection is confirmed, the microcontroller will inform the control unit instantly by sending a text about a fire incident.

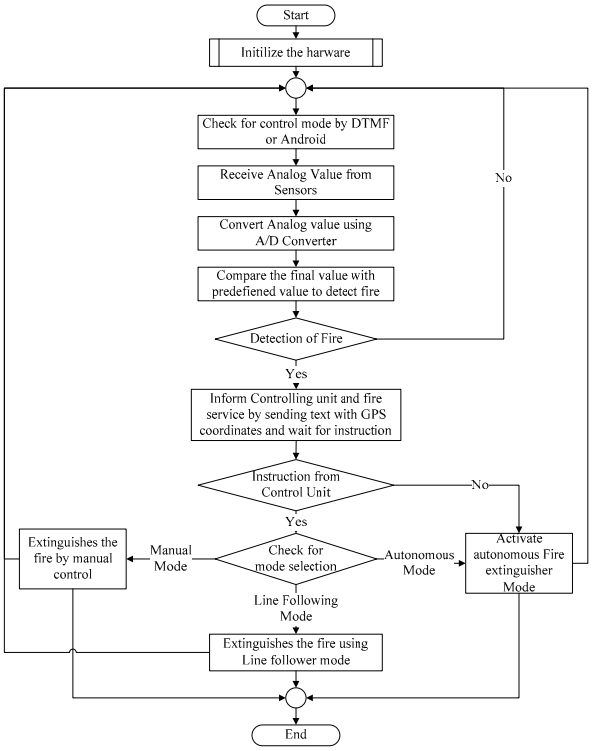


Fig. 7. Flow chart of Programming for fire extingusher robot.

An instruction will be delivered with the location coordinates to fire service to take immediate action from the control unit. The control unit will check mode selection by the user. There are three modes which can be selected by the user. The modes are manual control by DTMF, extinguish the fire by following the line tracker or fully autonomous mode.

If no action is taken from the user within a certain time period, the control unit will select the autonomous mode. Then the robot will be controlled fully automatically and will go for extinguishing the fire according to the information provided by the control unit. Automatic control mode can also be selected by the user. If manual mode is selected the user will have the ability to control the robot manually using DTMF remote or Android phone. This process will be continued and the sensors will be ready for scanning of fire source continuously for all the time.

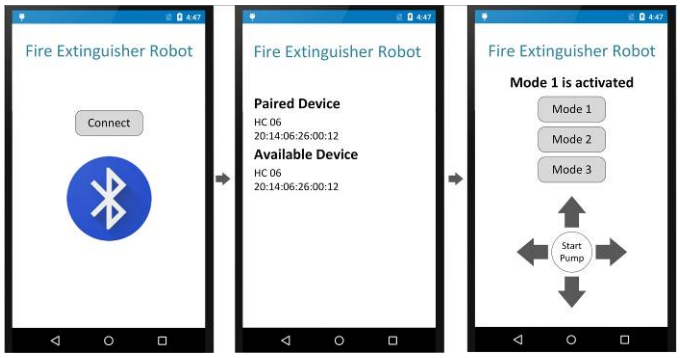


Fig. 8. Proposed User Interface of android app to control the robot.

V. Result Analysis and Discussion

The developed robot has been tested to evaluate performance analysis as well as to demonstrate the ability to extinguish the fire. The robot has shown quite good performance to detect and extinguish the fire using different sensors and developed fire extinguishing unit. Figure 9 and figure 10 illustrate the front view and the top view of the robot.

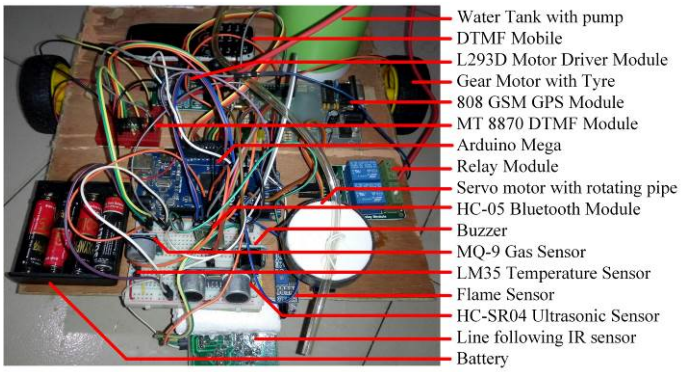


Fig. 9. Front view of the developed fire extinguisher robot.

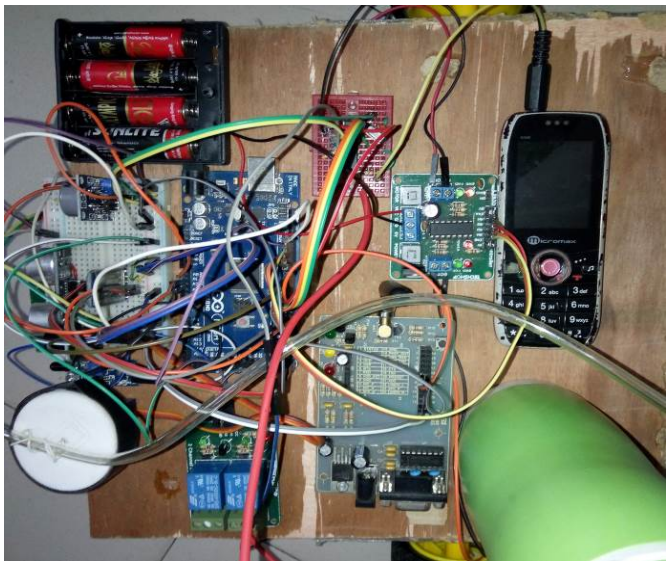


Fig. 10. Top view of the developed fire extinguisher robot.

A. Fire Detection using Multiple Sensor

Temperature sensor LM35 is capable of sensing temperature between +2 ºC and +150ºc [11]. The programming of microcontroller has determined a range of parameters for normal room temperature and abnormal room temperature during fire accident. The temperature has been raised using a lighter for testing purpose and it has been noticed that the sensor was able to detect the sudden rise of the temperature. Since the output voltage is the function of the temperature, it was able to get Celsius degree unit by analysis the output voltage of the sensor.

The smoke sensor has the ability of detection smoke using the method of making temperature cycle high and low. Smoke generally contains carbon and during the testing, smoke has been detected by measurement of output signal within one or two complete heating period. SnO_2 is the sensitive substance of smoke sensor with characteristics of lower conductivity in clean air and sensor's conductivity was more significant along with the rising of gas concentration [12].

Flame sensor has shown the ability to detect the source of fire flame within the range of a wavelength 760nm-1100 nm. It has been observed that the detection distance of flame has shown better performance within the range of to 25cm to 90 cm whereas standard range is defined as 20 cm to 100 cm.

B. Modes of operation

The program of microcontroller has been developed in such a way if any two sensors could sense abnormal parameters, the robot would send an SMS to its user or a notification to the paired android phone. The user can activate the whole system by using DTMF or android app. Band pass filter along with a digital decoder function for DTMF are capable of to decode DTMF tone [15]. When the control keys of the mobile phone has been pushed, a DTMF signal was transmitted to the microcontroller. The keypad buttons of mobile phone generated two high and low tones of different frequencies. Applying the decoder from the DTMF signal equivalent binary code has been attained during testing.

The robot can be activated by pressing number 1 key of the mobile. The user has the ability to select three modes of operation by pressing 2, 3 and 4. Number 2 key has been assigned to select the line the following mode. Number 3 has been used to set autonomous mode and number 4 has been able to select manual control mode using the DTMF phone.

1) Mode A (Line the following method): In the case of the line the following method, the robot has followed the predetermined path defined by a black line for detection of flame. The robot has searched to detect flame and whenever any flame has been detected by the robot, it has stopped the movement to activate the fire extinguisher unit. The fire extinguisher unit has sprayed the water using a DC motor from the water tank and servo motor has made ensured that the water could reach every possible spot of fire by rotating the water flowing pipe 10^{o} to 170^{o}. Table I shows the different position of the robot during line the following mode.

TABLE I. Position of Robot during Line Following Mode

Sensor 1	Sensor 2	Position of Robot
0	0	Robot' is out of track
0	1	Robot's left is out of track
1	0	Robot's right is out of track
1	1	The robot is on completely on track.

2) Mode B (Autonomous Mode): When the autonomous mode has been selected, all the operation of the robot has been performed by the robot autonomously. The robot has employed ultrasonic sensor for obstacle detection during autonomous mode.

3) Mode C (Manual Mode): The user has been able to control the robot by selecting manual mode by using DTMF remote or android app. 5, 6, 7 and 8 buttons of the DTMF mobile have been used to move the robot manually

respectively right, left, forward and backward position in any place. Once the incident place has been detected, the DC motor pump has started as soon as the detection of fire. The robot has stopped spraying water as well as the pump when the sensors have stopped sending abnormal parameters of fire detection.Table II represents the different DTMF buttons, output decoder, and action.

TABLE II. DTMF BUTTONS, OUTPUT DECODER, AND ACTION

Button	Output of Decoder	Action
1	0x01 (00000001)	Turn On the robot
2	0x02 (00000010)	Set mode A (Autonomous mode)
3	0x03 (00000011)	Set mode B (Line Following Mode)
4	0x04 (00000100)	Set Mode C (Manual Mode)
5	0x05 (00000101)	Forward
6	0x06 (00000110)	Backward
7	0x07 (00000111)	Left
8	0x08 (00001000)	Right
9	0x09 (00001001)	Spray Water

V. CONCLUSION

Autonomous fire extinguisher robot is just not an idea now, rather it is demand in this era of automation engineering. An autonomous robot can work for extinguishing fire simultaneously with the human without requiring any kind of command and also can work individually where there is life risk of a human. Even a robot can be more efficient and quick than human. Besides it takes some time to inform fire service team and then their arrival to the spot. But if there is an autonomous fire extinguisher robot in every factory or shop or house, the loss of life and assets can be reduced significantly as the robot will go for action as soon as it detects any kind of symbol of fire.

The robot has been designed with three types of sensors such as a flame sensor, smoke sensor, and temperature sensor whereas one or maximum two types of sensors was used to implement such kind of robot in the previous development[4]-[7]. As a consequence, the detecting of fire will be more precise and quick. Moreover due to the use of three kind of sensors, the robot will not stop spraying water until all three sensors respond negatively so that the fire cannot break out again some time later as it is a common nature of fire that it can turn on anytime if a flammable source is available in the incident spot. Additionally, use of multiple control ability will make the robot more reliable. However, the robot can be modified while implementing the manufacturing process in future.

REFERENCES

[1] T. Khoon, P. Sebastian and A. Saman, "Autonomous Fire Fighting Mobile Platform", Procedia Engineering, vol. 41, pp. 1145-1153, 2012.

[2] U. Prasanna and M. Prasad, "Automatic Fire Sensing and Extinguishing Robot Embedded With GSM Modem", International Journal of Engineering and Advanced Technology, vol. 2, no. 4, 2013.

[3] J. Setiawan, M. Subchan and A. Budiyono, "Virtual Reality Simulation of Fire Fighting Robot Dynamic and Motion", in International conference on intelligent unmanned systems., Indonesia, pp. 57-63, 2016.

[4] W. Dubel, H. Gongora, K. Bechtold, and D. Diaz, "An autonomous firefighting robot", Course EEL 566C Project Report,University of Florida, Dept. Electrical and Computer Engineering, 2003.

[5] A. Dhumatkar, S. Bhiogade, S. Rajpal, D. Renge and P. Kale, "Automatic Fire Fighting Robot", International Journal of Recent Research in Mathematics Computer Science and Information Technology, vol. 2, no. 1, pp. 42-46, 2015.

[6] S. Mathew, G. Sushanth, K. Vishnu, V. Nair and G. Kumar, "Fabrication of Fire Fighting Robot", International Journal of Innovation and Scientific Research, vol. 22, no. 2, pp. 375-383, 2016.

[7] J. Undug, M. Arabiran, J. Frades, J. Mazo and M. Teogangco, "Fire Locator, Detector and Extinguisher Robot with SMS Capability", 2015 International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), 2015.

[8] S. Dearie, K. Fisher, B. Rajala and S. Wasson, "Design and construction of a fully autonomous fire fighting robot", Proceedings: Electrical Insulation Conference and Electrical Manufacturing and Coil Winding Conference (Cat. No.01CH37264), 2001.

[9] I. Verner and D. Ahlgren, "Education Design Experiments in Robotics", 2006 World Automation Congress, 2006.

[10] "Arduino – Arduino Board Mega2560", Arduino.cc [Online]. Available: www.arduino.cc/en/Main/ArduinoBoardMega2560.

[11] "LM35 Precision Centigrade Temperature Sensors", www.ti.com, 2016. [Online]. Available: www.ti.com/lit/ds/symlink/lm35.pdf

[12] "Carbon Monoxide & Flammable Gas Sensor MQ-9", www.pololu.com, 2016. [Online]. Available:www.pololu.com/file/0J314/MQ9.pdf.

[13] "Ultrasonic Ranging Module HC - SR04",www.micropik.com, 2016. [Online]. Available: www.micropik.com/PDF/HCSR04.pdf.

[14] "Sim 808 Hardware Design V1.00", www.adafruit.com, 2014. [Online]. Available:cdn-shop.adafruit.com/datasheets/SIM808_Hardware+ Design _V1.00.pdf.

[15] Yun Chan Cho and Jae Wook, “Remote robot control system based on DTMF of mobile phone”,6th IEEE International Conference on Industrial Informatics, 2008 (INDIN 2008), pp. 1441 – 1446, 2008.

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

Stroke Matching Based Approach to Recognize Bangla Offline Conjunct Characters

Shamima Nasrin, Tasneem Zerin, Mohammad Mahadi Hassan
Department of Computer Science & Engineering, International Islamic University Chittagong, Chittagong, Bangladesh
Email: shimu.cse69@gmail.com, tasneem.zerin.chy@gmail.com, mahadi_cse@yahoo.com

***Abstract*— The main target of the approach presented in this paper is to recognize offline Bangla conjunct characters by Stroke Matching Based Approach. At first, the approach simplifies the word into strokes. Later it matches these strokes with the one that is stored in the dataset. The approach reduces the noise and then works over the strokes. Up to now a lot of recognizers are available to recognize BOC (Bangla Offline Characters). But a recognizer to recognize Bangla Conjunct letters is rare. Here in this research work Stroke Matching Based Approach modifies the existing strokes of Bangla Conjunct Letter to identify it. The recognition of the characters would be done by using neural network system with the matching of strokes of the characters.**

Keywords— Offline characrters; Bangla conjunct characters; Stroke Matching; Artificial Neural Network(ANN);

I. INTRODUCTION

In the natural language rank list Bangla owned its position as number seven as a spoken language since most of the native people of the world communicate using this language. It is originated from Indo-European [1] language. It contains 50 alphabets including 11 vowels that is Shoroborno and 39 consonants that is Banjonborno. And 253 conjunct characters composed of 2, 3 or 4 consonants. Among them 200 characters are formed of 2, 51 characters are composed of 3 and 2 characters are composed of 4 consonants [2]. Bangla language also contains 10 vowel modifiers that is known as Kar, 7 consonant modifier that is known as Fala. Till now Bangla handwritten Pattern Recognition, Character Recognition, Text Summarization etc are not quite rich. And the important part is, in most research, recognition of conjunct letters of Bangla language has been skipped due to its complex patterns. As a result, in many cases, conjunct letters are not clearly recognized. So along with the 50 letters of Bangla, it is also important to recognize them so that they can be recognized easily are stored so that it might turn into a great help for documentation of soft copies.

Most of the offline character recognition are performed using Optical Character Recognition Methods. But the process and way of working is way too hard and lengthy. Though the target of every method is to find a better solution in the character recognition, but every time came with a challenge. Therefore, for the character recognition, the proper segmentation of characters from the text is needed which is difficult because of losing essential data. Bangla Characters also face the same segmentation problem similar to ANN based approaches. So, to identify conjunct letters of Bangla Language Stoke Matching Base Pattern Recognition is established.

This thesis paper work is based on the recognition of Bangla Offline Character Recognition by matching the strokes from an image. The recognition of Bangla characters and text works have been done earlier using different technologies like Optical Character Recognition (OCR), Neural Network , SVM etc. The new approach is Stroke Matching based approach to recognize Bangla Offline Characters for recognizing them by machines. The chapters of this paper represent a go through over the topic in details along with the description of work.

The rest of the paper is organized as follows. Section II is the summarization of some related works. Section III gives the description of methodology of our system. Section IV describes feature extraction of our proposed system. Section V gives the description of experimental analysis and Section VI concludes the paper.

II. LITERRATURE REVIEW

A great amount of work has been done with Offline Character Recognition. The relevant works are briefly discussed below.

An approach proposed in [3] is on the basis of view-based algorithm, where only upper and lower views are analyzed for characters. The character is divided vertically in identical segments depending on the number of points need to be obtained. The calculation of y coordinates decides three element characteristic vector which describes the given character. The found two vectors combining with their two values describes the aspect ratio of the picture which transforms into 8 element vector representing the final character. The accuracy of the proposed scheme was 74.166% and accuracy of recognition is 75%.

Another research work which studies with the Stroke Segmentation and Recognition from Bangla Online

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

Handwritten Text, published in International Conference on Frontiers in Handwriting Recognition, 2012. Here the stroke (collection of points from pen down to pen up) of writing a character is analyzed and Support Vector Machine (SVM) classifier is used for recognizing segmented characters.[3] Correct Segmentation rate 97.89 % and Stroke recognition rate 97.68% [4].

This research work in [5] is on recognition of Bangla Basic character and digits by using convex hull method. They used Graham Scan algorithm for computing the convex hull of numeric pattern. In this research work they used simple 125 features based on different bays and lake attributes of convex hull using MLP based classifier. Recognition rate of this system is 76.86% from 10000 sample of handwritten Bangla text and 99.45% from 12000sample of Bangla numerals.

The proposed approach in this paper is an offline based character recognition system using multilayer feed forward neural networking. A new method "Diagonal based feature extraction" is introduced in this paper [6]. Every character Image of size 90x60 pixels is divided into 54 equal zone which is of 10x10 pixel in size. The extracted zone contain 19 diagonal lines and 19 sub features that obtained from each zone. The recognition system has been implemented using neural network training tool.

This research work proposed a recognition system of printed Bangla text. In this approach the text image is segmented into individual characters and then refined and converted to a designated M x N matrix [7]. Heuristic method is used to obtain better accuracy. Characters are separated into reasonable number of groups by using Heuristic method. The accuracy rate is nearly 75% for this system.

III. STROKE MATCHING BASED APPROACH

The main target of the approach represented in this paper is to recognize off-line Bangla conjunct characters. At first, the approach simplifies the letter into set of strokes. Then it matches these strokes with the one that is stored in the dataset. The approach reduces the noise and then works over the strokes. Up to now a lot of recognizers are available to recognize BOC (Bangla Offline Characters). But a recognizer to recognize Bangla Conjunct letters is rare. SM (Syntactic Method) is one of those popular methods that used to recognize BHC. Here in this research work Stroke Matching Based Approach modifies the existing strokes of SM and uses it to recognize Bangla Conjunct Letters. The proposed system consists of nine steps. The Flowchart of the system is given in Figure. 1.

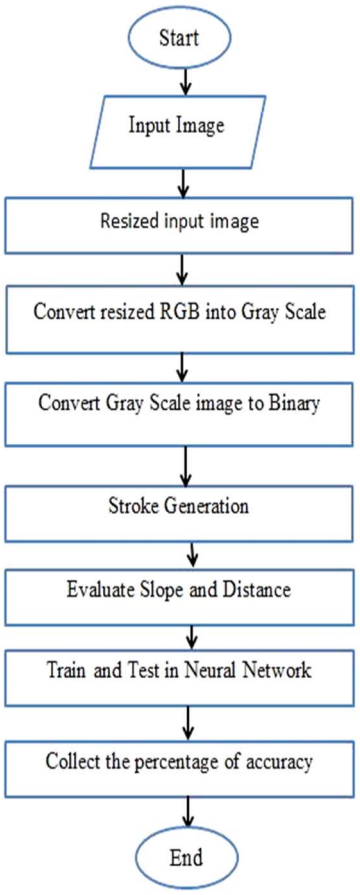


Figure 1. Flowchart of the Proposed System.

A. Input Characters

The input characters are inserted as image where the patterns are recognized to identify the character by using Stroke Matching Based Text Recognition which mainly exploits Modified Syntactic Method where it provides a capability for describing a large set of complex patterns by employing small set of simple pattern in strokes.

B. Stroke Generation

After forming the digital 0/1 matrix, the most important step is to contour tracing which means to convert the multi-pixel line into the single pixel line. That means it will filter the line to get a single formed line. The next step is to generate the straight line that is generating the strokes. The Stroke Matching Character Recognizer (SMCR) proposes a to develop an algorithm where offline characters and stroke generation works simultaneously. The pseudo code of the process is shown in *Algorithm 1*.

Algorithm 1: Pseudo code for stroke generation:
***Step 1*:** Convert the image RGB to Binary.
***step 2*:** Scaling the image into 120x120 pixel.
***Step 3*:** Apply thinning algorithm.
***Step 4*:** Search '1' in the digital matrix, and set the starting point at this position.
Step 5: Detect the horizontal and vertical line of the image
Step 6: Find the start point and end point of these lines to count them as stroke.
Step 7: Remove the horizontal and vertical lines and go to *Step 6* to evaluate the other strokes.
Step 8: Count slopes and distances between the regarding points.

C. *Defined Characters Stored In The Dataset*

Every character is defined separately along with their slope and distance between two points in the dataset so that it can be matched with characters that are in the image. Due to this, the confliction with any other character is reduced to low.

D. *Recognition*

The matching or recognition process includes matching the letter of the image with the one that is stored in the dataset. The process can only be executed after the completion of stroke generation and line drawing.

E. *Decision*

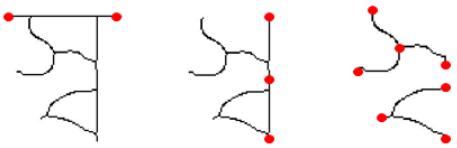
After matching the letter with the dataset letter it comes to an ending decision with a result of percentage of matching. By this the identification process ends here.

IV. FEATURE EXTRACTION

A. *Thinning*

The thinning algorithm reduces the data size where it extracts the information regarding the shape. It allows working more accurate in recognition. There are different types of methods for thinning. After the process of thinning the horizontal and vertical lines are removed to count the starting and ending point of the strokes.

B. *Counting Strokes*



(a) Horizontal Stroke (b) Vertical Stroke (c) Other stroke

Figure 2. Stroke points are showed with red dots in figure (a), (b) and (c)

Algorithom-2 : Algorithm for counting Strokes is given below

```
1.FOR i=2 to (Image,1)-1
2.    FOR j=2 to (Image,1)-1
3.       Pix= certain Pixel of Image
4.       IF (Pix(1)==0) && (Multiplication of Pixel
             position==0)
5.          IF(summation of Pixel Position >= Pix)
6.             SET value of  i = i
7.             SET value of  j= j
8.             SET count= count+1
9.          END
10        END
11.    END
12.END
```

C. *Calculating slope*

Slope of a stroke can be determine by the coordinates, e. g. (x1,y1) consider as the starting point and (x2,y2) as the ending point of a certain stroke .

- Slope, m = $tan^{-1}(y_2 - y_1)/(x_2 - x_1)$

D. *Calculating Distance:*

Distance between point to point is measured by using Euclidian Method.

- Function Euclidean Distance = CalcDistance(x1, y1, x2, y2)
- Distance = sqrt((x2-x1)^2+(y2-y1)^2)

V. EXPEIMENTAL ANALYSIS

A. *Experimental Setup*

The recognition process has been worked on Intel Core i5-2.40 GHz processor of 4.0 GBytes of RAM running on Windows 7 operating system. The desktop application has been performed in MATLAB 2013. In recognition, for 4 conjunct characters of 'ত' are represented in a Tree 01 of Figure 4, a test set of 40(= 4 x 10) character where 10 is the number of sample image has been generated to consider herein. Test set of other characters of total 32 Trees. Since a proper benchmark dataset is not available, we had to make our own dataset where the number of sample image is 10 for each conjunct character.

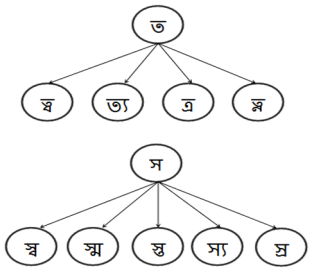


Figure 3. Sample tree of 2 characters

The character images were trained and tested in Neural Network to calculate the accuracy of performance. The result has been included in the Experimental Result part.

B. *Structure of Neural Network*

Neural networks are typically organized in layers. Layers are made up of a number of interconnected 'nodes' which contain an activation function. Patterns are presented to the network via the input layer, which communicates to one or more hidden layers where the actual processing is done via a system of weighted connection.

Here we worked over 50 conjunct characters. The feed forward neural network is trained in a supervised mode. The correct output that corresponds to the data is specified and known as target array. Our vector length is 20. So our input layer has 50 neurons with 20 combinations. So the input is 50×20 array. Since, the output layer depends on the number of characters trained, so in this case the output will be of 50 neurons. The suitable feed forward neural network with 50×20 array input and 50 outputs is shown in the Figure 4.

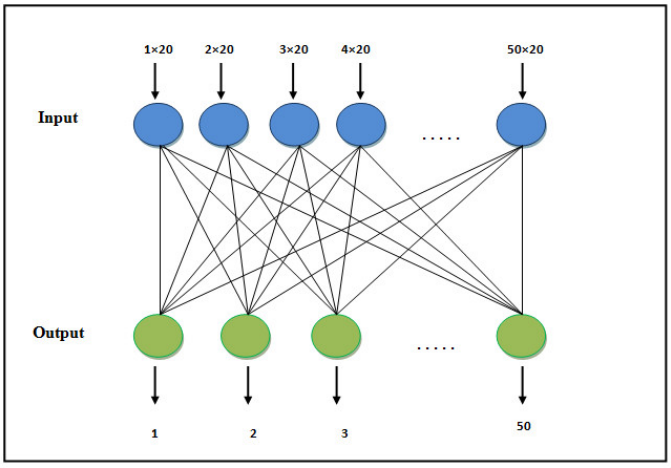


Figure 4: Design of Neural Network layers for Bangla Conjunct Character recognition.

C. *Experimental Results*

According to the Tree 01 of Figure 3 , the process at first takes input from scanned file with the extension of JPEG, Bitmap format. The below table shows the test result with recognition percentage.

TABLE I: TESTING PERCENTAGE USING NEURAL NETWORK

Characters	SAMPLES	RECOGNITION RATE
স্ব	10	98.89 %
স্র	10	91.10 %
স্ম	10	90.5 %
ব্য	10	91.13 %
প্ত	10	91.08 %
প্র	10	90.71 %
ষ্প	10	91.17 %
ষ্ম	10	90.79 %
ঙ্ক	10	90.75 %
স্প	10	91.39 %
ত্য	10	91.36 %
গ্র	10	90.74 %
ক্ত	10	90.18 %
ক্ষ	10	91.10 %

We have worked on 50 conjunct characters with 10 sample of every different characters. And the average training percentage is 98.73 %. And the average testing percentage of the characters are = 90.10 %. A Graphical User Interface is created to confirm the recognition. And for every character a fixed string has been set. Such as, 'SB' has been set for the character 'স্ব'.

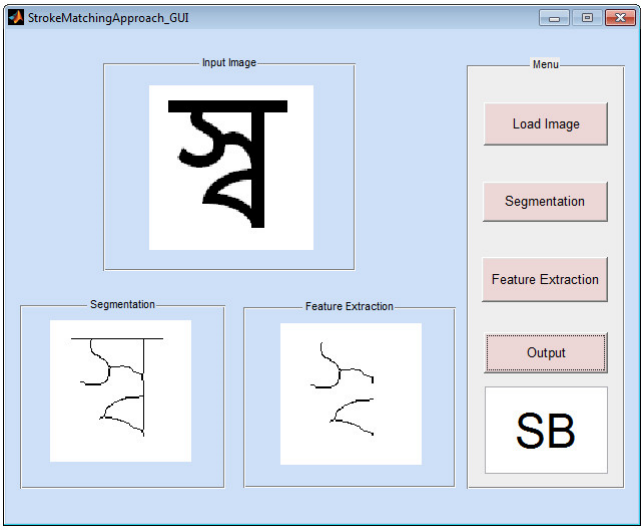


Figure 5: A GUI to recognize the character

D. *Discussion*

Handwritten character comes along with noises. So, stroke matching method at first removes the noise for the easy of character recognition. Then it collects the value points and slopes. It matches with the character stored in the dataset. To complete the recognition Pattern Recognition tool of Neural network is used.

VI. CONCLUSION

The proposed Stroke Matching Based approach is effective for Bangla Offline Character Recognition successfully. Stroke generation and matching completes the process without the loss of information. However, the stroke matching character recognition has some limitations. And after the overcome of maximum limitations the recognition rate stands up to 92.92%. A lot more improvements can be done by curve detection, proper character definition and to recognize all BOC.

REFERENCES

[1] Bhattacharya, Nilanjana, and Umapada Pal. "Stroke segmentation and recognition from Bangla online handwritten text." *Frontiers in Handwriting Recognition (ICFHR), 2012 International Conference on*. IEEE, 2012.

[2] Barman, Sumana, Amit Kumar Samanta, and Tai-hoon Kim. "Design of a view based approach for Bengali Character recognition." *Int. J. Advanced Science and Technology*. 2010. pp-49-62

[3] Mondal, T., et al. "On-line handwriting recognition of Indian scripts-the first benchmark." *Frontiers in Handwriting Recognition (ICFHR), 2010 International Conference on*. IEEE, 2010.

[4] Nasir, Mostofa Kamal, and Mohammad Shorif Uddin. "Hand Written Bangla Numerals Recognition for Automated Postal System." *IOSR Journal of Computer Engineering* 8.6 (2013): 43-48.

[5] Asaduzzaman, A. O. M., Md Khademul Islam Molla, and M. Ganjer Ali. "Printed bangla text recognition using artificial neural network with heuristic method." *Proceedings of International Conference on Computer and Information Technology*. 2002.

[6] Roy, Partha Pratim, et al. "A Novel Approach of Bangla Handwritten Text Recognition Using HMM." *Frontiers in Handwriting Recognition (ICFHR), 2014 14th International Conference on*. IEEE, 2014.

[7] Badsha, Md Alamgir, et al. "Handwritten Bangla Character Recognition Using Neural Network." *International Journal of Advanced Research in Computer Science and Software Engineering* 2.11 (2012): 307-312.

[8] Zibran, Minhaz Fahim, et al. "Computer Representation of Bangla Characters And Sorting of Bangla Words." *Proc. ICCIT*. 2002.

[9] Mashiyat, Ahmed Shah, Ahmed Shah Mehadi, and Kamrul Hasan Talukder. "Bangla off-line handwritten character recognition using superimposed matrices." *Proc. 7th International Conf. on Computer and Information Technology*. 2004.

[10] Hossain, SK Alamgir, and Tamanna Tabassum. "Neural net based complete character recognition scheme for Bangla printed text books." *Computer and Information Technology (ICCIT), 2013 16th International Conference on*. IEEE, 2014.

[11] Hasan, MA Mehedi, M. A. Alim, and M. Wahedul Islam. "A New Approach to Bangla Text Extraction and Recognition from Textual Image." *8th International Conference on Computer and Information Technology, ICCIT*. 2005.

[12] Hossain, SK Alamgir, and Tamanna Tabassum. "Neural net based complete character recognition scheme for Bangla printed text books." *Computer and Information Technology (ICCIT), 2013 16th International Conference on*. IEEE, 2014.

[13] Rabbani, Masud, et al. "A new stroke matching based approach to recognize Bangla handwritten text." *Computer and Information Technology (ICCIT), 2015 18th International Conference on*. IEEE, 2015.

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

Sentiment Analysis on Bangla and Romanized Bangla Text (BRBT) using Deep Recurrent models

Asif Hassan, Mohammad Rashedul Amin, Abul Kalam Al Azad, Nabeel Mohammed
Dept. of Computer Science and Engineering
University of Liberal Arts Bangladesh (ULAB)
Dhaka, Bangladesh

Abstract—**Sentiment Analysis (SA) is an opinion mining study analyzing people's opinions, sentiments, evaluations and appraisals towards societal entities such as products, services, individuals, organizations, events, etc. Of late, most of the research works on SA in natural language processing (NLP) are focused on English language. However, it is noted that Bangla does not have a proper dataset that is both large and standard. As a result, recent research works with Bangla in SA have fallen short to produce results that can be both comparable to works done by others in other languages and reusable for further prospective research. In this work, a substantial textual dataset of both Bangla and Romanized Bangla texts have been provided which is first of this kind and post-processed, multiple validated, and ready for SAimplementation and experiments. Further, this dataset have been tested in Deep Recurrent model, specifically, Long Short Term Memory (LSTM), using two types of loss functions – binary cross-entropy and categorical cross-entropy, and also some experimental pre-training were conducted by using data from one validation to pre-train the other and vice versa. Lastly, the results along with analysis are presented in this research.**

Keywords—Dataset; Bangla; Romanized Bangla; Sentiment Analysis; LSTM

I. INTRODUCTION

Bangla is spoken as the first languageby almost 200 million people worldwide, 160 million of whom are Bangladeshi [1]. Bangladeshi people are found to get increasingly involved in online activities such as - getting connected to friends and families through social media, expressing their opinions and thoughts on popular micro-blogging and social networking sites, sharing opinions and thoughts by means of comments on online news portals, doing online shopping through online marketplaces and other such applications. However, it is becoming increasingly harder for such businesses to monitor and analyze market trends, especially when it is done by analyzing the reaction of the customers on their products or services, due to less or no human-to-human interaction in such businesses. Moreover, the task of going through comments and reviews from each individual customers and figuring out the sentiments within is tedious and in some cases simply intractable, especially considering that - usually very high volume of data is generated very quickly in this day and age of digital connectivity. Therefore, application of automatedSentiment analysis (SA) [2] can play a vital role here for enhancing efficiency and productivity.

SA is widely employed as a machine learning application in many areas, and is known byby many other terms e.g. opinion extraction, sentiment mining, opinion mining, subjectivity analysis, emotion analysis, review mining, etc. Most of the research works found on SA are based on the English language, while Bangla SA is still at a formative stage. An interesting work by Das and Bandyopadhyay [2] on subjectivity detection included Bangla but it is not self-sufficient, as English is also needed. However, none of the works truly considered Bangladesh's perspective. We need to consider not just standardized Bangla, but Banglish (Bangla words mixed with English words) and Romanized Bangla. These three major types can again be loosely categorized in - good, standard, bad, wrong, totally wrong, particular to specific location (almost arcane), etc., depending on the level of clarity, grammatical correctness, meaningfulness, personal idiosyncrasies, impact of localization etc. Moreover, for the Romanized Bangla the added complexity is due to the variation in transliteration between people who know English well and those who do not[3]. The reason, that no clear standard is followed when 160 million Bangladeshi people write in any of the mentioned types, makes it all the more complicated and challenging to work with.

In the recent past, Deep Learning methods, specifically recurrent model-based deep learning models have enjoyed a lot of success in Natural Language Processing (NLP), compared to more traditional machine learning methods [4]. While there are other approaches to SA, in this research we will concentrate exclusively on deep learning based techniques. Our key contributions cover –

- A Data set of 10,000 Bangla and Romanized Bangla text samples, where each sample was annotated by two adult Bangla speakers
- Pre-processing the data in a way so that it is readily usable by researchers.
- Application of deep recurrent models on the Bangla and Romanized Bangla text corpus.
- Pre-train dataset of one label for another (and vice versa) to see if it gives better results.

The paper is organized as follows. In section 2, we discussed the background of our work and the works of others in the same field that inspired and helped us in a way. In section 3, we discussed in details about the dataset that we used for our experiments. Section 4 discusses the methodology and also includes the experimental setup for the deep recurrent models. Section 5 has all the discussion about various results found

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

from our experimentation, and lastly the article concludes with section 6.

II. Background

A. Sentiment Analysis

A key point of our work is Sentiment Analysis, on Bangla (and Romanized Bangla) language. Although the term "Sentiment Analysis" may have appeared for the first time in Nasukawa and Yi [5] , research works on sentiment appeared as early as in 2000 [6-8]. With advent of social media on internet e.g. Facebook, Twitter, forum discussions, reviews, and its rapid growth, we were introduced to a huge amount of digital data (mostly opinionated texts e.g. statuses, comments, arguments etc.) like never before, and to deal with this huge data the SA field enjoyed a similar growth. Since early 2000, sentiment analysis has become one of the most active research areas in NLP.

However, most of the works are highly concentrated on English language, favored by the presence of standard data sets. Standard datasets allow researchers to do their own experiments and compare their contributions with those of others. For the English language, an example of such a standard SA dataset is the IMDB Movie Review Data set, which contains 50,000 annotated (positive or negative movie review) movie reviews made by the viewers. This dataset was originally created by Maas, et al. [9] and since then has been used by a multitude of different studies.

A detailed survey paper [10] presented an overview on the recent updates in SA algorithms and applications, categorizing and summarizing total 54 articles that had been published till 2014. Godbole, et al. [11] collected opinions from newspaper and blogs, and assigned scores indicating positive or negative opinion to each distinct entity in the text corpus to do SA. In [12], they proposed and investigated a paradigm to mine the sentiment from a popular real-time micro-blogging service like Twitter, and they fashioned a hybrid approach of using both corpus-based and dictionary-based methods in determining the semantic orientation of the tweets.

B. Sentiment Analysis for Bangla

It is quite unfortunate that there is no standard collection of data, such as - the IMDB dataset, Twitter corpus etc. for Bangla texts. One effort for standardization came from an automatic translation of positive and negative words of SentiWordNet[13]. However, no corpus was created from this work, thereby limiting its usage to word level determination of sentiment, rather than the more complex natural language processing methods. Additionally, such simplified techniques do not consider the variety of ways in which people usually write, e.g. spelling mistakes, using colloquial terms etc.

A small dataset of Bangla Tweets were collected along with Hindi and Tamil by Patra, et al. [14], where the authors reported on the outcome of a shared Sentiment Analysis task of Indian languages. They used 999 Bangla tweets for training and 499 for testing. They did some post processing such as pruning of emoticons from the tweets and removal of duplicated posts. This data was annotated manually by native speakers. However, in terms of usability the dataset's small size is a limiting factor for modern deep learning techniques.

Another similar collection was done in [15], where 1400 Bangla Tweets were collected automatically. However, their dataset is not publicly available, and the size of the dataset is rather small.

A slightly larger corpus was collected, automatically annotated and manually verified byDas and Bandyopadhyay [2], as their collection was almost 2500 Bangla text samples from news items and blog posts. The uniqueness of their collection over the ones collected by others [14, 15] was the average size of 288 words of their samples, which is quite a bit larger than the 144 character Tweet limit.

With most of the other works proceeded in the similar way, the two biggest issues with the current state of affairs in Bangla SA research are - first and foremost, the absence of a standard and big enough dataset to compare against, which makes comparison of research work extremely difficult, and secondly, none of the Bangla SA research takes into account the very prominent practical aspect of the use of *Romanized Bangla*[3].

C. Deep learning

AI (Artificial Intelligence) has been traditionally done in two ways – i) Knowledge based, and ii) Representation learning based. Knowledge base approach to AI uses logical inference rules to reason about statements input by users. Cyc was one of the most famous of such projects [16]. The failure of knowledge based approach was the driving force into finding a way to give AI the ability to gather its own knowledge by extracting patterns or learning from the data – popularly known as Machine Learning. This new algorithm was based on representation of data or feature. That is, the system is given a number of features about the task in hand on which it will give a decision. Clearly if any of the features were wrong, it would mean wrong representation of the data and the system would not perform well. To rectify this situation representation learning based [17] algorithm was used. This algorithm gave better results than the manually tailored representation of data, and allowed systems to adapt to new tasks with ease. However, using this algorithm it was required that high level abstract features from the raw data were extracted without any error caused by misinterpretation due to the factors of variation, as there can be such factors (e.g. an accent in speakers speech) which would cause false representation in absence of highly sophisticated (human like) understanding. However, deep learning performed better with this issue, as it provides with complex representations expressed in terms of a number of other simpler representations.

D. Recurrent Neural Network

Recurrent Neural Network or RNN in short, has been widely used in speech recognition, handwriting recognition, natural language processing and others. Moreover, RNN is the

precursor to LSTM. While traditional neural networks failed to create a persistent model that would somewhat mimic the way our memory cells work for learning and remembering information, RNN – a class of ANN, has an interesting model design with a loop used as a feed-back connection which makes the information persistent[18, 19].The loop enables the flow of information from one step to the next.It is like there are multiple copies of same network, where a successor gets information from all the predecessors, connected in architecture that excels at processing sequential data.

E. Long Short Term Memory (LSTM)

While RNN's success was critical in speech and pattern recognition due to its ability to remember temporal dependencies, it was not without problems. RNNs were able to connect previous information to current task, only when the gap between the information was small. As the gap widened, RNNs started to perform poorly. Also, the depth and complexity of layers are increase, the vanishing gradient problem causes difficulty in training. Long Short Term Memory (LSTM)is an extension of simple RNNs, which reduce the vanishing gradient problem and can remember dependencies over larger gaps[20]. In 1997, Hochreiter and Schmidhuber introduced LSTM, where a memory cell had linear dependence of its present activity and its past activity. Input and outputgates were introduced to efficiently modulate input and output. However, the introduction of forget gates were crucial to effective modulation of the information flow between present and past activities[21, 22].

$$i_t = \sigma\left(W^{(xi)}x_t + W^{(hi)}h_{t-1} + W^{(ci)}c_{t-1} + b^{(t)}\right) \quad (1)$$

$$f_t = \sigma\left(W^{(xf)}x_t + W^{(hf)}h_{t-1} + W^{(cf)}c_{t-1} + b^{(f)}\right) \quad (2)$$

$$c_t = f_t \circ c_{t-1} + i_t \circ tanh\left(W^{(xc)}x_t + W^{(hc)}h_t + b^{(c)}\right) \quad (3)$$

$$o_t = \sigma\left(W^{(xo)}x_t + W^{(ho)}h_{t-1} + W^{(co)}c_{t-1} + b^{(o)}\right) \quad (4)$$

$$h_t = o_t \circ tanh(c_t) \quad (5)$$

Equations 1-5 capture the LSTM model where, σ is the logistic sigmoid function. *i, f, o* and *c* are the input gate, forget gate, output gate, and memory cell activation vectors, respectively.The process in LSTM includes three gating functions. Each memory cell c_t has its net input modulated by the activity of an input gate, and has its output modulated by the activity of an output gate. These input and output gates provide a context-sensitive way to update the contents of a memory cell. The forget gate modulates amount of activation of memory cell kept from the previous time step, providing a method to quickly erase the contents of memory cells.

III. DATASET DETAILS

Our dataset is called the BRBT dataset where BRBT stands for Bangla and Romanized Bangla Texts. This Bangla Sentiment Analysis (SA) dataset consists of total 9337 post samples. The dataset is unique because not only this is larger compared to others, but it also encompasses the till-now-ignored Romanized Bangla. Romanized Bangla is the Bangla written in English alphabets. Inclusion of Romanized Bangla in the dataset is paramount, because the ease of writing Bangla using any standard QWERTY keyboard (without a Bangla keyboard e.g. Bijoy® keyboard) and the simplicity of using English as base language for the posts, have popularized Romanized Bangla not just in personal messages and micro-blogs but also in Govt. sanctioned mass messages/announcements. The dataset is currently kept private for safe keeping and further improvement. However, it may be made available by personally contacting the owner/authors, and signing a consent form.

A. Data Statistic

Bangla texts holds 72% of whole textual data in the dataset while Romanized Bangla texts is the remaining 28%. There are –

- Total number of entries: 9337
- Bangla entries: 6698
- Romanized Bangla entries: 2639

B. Data Sources

Data were collected from various micro-blog sites, such as, Facebook, Twitter, YouTube etc, and some online news portal, product review panels etc. Following is the statistic of data sources -

- From Facebook: 4621
- From Twitter: 2610
- From YouTube: 801
- From online news portals: 1255
- From product review pages: 50

C. Post collection data processing

- *Removal of emoticons:*-emoticon, hash-tags were removed to give annotators an unbiased-text-only content to make a decision based on three criteria - positive, negative and ambiguous.

- *Removal of proper nouns:*-Proper nouns were replaced with tags to provide ambiguity. All text samples were collected from publicly available sources and did not reflect the opinion of the authors.

- *Manual validation (by native speakers):*- Collected data samples are manually annotated into one of three categories: positive (1), negative (0) and ambiguous (A). Each text sample was independently manually annotated by two different native Bangla speaking individuals for total two validations. Each annotator

validated the data without knowing decisions made by other. This ensures that the validations are unbiased and personal.

TABLE I. DATASET VALIDATION SAMPLES

Text Sample	Translation	1st Annotator	2nd Annotator
অনেক ভালো হয়েছে গান!	Very nice song!	Positive	Positive
মর্মান্তিক সড়ক দুর্ঘটনায় ৩ জন নিহত।	3 dead in a tragic road accident.	Negative	Negative
Chotobelarmodhur din gulokhub miss kori	Really miss the sweet childhood days	Positive	Negative
Symponyer set gulakemon?	How are Symphony mobile sets?	Positive	Ambiguous
আলো আলো তুমি কখনো আমার হবে না	Light, light, you'll never be mine	Ambiguous	Negative

D. Double Validation Analysis

Table2 gives shows the confusion matrix between the labels given by the two annotators.We can see that the annotators agreed on 75% of the texts samples, giving us a base-line of human level agreement for this data set. Not surprisingly, the greatest amount of disagreements arise on text samples which at least one of the annotators labelled as ambiguous.

TABLE II: CONFUSION MATRIX OF MANUAL ANNOTATIONS

First Validation	Second Validation		
	Positive	Negative	Ambiguous
Positive	2817	538	392
Negative	178	3864	404
Ambiguous	27	95	1022

IV. DATASET SETUP

The data was manually picked from various online micro-blog sites, product review panels, news portals etc. For tweets 'bn' parameters were used in the search option to access Bangla tweets only. There are over 10000 total Bangla and Romanized Bangla posts in the dataset.

We checked for empty rows or columns, missing annotation, proper <PN> tagging (for dataset with proper nouns replaced), proper categorization etc. The resultant dataset is now both unique and error-free in terms of the abovementioned flaws.

The entire data set was divided into Bangla and Romanized Bangla sections for convenience of future research. Scripts are available from the Data set's Github account to do the following:

- Converting textual data into tokens
- Saving the data as tuple ([data], [label1], [label2])
- Randomly shuffling the data
- Serializing each datasheets and splitting three sets from each and making them available for public to download and un-pickle to use them in their models.

For our experiments we applied the tokenizing, splitting, serializing scripts on the "full-text" (or unmodified texts column of the dataset with all the proper nouns, emoticons etc intact) also, hence creating additional sets of pickle files.

V. MODEL IMPLEMENTATION

Our dataset consists of three categories –

- Positive,
- Negative, and
- Ambiguous.

Depending on the dataset used and number of categories classified, we used three types of fully connected neural networks layer, which mainly differ by the number of nodes in the output layer (Fig. 1). One and two output nodes were used for categorizing between positive and negative sentiments, and three output nodes were used when ambiguous labels were also included.

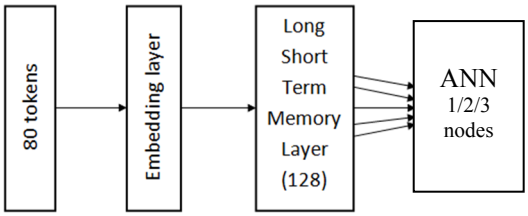


Fig. 1. Dense layer schematic

We used data for one validation set as pre-training for another validation set. More specifically, first we fit data from 1st validation in the model to pre-train for 2nd validation data – which is fit in the same model afterwards. Likewise, we fit data from 2nd validation to pre-train for 1st validation data. This sort of pre-training was to check whether it can be useful to pre-train on an independently sentiment analysis data even if the labels did not match.

VI. Experiments

A. Datase Preparation

Our model is based on Recurrent Neural Networks (RNN) – more specifically we used LSTM neural network. We used Keras' model-level library since it has all the required features to help us develop our deep learning model. We used Theano as the back-end for Keras. All our models are Keras Sequential models. First layer of the Sequential model is the Embedding layer. We used Embedding layer to implement the word to vector representation for the words in our dataset. We a***max_features***parameter as the input dimension argument for embedding layer. It signifies the total number of unique tokens returned by the tokenizer, which in turn means that max_features is also the vocabulary size (*input_dim*). The second layer is Long Short Term Memory (LSTM) with an internal state of 128 dimensions. The third is a fully connected layer with different activations for classification purposes.

In our experiments, the last layer has 1, 2 and 3 nodes, depending on the classification regime attempted. When attempting to classify only positive and negative sentiment, which are represented by 1 and 0 respectively, the final fully connected layer has been configured with 1 and 2 neurons. When a single neuron is used, the loss function employed was binary-cross-entropy. When 2 neurons were used, we used categorical-cross-entropy instead. With the inclusion of ambiguous labels, the number of classes increase to 3, for which we used 3 neurons with a softmax activation in the last layer. The loss function employed in this case was categorical-cross-entropy.

Even in this simply configuration, the number of parameters of the network is quite high and it is possible to easily overfit such models. To avoid this we used Dropout [23] rates of 20% between the Embedding and LSTM layer at training time..

B. Experiment model label tags

There are actually 36 unique experiments using the same LSTM model, depending on the dataset used, processing of texts, loss function used, processing of labels (annotations on data), and *input_dim* value for Embedding layer. However, it turns into a total of 72 experiments – one half of experiments where label 1 (1st validation) is used for pre-training, and the other half where label 2 (2nd validation) is used for pre-training. Following are the tags used in experiments and what they actually mean.

- Tags used for different types of dataset –

Dataset Type	Tag used in experimental labels
Bangla and Romanized Bangla (total)	BRBT
Bangla (only)	Bangla
Romanized Bangla (only)	RB

- Tags used depending on processing of texts/posts –

Processing of texts	Tag used in experimental labels
<PN> removed and other modifications	PN
Full texts (no modification)	FT

- Tags used based on loss function –

Loss function used	Tag used in experimental labels
Binary_crossentropy	bin
Categorical_crossentropy	cat

- Tags used based on Annotation data modification-

Annotation data modification	Tag used in experimental labels
Annotation value of 'A' removed (label along with data removed)	ra
Annotations value of 'A' converted to 2	ato2

- Tags used based on different type of max_features applied -

Max_features type	Tag used in experimental labels
Non-fixed, ranging from 20,000 ~ 40,000 depending on the dataset type and size	1
Value fixed at 500	2

VII. Results and Discussion

Highest accuracy was attained by Bangla dataset with categorical crossentropy loss, modified text, Ambiguous removed and non-fixed max_features, with 70% of accuracy – which is 20% more than chance for two category dataset. However, this experiment on BRBT dataset with categorical loss, modified text, ambiguous converted to 2, has a low accuracy score of 55% but for a three category it scores 22% more than chance (33%). Therefore, it is clear that most of experiment sets (dataset-wise, or PN-FT tag-wise, or loss function-wise, and label category-wise) scored above chance. However, none of the experiments with fixed max_features (vocabulary size for Embedding layer) scored well compared to the non-fixed variants.

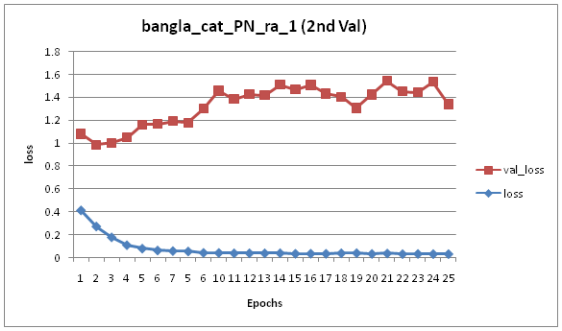


Fig. 2. loss-val_loss graph for bangla_cat_PN_ra_1 (2nd validation)

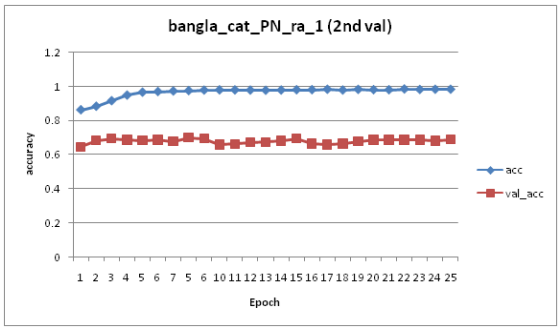


Fig. 3. acc-val_acc graph for bangla_cat_PN_ra_1 (2nd validation)

VIII. Conclusion

The goals of the research may be summarized as following:

1. Pre-processing the data in a way so that it is readily usable by researchers.
2. Application of deep recurrent models on a Bangla and Romanized Bangla text corpus.
3. Pre-train dataset of one label for another (vice versa) to prove its usefulness.

To meet the goals, a BRBT (Bangla and Romanized Bangla Text) dataset of total 9337 entries with 6698 entries for Bangla and 2639 for Romanized Bangla texts were pre-processed. Then dataset was split and serialized into training set, testing set and validation set.

For the experiments, LSTM which is a deep recurrent model was applied. There are total 32 different experiments based on the same model with only differences in dataset used, loss function applied, modification done (or not) on data (proper noun replaced with <PN> tags, duplication removal etc.) etc. While most of the experiments scored accuracy higher than chance in percentage, Bangla dataset with categorical crossentropy as loss function and non-fixed max_features for the embedding layer with "Ambiguous removed" scored highest with 78% in accuracy for 2 category (results compared from both pre-training set of experiments), and Bangla and Romanized Bangla dataset (modified text set) with categorical crossentropy loss, non-fixed max_features, and "Ambiguous converted to 2" scored highest with 55% in accuracy for 3 category.

The implementation of pre-training dataset of one label for another has showed that, even if the labels do not match it is useful to pre-train on an independently annotated SA data. From four experiments done from the alternate experiment set consistent result from 2nd validation data (label 2) have been observed.

IX. References

[1] Banglapedia. (August 30). *Bangla Language*. Available: http://en.banglapedia.org/index.php?title=Bangla_Language

[2] A. Das and S. Bandyopadhyay, "Subjectivity detection in english and bengali: A crf-based approach," *Proceeding of ICON,* 2009.

[3] S. Khan, "Convergence in spelling, and spell-checker for Romanized Bangla in computers and mobile phones," in *Informatics, Electronics & Vision (ICIEV), 2014 International Conference on*, 2014, pp. 1-5.

[4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature,* vol. 521, pp. 436-444, 2015.

[5] T. Nasukawa and J. Yi, "Sentiment analysis: Capturing favorability using natural language processing," in *Proceedings of the 2nd international conference on Knowledge capture*, 2003, pp. 70-77.

[6] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," in *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, 2002, pp. 79-86.

[7] S. Das and M. Chen, "Yahoo! for Amazon: Extracting market sentiment from stock message boards," in *Proceedings of the Asia Pacific finance association annual conference (APFA)*, 2001, p. 43.

[8] J. Wiebe, "Learning subjective adjectives from corpora," in *AAAI/IAAI*, 2000, pp. 735-740.

[9] A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts, "Learning word vectors for sentiment analysis," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, 2011, pp. 142-150.

[10] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Engineering Journal,* vol. 5, pp. 1093-1113, 2014.

[11] N. Godbole, M. Srinivasaiah, and S. Skiena, "Large-Scale Sentiment Analysis for News and Blogs," *ICWSM,* vol. 7, pp. 219-222, 2007.

[12] A. Kumar and T. M. Sebastian, "Sentiment analysis on twitter," *IJCSI International Journal of Computer Science Issues,* vol. 9, pp. 372-373, 2012.

[13] D. Das and S. Bandyopadhyay, "Developing Bengali WordNet Affect for Analyzing Emotion," in *International Conference on the Computer Processing of Oriental Languages*, 2010, pp. 35-40.

[14] B. G. Patra, D. Das, A. Das, and R. Prasath, "Shared task on sentiment analysis in indian languages (sail) tweets-an overview," in *International Conference on Mining Intelligence and Knowledge Exploration*, 2015, pp. 650-655.

[15] S. Chowdhury and W. Chowdhury, "Performing sentiment analysis in Bangla microblog posts," in *Informatics, Electronics & Vision (ICIEV), 2014 International Conference on*, 2014, pp. 1-6.

[16] D. B. Lenat and R. V. Guha, *Building large knowledge-based systems; representation and inference in the Cyc project*: Addison-Wesley Longman Publishing Co., Inc., 1989.

[17] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE transactions on pattern analysis and machine intelligence,* vol. 35, pp. 1798-1828, 2013.

[18] J. A. Bullinaria, "Recurrent neural networks," *Neural Computation: Lecture,* vol. 12, 2013.

[19] C. Olah, "Understanding LSTM Networks," ed, 2016.

[20] J. L. Elman, "Finding structure in time," *Cognitive science,* vol. 14, pp. 179-211, 1990.

[21] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation,* vol. 9, pp. 1735-1780, 1997.

[22] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," *Neural computation,* vol. 12, pp. 2451-2471, 2000.

[23] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research,* vol. 15, pp. 1929-1958, 2014.

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

Design and Simulation of a Novel Classification Framework for Separating Sentiment from Assorted Game Related Tweets

Syeda Tasmiah Islam
Computer Science & Engineering
University of Asia Pacific
Dhaka, Bangladesh
s.tasmiahislam@gmail.com

Ahmad Al-Sajid
Computer Science & Engineering
University of Asia Pacific
Dhaka, Bangladesh
ahmadalsajid@gmail.com

Molla Rashied Hussein
Computer Science & Engineering
University of Asia Pacific
Dhaka, Bangladesh
mrh.cse@uap-bd.edu

***Abstract*—Twitter has been exceedingly a ubiquitous phenomenon with a thriving appearance in almost every country in the world. In this modern epoch, game is one of the most popular genres to tweet. Furthermore, sentiment analysis is used to analyze the patent opinion of a person by mining text units i.e. tweets. In this paper, strategy based categories of virtual games were identified; top most countries according to users were detected; foremost usable language of the users was ascertained; and advertisement from assorted game related tweets were filtered to retrieve sentiment tweets exclusively through simulation. In light of these findings, a new gestalt of sentiment analysis of virtual game related tweets is accomplished.**

Keywords—Sentiment Analysis; Natural Language Processing; Naïve Bayes Classifier; Twitter; Games

I. INTRODUCTION

Online social media have acquired astonishing global augmentation and become a resounding fashion of the neoteric vogue by drawing attention from miscellaneous types of users. This is predominantly a digital platform that enriches the social norm overhead virtually. Social relations among people are being expanded and prolonged by content-sharing, social interaction, and also proper cooperation. Facebook, Twitter, LinkedIn, Google+ etc. are the most meticulous social network sites.

At the present time, online social media are getting more popularity for innovating and updating information into a dynamic forum incessantly [1]. An innate feature of many social network sites that differentiates them from mainstream news media sites is the information network evolved by the users following their preferred generators of information [2]. Twitter is one of the micro-blogging websites which was launched back in 2006. In fact, Twitter is a network of information constructed of one hundred and forty (140)-character messages, namely Tweets. This is the best platform to commence the peregrination for finding out and following several engrossing Twitter accounts to access resources that users desire recurrently [6].

Crafting a post or tweet is also another spontaneous matter. Eventually, Twitter is a swift content provider of real time updates. Some researchers confide that Twitter can be suitable for predicting particular and certain outcomes such as: Elections, Business Strategies, Political Protests and Challenges as well as Human Behavior and explanation of their momentary actions. Researchers apply different kinds of sorted schemes and processes for their investigations, where Sentiment Analysis is one of those investigations.

Sentiment Analysis is a comparatively recent developing sector which covenants with extracting and assenting user opinion automatically [7]. Sentiment Analysis prefers using the Natural Language Processing (NLP) tasks at numerous levels of granularity [7,8]. Generally, NLP, Machine Learning methods, Statistics are applied to recognize, mine the sentiment contents or substances of the targeted text units (blogs, user reviews, comments and so on). This analysis is mentioned as the terminology called Opinion Mining as well. For redacting the Sentiment Analysis or Opinion Mining, researchers need both structured (tweet metadata) and unstructured (tweet text) segment of data.

In this paper, design of a novel classification framework is prepared for separating sentiment from assorted game related tweets coalesced with advertisements; simulation of the developed framework is carried out; and analysis of the result is depicted to validate the accomplishment of the developed framework.

II. LITERATURE REVIEW

According to Giachanou and Crestani (2016), user-provided information is a powerful origin of public views and it can be vital for the diversified applications that want to comprehend and analyze the public opinion about certain incidence [1]. As users publish real time reaction about approximately everything on the micro-blogging website Twitter, the website generates contemporary and diversity in terms of challenges. Because of easy access to published posts, Twitter is considered as one of the largest datasets encompassing user generated information.

Holton et al. (2014) described that a survey conducted on Twitter has revealed a more discrete social orientation between the followers rather forwarding messages as Retweets [10]. Twitter characterized by some features it has conceded, such as:

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

Tweet: Creating a post in Twitter. Each tweet permits 140-charaters for sharing users' personal information, opinions, interests, news, views, photos, blogs etc.

Mention: Mentions in a tweet entails the post indicating a different user. Mentions have to be referenced to valid usernames. Mentioning in Twitter is much more convenient than the other social network websites like Facebook, although Facebook is coping up regarding this issue.

User: At first, a user has to be registered him/herself for tweeting with legitimate credentials. User has to be allotted a unique and valid username.

Hashtag: A hashtag is a metadata (data about data) tag used in micro-blogging services or social network websites. Hashtags are used to specify the concern of a tweet to a certain entity. #game, #ClashClans are some of the examples of hashtags.

Retweet: This is nothing but just forwarding or reposting an exact previous post. Retweeting is considered as a powerful tool for promoting and publicizing trendy scorching topics or contents.

Follower: It refers to the users who follow other users or interesting matters to get up-to-date news, views, information etc.

Privacy: Twitter provides the preference to the users to make their posts appear to everyone or solitary to the preferred followers in Twitter, as mentioned by Giachanou and Crestani (2016) in their paper [1].

Cambria et al. (2013) have chalked down Sentiment analysis into three major categories: knowledge-based procedures, statistical methodologies and hybrid approaches [11]. Knowledge-based procedures catalog both prominent and random words to particular emotions. On the other hand, statistical methodologies maintain the components from machine learning. The sentiment uses classifiers trained on both Twitter sentiment as well as reviews from the data sets. The larger part of Sentiment Analysis and Twitter Sentiment Analysis methods reveal sentiment based on a feature set [1]. The outcome will be more precise when the text is similar to the original training data. The most repeatedly used semantic features are sentiment words, concepts, negation and opinion words. Opinion words refer to phrases that are characterized as suggestive thoughts. Sentiment words comprise both positive and negative sentiment. According to Amati et al. (2014) and Aston et al. (2014), Twitter Sentiment Analysis (TSA) mostly depends on Sentiment Analysis using negation [14, 15].

However, game related tweets contain both sentiment and advertisement. Therefore, a classification framework is required to separate these two to retrieve sentiment for further analysis. Not much research has been undertaken in this regard, thus the work presented in this paper is unique in nature.

III. METHODOLOGY

Social network sites have revolutionized the way in which people interact with each other. Micro-blogging websites like Twitter is an impressive and prominent tool for obtaining new thoughts or insight. As random tweets had been collected from Twitter, different data types were retrieved. Generally, data are labeled (structured) but the text units are unlabeled (unstructured). For this reason, structured and unstructured data were analyzed separately, as shown in Fig. 1.

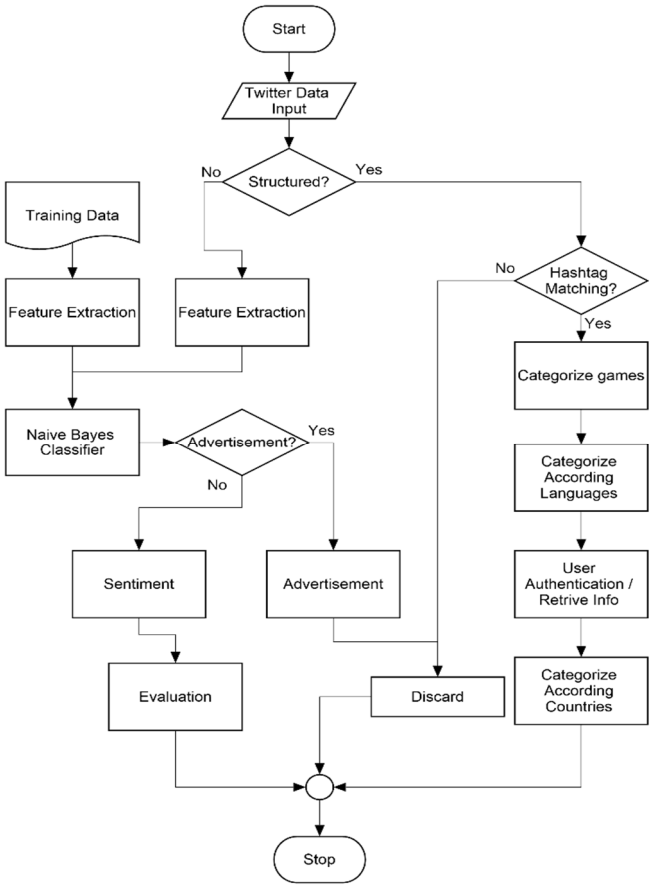


Fig. 1. Analytic approach of sentiment and advertisement classification

A. Structured data analysis:

Structured data is simple to sort. Sentiment analysis provides three types of approaches and one of these three is statistical method. The statistical method is chosen to run the simulation work. Statistical Natural Language Processing uses stochastic, probabilistic, and statistical methods to evolve some of the difficulties including extended sentences which are highly oracular when processed with realistic semantic and conveying numerous possible analyses. Structured data is mined for finding categories of games, countries with most users and major language used in Twitter.

Games Categories: Games are nothing but a calculated strategy or approach. Rules, competitors and achievements are must criteria for any kind of game. In this modern world, game is one of the best current affairs for posting in various social network sites like Facebook, Twitter etc. [16]. For the research, random tweets regarding games were collected from Twitter. Afterward, those tweets were arranged into different categories(as shown in Fig. 2). Such as:

Massively Multiplayer Online (MMO): This type of game is an online game which is susceptible of supporting

large numbers of players simultaneously in the same case. For example, World of Warcraft, Star Wars: The Old Republic, Planet Calypso etc.

Simulations: This kind of games attempts to copy different activities from real life in the form of a game for various motives as well as training, analysis, or prediction, e.g. Game Dev Story, Goat Simulator, Infinite Flight Simulator etc.

Adventure: For over 30 years, adventure games have been the most story-driven computer game group. The contestant plays a fantasy role in an episodic adventurous feat story, such as: Crashland, Dead Effect 2 etc.

Real-Time Strategy (RTS): Real-time strategy games have already gained the popularity of turn-based strategy computer games. The participants position and maneuver units and structures under their control to secure areas of the map and/or destroy their opponents' assets. Like, Total War Battles: Kingdom, Clash of Clans.

Puzzle: The Room 2, Monument Valley etc. are puzzle games that are refer to exercise the brain by fitting different pieces together.

Action: Adventure games are about exploration and big worlds to look through. Need For Speed Rivals, WARP. Action is probably the spacious and most isolated genre in gaming.

Browser games: A browser-based game or online game is played completely within a web browser rather than through a video game or other tackle. Roblox, Animal Jam, Sword Saga etc. are examples of web browser game.

Sports: An athletic recreation requiring physical strength and skill often of a competitive nature virtually. For example: FIFA 14, PES 2014 etc.

Shooter: The main fact of shooter game is that it can be from any perspective. Shooter games are an ample action genre where the player shoots target as a primary gameplay mechanic, such as: Hitman, Sniper, Modern Combat.

Role-Playing Game (RPG): Pokémon Go, 7 Mages, Desktop Dungeons etc. are the Role-Playing games.

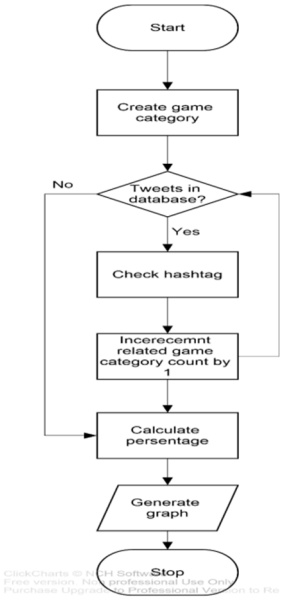


Fig 2: Game Categorization workflow

Language Comparison: To see which languages are mostly used, first a list of used languages in collected tweets was created. After that, the "lang" entity of every tweet was checked and the related language count was increased by one. When all the tweets were checked, the percentage was calculated and a graph was created to show the most used top five languages.

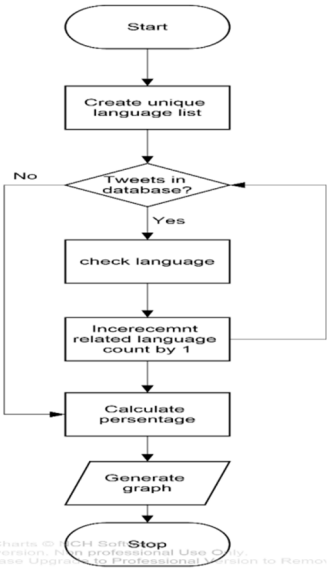


Fig 3: Language Comparison workflow

User Authentication: To get the authentic users data from collected tweets, a unique list of "user.id_str" collected from the tweets. Then every single user's "user" entity was searched from the database that contains tweets and stored to a new database.

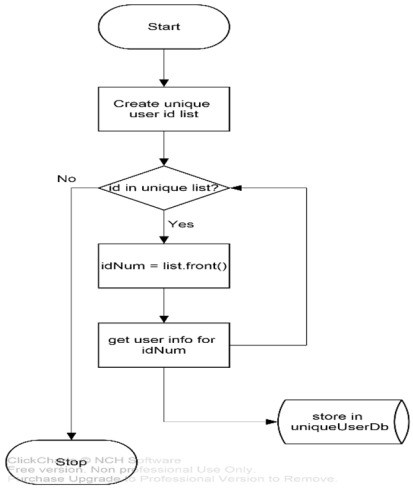


Fig 4: User Authentication workflow

Comparison of countries according to user count: After getting the authentic users from the tweets, a country list was created according to users' time zone or location. Then every user's location or time zone was checked to detect the country from the user was.

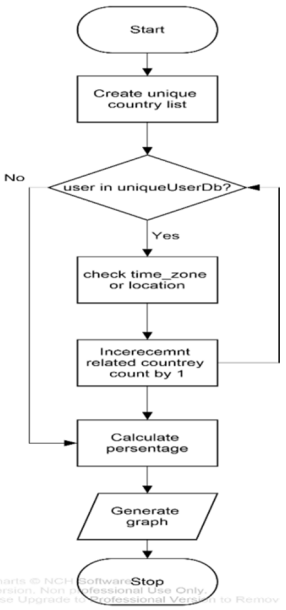


Fig 5: Comparison of countries in terms of users

B. *Unstructured data analysis:*

For the sentiment analysis of the unstructured data, NLP was conducted and Machine Learning process was used to detect advertisement for separating from the personal (sentiment) tweets.

Natural Language Processing (NLP): NLP enables computers to derive meaning from human or natural language input. It has some major tasks, namely Automatic condensation, generating a readable and understandable concise of a shred of text.

Oration Analysis: This semantic includes a number of connected tasks.

Machine Translation: Automatically interpret text unit from one human language to another. This is a complex process.

Natural Language Understanding (NLU): NLU involves the recognition of the intended semantic from the multiple realistic semantics that can be executed from a natural language expression.

Information Extraction: This is apprehensive in general with the release of semantic information from text. It covers the relationship extraction as well.

Machine Learning: Machine Learning is about learning structure from data. This is intimately related to computational statistics. This is sometimes mixed up with data mining. It has supervised and unsupervised approaches to learn. It provides some processes to work such as classification, clustering, regression, density estimation etc. Machine learning and Statistics are nearly related fields. In this research work, Naïve Bayes Classifier was used for classification.

Naïve Bayes Classifier: This classifier is one of the simplest probabilistic classifiers. This is based on Bayes' theorem with strong (naïve) independence postulations between the features. Naïve Bayes Classifier can be trained efficiently and effectively in a supervised learning environment comprising of several kinds of probability models.

Classification of game related tweets into sentiment or advertisements using Naïve Bayes Classifier: At first, a set of texts from random tweets were collected from the database. Then the tiny URLs (Uniform Resource Locators) were removed as they had no impact on the text. After that, they were tagged with two words, either 'advertisement', or 'sentiment'. They were saved in a CSV file. Then the Naïve Bayes classifier was trained to recognize the pattern of advertisements and sentiments with learning using the .csv (Comma Separated Values) data set. Then, every textual part of the tweets was tested according to this classifier to detect if the text was advertisement or not. After that, a comparison graph was plotted. A manually created testing set of data proved that the accuracy of the classifier was 76.02%.

@DuncanBannatyne @carolvorders Clash of clans th10 war base 2016 anti 2 star:A437 #ClashOfClans	advertisement
I found this image, could be useful for all #PokemonGO trainers! B475	advertisement
#GTAOnline #BossDonz CHOPPER CHOP	sentiment
@ChannelStarWars ESTONIAN ROuGE ONE POSTER!!! #starwars	advertisement
@MountainDew So I saw this across the store, you can say I'm Lazer focused...?? #badhumor #Titanfall2	advertisement
@RonCafc Epic In Every Sense of the Word.	advertisement
@Titanfallgame will split screen be in the game? #Titanfall2	advertisement
and I are adding glitter to the #StarWars universe! #sorrynotsorry @missingwords	sentiment
brand new gta5 glitch how to lower your car past competition suspension #GTAOnline #gta5	advertisement
Clash of clans th10 war base 2016 anti 2 star: B423 #ClashOfClans	advertisement
do give damn good trailer, and this is no exception B428s	sentiment
Estonian Rouge One poster!! #starwars #rougeone B252	sentiment
Even Grandmother Willow gets in on the action. #StarWars #ShipWars	advertisement
Final #starwarsrogueone trailer is out. Could this be the best #StarWars film ever? It looks just incredible!	advertisement
is doing a Flip Knife Blue Steel #Giveaway! Come Join!B20 #csgo #TurboNation	advertisement
is giving out a $20 Steam/Battlenet Halloween gift! Click this link B326 #halloween #warcraft #steam #giveaway	advertisement

Fig 6: A glimpse of the training dataset

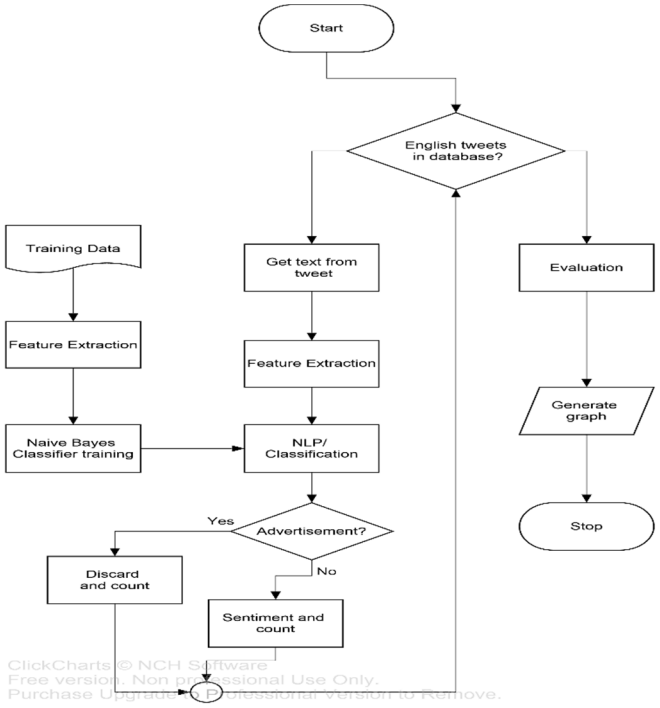


Fig 7: Naïve Bayes Classifier workflow

IV. RESULT

Among these aforementioned categories of total 77 games categorized into 10 categories, the most desired strategies of games were ascertained according to statistics which are tweeted most of the time on twitter. Fig. 8 shows that Massively Multiplayer Online games are the most exoteric and 43.25% people already tweeted about this on Twitter. As opposed to, Simulation category is in the lowest of the scale of popularity with 0.25% after puzzle category that stands with 0.35%. Pokémon go, Battle Champ and other Role-Playing games are in second position with 21.42% popularity. Thereafter, Sports (12.24%), Adventure games (9.05%), Action game (5.29%), Browser games (3.72%) are in the shallower position methodically. Real-Time strategy (2.14%) and Shooter (0.82%) are contiguous to the downcast position. Still these types of games are mostly tweeted.

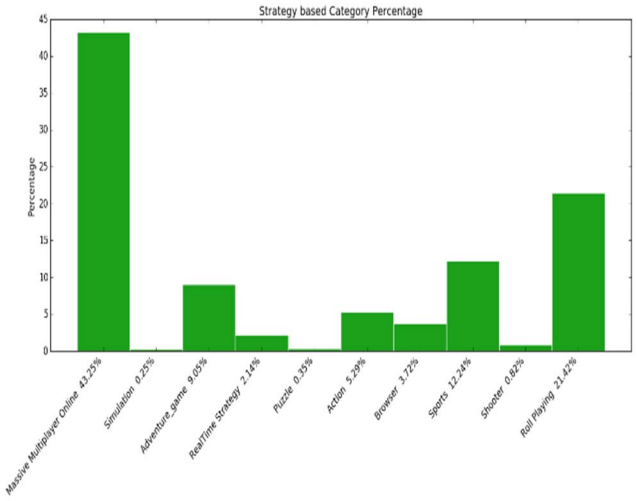


Fig 8: Strategy based popular game

Considering more games related tweet facts, top countries according to users is one of them. Among 97 countries and 188434 users, Fig. 9 depicts that among the countries with active users, United Kingdom (14.78%) is in the top most position. Moreover, the rest of countries have sundry users according to the statistical approach. For example, United States have only 7.80% of users who tweet on games. Ecuador (2.89%) and Russia (2.49%) are the lowest among the topmost countries. The rest of the countries are The Netherlands (5.96%), Spain (5.92%) and others to follow.

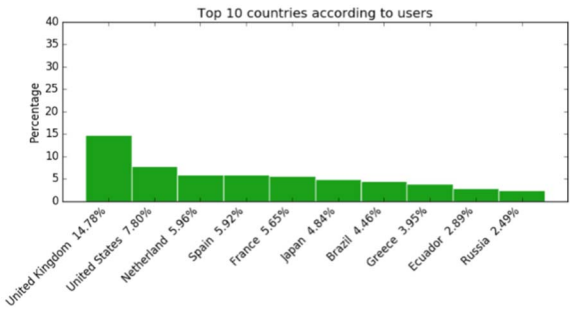


Figure 9: Top 10 countries according to the number of users who tweet regarding games

Language is another important segment of game related tweets. Dynamic users and spectators are reflected in the platform's language disruption. Fig. 10 portrays that English (73.75%) is on the foremost position as users' preferred language. Spanish (7.32%), French (5.17%), Japanese (4.86) and Portuguese (2.33%) trail far behind English.

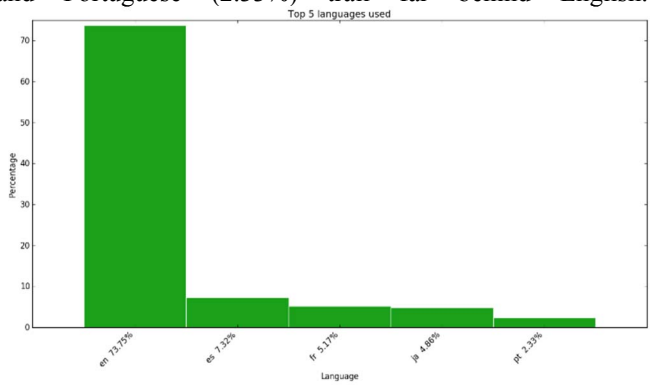


Figure 10: Top 5 Languages according to the number of users who tweet regarding game

Everyday about 500 million of tweets are tweeted. Among them all are not entirely similar in types. Some tweets are sentiment and rests of them are advertisement. Specially, the retweets are mostly identified as advertisements. For detecting advertisement and separating them from sentiments, NLP techniques were used.

Total tweets collected were 353,671 from 188,434 users, among which 329,316 were related to games. Among them, 246,281 were in English Language. With the relevant training data, applying Naïve Bayes Classifier, it was detected from the Fig. 11 that 66.33% of the tweets were advertisements and the rest of 34.67% were sentiment tweets.

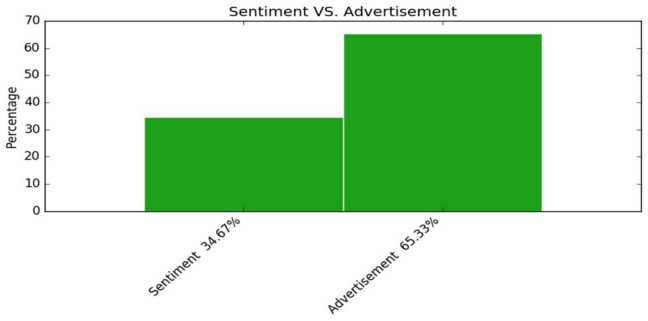


Figure 11: Advertisement vs. Sentiment tweets

V. ANALYSIS OF THE RESULT

From this outcome, it can be evolved that most of the game related tweets are about Massively Multiplayer Online (MMO). Rests of the games concerned tweets are slightly vicinal of MMO.

Most of the tweets are tweeted from the first world countries. So, it can be decided that they are the prime consumer of the gaming industry.

As English is the most widely used language in the world, most of the social network sites (SNSs) use English as their default Language. English has no close contender in any way.

Naïve Bayes Classifier is an old algorithm, but it shows good accuracy of 76.02% in this work. Naïve Bayes Classifier is confident to identify the advertisement from the sentiment with accuracy. As the training data set was generated manually, this result can be improved providing enhanced training data set.

VI. Limitations of the Work

After observing and classifying all these tweets, the countries or the time zones of some of the users cannot be distinguished simultaneously. As some users located their countries or time zones, several users gave either geolocation or time zones and some particular users did not give any of them during they tweet.

At the present time, some games are made with mixed categories. Therefore, separating them into specific types of categories gives inconsistent outcomes.

The accuracy of entire consequence relies on the training data set. Hence, improving training data set would give better outcome.

VII. Conclusion and Future Work

Twitter is a short broadcast medium of blogging. Recently the research interest for analyzing tweets is increasing rapidly as more people express their sentiment through Twitter. This research reveals an overview on the recent updates of game related tweets. This also elicits that game related advertisements can be detected and separated from sentiment. Research outcome can help shaping up the game development in particular.

Naive Bayes classifier is suitable for this task. The accuracy of the outcome is about 76.02%.

This research will be subsidiary for the start-up companies who what to deal with games business. Which games are more popular and should be developed, these types of decisions can be taken spontaneously if they follow this statistical survey. So, this paper can also predict the game business strategy.

To know the reaction of the people about game, the advertisements and the sentiments must be segregated. And here the first step to detect the advertisement and separate them from sentiments have completed. In future, further sentiment analysis regarding polarity, mood and dimensions will be accomplished.

References

[1] A. Giachanou, F. Crestani, "Like It or Not: A Survey of Twitter Sentiment Analysis Methods," ACM Computer Survey 49(2), June 2016, p. 28.

[2] Mahmoudreza, Przemyslaw, Isabel, Krishna, Manuel, "On the Efficiency of the Information Networks in Social Media," WSDM'16, San Francisco, CA, USA, February 22–25, 2016.

[3] R. Bosagh Zadeh, A. Goel, K. Munagala, and A. Sharma, "On the Precision of Social and Information Networks," In Proceedings of the first ACM Conference on Online Social Networks, ACM, 2013, pp. 63–74.

[4] N. A. Christakis and J. H. Fowler, "Social Network Sensors for Early Detection of Contagious Outbreaks," PloS one, 5(9): e12948, 2010.

[5] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, a Social Network or a News Media?" In Proceedings of the 19th International Conference on World Wide Web, 2010, pp. 591–600.

[6] Getting started with Twitter [Online]. Available: https://support.twitter.com/articles/215585.

[7] J. Spencer and G. Uchyigit, "Sentimentor: Sentiment Analysis of Twitter Data," School of Computing, Engineering and Mathematics University of Brighton, Brighton, BN2 4GJ [Online]. Available: http://sentimentor.co.uk .

[8] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, R. Passonneau, "Sentiment Analysis of Twitter Data," Department of Computer Science, Columbia University, New York, NY 10027 USA, 2011.

[9] A. Bermingham and A. Smeaton, "Classifying Sentiment in Microblogs: is Brevity an Advantage?", ACM, 2010, pp. 1833–1836.

[10] A. Holton, K. Baek, M. Coddington, C. Yaschur, "Seeking and Sharing: Motivations for Linking on Twitter," Communication Research Reports 31 (1), 2014, pp. 33–40.

[11] E. Cambria, B. Schuller, Y. Xia, C. Havasi, "New Avenues in Opinion Mining and Sentiment Analysis," IEEE Intelligent Systems 28 (2), 2013, pp. 15–21.

[12] R. Stevenson, J. Mikels, T. James, "Characterization of the Affective Norms for English Words by Discrete Emotional Categories," Behavior Research Methods 39 (4), 2007, pp. 1020–1024.

[13] S. M. Kim, E. H. Hovy, "Identifying and Analyzing Judgment Opinions," Proceedings of the Human Language Technology / North American Association of Computational Linguistics Conference (HLT-NAACL), New York, NY, 2006.

[14] G. Amati, M. Bianchi, and G. Marcone, "Sentiment Estimation on Twitter," In Proceedings of the 5th Italian Information Retrieval Workshop (IIR'14), 2014, pp. 39–50.

[15] N. Aston, J. Liddle, and W. Hu, "Twitter Sentiment in Data Streams with Perceptron," Journal of Computer and Communications, 2014, pp. 11–16.

[16] The Most Engaging Topics on Social Media in 2015 (so far) [Online]. Available: http://www.mavrck.co/2015s-most-engagingtopics-on-social-so-far/

[17] X. Hu, L. Tang, J. Tang and H. Liu, "Exploiting Social Relations for Sentiment Analysis in Microblogging," In Proceedings of the 6th ACM International Conference on Web Search and Data Mining (WSDM'13), ACM, New York, NY, 2013, pp. 537–546.

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

Bangla Handwritten Digit Recognition Using Autoencoder and Deep Convolutional Neural Network

Md Shopon
Department of Computer Science and Engineering
University of Asia Pacific
Email: shopon.uap@gmail.com

Nabeel Mohammed
Department of Computer Science and Engineering
University of Liberal Arts Bangladesh
Email: nabeel.mohammed@ulab.edu.bd

Md Anowarul Abedin
Department of Computer Science and Engineering
University of Liberal Arts Bangladesh
Email: anowarul.abedin@ulab.edu.bd

***Abstract*—Handwritten digit recognition is a typical image classification problem. Convolutional neural networks, also known as ConvNets, are powerful classification models for such tasks. As different languages have different styles and shapes of their numeral digits, accuracy rates of the models vary from each other and from language to language. However, unsupervised pre-training in such situation has shown improved accuracy for classification tasks, though no such work has been found for Bangla digit recognition. This paper presents the use of unsupervised pre-training using autoencoder with deep ConvNet in order to recognize handwritten Bangla digits, i.e., 0 - 9. The datasets that are used in this paper are CMATERDB 3.1.1 and a dataset published by the Indian Statistical Institute (ISI). This paper studies four different combinations of these two datasets - two experiments are done against their own training and testing images, other two experiments are done cross validating the datasets. In one of these four experiments, the proposed approach achieves 99.50% accuracy, which is so far the best for recognizing handwritten Bangla digits. The ConvNet model is trained with 19,313 images of ISI handwritten character dataset and tested with images of CMATERDB dataset.**

I. INTRODUCTION

This paper primarily aims at recognizing handwritten Bangla numeral digits. Bangla is the mother language of Bangladesh and the 7th most widely spoken language in the world [1]. There are more than 200 million native Bangla speakers. It is the official language of Bangladesh and several Indian states including West Bengal, Tripura, Assam and Jharkhand [2]. As application of optical character recognition (OCR) is widespread in these regions, recognizing handwritten Bangla digits is becoming more important [3].

Recognizing Bangla handwritten digits is more challenging than recognizing English digits because of their critical shapes and varied sizes. However, this is a classical image classification problem and datasets are important for this purpose. This constitutes another major challenge to the recognition of Bangla digits compared to that of English digits; there are very few datasets available in Bangla. CMATERDB 3.1.1 and ISI handwritten character dataset [4] are the most notable among them. The largest dataset of English handwritten digits is MNIST consisting 60000 images [5], whereas ISI, the largest Bangla dataset, has only 23299 images [6]. The best accuracy achieved on MNIST dataset so far is 99.79% [7], which is very close to how human would recognize. However, such excellent result is unprecedented in Bangla digit recognition.

Many research works were conducted on Bangla handwritten digits recognition [8] - [13]. C. L. Liu and C. Y. Suen proposed a benchmarked model with a reported accuracy rate of 99.4%. They worked on the ISI numeral dataset using advanced normalization techniques and gradient based feature extraction [6]. In [10], a hierarchical bayesian network was proposed to classify the images with an accuracy of 87.5%. This model was based on George and Hawkins' original implementation stated in [14]. A quad tree based feature set was used in [12]; the accuracy rate of this model was 93.38%. In [11], sparse representation classifier produced an accuracy rate of 94%. [13] worked with local binary pattern for classifying Bangla digits resulting an accuracy of 96.7%.

Neural networks are very versatile for the classification problems. [4] is one of the initial studies employing end-to-end neural network training in this context. A Le-net like architecture was used in [4] to classify Bangla digits of ISI numeral dataset. They achieved an accuracy rate of 98.375% on the test set. Recently, [15] used similar neural network on the ISI dataset. As it is a comparatively small dataset, they augmented the training images by rotating each image by (5°, 10°, 20° and 30°). However, their augmentation process was almost equivalent to hand fitting the images, as the change in rotation angles was done individually and then tested for validation accuracy. The best accuracy 98.98% was achieved only when the images were rotated by 10°. By doing so, the images were fine-tuned in such a way that the model learns the specific features which were applicable to the test set.

Convolutional neural networks (CNN) are well recognized models for handwritten digit recognition. [16] surveys different such approaches, but all of them are for English digits recognition [17]. The result discussed in [7] was also achieved using ConvNet. For Bangla digit recognition, this study finds that [15] is the only work that uses ConvNet for supervised learning. However, unsupervised pre-training along with supervised ConvNets has not been tried in any Bangla digit recognition work so far.

Autoencoders are used for unsupervised learning; An autoencoder has an input layer, a set of hidden layer and an output layer. In [19], denoising autoencoders were used for

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

extracting features from images. In [20], content based image retrieval was done with the help of deep autoencoders. [21] presented a detailed explanation of the efficacy of unsupervised pre-training for supervised learning. [22] applied autoencoders as a pre-training method on the MNIST dataset and achieved very competitive error rates. No such attempt can be found for Bangla character recognition. Thus, this paper uses autoencoders and deep ConvNet for classifying images of two widely used standard datasets - CMATERDB and ISI numeral dataset are used for the experiments.

The rest of the paper is organized as follows: Section II contains the background on ConvNet and autoencoders. Section III introduces our proposed approach, details of our used dataset and necessary pre-processing. Section IV describes different experiments we did and the methods followed for them. Section V discusses our results and the impact. Finally, conclusions are made in Section VI.

II. Background

A. ConvNet

Convolutional neural network [23] or ConvNet is a special kind of ANN (Artificial neural network). ConvNets have learnable weights and biases. Just like ANN, ConvNets are trained with the backpropagation algorithm though they have a different architecture from traditional multi layer perceptrons. There are many advantages of using ConvNets as an image classification tool, mainly due to the fact that they can be used as a feature learner and data classifier simultaneously. [16] showed that ConvNet based approaches outperform other more traditional techniques, e.g., SVM, KNN etc.

MNIST is one of the most popular dataset for handwritten digit classification. The best result for this dataset so far is 99.79% which was done by regularizing the neural networks using dropconnect [7]. Other than classification ConvNets are used for many different purposes, such as face recognition, speech recognition, natural language processing etc [23]. Le-Net5 is well known ConvNet architecture developed by Yann LeCun in 1998 [18]. Le-Net5 architecture was able to successfully classify digits and recognize the handwritten digits on bank cheques.

The first layer of a ConvNet is a Convolution layer with one or more kernels of fixed size. The input presented to these networks are usually images, and the kernels convolved with the input. The output of the convolution operations are passed through non-linearities, i.e., the Rectified Linear unit (ReLu) [24].

Pooling layers are also applied to outputs of the convolution operations. These are methods to reduce the data size, and can be done through averaging over a fixed region, choosing the maximum from a fixed region, or any other chosen method.

Dense layers are also known as fully connected layers which are used in the last stage of ConvNet. It connects the ConvNet to the output layer and constructs the expected number of outputs. For calculating the spatial dimensions of a ConvNet the following formula is used:

$$W_{out}(i) = 1 + \frac{W_{in} - R + 2P}{S}$$

Here i is the i^{th} input dimension, P is the value of padding, R_i is the receptive field and S is the value of stride.

B. Autoencoder

The functionality of an autoencoder is simply to output the input which it receives, at times, a transformed version of the input. This is not a difficult operation under normal circumstances. The utility of autoencoders derives from the fact that the input data is first transformed into a low dimensional representation (the encoding part). This low dimensional representation is then used to regenerate the input (the decoding part). This scheme enables autoencoders to learn useful representations of the input data without needing any labels a-priori.

Autoencoder can be implemented using various types of neural networks, e.g., ANN, ConvNet etc., depending on the desired representation scheme. When used as a pre-training method, the autoencoder is first trained in an unsupervised manner, enabling the encoding part of the autoencoder to adjust its weights to output useful low dimensional representations of the data. These weights are then combined with further layers and trained in a supervised manner. This approach has yielded good results for multiple classification tasks [21].

III. Proposed Approach and Dataset

A. Proposed Approach

As we discussed in Section I and II, results of recent studies on image classification problems using convolutional neural networks are very promising; traditional MLP is not sufficient for this purpose. So we propose to use Deep ConvNet that consists of more than one hidden layer. Figure 1 shows the proposed approach following a discussion about its parts.

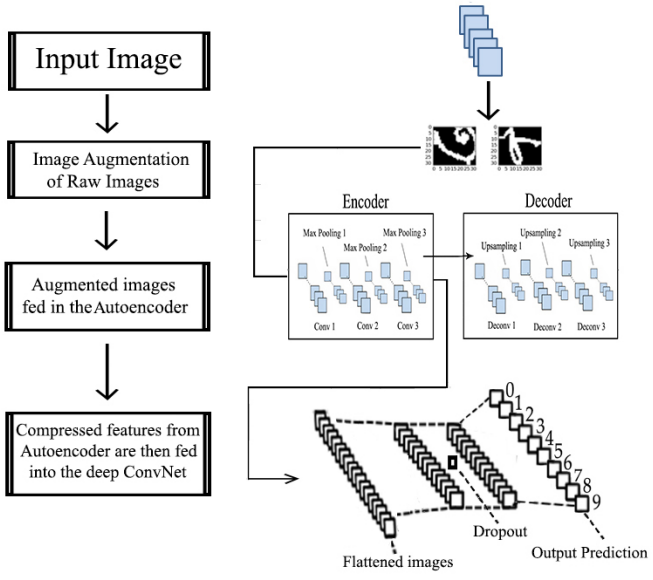


Fig. 1: Diagram of Our Proposed Approach

The encoder consists of 3 convolutional layers, each followed by a 2×2 max pooling layer. All three layers of the

encoder has 32 3×3 kernels. In between the layers dropout of 25% was used to reduce overfitting. The decoder has a similar architecture with each convolutional layer having 5 neurons, instead of 32. All the layers have RelU activation. Figure 2 shows the architecture of the autoencoder.

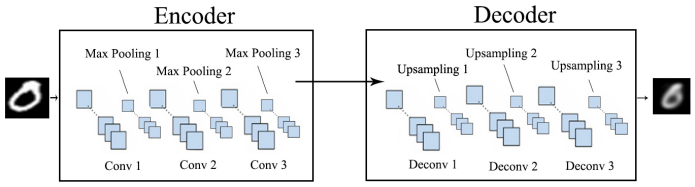


Fig. 2: Architecture of the Autoencoder

Figure 3 shows us the architecture of the convolutional model, where the first two convolutional layers are taken from the encoding part of the autoencoder, thus leveraging the weights learnt during unsupervised pre-training.

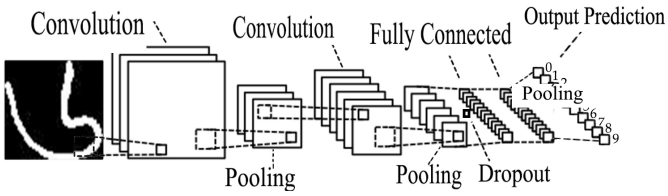


Fig. 3: Architecture of ConvNet Model

Bengali Digit	English Digit	CMATERDB	ISI Numeral
১	1		
২	2		
৩	3		
৪	4		
৫	5		
৬	6		
৭	7		
৮	8		
৯	9		
০	0		

Fig. 4: Digits for CMATERDB and ISI Numerals

B. *Dataset and Preprocessing*

Deep learning is a data-driven field and requires large data sets. We worked with two datasets with our proposed approach - CMATERDB and ISI dataset. CMATERDB consists a total of 6000 images. We divided the images into 4200 and 1800 for training and testing purpose. The ISI numeral dataset consists a total of 23,299 images. The dataset was divided into 19,313 and 3986 for training and testing purpose respectively.

Images of CMATERDB are 32×32 pixels each. The images were preprocessed before feature extraction in order to normalize the features. They were first converted into grayscale images. The images initially had white(0) background and black(255) foreground. They were converted into black background and white foreground for our experiments.

ISI numeral dataset also had the same properties except their image pixels was arbitrary. They were first reshaped into 32x32 pixels. After that, the same operations were done against this dataset just as done in CMATERDB. Figure 4 Shows the images from both the dataset after pre-processing.

To enhance the training sets, each image was randomly rotated between 0° and 50°. Each image was also shifted vertically by a random amount between 0 and 6 pixels. Horizontal shifts were also done in a similar range. These augmentations increased the size of the training sets significantly, allowing for reduced risk of overfitting. Figure 5 shows some samples of the augmented images

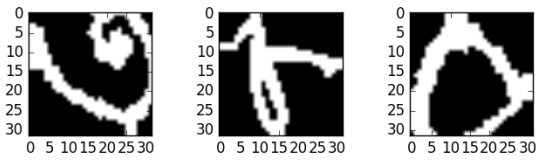


Fig. 5: Augmented images

IV. Experiments

The proposed model was trained in three different configurations: SCM stands for the first model we made which is Simple Convolutional Model, SCMA stands for Simple Convolutional Model with Augmented images and ACMA stands for Autoencoder with Convolutional Model with Augmented images. All the three configurations were trained and tested against CMATERDB and ISI numeral dataset, and cross-validated with each other. The parameter size is 81053 for both SCM and SCMA; for ACMA, it is 85770.

The autoencoder was trained for 40 epochs with the RMSProp optimiser using binary cross entropy as the cost function. It is also possible to use mean-squared-error as the loss function, however that was not attempted in this study. After the autoencoder is trained, the learnt weights are reused in the ConvNet used for supervised learning. The layers were not frozen at any stage and training continued to update the

weights from the first epoch of supervised training. Supervised training was done for 120 epochs, with the RMSProp optimiser using categorical-cross-entropy as the loss function.

V. Results and Analysis

For both datasets, ACMA gave the best accuracy rates, even in cross-dataset testing, demonstrarting the efficacy of unsupervised pre-training for this purpose.

Table 1 and 2 show the previously published results on the CMATERDB and ISI Character datasets. Table 3 summarises the accuracy rates obtained in different experiments. For the ISI character data set, the results achieved by our ACMA configuration is competetive at 98.29% accuracy. Infact, among all the ConvNet-based studies, our reported result is second only to that of [15], and we already have discussed the flaw in their approach in Section I. When an ACMA model is trained on the much smaller CMATERDB and tested on the ISI dataset, the results are still an impressive 97.29%, which is comparable to some previously published results shown in Table 2. To reinforce the usefulness of supervised pre-training, we can see that ACMA gives an improvement of $.47\%$ over SCMA with the same dataset augmentation in place.

TABLE I: Past results on CMATERDB

Work	Accuracy
Haider Adnan Khan et al. [11]	94%
Basu et al. [25]	95.1%
Hassan et al. [13]	96.7%
Basu et al. [26]	97.15%
Sarkhel el at. [27]	98.23%
Das et al. [28]	98.55%

TABLE II: Past results on ISI

Work	Accuracy
Nasir and Uddin [29]	96.80%
Wen and He [30]	96.91%
Das et al. [31]	97.70%
Akhnad et al. [32]	97.93%
Bhattacharya and Chaudhuri [4]	98.20%
CNNAP [15]	98.98%

TABLE III: Accuracy in different experiments

Dataset	SCM	SCMA	ACMA
Train:CMATERDB, Test:CMATERDB	96.59%	97.29%	98.61%
Train:ISI, Test:ISI	97.02%	97.43%	98.29%
Train:CMATERDB, Test:ISI	94.69%	96.82%	97.29%
TRAIN:ISI, TEST:CMATERDB	98.26%	98.76%	99.50%

When ACMA is trained on the CMATERDB and tested on CMATERDB, the accuracy rate achieved is 98.61%, which is better than the previously published results on this dataset as shown in Table 2. The accuracy rate of 99.50% on CMATERDB is obtained when ACMA is trained on the ISI dataset. As far as we could ascertain, this is the best reported result on the dataset, even better than the 98.61% accuracy achieved in this study. This is particularly interesting because the characteristics of the images in the two datasets are quite different, with the ISI images being smoother and less blocky compared to the CMATERDB images. This leads us to conclude that while the autoencoder used in this study was trained to reproduce images of each dataset individually, the representations learnt by this method generalise across different types of images.

VI. Conclusion

This paper presents the implementation of unsupervised pre-training, using an autoencoder, as a pre-cursor to supervised training for Bangla handwritten digit recognition. To demonstrate the effectiveness of this approach, we tested with three different training configurations across two different standard Bangla character datasets. For the test images of the ISI numeral dataset, the proposed approach achieves accuracy rates comparable with previous works. For the CMATERDB, it achieves an accuracy rate of 99.50%, which is the best reported result on this dataset so far. Apart from demonstrating the utility of unsupervised pre-training in the context of Bangla digit recognition, our results also indicate that such pre-training can be useful even when the data sets are independently (and blindly) collected. For every experiment, the model with autoencoder and ConvNet gives better accuracy rate than the model with only ConvNet. The proposed approach achieves state-of-the art results for CMATERDB and very good results for the ISI dataset. It is worth mentioning that previous studies on Bangla handwritten digit recognition rarely reported on these two datasets together. Future studies can explore whether pre-training on larger datasets, which are not specific to Bangla, can help achieve better results to be practically useful.

VII. Acknowledgment

This project is supported by the ICT Division, Ministry of ICT, Government of the People's Republic of Bangladesh (Project ID: 56.00.0000.028.33.066.16-731)

References

[1] Singh, V.K., "Most Spoken Languages in the world", 2012. [Online]. Available: goo.gl/fhTq2S. [Accessed: 20- Oct- 2016].

[2] Chatterji, S. K., "The origin and development of the Bengali language", 2002.

[3] Pal, U., and Chaudhuri, B. B., "OCR in Bangla: an Indo-Bangladeshi language." In Pattern Recognition, 1994. Vol. 2-Conference B: Computer Vision and Image Processing., Proceedings of the 12th IAPR International. Conference on, vol. 2, pp. 269-273. IEEE, 1994.

[4] Bhattacharya, U., and Chaudhuri, B. B., Handwritten numeral databases of indian scripts and multistage recognition of mixed numerals, In IEEE transactions on pattern analysis and machine intelligence, 31(3) on pp.444-457. IEEE,2009.

[5] LeCun, Y., Cortes, C., and Burges, C. J., "The MNIST database of handwritten digits.", 1998. APA

[6] Liu, C. L., and Suen, C. Y., A new benchmark on the recognition of handwritten Bangla and Farshi numeral characters, In Proc. 11 th ICFHR, 2008.

[7] Wan, L., Zeiler, M., Zhang, S., Cun, Y. L., and Fergus, R. , Regularization of neural networks using dropconnect, In Proceedings of the 30th International Conference on Machine Learning (ICML-13), pp. 1058-1066. 2013.

[8] Hossain M.Z., Amin M.A. and Yan H., "Rapid feature extraction for BangIa handwritten digit recognition", In Machine Learning and Cybernetics (ICMLC), 2011 International Conference on, vol. 4, pp. 1832-1837. IEEE, 2011.

[9] Sazal, M. M. R., Biswas, S. K., Amin, M. F., and Murase, K., Bangla Handwritten Character Recognition using Deep Belief Network, In Electrical Information and Communication Technology (EICT), 2013 International Conference on, pp. 1-5. IEEE, 2014. December 2013.

[10] Xu, J. W., Xu, J., and Lu, Y, Handwritten Bangla digit recognition using hierarchical Bayesian network, In Intelligent System and Knowledge Engineering, 2008. ISKE 2008. 3rd International Conference on, vol. 1, pp. 1096-1099. IEEE, 2008.

[11] Khan, H. A., Al Helal, A., and Ahmed, K. I., "Handwritten Bangla digit recognition using Sparse Representation Classifier," In Informatics, Electronics and Vision (ICIEV), 2014 International Conference on, pp. 1-6. IEEE, 2014.

[12] Roy, A., Mazumder, N., Das, N., Sarkar, R., Basu, S., and Nasipuri, M.,"A new quad tree based feature set for recognition of handwritten bangla numerals." In Engineering Education: Innovative Practices and Future Trends (AICERA), 2012 IEEE International Conference on, pp. 1-6. IEEE, 2012.

[13] Hassan, T., and Khan, H. A. "Handwritten Bangla numeral recognition using Local Binary Pattern," In Electrical Engineering and Information Communication Technology (ICEEICT), 2015 International Conference on, pp. 1-4. IEEE, 2015.

[14] George, D., and Hawkins, J. A Hierarchical Bayesian model of Invariant Pattern Recognition in the Visual Cortex, In Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005., vol. 3, pp. 1812-1817. IEEE, 2005.

[15] Akhand, M.I.H.R.I. M.A.H., Ahmed, M., Convolutional neural network training with artificial pattern for bangla handwritten numeral recognition, ICIEV, vol. 1, no. 1, pp. 16, 2016.

[16] Kamavisdar, P., Saluja, S., and Agrawal, S. "A Survey on Image Classification Approaches and Techniques", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 1, pp. 1005 1009, January 2013

[17] Bottou, L., Cortes, C., Denker, J. S., Drucker, H., Guyon, I., Jackel, L. D., and Vapnik, V., "Comparison of classifier methods: a case study in handwritten digit recognition." In International conference on pattern recognition, pp. 77-77. IEEE Computer Society Press, 1994.

[18] LeCun, Y., "LeNet-5, convolutional neural networks.", 2015. [Online] Available: http://yann.lecun.com/exdb/lenet. [Accessed: 20- Oct- 2016]

[19] Vincent, P., Larochelle, H., Bengio, Y., and Manzagol, P. A., "Extracting and composing robust features with denoising autoencoders." In Proceedings of the 25th international conference on Machine learning, pp. 1096-1103. ACM, 2008.

[20] Krizhevsky, A., and Hinton, G. E., "Using very deep autoencoders for content-based image retrieval." In ESANN. 2011.

[21] Erhan, D., Bengio, Y., Courville, A., Manzagol, P. A., Vincent, P., and Bengio, S., "Why does unsupervised pre-training help deep learning?," Journal of Machine Learning Research,pp. 625-660, February 2010.

[22] Huang, F. J., Boureau, Y. L., and LeCun, Y., "Unsupervised Learning of Invariant Feature Hierarchies with Applications to Object Recognition," In 2007 IEEE conference on computer vision and pattern recognition, pp. 1-8. IEEE, 2007.

[23] LeCun, Y., and Bengio, Y. "Convolutional networks for images, speech, and time series." The handbook of brain theory and neural networks 3361.10 (1995): 1995.

[24] Nair, V., and Hinton, G. E. "Rectified linear units improve restricted boltzmann machines." Proceedings of the 27th International Conference on Machine Learning (ICML-10). 2010.

[25] Basu, S., Sarkar, R., Das, N., Kundu, M., Nasipuri, M., and Basu, D. K., Handwritten bangla digit recognition using classifier combination through ds technique, In International Conference on Pattern Recognition and Machine Intelligence, pp. 236-241. Springer Berlin Heidelberg, 2005.

[26] Basu, S., Das, N., Sarkar, R., Kundu, M., Nasipuri, M., and Basu, D. K. , A novel framework for automatic sorting of postal documents with multi-script address blocks, Pattern Recognition, vol. 43, pp. 3507- 3521, 2010.

[27] Sarkhel, R., Das, N., Saha, A. K., and Nasipuri, M., A multi-objective approach towards cost effective isolated handwritten bangla character and digit recognition, Pattern Recognition, vol. 58, pp. 172189, 2016

[28] Das, N., Reddy, J. M., Sarkar, R., Basu, S., Kundu, M., Nasipuri, M., and Basu, D. K., A statisticaltopological feature combination for recognition of handwritten numerals, Applied Soft Computing, vol. 12, no. 8, pp. 24862495, 2012.

[29] Nasir, M. K., and Uddin, M. S., Hand written bangla numerals recognition for automated postal system, IOSR Journal of Computer Engineering, vol. 8, no. 6, pp. 4348, 2013.

[30] Wen, Y., and He, L. , A classifier for bangla handwritten numeral recognition, Expert Systems with Applications, vol. 39, no. 1, pp. 948 953, 2012.

[31] Das, N., Sarkar, R., Basu, S., Kundu, M., Nasipuri, M., and Basu, D. K., A genetic algorithm based region sampling for selection of local features in handwritten digit recognition application, Applied Soft Computing, vol. 12, no. 5, pp. 15921606, 2012.

[32] Rahman, M. M., Akhand, M. A. H., Islam, S., Shill, P. C., and Rahman, M. H., Bangla handwritten character recognition using convolutional neural network, I.J.Image, Graphics and Signal Processing(IJIGSP, vol. 7, no. 3, pp. 4249, 2015.

5

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

Pseudo Random Sequence over Finite Field using Möbius Function

Fatema Akhter
Jatiya Kabi Kazi Nazrul Islam University, Bangladesh
Email: fatema.kumu02@gmail.com

Yasuyuki Nogami
Okayama University, Japan
Email: yasuyuki.nogami@okayama-u.ac.jp

***Abstract*—Pseudo random sequences play an important role in cryptography and network security system. This paper proposes a new approach for generation of pseudo random sequence over odd characteristic field. The sequence is generated by applying a primitive polynomial over odd characteristic field, trace function and möbius function. Then, some important properties of the newly generated sequence such as period, autocorrelation and cross-correlation have been studied in this work. The properties of the generated sequence are evaluated on various bit length of odd characteristics. Finally, the experimental results are compared with existing works which show the superiority of the proposed sequence over existing ones.**

***Index Terms*—Finite field; primitive polynomial; trace function; möbius function**

I. Introduction

Pseudo random sequences with long period and high autocorrelation have many applications in spread spectrum communication, mobile communication and radar engineering [1]–[6]. Their applications are also significant in network security due to their cryptographic properties like linear complexity [7]–[9]. During the last two decades, random sequences such as Maximum Length Sequence [10], Legendre Sequences [11]–[13] and Sidelnikov Sequences [14]–[16] have been studied by many researchers for their importance in multi user communication, cryptography and network security. Primitive polynomials, trace and odd characteristics have become important parameters in these researches as they determine the scope and feature of the pseudo random sequences. Several methods and techniques have been studied in order to achieve true randomness. Unfortunately they are still in research to achieve the goal.

This paper proposes a new approach for generation of pseudo random sequence with a primitive polynomial, trace function, and möbius function over finite field of odd characteristic. Pseudo random sequences have some important properties such as period, autocorrelation, cross-correlation and linear complexity that characterize the applications of the sequence in communication or security. Unfortunately, it is familiar that these important properties cannot be proved theoretically. However, if they are defined over a finite field, $\mathbb{F}_{p^m}$, sometimes they can be proved [17]. If security is the major concern, it is said that period, autocorrelation and linear complexity of the sequence is preferred to be large while cross-correlation is preferred to be low. Therefore it is very important that these features must be studied while generation of the sequences, in particular for security use. This study proposes a new approach for generation of pseudo random sequence and then, presents the period, autocorrelation and cross-correlation of the generated sequence. Other properties e.g. linear complexity is not studied in this study, which will be in the future work.

In the extension field of $\mathbb{F}_{p^m}$ let, p be an odd characteristic and m be the degree of the primitive polynomial $f(x)$. $f(x)$ is primitive meaning that a maximum length sequence over $\mathbb{F}_{p^m}$ can be generated from the polynomial. Let ω be its zero, that becomes a primitive element in $\mathbb{F}^*_{p^m}$, then the sequence $\mathcal{S} = \{s_i\}$, $s_i = \mathrm{Tr}\left(\omega^i\right)$, $i = 0, 1, 2, 3, ..., (p^m - 2)$ becomes a maximum length sequence of period, $L = p^m - 1$, where $\mathrm{Tr}\left(\cdot\right)$ is the trace function over $\mathbb{F}_p$ [17]. In this proposed work, first a primitive polynomial is generated over odd characteristic, then a maximum length vector sequence is generated from the primitive polynomial, then trace function is applied on the vector sequence to map the vectors to scalars and finally, möbius function is applied on the scalars to generate the pseudo random sequence. Throughout this paper, let random sequence mean three values $\{-1, 0, +1\}$ corresponding to möbius function, p denotes an odd prime number and $\mathbb{F}_{p^m}$ denotes the finite field over p, where $m \geq 2$ is a positive integer representing extension degree. $\mathbb{F}^*_{p^m}$ is the the multiplicative group of $\mathbb{F}_{p^m}$ i.e. $\mathbb{F}^*_{p^m} = \mathbb{F}_{p^m} - \{0\}$. This paper presents all the terminologies and theorems based on the odd characteristic prime field $\mathbb{F}_{p^m}$. However, any arbitrary field can be adaptable as the base field.

The paper is organized as follows: Section II describes the preliminary studies for proposed approach with the necessary definitions and algorithms needed for implementation. Section III describes the proposal of pseudo random sequence with some examples. Section IV presents the experimental results and describes the period, autocorrelation and cross-correlation properties of the generated sequence. Finally, section V concludes the paper with some future recommendations.

II. Preliminaries

This section briefly describes the basic terminologies and mathematical fundamentals such as primitive polynomial, trace function, möbius function, autocorrelation and cross-correlation of pseudo random sequence.

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

A. Primitive Polynomial

A primitive polynomial [18], [19] of a finite field is a polynomial that can generate all the elements of the field. The multiplicative group, $\mathbb{F}_{p^m}^*$ of non-zero elements of $\mathbb{F}_{p^m}$ is cyclic and can be generated from a primitive element. Let g be a generator of $\mathbb{F}_{p^m}^*$ with elements $\{1, g, g^2, ..., g^{p^m-2}\}$, then every non-zero element is represented as g^i for $i = 0, 1, .., p^m - 2$. The element g^i is primitive if and only if $\gcd(i, p^m - 1) = 1$. $f(x)$ is a primitive polynomial if and only if the field element x generates the cyclic group of all non-zero field elements of the finite field $\mathbb{F}_{p^m}^*$. Therefore, $f(x)$ is primitive polynomial if and only if

1) $x^{p^m-1} \equiv 1 (\text{mod } f(x))$,
2) $x^k \not\equiv 1 (\text{mod } f(x))$ for $1 \leq k \leq p^m - 2$.

Property 1: Let g be a generator of $\mathbb{F}_{p^m}^*$, $g^{(p^m-1)/(p-1)}$ becomes a non-zero element in the prime field $\mathbb{F}_p$ and is also a generator of $\mathbb{F}_p^*$.

(Proof) Since g is a generator of $\mathbb{F}_{p^m}^*$ with order $p^m - 1$, if i is a non-negative integer, the order of g^i is given by

$$\frac{p^m - 1}{\gcd(i, p^m - 1)}. \tag{1}$$

Then, the order of $g^{(p^m-1)/(p-1)}$ becomes $p - 1$. Therefore, $g^{(p^m-1)/(p-1)}$ is a generator of $\mathbb{F}_p^*$.

B. Trace Function

Trace function [20] $\text{Tr}(\cdot)$ is used to map an extension field element to prime field element. Let X be an extension field element of $\mathbb{F}_{p^m}$, this can be map to prime field element x of $\mathbb{F}_p$ as

$$x = \text{Tr}(X) = \sum_{i=0}^{m-1} X^{p^i}. \tag{2}$$

The trace function has the linearity on $\mathbb{F}_p$ when $a, b \in \mathbb{F}_p$ and $X, Y \in \mathbb{F}_{p^m}$ as follows:

$$x = \text{Tr}(aX + bY) = a\text{Tr}(X) + b\text{Tr}(Y). \tag{3}$$

Property 2: For each $i = 0, 1, 2, ..., p - 1 \in \mathbb{F}_p$, the number of elements in $\mathbb{F}_{p^m}$ whose trace with respect to $\mathbb{F}_p$ becomes i is given by p^{m-1}.

(Proof) Elements in $\mathbb{F}_{p^m}$ are the roots of $x^{p^m} - x$. It is factorized over $\mathbb{F}_p$ as follows

$$x^{p^m} - x = \prod_{i=0}^{p-1} (\text{Tr}(x) - i). \tag{4}$$

Since the degree of $\text{Tr}(x)$ is p^{m-1} and $\text{Tr}(x)$ does not have any duplicate root, the property is shown. Therefore, the number of elements in $\mathbb{F}_{p^m}^*$ whose trace becomes i is p^{m-1}.

C. Möbius Function

Möbius function [21], denoted as $\mu(\cdot)$ can translate any natural number to any of the values, $\{-1, 0, +1\}$. For any $n \in \mathbb{N}$, where $\mathbb{N}$ is natural number, möbius function $\mu(\cdot)$ is defined as:

$$\mu(n) = \begin{cases} 1, & \text{if } n = 1, \\ (-1)^k, & \text{if } n \text{ is a product of } k \text{ distinct primes}, \\ 0, & \text{if } n \text{ is divisible by square of a prime.} \end{cases} \tag{5}$$

This paper uses the möbius function to translate the prime field element x of $\mathbb{F}_p$ to the sequence element $s_i \in \{-1, 0, +1\}$. The möbius function takes only natural number, $n \in \mathbb{N} : n > 0$ and has no definition for $n = 0$. As 0 is an element in prime field $\mathbb{F}_p$, this paper modifies the definition of möbius function on prime field $\mathbb{F}_p$ as follows:

$$\mu(n) = \begin{cases} 1, \text{if } n = 1, \\ (-1)^k, \text{if } n \text{ is a product of } k \text{ distinct primes}, \\ 0, \text{if } n = 0 \text{ or } n \text{ is divisible by square of a prime.} \end{cases} \tag{6}$$

For $n \leq 11$, the möbius function values are:

n	0	1	2	3	4	5	6	7	8	9	10	11
μ	0	1	-1	-1	0	-1	1	-1	0	0	1	-1

D. Autocorrelation

This paper presents the pseudo random sequence as: $\mathcal{S} = \{s_i\}$, for $i = 0, 1, 2, 3, ..., (L - 1)$, where $s_i \in \{-1, 0, +1\}$ and $L = p^m - 1$ is the period of the sequence, i.e. $s_i = s_{i+L}$. Then, the autocorrelation, sometimes called the periodic autocorrelation, $\mathcal{R}_{\mathcal{S}}(x)$ of the sequence $\mathcal{S}$ after shifted x position is defined as follows:

$$R_S(x) = \sum_{i=0}^{L-1} s_i s_{i+x}. \tag{7}$$

for all $x = 0, 1, ..., L-1$, where s_i and s_{i+x} are the ith element and the x position shifted element of the generated sequence $\mathcal{S} = \{s_i\}$.

E. Cross-correlation

Let $\mathcal{S} = \{s_i\}$ and $\mathcal{T} = \{t_i\}$ be two random sequences with same length, $L = p^m - 1$, then this paper defines the cross-correlation, $\mathcal{C}_{\mathcal{S},\mathcal{T}}(x)$ of $\mathcal{S}$ and $\mathcal{T}$ as follows:

$$\mathcal{C}_{\mathcal{S},\mathcal{T}}(x) = \sum_{i=0}^{L-1} s_i t_{i+x}. \tag{8}$$

for all $x = 0, 1, ..., L - 1$, where s_i and t_{i+x} are the ith element of sequence $\mathcal{S}$ and x position shifted ith element of $\mathcal{T}$ respectively.

III. Proposed Pseudo Random Sequence

This section describes the proposed pseudo random sequence using primitive polynomial, trace function and möbius function. Then, this section describes the implementation procedure of the proposed psudo random sequences with some examples.

A. Proposal of Pseudo Random Sequence

Let ω be a primitive element in the extension field $\mathbb{F}_{p^m}$. The proposed pseudo random sequence $\mathcal{S} = \{s_i\}$ is presented as:

$$\mathcal{S} = \{s_i\},\ s_i = \mu\left(\mathrm{Tr}\left(\omega^i\right)\right) \tag{9}$$

for $i = 0, 1, 2, 3, ..., (p^m - 2)$, where $\mathrm{Tr}(\cdot)$ represents the trace function and $\mu(\cdot)$ represents the möbius function.

The period of the generated sequence and the number of peaks in autocorrelation are given below:

- Period, $L = p^m - 1$
- Number of peaks in autocorrelation = $p - 1$

It is noted that the period of the generated sequence and the number of peaks of autocorrelation are given from the experimental results. The theoretical proof is not studied in this paper, which will be in our future work.

B. Implementation Procedure

This section briefly describes the implementation procedure of the proposed pseudo random sequence. The implementation algorithm of the proposed pseudo random sequence is given below:

Pseudo Random Sequence Generation

① Generate primitive polynomials over finite field $\mathbb{F}_{p^m}$. There are $n = \dfrac{\varphi(p^m - 1)}{m}$. primitive polynomials, where $\varphi(\cdot)$ represents the *euler totient function*.
② Select one of the primitive polynomials generated in the above step. This is the generator primitive polynomial, $f(x)$.
③ From the primitive polynomial, $f(x)$, generate all the vectors over finite field $\mathbb{F}_{p^m}$. There are $2^m - 1$ vectors in the finite field $\mathbb{F}_{p^m}$.
④ For each vector $X^i \in \mathbb{F}_{p^m}$ for $i = 0, 1, 2, ..., p^m - 2$ do the followings:
   a) Map X^i to prime field element $x_i \in \mathbb{F}_p$ using the trace function defined in Eq. (2) as follows: $x_i = \mathrm{Tr}\left(X^i\right)$
   b) Translate prime field element $x_i \in \mathbb{F}_p$ to pseudo random sequence element $s_i \in \mathcal{S}$ using the möbius function defined in Eq. (6) as follows: $s_i = \mu(x_i)$.
⑤ Finally output the pseudo random sequence $\mathcal{S} = \{s_i\}$.

C. Examples of the sequence

Let $p = 3$, $m = 2$ and $f(x) = x^2 + x + 2$ be the primitive polynomial over $\mathbb{F}_{3^2}$. In this case,

Tr	0	1	2
μ	0	1	-1

Then, the generated pseudo random sequence is:

$$\mathcal{S} = \{-1, -1, 0, -1, 1, 1, 0, 1\}. \tag{10}$$

Let $p = 5$, $m = 2$ and $f(x) = x^2 + x + 2$ be the primitive polynomial over $\mathbb{F}_{5^2}$. In this case,

Tr	0	1	2	3	4
μ	0	1	-1	-1	0

Then, the generated pseudo random sequence is:

$$\begin{aligned}\mathcal{S} = \{&-1, 0, -1, 0, 1, 0, 0, -1, 0, 0, -1, -1,\\ &-1, 1, -1, 0, 0, 1, 1, -1, 1, 0, -1, -1\}.\end{aligned} \tag{11}$$

IV. Results and Discussions

This section describes the features of proposed pseudo random sequence such as period, autocorrelation, cross-correlation and compare with some popular existing works.

A. Period of Sequence

The proposed pseudo random sequence, $\mathcal{S} = \{s_i\}$ has the period of $L = p^m - 1$ for finite field $\mathbb{F}_{p^m}$. For $p = 5$ and $m = 3$, there are 20 primitive polynomials. In this paper, 2 primitive polynomials are used to generate squences. The generated sequence and the period of the sequence for each primitive polynomial is given below:

- $f_1(x) = x^3 + x^2 + x + 3$
  Period, $L = 124$

$$\begin{aligned}\mathcal{S} = \{&-1, 0, 0, -1, 1, 0, 1, -1, 0, 0, 0, 1, 0,\\ &-1, 0, -1, 1, 1, 0, -1, 1, 0, -1, 0, -1,\\ &0, 1, 1, 1, 0, 1, 1, -1, -1, 1, -1, -1,\\ &-1, 0, 0, 0, -1, -1, 0, 0, 0, 1, -1, -1,\\ &-1, 0, -1, 0, 1, -1, 1, -1, -1, -1, -1,\\ &0, -1, -1, 1, 1, -1, 0, 1, 0, -1, 0, 0, 1, 0,\\ &0, -1, 0, -1, 0, 0, 1, -1, 0, 0, -1, 1, -1,\\ &1, 0, 0, 0, 0, 0, 0, -1, -1, 0, -1, -1, -1,\\ &1, 0, 0, -1, -1, 0, 1, 0, 0, -1, -1, -1, 1,\\ &-1, 0, 0, -1, 0, -1, -1, -1, -1, 0, -1\}.\end{aligned} \tag{12}$$

- $f_2(x) = x^3 + 3x^2 + x + 2$
  Period, $L = 124$

$$\begin{aligned}\mathcal{S} = \{&-1, -1, -1, 1, 1, -1, 1, -1, 1, -1, -1,\\ &0, 0, 0, 0, 1, 0, 1, 0, 0, 1, -1, -1, 0, 0, -1,\\ &1, 0, 0, -1, 1, 0, 1, 1, -1, -1, 1, -1, 0, -1,\\ &1, 1, 0, -1, -1, -1, -1, 0, -1, 0, -1, -1,\\ &0, 1, -1, 0, 1, -1, 0, 0, 0, -1, -1, -1, -1,\\ &0, 0, -1, 0, -1, 0, -1, -1, 0, 1, 1, 1, 0, 0,\\ &0, 0, 1, 0, -1, -1, 1, 0, -1, 0, 0, 0, -1, 0,\\ &1, 0, 0, -1, -1, 0, -1, 1, -1, 0, 0, 0, -1,\\ &-1, -1, -1, 0, -1, 0, -1, -1, 1, 0, -1,\\ &0, 0, -1, 0, 0, 1, -1\}.\end{aligned} \tag{13}$$

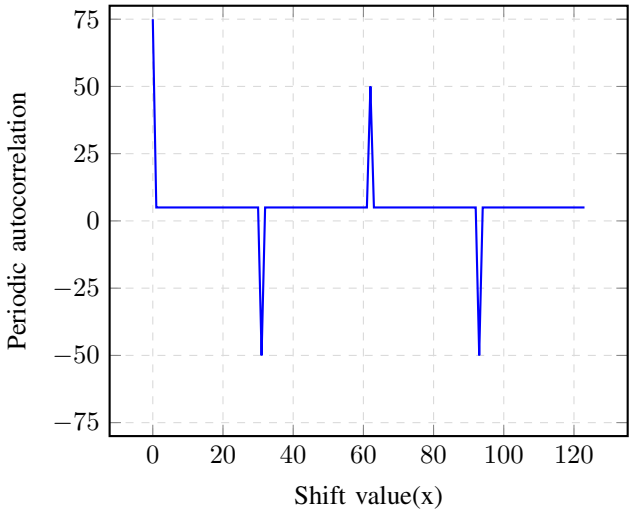


Fig. 1: Autocorrelation for the sequence, $p = 5$ and $m = 3$.

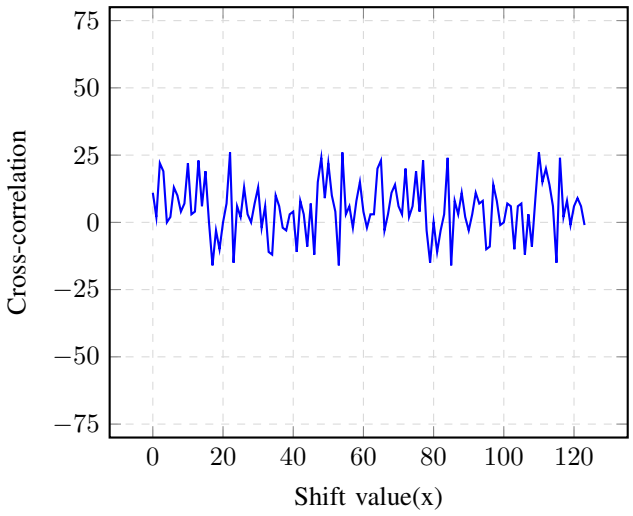


Fig. 2: Cross-correlation between sequence, $\mathcal{S}$ and $\mathcal{T}$.

B. Results of Autocorrelation

The periodic autocorrelation of the sequence represented at Eq. (12) is presented in Eq. (14) and shown in Fig. 1.

$$\mathcal{R}_{\mathcal{S}}(x) = \begin{cases} 75 & \text{if } i = 0 \\ -50 & \text{else if } i = 31, 93 \\ 50 & \text{else if } i = 62 \\ 5 & \text{otherwise} \end{cases}. \quad (14)$$

C. Results of Cross-correlation

To study the cross-correlation property of the proposed sequence, sequence $\mathcal{S} = \{s_i\}$ is generated from $f_1(x)$ while sequence $\mathcal{T} = \{t_i\}$ is generated from $f_2(x)$. The cross-correlation of the two sequence is shown in Fig. 2.

D. Comparisons with Existing Approaches

The proposed random sequence $\mathcal{S} = \{s_i\}$ has the period, $L = p^m - 1$ over odd characteristic $\mathbb{F}_{p^m}$. This is the maximum period that can be achieved on $\mathbb{F}_{p^m}$ and is equal to the length of maximum value sequence. The proposed sequence generated using möbius function has not been reported anywhere in the literature yet, where our goal is to generate random sequence with maximum period and autocorrelation value. The original Legendre sequence has the length of p which is extended by [17], [22], [23] later. Sequence generated in [22] has a period of length, $L = \frac{k(p^m - 1)}{p - 1}$[1]. A comparison result with [22] is shown in Fig. 3. The proposed sequence has larger period than that of [22] for various bit length of characteristic.

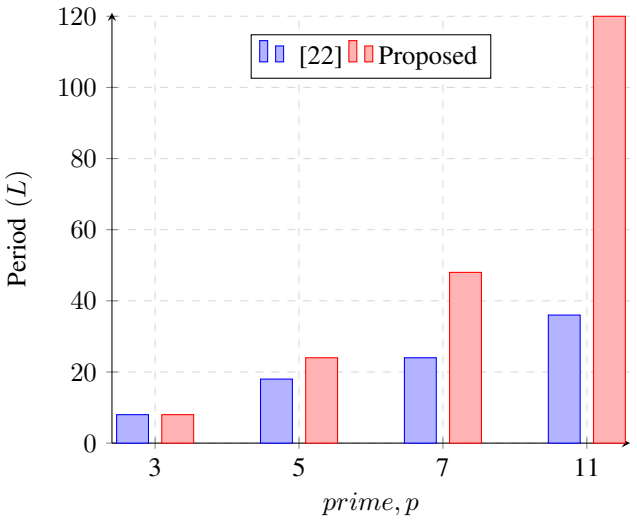


Fig. 3: Comparison of period between [22] and proposed sequence.

V. Conclusion

This paper proposes a new approach for generation of pseudo random sequence using primitive polynomial over odd characteristic field $\mathbb{F}_{p^m}$, trace function and möbius function. Appropriate algorithms and theorems are presented to clarify the proposed approach through parameterization and examples. Then, this paper presents some important features of the newly generated sequence such as period, autocorrelation and cross-correlation for $\mathbb{F}_{5^3}$. The results are finally compared with some popular approaches for the effectiveness of the proposed sequence. The comparison results show the superiority of the proposed sequence over existing ones in terms of period, autocorrelation and cross-correlation. In future, the proposed approach can be extended in several ways. Firstly, the sequence can be binarization in order to generate binary sequence. Then some other important properties such as linear complexity, distribution of bit pattern can be studied.

[1]Proposed $\mathcal{S} = \{s_i\}$ takes only 3 values, $s_i \in \{-1, 0, 1\}$, this paper considers only the existing work with 3 values, i.e. $k = 3$ in [22] for comparison.

ACKNOWLEDGMENT

The authors would like to express their greatest gratitude to the anonymous reviewers for their constructive feedback and critical suggestions that helped significantly to elicit the utmost technical attribute of this research work.

REFERENCES

[1] P. Banerjee, G. Raj, K. Attrey and M. Kaushik, "Usefulness of truncation of full length pseudo random sequence for CDMA communication," 2nd International Conference on Signal Processing and Integrated Networks (SPIN), Noida, pp. 819-822, Feb. 2015.

[2] Pankaj, A. K. Singh and B. S. Bora, "Design of enhanced pseudo-random sequence generator usable in GSM communication," International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), pp. 530-534, Mar. 2016.

[3] O. Reyad and Z. Kotulski,"Pseudo-random sequence generation from elliptic curves over a finite field of characteristic 2," Federated Conference on Computer Science and Information Systems (FedCSIS), Gdansk, pp. 991-998. Sept. 2016.

[4] A. Rui and Z. Xuefeng, "Pseudo-random sequence generating method based on high dimensional cat map," Computer Engineering and Applications, Oct. 2016

[5] M. K. Simon, J. K. Omura, R. A. Scholtz and B. K. Levitt, "Spread spectrum communications handbook," Computer Science Press, McGraw-Hill, New York, Apr. 1994.

[6] P. Fan, and M. Darnell, "Sequence design for communications applications," John Wiley & Sons, New York, 1996.

[7] T. W. Cusick, C. Ding and A. Renvall, "Stream ciphers and number theory," Elsevier, Amsterdam, Apr. 1998.

[8] X. Du. and Z. Chen, "Linear complexity of quaternary sequences generated using generalized cyclotomic classes modulo $2p^*$," IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, vol. E94-A, no. 5, pp. 1214 - 1217, May 2011.

[9] Y. T. Kim, M. K. Song, D. S. Kim, and H. Y. Song, "Properties and cross-correlation of decimated sidelnikov sequences," IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, vol. E97-A, no. 12, pp. 2562 - 2566, Dec. 2014.

[10] S. W. Golomb, "Shift register sequences," Holden-Day, San Francisco, 1967.

[11] N. Zierler, "Legendre Sequence," Massachusetts Inst. of tech. (MIT) Lexington Lincoln La, Lincoln Publications, 1958.

[12] J. S. No, H. K. Lee, H. Chung, H. Y. Song and K. Yang, "Trace representation of legendre sequences of mersenne prime period," IEEE Transactions On Information Theory, vol. 42, no. 6, pp. 2254 - 2255, Nov. 1996.

[13] C. Ding, T. Hesseseth, and W. Shan, "On the linear complexity of legendre sequences," IEEE Transactions On Information Theory, vol. 44, no. 3, pp. 1276 - 1278, May 1998.

[14] N. Y. Yu and G. Gong, "New construction of M-ary sequence families with low correlation from the structure of sidelnikov sequences," IEEE Transactions On Information Theory, vol. 56, no. 8, pp. 4061 - 4070, Aug. 2010.

[15] M. Su and A. Winterhof, "Autocorrelation of legendre sidelnikov sequences," IEEE Transactions On Information Theory, vol. 56, no. 4, pp. 1714 - 1718, Apr. 2010.

[16] Y. T. Kim, D. S. Kim and H. Y. Song, "New M -ary sequence families with low correlation from the array structure of sidelnikov sequences," IEEE Transactions On Information Theory, vol. 61, no. 1, pp. 655 - 670, Nov. 2014.

[17] Y. Nogami, K. Tada and S. Uehara, "A geometric sequence binarized with legendre symbol over odd characteristics field and its properties," IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences., vol. E97-A, no. 12, pp. 2336 - 2342, Dec. 2014.

[18] T. Hansen, and L. M. Gary, "Primitive polynomials over finite fields," Mathematics of Computation vol. 59, no. 200 pp. 639-643, 1992.

[19] E. W. Weisstein, "Primitive polynomial," From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/PrimitivePolynomial.html.

[20] A. A. Nechaev, and A. K. Alexey, "Trace-function on a Galois ring in coding theory," International Symposium on Applied Algebra, Algebraic Algorithms, and Error-Correcting Codes. Springer Berlin Heidelberg, 1997.

[21] M. Abramowitz, and I. A. Stegun, "The Möbius function," Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing. New York: Dover, pp. 826, 1972.

[22] H. Ino, Y. Nogami, N. Begum, S. Uehara, R. M. Zaragoza and T. Kazuyoshi, "A consideration on crosscorrelation of a kind of trace sequences over finite field," 3rd. International Symposium on Computing and Networking (CANDAR), pp. 484 - 486, Dec. 2015.

[23] Y. Nogami, K. Tada and S. Uehara, "A binarization of geometric sequences with legendre symbol and its autocorrelation," The 6th International Workshop on Signal Design and Its Applications in Communications, pp. 28-31, Oct. 2013.

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

Transparency- A Key Feature Integration in Existing Privacy Frameworks for Online User

[1]Seraj Al Mahmud Mostafa, [1]Sheak Rashed Haider Noori and [2]Saujanna Jafreen
[1]Department of Computer Science and Engineering, [2]Department of Natural Sciences
[1,2]Faculty of Science and Information Technology
Daffodil International University, Dhaka-1207, Bangladesh
{seraj.mostafa@diu, drnoori@daffodilvarsity, jafreen.ns@diu}.edu.bd

***Abstract*—Privacy in Online Social Network (OSN) is a very important issue in users' data sharing. Users' often share their personal information in different forms such as public, private, or in groups for developing networks and/or sharing their views. OSN applications collect users information with their permission by offering social networking services. In some cases OSN uses users data for personalization, ads/campaigning, and service quality analysis purposes. In worst cases, OSN can share information with third parties for legal and/or business purposes. In a way, there are potential risk of privacy failure during data sharing might happen. This risks are mostly underestimated or ignored by the OSN users due tricky and hard to understand privacy policy information. In this paper, we put emphasis on understanding transparency as key ingredient for privacy policy development, and have proposed to integrate Transparency as key feature to the current privacy frameworks to ensures that users have understanding what kind of data they are sharing and how it is managed over the OSN.**

***Index Terms*—Privacy; Transparency; Online Social Network; OSN; Data Sharing; Social Services; Online Services;**

I. Introduction

Internet has changed the pattern of human lifestyle significantly. People adopt new ways of interactions in this obvious change, such as online business, reading newspaper, watching movies, maintaining social communications, shopping and so on. In general, users have their own reasons to be online and they find them comfortable in fulfilling their needs. In order to meet the demand of the consumers, many new services are being evolved. In this process of development, there are many social network based applications, offering exciting services to the online users. For example, people search jobs and maintain professional network by creating a profile in professional network e.g., LinkedIn. Similarly, effective social network is being maintained by sharing ideas, thoughts, pictures, and videos by using social media such as, Facebook and Google+. Personalized recommender system based services are also being popular, such as Spotify, Netflix, Amazon and YouTube.

The consumers information is collecting by the OSN itself, third party add-on or apps by providing some free services (e.g., games, books, TV shows, etc). Exchanging of this information is allowed through users permission where users are mostly unaware or they simply ignore the privacy matter of their data. In most cases, users have no idea whether their data is shared publicly or with third party without the owner consent. In fact, users have no control on their personal data once they are exposed.

Ease of use, speed and low cost make internet the most popular means of communication. Due to that, users are in need of revealing information to get desired services of OSN which is a trade-off between functionality and user privacy, and lose their private data which is not desirable. In this paper we addressed the above discussed issues with following specific research questions:

1. Do user know what is happening to their shared data?

2. Do user trust the existing OSN service providers and their third parties?

3. Are the current frameworks transparent enough in terms of privacy?

In the following sections we analyze the above questions and propose the transparency feature that addresses the lack of existing frameworks. Section II listed OSN related privacy frameworks and section III discusses the key feature. In Section IV a prototype implementation of our proposed framework is explained. A general discussion took place in section V followed by conclusion in section VI.

II. OSN Related Privacy Frameworks

Various services are offered by OSN as well as their third party applications such as, candy crush, clash of clans, angry birds on Facebook. Generally, these services or applications are freely offered to the online users. As the online users are massive in number and continuously increasing, different online portals are also extending their various services extensively. The online users enjoy these free services in an exchanging of their personal information. However, the offered services are mostly user centric where the service provider collects user information and their behavioral data. In most cases the user does not have control over their data due to lack of transparent data and service sharing policies of the existing privacy frameworks. Guo and Chen proposed a framework to help the users to know their optimal privacy setting with a sensitivity value for a particular level of privacy concern [1]. This feature allows user to find whether they have the

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

desired level of protection on their data. Besmer and Lipford examined motivations, intentions, and concerns of user when they engage with applications, as well as their perceptions of data sharing [2]. Wills and Krishnamurthy claimed that users personal identifiable information is leaked through OSN sites and is shared with third-party aggregators and advertising partners [3]. Leon et al., examined how online users privacy is violated [4]. They calculated the time length of data held by other parties, user accessibility to modify data, range of websites for advertising and the websites collecting user data. Recent research on information privacy is increasingly focusing on privacy leakage and security matters associated with third-party apps on different mobile platforms [5], [6], [7] including Android, Symbian and iOS. Synnes K., et al. [8] and J. Rana [9] pointed the complexities over OSN based on user communication pattern. Now a day people are mostly on mobile devices to be online where the device information are also collected e.g., microphone, camera, wifi, sms, photos/media along with IP and MAC addresses. These policies are also prone to vulnerabilities in terms of privacy during data sharing. However, none of the frameworks provide transparency mechanism that allow user to get a clear idea of their data sharing purposes. Therefore, our approach is to add the transparency feature to the existing frameworks that would clearly explain the transparency between user and service provider.

III. Transparency as a key feature for Privacy Frameworks

The main purpose of this research is to add transparency as a key feature in the privacy framework where both the users and service providers can have a clear understanding of shared items. In this framework transparency will ensure the features like, **(i)** what data is taken for availing a particular service, **(ii)** sending alerts to user if personal data is shared with unknown parties, and **(iii)** having the control on personal data even after sharing. We decided to add these key features from the findings of our survey by analyzing the user data sharing behavior over the internet. We conducted an online survey which carried out in November 2014 over a four week long period, with a total of 1150 participants between the age of 18 and 55 years in a South-east Asian country. The focus of the survey was to know the user behavior on sharing personal data while using internet based services. The survey consisted of 40 primary questions with 22 demographic questions including 101 variables. In order to ensure the respondent privacy, participation in the follow-up sample was completely optional. Additionally, providing an email address was not required and cookies were not used. We had a wide range of queries to our survey participants regarding internet usage, their privacy concern, data collection and usability by different online portals and the importance of protecting private data while sharing with others. We present the following case studies that actually focus on the user data sharing behavior.

	Yes	No	Don't know	Average	Responses
Global Internet companies (e.g. Google, Microsoft, Amazon)	475 42.6%	391 35.1%	248 22.3%	0.0	1,114
Social media sites (e.g. Facebook, Twitter, Instagram)	406 36.4%	442 39.6%	267 23.9%	0.0	1,115
National Internet companies (e.g. bikroy.com)	232 20.8%	478 42.9%	403 36.2%	0.0	1,113

Fig. 1. User trust level on different media

	Never	At least once in a year	At least once a month	At least once a week	Once a day	Multiple times per day	Average	Responses
Facebook	31 2.7%	26 2.3%	14 1.2%	37 3.3%	100 8.8%	924 81.6%	0.0	1,132
GMail/Yahoo Mail	57 5.0%	32 2.8%	111 9.8%	168 14.9%	250 22.1%	513 45.4%	0.0	1,131
Mobile Phone Calls	18 1.6%	11 1.0%	4 0.4%	4 0.4%	26 2.3%	1,069 94.4%	0.0	1,132

Fig. 2. User availability on different media

Case study 1: **Do User know what is happening to their private data?** We posed two questions to determine the users understanding of the OSNs policies. For the questions ”Did you know Facebook has announced it will start collecting information about what you do in other partner websites and mobile apps, and will use that information to show you on Facebook ads tailored to your interests?” and ”Did you know Google+ can use the pictures you upload to Google+ to advertise its own products and services?”, we found, 65% of the users are unaware about their personal data usage, around 56% users care for their data whereas 32% users do not care on privacy at all.

Case study 1 findings: Most of the online users do not know how their private data is shared among OSN and third parties. Research shows, 65% of users have no idea on how their data is shared and 56% among the users care about their privacy.

Case study 2: **Do User Trust Existing OSN services in terms of Privacy?**

In this part we focused on user trust level on OSN services. From the question ”Do you trust that the following service providers do not misuse the data you leave when you register and use their services and apps? 1, we found different view points from the user like, 42% people trust the global internet companies whereas more than 57% are not well concerned. Interestingly, 63% of the users do not trust the most popular online social sites but rest of them do. 80% of the users feel insecure in trust level which is remarkably higher.

Case study 2 findings: A significant number of users have put their thoughts on the trust level where maximum users cannot rely on the OSN sites, meanwhile few of them do.

Case study 3: **Are the Online Interactions Transparent?** When ”How often do you use the following tools and services?” was asked we got more than 81% respondents spending their time on Facebook 2 and a big numbers of third party apps are very active in this platform, so we took Facebook users as our test case. Another question was ”From the standpoint of personal privacy, please indicate to what extent you agree with each of the following statements” 3.

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Average	Responses
It usually bothers me when websites and mobile apps automatically collect data about how I use them, such as what I click on	51 4.6%	215 19.2%	185 16.6%	472 42.3%	194 17.4%	0.0	1,117
It bothers me to give personal information, such as name and email, to so many websites and mobile apps	49 4.4%	151 13.6%	210 18.9%	506 45.4%	198 17.8%	0.0	1,114
A good online privacy policy should have a clear and easily noticeable disclosure	34 3.1%	104 9.3%	125 11.2%	496 44.6%	354 31.8%	0.0	1,113
It is very important to me that I am aware and well informed about how my personal data will be used	70 6.3%	60 5.4%	131 11.8%	471 42.3%	381 34.2%	0.0	1,113
I would like to have control over what personal data websites and mobile apps can collect about me	69 6.2%	55 5.0%	148 13.3%	451 40.7%	386 34.8%	0.0	1,109
For me, it is important to have control over how my personal data is used by websites and mobile apps	27 2.4%	107 9.7%	154 13.9%	452 40.9%	366 33.1%	0.0	1,106
I would like to have control over who can access the data websites and mobile apps collect about me	21 1.9%	55 5.0%	145 13.2%	528 47.9%	353 32.0%	0.0	1,102
For me, it is important to have the possibility of permanently deleting any personal data previously collected by a website or mobile app	34 3.1%	121 10.9%	152 13.7%	445 40.2%	355 32.1%	0.0	1,107
I would like to have control over what personal data websites and mobile apps can collect about me	69 6.2%	55 5.0%	148 13.3%	451 40.7%	386 34.8%	0.0	1,109

Fig. 3. User view on data sharing policies

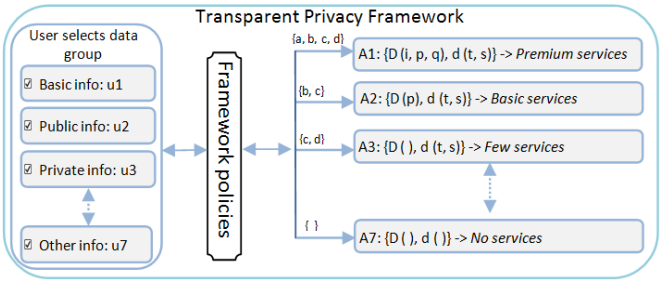


Fig. 4. Proposed Transparent Data Privacy Framework

Around 63% of the users do not want to be interrupted by any external services. On the other hand 76% users are aware of their data and the same percentage among them wants a transparent agreement on the data sharing. It is also seen that approximately 80% of the respondents want a control over their data during sharing and also need the right of accessibility on their data stored by other parties.

Case study 3 findings: Those who are aware of their privacy, they do not want to face the lack of security to their personal data and they want a clear visibility of their shared data.

At the end of the survey, we came up with a conclusion that, the OSN user enjoying services are not completely aware of privacy. Those who are aware have no control on their shared data and they are quite worried with existing privacy framework. Considering all these issues the proposed feature explained in the following chapter to assure privacy with transparency for sharing data and services.

IV. Proposed Transparency Feature

The previous sections discussed the transparency feature with a goal of transparent data sharing. This section describes the data sharing policies step by step to ensures transparency between user and service provider:

Step 1: The data items are divided into strong data set ***D*** and weak data set ***d***. Strong data sets are the mandatory field to get premium services and weak data sets are less priority data to avail filtered services. Strong and weak data set are defined by the service provider based on their feature and services they offer. User must agree on those definitions with multiple options if they want to enjoy a service. On the other hand user information is grouped into multiple layers such as: ***u1, u2, u3, u4*** etc. 4 that user can select during availing any service. So, the types of feature a user can receive depend on what data sets are provided to the service provider. As an example, if all the data set of a user are selected like: ***u1, u2, u3, u4*** in the ***D*** category defined by the service provider, then the user is able to enjoy a premium ***A1*** service but if the user decides to select few data set in ***d*** category, the user will avail limited feature that is a basic ***A2*** service. Basically, it is a one-to-one relation between every data set and a feature where sharing behavior is transparent from both sides.

Step 2: There might be a chance of misuse of personal data once they are shared to the OSN. Let us assume that, an user ***X*** conserving private data with service provider ***SP*** and a movie portal ***TP*** is providing a movie service to ***X*** through ***SP***. So, there is a policy between ***SP*** and ***TP*** to share users data. On the other hand, ***TP*** is connected to some external channels for movie broadcasting to those users. The question is, how safe are the users data when there is a chance of data sharing through another external link. In this case, our proposal is to provide an alert both to the ***X*** and to the ***SP***. To do so, ***TP*** must contain a verified phone number/email id to send notification to both ***X*** and ***SP*** in case of any misuse of data during any offer through an external link.

Step 3: Once ***X*** came to know that his/her data is being used by some other unknown parties M, the user must need a control to stop the chance of misuse. In this case, if an alert ***S/E*** received by ***X***, immediately it would be able to take action ***Y*** to break that external link by replying the sms or email to the ***SP***.

The data exchange over OSN is quite sensitive issue in terms of privacy maintenance. The steps we discussed above are now implemented by the set of logic rules maintaining the social data privacy standards. The rules for social data privacy have been discussed in different policy bodies such as RBAC, EPAL, P3P, and XACML in terms of user information sharing. We followed XACML [10] (eXtensible Access Control Markup Language) in the proposed framework that defines the set of target rules for resources, subjects, actions, and environment.

Definition 1: A policy p can be defined as a tuple (r, A, D, d) where $r \in R$ identifies the resource the policy refers to; $A \in A$ is a set of permitted actions on the resource r; D is a set of strong data set and d is a set of weak data set. both expressed as logic literals.

Definition 2: A weak or strong data set is a tuple (u, c) where $u \in W$ is a user and c is a logic literal.

Definition 3: The Privacy function pf is defined as $pf : R- > 2O$ and associates a resource to a set of private data.

For the sake of readability, we show the policy, ρ = (r, A1,..., An, D1, D2,..., Dn, d1, d2,..., dn) in the following syntactical representation:

ρ P

```
{
      strong-data: D1, D2, ..., Dn
      weak-data: d1, d2,..., dn
      resource: r
      can-do: A1
      can do: A2
      can do: ... An
}
```

Here, if the user allow access to either strong or weak data set then the service can be provided with all options which is a customized service according to the users data access permission. To be more specific, if user have following data:

```
D(i): Date of birth
D(p): Email
D(q): Phone number
d(s): Sex
d(t): City
```

The service providing policy between the user and the service provider in terms of interaction can be referred as follows:

```
{D(i),(p),(q)} U {d(s)} -> A1
{D(i)} U {d(t),(s)} -> A2
A1: Premium service
A2: Basic service
```

The logic literals are:

```
X:  is a user
i:  all the properties of the user (e.g.
    email, phone, dob, sex, city)
j:  are selected group of data item
X[i]:  all the data of an user
X[i(j)]: some selected data item of the
    user (j may be equal to i, or j can
    have some data from the set of i)
SP: Service provider
TP: Service offering third party (user
known)
M:  channel for broadcasting services (M
    could be any other party which user do
    not know)
E: Email alert
S: SMS alert
Y: Stop sharing data
C: Cancel offered service
}
```

ρ Q

```
{
   if X[i(j)] shared to any M
   then X receives S from TP
        if (X do Y with TP)
         {
            TP do C to user
         }
        SP receives E/S from TP
        SP do Y of X with TP
 }
```

Therefore, by using the above policies in the proposed Feature, a user is offered services with different feature enabled/disabled according to the user shared data. On the other hand, user has the freedom of choosing his/her data from weak and strong data sets to avail the services they need from the service providers.

V. Discussion

Currently people have a virtual identity in addition to their real life. Most of the real things have another form in virtual world, even a signature. User have options to do more in virtual world compared to their real world and due to that they are sharing information instead of desired services they need. This information could be a huge data set for a service provider which they can use for any special event to serve their own purposes without user consent. However, there will always be a necessity of information exchange being online and this rate will be increasing day to day. Due to that, we have to restructure the privacy framework by adding additional feature to handle upcoming variances and complexities.

VI. Conclusion

In this paper, we have proposed a novel transparent data sharing feature. In order to do that, we have conducted a survey to understand the user awareness, willingness and behavior over the internet. Through this Feature user will have control on what information to give and what to not, and enjoy a complete services with all option available upon approval of all the information. The service provider can customize their services based on the information received from the users. They also can propose different level of services, such as a service with limited features, a service for a fixed term, and additional services with additional information claim. Our survey results have already explained the real scenarios where we have seen user will share personal information even after a chance of misuse of private data. Taking these users online behavior in our concern the Feature is proposed through policies where both the user and service provider have control and freedom over their data and services with alternative options in a transparent way.

REFERENCES

[1] S. Guo and K. Chen, Mining privacy settings to find optimal privacy-utility tradeoffs for social network services, in Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom), IEEE, 2012, pp. 656665.

[2] Besmer, A. and Lipford, H. Users' (Mis)conceptions of Social Applications. In Proc. GI 2010, ACM Press (2010), 63-70.

[3] Krishnamurthy, B. and Wills, C. On the Leakage of Personally Identifiable Information Via Online Social Networks. In Proc. WOSN 2009, ACM Press (2009), 7-12.

[4] Pedro G. Leon, Blase Ur, Yang Wang, Manya Sleeper, Rebecca Balebako, Richard Shay, Lujo Bauer, Mihai Christodorescu, and Lorrie Faith Cranor. SOUPS '13: Proceedings of the 9th Symposium on Usable Privacy and Security, July 2013, ACM.

[5] Anderson, J., Bonneau, J., and Stajano, F. Inglorious Installers: Security in the Application Marketplace. In Proc. WEIS 2010 (2010).

[6] Felt, A.P., Ha, E., Egelman, S., Haney, A., Chin, E.,and Wagner, D. Android Permissions: User Attention, Comprehension, and Behavior. In Proc. SOUPS, ACM Press (2012).

[7] Howell, J. and Schechter, S. What You See is What They Get: Protecting Users from Unwanted Use of Microphones, Camera, and Other Sensors. In Proc. Web 2.0 Security and Privacy (2010).

[8] Synnes, K., Kranz, M., Scheln, O., Rana, J. (2012). Social Interaction for Digital Cities.

[9] Rana, Juwel. "Improving group communication by harnessing information from social networks and communication services." Department of Computer Science, Electrical and Space Engineering, Lule University of Technology, Licentiatavhandling Lule: Lule Tekniska Universitet (2011).

[10] eXtensible Access Control Markup Language (XACML), Version 2.0, OASIS Standard, Feb 1, 2005.

[11] Good, N., Dhamija, R., Grossklags, J., Thaw, D., Aronowitz, S., Mulligan, D., and Konstan, J. Stopping Spyware at the Gate: A User Study of Privacy, Notice and Spyware. In Proc. SOUPS 2005, ACM Press (2005), 43-52.

[12] Kun Liu, Evimaria Terzi A Framework for Computing the Privacy Scores of Users in Online Social Networks, 2009 Ninth IEEE International Conference on Data Mining.

[13] Anderson, J., Bonneau, J., and Stajano, F. Inglorious Installers: Security in the Application Marketplace. In Proc. WEIS 2010 (2010).

[14] Agrima Srivastava, G Geethakumari A Framework to Customize Privacy Settings of Online Social Network Users, 2013 IEEE Recent Advances in Intelligent Computational Systems (RAICS)

[15] Hull, G., Lipford, H., and Latulipe, C., Contextual gaps: Privacy issues on Facebook. Ethics and Information Technology 13,4 (2010), 289-302.

[16] Wang, N., Xu, H., and Grossklags, J. Third-Party Apps on Facebook: Privacy and the Illusion of Control. In Proc. CHIMIT 2011, ACM Press (2011).

Learning from Tagore

Shimul Hassan, Nabeel Mohammed and Sifat Momen
Department of Computer Science and Engineering,
University of Liberal Arts Bangladesh,
House No. 56, Road No. 4/A, Satmasjid Road, Dhaka 1209, Bangladesh
(email: {shimul.hassan.cse, nabeel.mohammed, sifat.momen}@ulab.edu.bd)

***Abstract*—This paper presents the use of a generalized learning technique to automatically generate Bangla poetry. We have trained a Long Sort Term Memory (LSTM) based recurrent neural network model on 350 poems written by Rabindranath Tagore to inspect if the recurrent neural network learns to generate poems. Using the technique described in this paper, we are able to generate poems using seed texts provided as input. The generated poems display some interesting patterns, such as rhythm as well as limited semantic and emotional contexts. To the best of our knowledge, this is the first work on automatic Bengali poetry generation using such models.**

Keywords—Bangla poetry, LSTM, machine learning, poetry generation, recurrent neural network, Tagore

I. INTRODUCTION

One of the dreams of AI practitioners is to create machines that can think and/or act like humans. Machines that do so are often considered to be intelligent [1]. In the AI world, there has been a long history of trying to imitate human behavior by machines [2]. Any output generated by a machine that cannot be easily distinguished by a human of whether it was actually generated by a human being or a computer program is said to have passed the Turing test and therefore can be called intelligent. Thus the ability to fool a human being by a machine is often considered to be the hallmark of an intelligent machine.

Although there is no clear agreement amongst researchers about what an intelligent task can be, it should not be surprising that an intelligent task contains some elements of creativity. Ability to write poetry has elements of creativity, and therefore can be considered as an intelligent task. In fact, that is the reason why poets in different societies are often considered to be intellectual.

Bengali (ethnonym Bangla) is spoken by about 250 million natives and 300 million people around the world. It is the national language of Bangladesh and the second most spoken language in India. With such huge amount of people communicating in this language, it turns out to be the 7th most spoken language in terms of the native speakers and the 10th largest spoken language worldwide. The language originates from Sanskrit and belongs to the Indo-Aryan and also the Indo-European language family. The language has deep connection to its literature with its earliest works appearing in the 10th and the 12th centuries. Bangla language has been dominated by numerous legendary literates across different centuries – thus resulting in a language with rich literature.

This paper presents a first step towards using a generic learning tool to learn, and then generate Bangla poetry automatically. In this paper, we propose a generic machine learning mechanism to auto-generate Bengali poems using recurrent neural networks (RNN). However, in this paper, we do not test whether it passes the Turing test and hence we do not yet claim that our proposed technique is intelligent. We train recurrent neural network with 350 poems of Rabindranath Tagore (a Nobel Laureate and a legendary Bengali author who re-shaped Bangla literature and music) to see if the network learns to generate poems containing certain texts (that are given as inputs).

The rest of the paper is organized as follows: In section II, we discuss background work related to this paper. Following this, we discuss the research methodology and experimental setup in Section III. Experimental results are outlined and discussed in section IV. Finally, we conclude the paper with remarks on our future research endeavor in section V.

II. BACKGROUND

The effort of generating automatic poems have been there from as long as that of the 1970s. For some notable works in the area of automatic poetry generation, readers are encouraged to see the works listed in [3-9]. There have been few work in generating poetries automatically in English and other languages. However, automatic poetry generation in Bengali did not get much attention. To the best of our knowledge, we found only one paper which looked into the case of poetry generation in Bangla [3]. In their work, Das and Gambäck [3] looked into the algorithm of creating rhythms in poetries – thus making the poetries rhythmic. However, what their work lacks in is that there is no creation of semantic structures in their poems. So, although the poems become rhythmic, they lack in semantic and emotional context.

A. Recurrent Neural Network

A Recurrent Neural Network (RNN) is generally used in sequence learning where it performs the same calculations on every element of a sequence [10]. RNN's hold an internal state where it attempts to holds information about calculations done so far. RNN can operate on every sequence of vector. Figure 1 shows the structure of a RNN.

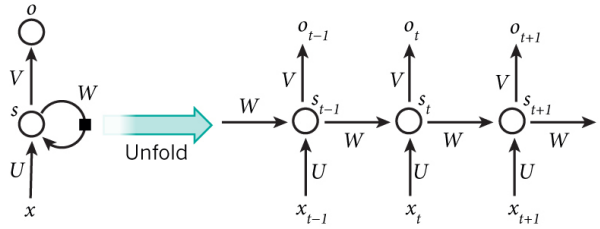


Figure 1: Recurrent Neural Network [15]

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

In figure 1, x_t is input at time step t and s_t is the cell of the network at time step t. We calculate s_t from the output of the previous state and the input from the current time step as illustrated in equation 1.

$$S_t = f(Ux_t + Ws_{t-1}) \quad (1)$$

where f is a nonlinear function such as *tanh* or sigmoid function and h_t is the output at step t simply calculated as a function of s_t. Traditional neural network uses different parameters in every steps but RNNs shares the same parameters in every time steps.

B. Long Short Term Memory (LSTM)

RNNs have proven to be effective at sequence learning, provided the sequences are not too long. LSTM, a variant of RNN, was proposed [11] and got more attention due to its ability to generate better outcomes. LSTM was designed to solve the difficulties faced in training RNNs – particularly the vanishing gradient problem [12]. LSTM introduced a more complicated way of storing the memory cell state. The value of this memory cell state as well as the output of the LSTM unit depends on some internal gates namely the input gate (i), the output gate (o) and the forget gate (f). A mathematical description of the LSTM model is shown in equations 2 – 6.

$$i_t = \sigma\left(W^{(xi)}x_t + W^{(hi)}h_{t-1} + W^{(ci)}c_{t-1} + b^{(t)}\right) \quad (2)$$

$$f_t = \sigma\left(W^{(xf)}x_t + W^{(hf)}h_{t-1} + W^{(cf)}c_{t-1} + b^{(f)}\right) \quad (3)$$

$$c_t = f_t \circ c_{t-1} + i_t \circ tanh\left(W^{(xc)}x_t + W^{(hc)}h_t + b^{(c)}\right) \quad (4)$$

$$o_t = \sigma\left(W^{(xo)}x_t + W^{(ho)}h_{t-1} + W^{(co)}c_{t-1} + b^{(o)}\right) \quad (5)$$

$$h_t = o_t \circ tanh(c_t) \quad (6)$$

C. Character level LSTM based model

A recurrent neural network and its variants are found to be very useful in sequence modeling. A character level RNN is trained on a text corpus a single character at a time and can be used for text prediction character by character. In recent time, Karpathy and colleagues [14] used LSTM-based recurrent neural network to train on the collected works of Shakespeare, the Linux kernel etc… The results of the work quite clearly show two things: (1) such models can learn document structure and reproduce such structure faithfully, and (2) at times these models also seem to learn semantic structure from the training texts. To reiterate - the amazing achievement of these models is that they learnt such high-level concepts even though they were trained on a character by character basis.

Instead of using a traditional RNN, our model uses a stacked LSTM structure which has two LSTM layers in sequence followed by a single softmax layer. Figure 2 shows a schematic diagram of the structure.

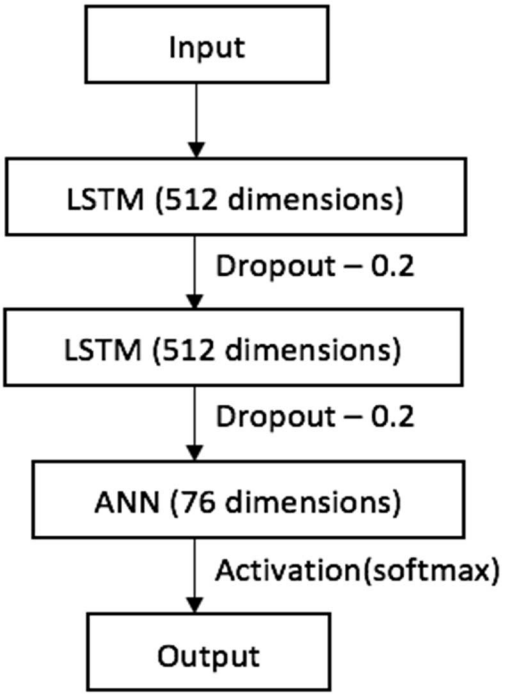


Figure 2: Schematic diagram of the model used

1. Experimental Setup

The aim of this paper is to automatically generate Bangla poems with generalized learning mechanisms. As mentioned above, our learning tool is an LSTM based recurrent neural network model which we want to train using the smallest possible unit of text (characters). To do so, we needed a corpus of Bangla poetry. For this initial foray in this area, we have chosen to use a collection containing 350 poems written by Rabindranath Tagore. Each poem was concatenated to each other without their titles. In total, there were 1593675 characters which were made of 76 unique Unicode characters. We kept the poems as it was; including the spacing.

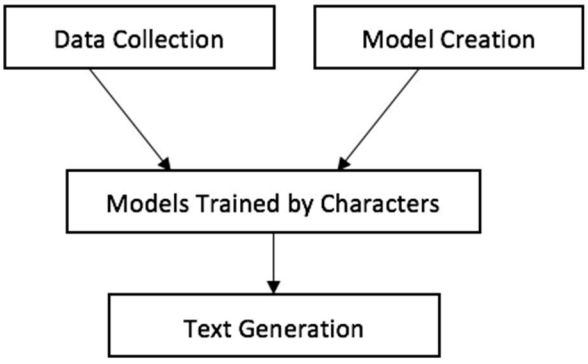


Figure 3: Research Methodology Diagram

D. Training

We trained the model for 33 epochs, which means that the model went through the dataset 33 times. As character prediction is essentially a classification problem, we used categorical cross-entropy as our loss function and rmsprop as the optimizer [16] [17].

E. Testing

The only way to evaluate machine-generated poetry is to generate poems and do an evaluation. We have chosen to do the evaluation in a subjective manner mainly because not all poems can be constrained to rhyming. Figure 4 shows the seeds texts use in [3] and the ones we have chosen to use. Our set includes the two seed texts used in [3] and we have expanded on this set by including 5 new ones.

III. RESULTS

The results shown in this sections comprise mainly of Bangla poems. For an international audience, a phonetic transliteration of each Bangla poem is given in Appendix A. The seed texts used for our evaluation are shown in Figure 4.

Seed texts used in [3]	
পালকী চলে	স্তব্ধ গাঁয়ে

Seeds texts used in this work	
আজি হইতে শতবর্ষ	পরানের আকুলতা
পালকি চলে	আলোর মূর্ছনা
স্তব্ধ গাঁয়ে	হারানো দিগন্ত
বাঁশির সুরে	

Figure 4: Seed Texts

Figures 5-8 shows the evolution of the quality of the poetry generated by the same model at different epochs using the same seed text (Aji hote shoto borsho pore). The generated poem shown in the first sample do not display and semantic or structural cohesion. This is not surprising as the model has only seen the data once.

আজি হতে শতবর্দ্বান্তাদির ক্রান্ত্বার্তার্ত্র প্রান্ত্বান্তার্তার্ত্রার্তার্দ্রার্তার্তত্বরত্ত্রান্ত্বার্ত্রার্ত্রান্ত্বান্তাত্ত্রার্শত্রান্তার্ত্র?

Figure 5: Output at Epoch 1

Figure 6, shows the generated poem for the same seed text after 5 epochs. It is immediately evident that the model has learnt some structure as well as some common spelling. It is also noticeable that the model has achieved some rhythm (lines 3 and 4).

আজি হতে শতবর্ষকার
স্বপ্ন হতে না করি বিশানীত কানে কালায় নাহি না কানে
বিহাসি কালো বালিয়া যায়
কাঁদি নাহি সাজা করি কালায়
স্বপনের সাথা কর?

Figure 6: Output at Epoch 5

Figure 7 shows the generated samples at epoch 20 and we see more words being spelled correctly. It is also interesting to note that the seed text expresses an element of time (translation - a century from today), and the poem has other words which can be associated with the concept of time e.g. ক্ষণিক (for an instant)

আজি হতে শতবর্ষ করি নাহি করি বালিকা ।
ক্রমে ক্ষণিক্ষ করি করি করি সব নিত্য নিরালিত জাগিয়ে যায় গাঁথি হয় না কিছু নাই ,
সেথা সুদূর বিস্ময়--
কাঁদিল না কি আর সাজান্তরে
সে?

Figure 7: Output at Epoch 20

At epoch 33, the improvements are much less noticeable. However, a majority of the words used in the poem actually belong to Bangla at this point.

আজি হতে শতবর্ষ হয়ে যেন নাই করে আর কিছু নাহি জানি নাই দেখি ।
স্বপ্ন মঙ্গলের কথা অমৃতসমান,
গাহিয়া কাঁদে গেল বুড়ি ,
কত সুখের সাথে অনন্তরাজে।
বিশ্বের মতো আলোকের ?

Figure 8: Output at Epoch 33

Figures 9 – 16 show some more example poems generated by the model, after it was trained for 33 epochs. The first two seed texts (Figure 9-10) were taken from the work of Das and Gambäck [3]. While their work only produced a single line of

poetry for a given seed text (see figures 9 and 10), the RNN-based model used in this study can produce much larger poems quite easily. Also, the work in [3] concentrated on generating poetry with rhythm. While this is a commendable approach, any aficionado of Bangla poetry will attest to the fact that rhythm is not the ultimate hallmark of good poetry. For example, the sample shown in Figure 11, uses the word for a palanquin (which in this culture is often associated with a new bride going to her in-laws) with seeing someone in a blessed dream. The poem generated in Figure 14 was seeded with a text with an approximate meaning 'cadence of light'. The generated poem creates a story of someone who is enthralled with it. Each of the other samples exhibit similar interpretive context.

Given
পালকী চলে !
'Palanquin moves!'

Generated
দুলকি চালে
'Trot pace'

Figure 9: Poem 1 generated by Das et al. [3]

Given
স্তব্ধ গাঁয়ে
'Stunned village'

Generated
রুদ্ধ দ্বারে
'Cloggy doors'

Figure 10: Poem 2 generated by Das et al. [3]

Given
পালকি চলে
'Palanquin moves!'

Generated
পালকি চলে পালকিয়া তার বাঁশির মালা খুলি নারি কাজে ।
সে যে রাত কিড়িয়ে দেখি যায় নাই করে,
আমি বলি , বলিতে পারে না আমি
তারে দেখিতে চায়
সেই সুদূর স্বপ্ন মঙ্গলে।
'Palanquin goes with the girl put of her staled necklace'
'She is wondering,is it night'
'I am telling, I can not tell
I wish to see him in my long cherished dream.

Figure 11: Some Generated Poems (1)

Given
তব্ধ গাঁয়ে
'Stunned village'

Generated
স্তব্ধ গাঁয়ে পায় না হয় না কী ক্ষয়,
যত কিছু নাহি জানি ক্ষণিকিত সবে
অনন্ত পাতারাশি
সামনা করি সহিত নাহি কানা,
যেখানে নিরখিনী নিজ প্রাণ যে সব মানিত না।
আমি বলি, নির্মম?

Translation: Is there no corrosion on stunned village,
when we do not know anything we do work like blind,
where Nirkhini did not believe
I think that was heartless

Figure 12: Some Generated Poems (2)

Given
পরানের আকুলতা
'Ardorous heart'

Generated
পরানের আকুলতা ত্বরাত্রি আলোক সেথা তারা এসেছে আমি রাজবালায়
সে যে রাত কিড়িয়ে দেখি সে কিছু নয় ,
করিছে সুদূর অনন্তসাগর সাথে কথা নাই কি জানে।

Transliteration: 'They have come in early night with the
light of ardorous soul
I speak to myself, I am trying to find on darkness
but
there was nothing
Amasses all the words and nothing to say? '

Figure 13: Some Generated Poems (3)

Given
আলোর মূর্ছনা
'Cadence of light'

Generated
আলোর মূর্ছনা পাতাগুলির মতো
সে কিছু বা মোহে মানিয়া বসি নিশারাত
করিছে সুদূর সারাদিন সুখের সাথে নির্দয় নির্দ্দেশি পাতাগুলির মতো
সে কিছু কথা বলিতে গেল গেয়ে

'The cadence of light is like leaves
He waited whole night for fascination of getting something
He does something special all day with happiness
by cruel dictates like missing leaves
He wanted to say something by singing'

Figure 14: Some Generated Poems (4)

Given
হারানো দিগন্ত
'Missing Horizon'

Generated
হারানো দিগন্ত সে সব গিরিশির করি সুখে কিছু নয়নির মাঝে
সন্ধ্যাবেদিকে নির্জন উচ্ছলিল অবকাশ
অনন্ত আকাশে ত্যাসে মানবে রাজবালায়
সে কিছু বা পাই নাহি করে,
আমি বলি,
Translation: 'Happiness like missing horizon
feeling like top of the mountain
something like enjoyment with the eyes at lonely evening
Siting under the eternal sky find apathy to do anything
I say'

Figure 15: Some Generated Poems (5)

Given
বাঁশির সুরে
'Melody of Pipe'
Generated
বাঁশির সুরে পাখি গেল গাঁয়ের কালে কানে আসি
সে যে রাজবালায় জাগি এসেছি আবার বাঁশির মালা দ্বারে বাঁধিবে না আমি বলি ,
কত কাল করে হোক নাহি করিয়া দেখিতে চায়।
'A bird goes on the village with making melody of pipe
He was awakened talking again he will not inhibit
with the illusion of pipe
want to watch like for ever '

Figure 16: Some Generated Poems (6)

IV. CONCLUSION

This paper looks into the designing of a machine using LSTM to generate automatic poems. So far, we have used 350 poems from Tagore's preeminent collection to train the neural network. We were able to create automatic poems that were not only rhythmic but also display limited context - both semantic and emotional. For our future work, we would like to increase the dataset by taking poems from different poets to see if that generates poetries that have more semantic and emotional context. Furthermore, we plan to create a Turing test for Bangla poems to see how well people can distinguish whether the poem was generated by a machine or a human.

REFERENCES

[1] S.J. Russell, P. Norvig, J.F. Canny, J.M. Malik, D.D. Edwards, "Artificial Intelligence: A modern approach", Prentice Hall, vol. 2, 2003.

[2] N.Sharkey and A.J.C. Sharkey, "Artificial Intelligence and natural magic", Artificial Intelligence Review, vol. 25, no. 1-2, pp: 9 – 19, 2006.

[3] A. Das and B.Gambäck, "Poetic machine: Computational creativity for automatic poetry generation in bengali", 5th International Conference on computational creativity, ICCC. 2014.

[4] S.Colton, J. Goodwin, and T. Veale. "Full face poetry generation." Proceedings of the Third International Conference on Computational Creativity. 2012.

[5] E.Greene, T.Bodrumlu, and K.Knight. "Automatic analysis of rhythmic poetry with applications to generation and translation." Proceedings of the 2010 conference on empirical methods in natural language processing, Association for Computational Linguistics, 2010.

[6] M.T.Wong, and A.H.W. Chun. "Automatic haiku generation using VSM." WSEAS International Conference. Proceedings. Mathematics and Computers in Science and Engineering, Eds. Qing Li, S. Y. Chen, and Anping Xu. No. 7, World Scientific and Engineering Academy and Society, 2008.

[7] H.Oliveira, "Automatic generation of poetry: an overview." Universidade de Coimbra, 2009.

[8] P.Gervás, "An expert system for the composition of formal spanish poetry." Knowledge-Based Systems 14.3 : 181-188, 2001.

[9] Díaz-Agudo, Belén, Pablo Gervás, and Pedro A. González-Calero. "Poetry generation in COLIBRI." European Conference on Case-Based Reasoning. Springer Berlin Heidelberg, 2002.

[10] P.J. Werbos, "Generalization of backpropagation with application to a recurrent gas market model." *Neural Networks* 1.4 (1988): 339-356.I.S. Jacobs and C.P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.

[11] S.Hochreiter, and J.Schmidhuber. "Long short-term memory." Neural computation 9.8: 1735-1780, 1997.

[12] Y.Bengio, P.Simard, and P.Frasconi. "Learning long-term dependencies with gradient descent is difficult." IEEE transactions on neural networks 5.2: 157-166, 1994.

[13] J.Chung., C.Gulcehre, K.Cho, and Y.Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling.", 2014.

[14] A.Karpathy, J.Johnson, and L.Fei-Fei. "Visualizing and understanding recurrent networks." *arXiv preprint arXiv:1506.02078*, 2015.

[15] D.Britz, "Recurrent Neural Networks Tutorial, Part 1 – Introduction to RNNs"2015.[Online].Available:http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/

[16] Tieleman, T. and Hinton, G., 2012. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural Networks for Machine Learning*, *4*(2).

[17] Liu, C., Current Topics in Artificial Intelligence: Optimization.

APPENDIX A

Transliteration of Figure 11:
Palki chole palkia tar bashir mala khuli nari kaje
She je rat kiria dekhi jay nai kore
Ami boli, bolite pare na ami
Tare dekhite chay
Shei shudur shopno mongole

Transliteration of Figure 12:
stobdho gaye pay na hoy na ki khoy,
Joto kichu nahi jani khonikito shobe
Ononto patarashi
Shaman kori shohit naki kana,
Jekhane nirkhini nij pran je shob manito na.
Ami boli nirmom?

Transliteration of Figure 13:
Poraner akulota toratri aloke shetha tara esheche ami rajbalay
she je rat kiria dekhi she kichu noy,
koriche shudur onontoshagor shathe kotha nai ki jane.

Transliteration of Figure 14:
Alor murchona patagulir moto
she kichu ba mohe mania boshi nishirat
koriche shudur sharadin shukher shathe nirdoy nirdeshi patagulir moto
she kichu kotha bolite gelo geye.

Transliteration of Figure 15:
harano digonto she shob girishir kori shukhe kichu noyonir majhe
shondhabedike nirjon ucholil obokash
ononto akashe tashe manobe rajbalay
she kichu ba pai nahi kore,
ami boli ,

Transliteration of Figure 16:
bashir shure pakhi gelo gayer kale kane ashi
she je raajbalai jagi ashechi abar bashir mala
dare badhibe na ami boli
koto kaal kore hok nahi koria dekhite chai

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

Memory Forensics Tools: Comparing Processing Time and Left Artifacts on Volatile Memory

Khaleque Md Aashiq Kamal*, Mahmoud Alfadel* and Munawara Saiyara Munia†
*King Fahd University of Petroleum and Minerals
Dhahran, Saudi Arabia
Email: aashiqkamal@gmail.com , mahmood_1411@hotmail.com
† Ahsanullah University of Science and Technology
Dhaka, Bangladesh
Email: munawaramunia@gmail.com

***Abstract*—Digital investigation is becoming an increasing concern. Many digital forensic tools are being developed to deal with the challenge of investigating digital crimes. Acquisition of volatile memory is one of the vital steps of digital forensics process. Passwords data, indications of digital forensics methods, memory malware may be contained in volatile data which may overlooked by the investigator. The Success of memory acquisition mainly depends on the effectiveness of the memory acquisition tool. This paper compares memory forensics tools based on processing time and left artifacts on volatile memory. Furthermore, we examined how the processing time of the tools varies in terms of different volatile memory size. In order to conduct this work, we use the following tools: FTK Imager, Pro Discover,Nigilant32, Helix3(dd), OSForensics and Belkasoft RAM Capturer. The results show that Belkasoft RAM Capturer has the least amount of left artifacts, and it has also the lowest processing time. Moreover, this work concludes that tested tools are significantly different based on left artifacts on the volatile memory with 95% confidence level. Also, statistically, increasing the memory size x times does not increase the processing time x times of the tools.**

***Index Terms*—Forensics tools, acquisition tools, volatile memory, memory acquisition.**

I. Introduction

Hacking is a critical problem in present world. Some hackers hack for their malicious goals and some do for fun. FBI reported in 2008 that Internet fraud had cost dollars 264.6 millions. So one of the major challenges in current time is the investigation of the crime in computer technology. [1] Digital forensics is an important part of approximately every investigation. Digital Forensic software tools are applied regularly by the experts in several levels. Government organizations, military forces and other public, private organizations are using these tools. Expansion in forensic study, software tools, and procedure over the last decade has been very victorious and numerous in control situation now depend on these tools on an usual basis without realizing it . [2] In addition to criminal investigation, these tools are used for the purposes of maintenance, debugging, data recovery, and reverse engineering of computer systems in private settings.

In [2] described six categories for digital forensics research: evidence modeling, network forensics, data volume, live memory acquisition, media types, and control systems. Acquisition of live memory data means RAM will allow digital investigators to get evidence that will not be found in a hard drive investigation.

This work observes present work in the acquisition of live memory artifacts and assesses the impact of using specific tools in the capturing of live memory based on left memory artifacts and processing time.

The rest of the paper is organized as follow: Section II is reviewing the background information related to our work. In section III we are describing on the impact of acquisition tools on memory. Next section is describing the related works to digital forensics tools. In section V, we will go thorough the digital forensics investigation approach. Section VI is addressing the problem statement of our research. Next section provides our research approach. Section VIII is providing our experiment and results.Finally Section IX shows statistical analysis.

II. Background

A. Digital Forensics

The main idea of digital forensics originated from the general forensic idea of criminal work. But it is not exactly same to the general forensic term. It needs support expertise to examine electronic data records that are not physically substantial. According to[3], Computer forensics has two steps: forensic in the occurrence area and laboratory forensic analysis . Forensic work in the occurrence area is carried out to defend what proof of the occurrence keeps from the event of the occurrence and to recognize potential evidence that may help the examination. Moreover, the proof is reserved in an evidence storage or drived to a specialized forensic lab for deeper analysis. Finally, the analyzed end results are presented to the court who are in charge for the investigation. Moreover , all forensics related to digital data belongs to the category of digital forensics. Computer forensics is the art of utilizing susceptible process and measures to keep, recognize, dig out, trace, and understand digital media facts and to analyze the basis of those proof[4].

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

B. Digital Evidence

Conventional forensics depends on proof, the accurate data employed in criminal events, to verify that the defendant consign a crime. There are mainly three types evidence: (1) evidence based on observation, (2) evidence based on object (3) evidence based on document. [3] Moreover, Digital evidence is totally different to traditional evidence. We can not examine digital evidence using only open eye. We depend on electronic technology to understand the evidence. Most of the important evidence might be included in the RAM memory.

C. Importance of RAM data analyzing

To understand the importance of examining RAM , it is essential to explain wherever this information are to be originated and what data does it contain. According to K. Hausknecht et al.[5], all the information of computer system necessary to visit the ram at any time of process. Data will be kept in RAM as well as on the systems HDD. But the key distinction is that RAM shows us what was occurred on the PC at any fix point of time. It is essential to remark that there is enormous of data which are never stored on the HDD but it is available in RAM such as internet data. According to the same authors, we can get a lot of information from RAM data. It mainly depends on the system and the OS. Most significant information are: processes data, network data, information on open files, user data , registry entries, drivers used, processes and data which are hidden, temporal data, used dlls, information about opened session etc.

D. Memory Artifacts

Memory artifacts mean footprint on the memory. While any application runs on a machine, it occupies a portion of memory. During the run time of the application, it leaves footprint on volatile memory.

E. Memory acquisition methods

To analyze a RAM, an investigator should have a tool to acquire the RAM image. There are mostly two methods used for capturing RAM: software and hardware method. According to authors [5] Software method is a commonly used to seize the RAM and it is advised to do if the system is live. The main problem of software method is that delinquent can cover its information from such acquisition tools while it is enabled. Another problem is that when these tools are run , they left artifacts on the memory. It can violate important information of the RAM. To recover these issues, investigator can use hardware methods but this a very expensive way to acquire RAM image.

F. Tools for memory acquisition

There are commercial tools for acquiring RAM such as ProDiscover, Helix3 , FTK Imager. There are non-business software tools such as dd, memdump or Dumpit etc. According to authors [2],among computer forensics expertises, the most excellent way for resolving the reporting issue is to purchase any of these tools from market. But, this just works for financed associations. And there are numerous conditions in where business commercial software tools get failed and specialist needs to depend on open source software. While some of these tools are of very good quality, other tools are poorly recognized, out dated, and even discarded.

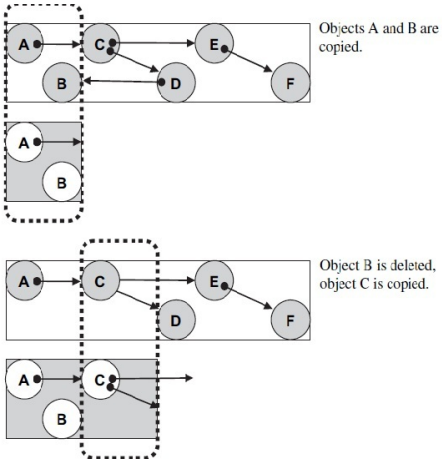


Fig. 1. Acquisition of the memory image from a live system

III. Impact of acquisition tools on memory

Any tool cannot guarantee that other systems act would get leave for the period of capture a memory ; it means system action is essential for the software to implement. According to [6] while the image of memory is captured, it is as well altered by other processes which are running parallelly and self-sustainability is hampered because of the chronological character of the memory capturing process. Moreover, there is a chance that the suspected executor may have a process in execution to destroy information in memory with parallel by the dump of memory.

As a result, we can say that a dump of memory is captured in parallel with the system action as a technique that produce non-self-sustainability data, and potentially loses data of interest. While this data may expose valuable information, it is less useful from a forensic point of view than data that is self-consistent and therefore more readily interpreted.

Fig. 1. shows that how the memory image is captured in a live system and new processes effect the acquisition technique.

As forensics tools executing on the live machine under examination will change the content of the memory whose condition is being acquiesced, their use is error in terms of proof acceptability. So, measuring the extent of the volatile memory changes by causing of running a live forensic tools becomes more and more important.

IV. Related work

All the digital forensic tools that will be used during any investigation should be tested before using them due to the potential changes that can occur in original data. Several works have been done by the forensics researchers to evaluate or

illustrate the features, quality ,drawbacks of the forensics tools. The study of [7] highlighted a number of usability issues which needs to be considered for comparing forensics tools. User interface,level of expertise, training needs ,reporting and documentation etc. According to them, people are more comfortable with GUI interface than to command line interface tools.

Seo et al. [8] has described several dumping techniques like WinDD, process dump using task manager for windows, dd(linux) , MDD etc. It is very difficult to dump pure memory using WinDD and a collection of more than 4GB of RAM can not be done by MDD.

Okolica et al. [9] has introduced a self contained tool named CMAT(Compiled memory analysis tools). It can extracts environmental and activity data from the live memory. CMAT parses a memory dump to find active, inactive and hidden processes as well as system registry information. It then compiles live response of forensic information from these processes and registry files and assembles it into a format suitable for data correlation.

On the other hand, Wazid et al. in [1] has categorized the tools in several field. Computer forensics tools,memory forensics tools, Network forensics tools, mobile phone forensics tools, database forensics tools. They categorized windowsSCOPE and memoryze as memory forensic tools. Both are used for windows. Memoryze is used for live system and windowsSCOPE performs network wide live memory forensic.

Manson et al. [10] ,has compared another three tools-FTK imager,Encase,Autopsy. According to them, Encase imports image fastest but it is difficult to use for almost all. On the other hand, FTK Imager is easy to use to basic computer user and autopsy is easier to linux user. And according to the author [7],FTK is similar to the windows classic tree view. Thus it has more intuitive interface than other tools.

In [11] Carvajal et al. have focused on digital forensic tools that collect evidence from RAM which contains volatile data such as network connections, logged users, processes, etc. In their paper they compared six forensic tools including: FTK Imager, Pro Discover, Win32dd, Nigilant32, Memoryze, and Helix3 (dd).They evaluated tools based on user interface, reporting, processing time, training, and leaving fingerprints or artifacts. They investigated GUI types of tools leaves more artifacts than the command line types of tools.

The authors of [12] have focused on digital forensic tools that collect evidence from RAM by using fmem and dd tool. But According to the authors Fmem and dd tools are successful in linux but in android, its performance is poor. This is mainly due to some of the android kernel security mechanism that prevent the kernel to load external code.They introduced a tool named lime. It is an open source tool to capture memory from android phones. It can capture through the local SD card.

Another method of capturing android memory is DDMS. Leppart et al. [13] used DDMS. But, it can not acquire complete physical memory. It only can dump single process heap contents. That's why it uses a lot of memory data. To overcome the drawback, Zhou et al. [14] used lime to capture the full memory.

However software method is comparatively cheaper and easier to other imaging process, the method has some drawback. According to [15] while these tools are executed on the machine, valuable forensics information might be deleted before it is preserved.

V. Digital Forensics Investigation Approach

Digital forensics investigation follows an idle approach for effective result. Mostly investigation process follows the phases as mentioned below. [16] This phases are not restricted to follow but following those phases might bring efficient result of investigation.

- ***Collection***
  - This step handles the gathering of a variety of possible basis of digital proof e.g. RAM data, Mobile data,Hard Disk information etc.
- ***Identification***
  - This step focuses on the identification of digital proof. The evidence are labeled in this step.
- ***Acquisition***
  - Decode to the origin of e-proof evidence from the different sources which are acquiesced.
- ***Preservation***
  - In this step, It gives emphasis of using the satisfactory procedures that make sure the reliability and the genuineness of proof evidence.
- ***Examination and Analysis***
  - It involves the actions such as search, filter, detection and investigation/assessment for relevant weight of the acquiesced e evidence.
- ***Reporting***
  - In this step, the final result and finding of investigation process is reported in details.

In this work our focus is on third phase of digital forensics investigation process which is acquisition.

VI. Problem Statement

This work is trying to investigate which memory acquisition tools leave less memory artifacts while it is executed. Moreover, the relation between the size of volatile memory and the processing time of capturing volatile memory image will be evaluated. Some tools like: FTK Imager, Pro Discover, Nigilant32,Helix3 (dd),OSForensics,Belkasoft RAM Capturer will be used in this work to investigate this relation. Through this research work we are trying to answer the following questions:

- Can we differentiate those tools based on the amount of left artifacts in the volatile memory and processing time of capturing ?

- What is the relation between the size of volatile memory and the processing time of capturing volatile memory image?

- What is the statistical influence of multiple runs of tools in multiple environments ?

Finally, we will evaluate our investigation result based on statistical analysis.

From our research questions we formulate two set of hypothesis to be tested.

- $\mathbf{H_a0}$: The difference of left artifacts amount on volatile memory of the tools is not significant statistically.

  $\mathbf{H_a1}$: The difference of left artifacts amount on volatile memory of the tools is significant statistically.

- $\mathbf{H_b0}$: Increasing the memory size x times increases the processing time of the tools x times.
  $\mathbf{H_b1}$: Increasing the memory size x times does not increase the processing time of the tools x times.

VII. Approach

To come up with the solution of our problems, we need a dependable approach to implement our work. To test our experiment, two different environments setup were needed to gather deeper data from the variation of source. Here, we can divided our work in two steps:

- Experimental phase: It starts with the installlation of selected tools and ends with recording our desired data. Fig. 2 shows all the steps of our experimental phase.
- Statistical Analysis: In this phase, we will evaluate our experimental results based on statistical analysis.

VIII. Experiment and Results

In this section we are describing how we have done our experiment. Moreover, our experiment results will be illustrated. To formulate the experiments, our examined tools are collected from the authentic sources and minimum requirement, characteristic has been observed.

The acquisition of RAM using tools is the first goal of this work. Six tools have been used to acquire memory image from the two differnt experimental setup. Table I shows the list of the tools. We used windows 7 operating system as a guest on virtual machine named oracle virtual box. For finding the memory artifatcts by the acquisition tools, we used all the listed tools in the virtual environment which has memory of 1024 MB.

We have run all the tools three times to observe the variation of result.During the capturing of memory, all variates values of physical and virtual memory have been recorded. The detail of both of the examined environment is given below.

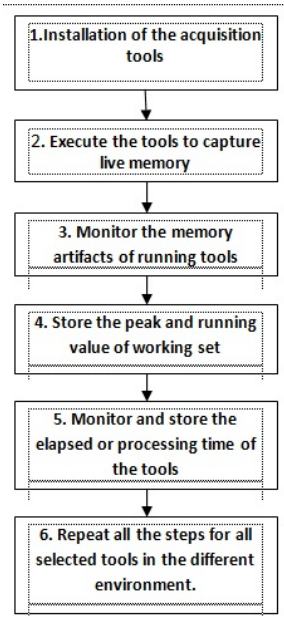


Fig. 2. Steps of the experimental phase

TABLE I
The List of examined tools

Name of the tools	Developer	Type
FTK Imager	Access Data	Commercial
Nigilant32	Agile Risk Management	Commercial
ProDiscover	Technology Pathway	Commercial
Helix3(dd)	E-fense Commercial	Commercial
OSForensic	PassMark Soft	Commercial
Belkasoft RAM Capturer	Belkasoft Forensics	Commercial

- ***Experimental Setup 1***
  - Model: HP NOTEBOOK PC
  - Operating System (guest): Windows 7
  - Operating System (host machine): Windows 7
  - RAM : 1024 MB
  - Processor: INTEL PENTIUM DUAL CORE 2.40 GHZ
  - Internet Status : Not Connected
- ***Experimental Setup 2***
  - Model: HP NOTEBOOK PC
  - Operating System (guest): Windows 7
  - Operating System (host machine) : Windows 7
  - RAM : 512 MB
  - Processor: INTEL PENTIUM DUAL CORE 2.40 GHZ
  - Internet Status : Not Connected

There are varieties of artifatcts in KB size between the tools. The peak and the average value are given in Table II.

TABLE II
MEMORY ARTIFACTS ON PHYSICAL MEMORY

Name of the tools	Physical Memory(Peak Value in KB)	Physical Memory(Average Value in KB)
FTK Imager	33660	32981
Nigilant32	28316	25485
ProDiscover	30984	28181
Helix3(dd)	24584	23541
OSForensic	32380	29531
Belkasoft RAM Capturer	5248	4934.2

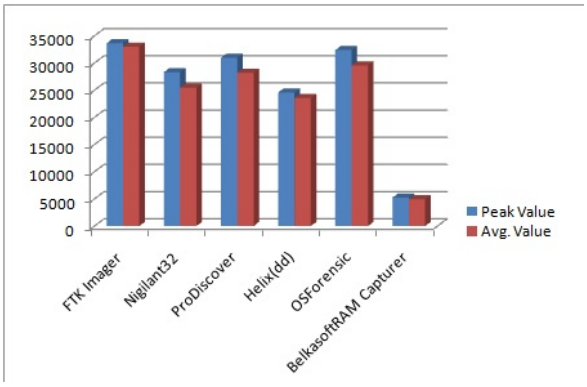


Fig. 3. Physical Memory peak and average value(KB)

Fig.3 graphically displays the differences between acquisition tools based on the left artifacts criteria in physical memory.

Also,the utilization of virtual memory (peak and average value) during execution has been listed in Table III.

From Table II, we can say Belkasoft RAM Capturer has the least memory artifacts while FTK Imager has the most. In addition, Table IV shows the processing time of the tools on both experimental environment. As we notice, Belkasoft RAM Capturer tool has the lowest processing time to capture the memory. Our second goal is to find the relationship between the processing time of memory acquisition and the memory size of the machine. We run those tools on the two different environment by changing the virtual machines memory size.

Table IV is showing the ratio of processing time in both environment. Moreover, Its shows the average of the ratio

TABLE III
MEMORY ARTIFACTS ON VIRTUAL MEMORY

Name of the tools	Virtual Memory(Peak Value in KB)	Virtual Memory(Average Value in KB)
FTK Imager	17428	17250
Nigilant32	15628	12994
ProDiscover	15836	14106
Helix3(dd)	19244	18949
OSForensic	152808	149606
Belkasoft RAM Capturer	1804	1762.5

TABLE IV
PROCESSING TIME OF THE COMAPARED TOOLS

Name	Processing time (seconds) 512MB	Processing time(seconds) 1024MB	Ratio
FTK Imager	58	70	1.206
Nigilant32	85	101	1.188
ProDiscover	95	177	1.86
Helix3(dd)	45	80	1.77
OSForensic	49	70	1.438
Belkasoft RAM Capturer	37	56	1.51
Average of Ratio			1.494

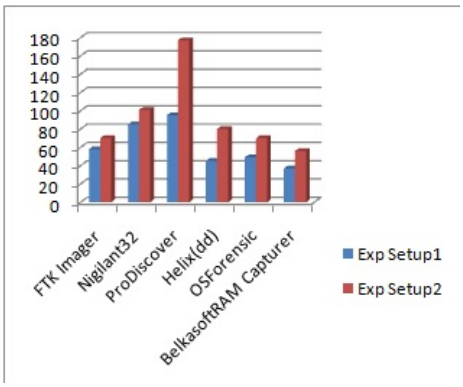


Fig. 4. Processing time in the both experimental setup(in seconds)

of processing time in experimental setup 1 and experiment setup 2 which is 1.494. But, the ratio of memory size in both set up is exactly 2 (1024/512=2). From this relation we can say, increasing the memory size x times does not increase the processing time x times. Fig. 4 is graphically shows the processing time in both experimental setup.

IX. STATISTICAL ANALYSIS

This section is describing the statistical analysis on our experimental results. To validate our results we have done statistical analysis techniques. For $\mathbf{H_a}$, we have applied Analysis of Variance technique(ANOVA), and Confidence Interval(CI) for $\mathbf{H_b}$.

A. Statistical Analysis of Left Artifacts on Volatile Memory

Anova is a technique that is used to differentiate multiple factors in order to conduct analysis of variance. The main objective of Anova is to find statistically that one alternative is better than to other alternative or not. In this work,we analyze the results of left memory artifacts of the acquisition tools. Anova does the work as following:

- it separates total variation observed in a set of measurements into:
  - (1) Variation **within** one system.
    - Due to random measurement errors
  - (2) Variation **between** systems
    - Due to real differences + random error

TABLE V
OUTCOME OF ANOVA ANALYSIS

Name	Value
F-tabulated	2.437692635
F-calculated	6.816682225

The main idea behind Anova is to check the variation(2) if it is statistically>variation(1)?
To get much deeper understanding on how to use and apply Anova, we looked in the book published in 1991 by Jain's [17]. The book(The Art of Computer Systems Performance Analysis) explains thoroughly how to use Anova and calculate these two variations. Here, we show the basic measurements we need to come up with Anova analysis technique. The technique applied in this work uses F-test which is decompressed to F-computed and F-table. F-table is obtained from a specific table as mentioned in the book. F-computed needs to be calculated as follows:

$$F = \frac{S_a^2}{S_e^2}$$

Where in general :

$$S_x^2 = \frac{SS_x}{df}$$

df =degrees of freedom
Sa =the mean square value for SSA
Se =the mean square value for SSE
α =the significant level needed to be obtained
SS_e =calculated by variation due to errors in measurements
SS_a =calculated by variation due to effects of alternatives.

After calculating the previous part(F-computed and F-table), we need to compare between both of the values in order to take such the following decisions:
If F-computed is greater than F-table, we have $(1 - \alpha) * 100\%$ confidence that differences among alternatives(modes) are statistically significant.

As shown in Table V, F-calculated(it is calculated based on our experimental results of Table II and III) is clearly greater than the F-tabulated.

As a result, we can come to say that our tested acquisition tools are significantly different with 95% confidence interval based on left artifacts on the volatile memory. So, we reject H_a0 and accept H_a1 .

B. Statistical Analysis of Processing Time

According to the experiment, increasing the memory size x times do not increases the processing time x times. To validate it we have done confidence interval(CI) [17] for the average accuracy. As shown in Table VI, the mean of processing time ratio(1.49) is located between upper bound(C1) and lower bound(C2). Consequently, $\mathbf{H_b0}$ is rejected, and $\mathbf{H_b1}$ is accepted. The outcome of the statistical analysis validates the stated results previously.

TABLE VI
CONFIDENCE INTERVAL OF PROCESSING TIME RATIO

Confidence Interval for the Average Ratio	
Mean	1.495
SD	0.279
Confidence Coefficient	1.96
Error Margin	0.223
Upper Bound C1	1.72
Lower Bound C2	1.27

X. CONCLUSION

The field of digital forensics is progressing day by day to enrich the area of security domain. New advance forensics tools are existence to investigate occurrence. But the threat to forensics is also becoming equipped with anti forensics tools which can remove digital evidence or can make delay in the evidence capturing process.As a result , efficient forensics tools are necessary to maintain a proper investigation process. This research examined six highly used memory forensics tools based on left artifacts and processing time. According to our experiment result and analysis, we can conclude that BelkasoftRAMCapturer has least memory artifacts while it is executed on any machine. As a result, it will erase less system information from the volatile memory which are very much important for digital forensics investigation. On the other hand, FTK Imager has the highest memory artifacts which may erase valuable evidence from the volatile memory. For processing time, BelkasoftRAM Capturer has the lowest while ProDiscover has the highest to capture memory for the digital investigation process. Moreover, we can say that if the memory size is increased then the processing time of capturing memory of those acquisition tools may not be increase rationally.

REFERENCES

[1] M. Wazid, A. Katal, R. Goudar, and S. Rao, "Hacktivism trends, digital forensic tools and challenges: A survey," in *IEEE Conference on Information & Communication Technologies (ICT),2013*. IEEE, 2013, pp. 138–144.
[2] S. L. Garfinkel, "Digital forensics research: The next 10 years," *digital investigation*, vol. 7, pp. S64–S73, 2010.
[3] I.-L. Lin, Y.-S. Yen, and F.-Y. Leu, "Research on comparison and analysis of the defsop, nist cell sop, and iso27037 sop," in *Eighth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS),2014*. IEEE, 2014, pp. 511–516.
[4] I.-L. Lin, H.-C. Chu, W.-N. Wu, and C.-P. Chang, "To construct the digital evidence forensics standard operation procedure and verification on real criminal cases-take linux/unix system as an example," *The 10th Cyberspace2008 Cybersecurity, Cybercrime and Cyberlaw*.
[5] K. Hausknecht, D. Foit, and J. Buric, "Ram data significance in digital forensics," in *38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO),2015*. IEEE, 2015, pp. 1372–1375.
[6] E. Huebner, D. Bem, F. Henskens, and M. Wallis, "Persistent systems techniques in forensic acquisition of memory," *Digital Investigation*, vol. 4, no. 3, pp. 129–137, 2007.
[7] H. Hibshi, T. Vidas, and L. Cranor, "Usability of forensics tools: a user study," in *Sixth International Conference on IT Security Incident Management and IT Forensics(IMF),2011*. IEEE, 2011, pp. 81–91.

[8] J. Seo, S. Lee, and T. Shon, "A study on memory dump analysis based on digital forensic tools," *Peer-to-Peer Networking and Applications*, pp. 1–10, 2013.

[9] J. Okolica and G. Peterson, "A compiled memory analysis tool," in *Advances in Digital Forensics VI*. Springer, 2010, pp. 195–204.

[10] D. Manson, A. Carlin, S. Ramos, A. Gyger, M. Kaufman, and J. Treichelt, "Is the open way a better way? digital forensics using open source tools," in *40th Annual Hawaii International Conference on System Sciences, 2007. HICSS 2007.* IEEE, 2007, pp. 266b–266b.

[11] L. Carvajal, C. Varol, and L. Chen, "Tools for collecting volatile data: A survey study," in *International Conference on Technological Advances in Electrical, Electronics and Computer Engineering (TAEECE), 2013*. IEEE, 2013, pp. 318–322.

[12] J. Sylve, "Android mind reading: memory acquisition and analysis with dmd and volatility," *Retrieved from*, 2012.

[13] J. Sylve, A. Case, L. Marziale, and G. G. Richard, "Acquisition and analysis of volatile memory from android devices," *Digital Investigation*, vol. 8, no. 3, pp. 175–184, 2012.

[14] F. Zhou, Y. Yang, Z. Ding, and G. Sun, "Dump and analysis of android volatile memory on wechat," in *IEEE International Conference on Communications (ICC), 2015*. IEEE, 2015, pp. 7151–7156.

[15] S. Vömel and F. C. Freiling, "A survey of main memory acquisition and analysis techniques for the windows operating system," *Digital Investigation*, vol. 8, no. 1, pp. 3–22, 2011.

[16] O. K. Appiah-Kubi, S. Saleem, and O. Popov, "Evaluation of some tools for extracting e-evidence from mobile devices," 2011.

[17] R. Jain, *The art of computer systems performance analysis*. John Wiley & Sons, 2008.

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

Design and Implementation of Smart Attendance Management System Using Multiple Step Authentication

Dhiman Kumar Sarker,
Department of Electrical and Electronic Engineering
Chittagong University of Engineering and Technology
Chittagong-4349, Bangladesh.
sagor.eee.cuet@outlook.com

Nafize Ishtiaque Hossain, Insan Arafat Jamil
Department of Electrical and Electronic Engineering
Chittagong University of Engineering and Technology
Chittagong-4349, Bangladesh
nafize.ishtiaque@gmail.com, ankon.arafat@hotmail.com

Abstract— **In the traditional attendance system of Bangladesh, the teachers either call the name or identity number of the students to which the students respond or pass the attendance sheet to the students to sign. With the increase of the number of students in the last two decades, the difficulties in attendance management system has increased remarkably. Again, in case of passing attendance sheet to the students, some students sign multiple times and proxy attendance is taken. These two systems are very time consuming. To overcome these inconveniencies, this paper represents a smart attendance system prototype. In this paper radio frequency identification, biometric fingerprint sensor and password based technologies are integrated to develop a cost effective, reliable attendance management system. A desktop application is developed in C# environment to monitor the attendance system.**

Keywords— **RFID, Biometric fingerprint sensor, password, C# language**

I. INTRODUCTION

The rate of literacy is one of the important parameters for the development of a country. The rate and the standard of education in Bangladesh in past few decades have been increased remarkably. The standard of education depends on the percentage of student attendance in class room, student result, quality of teaching etc. In modern teaching system, instructors allot a certain percentage of marks for attendance in classroom. As the number of students is increasing in classroom more, more time is required for traditional attendance systems. The systems are also faulty in many cases. The late comer students do not get attendance in traditional attendance system as a punishment. Proxy attendance is also a problem in present attendance system. In this paper, a three layer identification system is proposed. The attendance management system is also involved in developing a C# application for constant visualization of attendance. In the section II the literature review, in section III the system architecture and identity verification algorithm, in section IV software development, in section V performance analysis, in section VI comparative analysis and in section VII future work is described.

II. LITERATURE REVIEW

A real time keypad based attendance system with the ability to collect data form EEPROM to a Raspberry Pi server from different wireless sensor node is developed previously [1]. A matric card and barcode scanner based wireless attendance system is proposed by some researcher [2]. Motion sensor and RFID based attendance system with Web ASP.net application is shown in some research [3]. An RFID based attendance and monitoring system which is connected to a database using SQL is developed by researchers at Notre Dame University [4]. In literature [6], RFID is interfaced with Arduino and the data is sent to web for laboratory attendance. A creative system which is based on secret code generation with MD5 algorithm for each student daily integrating near field communication technology (NFC) is also developed [7]. This work also developed desktop, web and android application. Near field communication technology is also used in [8], to check the attendance of students in Kasetsart University at Si RaCha campus. An excellent system combined with NFC and biometric fingerprint identifier is used for the development of an electronic voting machine recently [5]. An online attendance system where automatically warning letter and graph charts are generated has been proposed by some researchers [9]. Web based attendance system is offered in literature [11] and [12]. The authors of [13] and [14] proposed the integration of biometric fingerprint sensor with zigbee for attendance system. Researchers of University Teknologi MARA Malaysia also developed barcode type attendance [15] which is very much similar to the system describe previously [2].

III. SYSTEM ARCHITECTURE

In this system the main processing unit is arduino mega 2560. The developed system consists of a fingerprint sensor, RFID reader and tag, 4X4 keypad, 16X2 LCD display and a personal computer. The brief description of the architecture of the system and verification algorithm are given below.

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

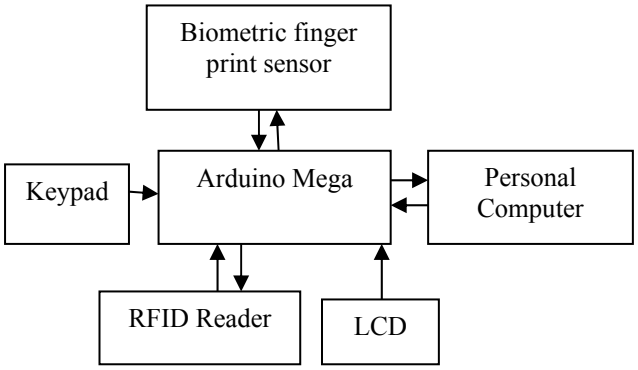


Figure 1. System Architecture

a) RFID reader and Tag: RFID stands for Radio Frequency Identification. The modern RFID systems are better than the barcode system and system which employs near field communication (NFC). The RFID system widely used in access control, goods identification, passport, tracking vehicle, people etc. RFID tags are three types which includes active, passive and battery powered. The passive tag does not need any external power and this type of tag is used to develop this system. The RFID readers are also three types namely active reader passive tag, active reader active tag and passive reader active tag. The first type of reader is used in this work. At first RFID reader sends a specific radio frequency when the RFID tags are brought near it. As the tag is exposed to the specific radio frequency, it will respond and send a unique tag number to the RFID reader. The RFID reader and tag is interfaced with the processing unit via UART communication.

Since the reader is interfaced with the processing unit, unregistered tag number will be detected. Here Grove - 125kHz RFID reader is used.

b) Biometric finger print sensor: The method of recognizing a person's physiological or behavioral characteristics is known as biometrics. Some unique physical creed of human being includes fingerprint, iris, voice, facial pattern etc. Biometric finger print sensor GT511C1R by ADH tech is used to develop this system. The sensor has 32 bit ARM cortex M3 core. Its maximum identification time is less than 1.5 seconds. The DSP chip inside the sensor identifies the finger print using SmakFinger3.0 Algorithm. At first the finger prints need to be registered to the finger print sensor. This process is known as enrollment. Enrollment of finger print sensor is done with an input ID. For the enrollment of two successive finger print to the sensor, the sensor enrolls the first finger print and waits for the the removal of the finger from the sensor. After the removal of finger the sensor selects another ID for which it will enroll the next finger print.

For the verification of the finger print, the host at first sends the verify command. After receiving the verification command, the finger print sensor scans the finger print. Then after processing, the sensor sends the ID number i.e. acknowledgement byte to the host device if the fingerprint is registered previously. Otherwise the sensor will not send acknowledgement byte to the host device.

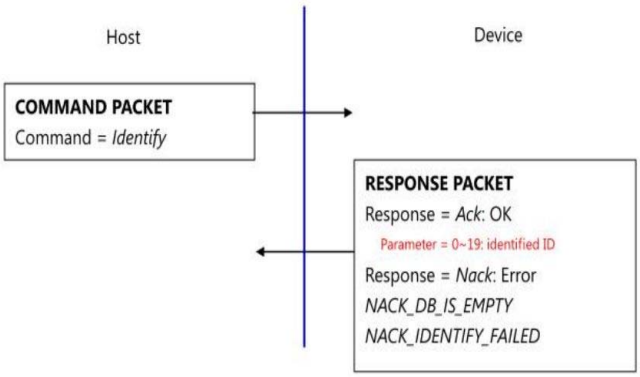


Figure 2. Finger print identification [16]

c) Identity Verification:

The identity verification of the system requires three layers of authentication. If any one of the authentication is failed then the next authentication step will never execute and the system will terminate its operation. The three layer identity verification is accomplished by the algorithm which is described in the following figure 2.

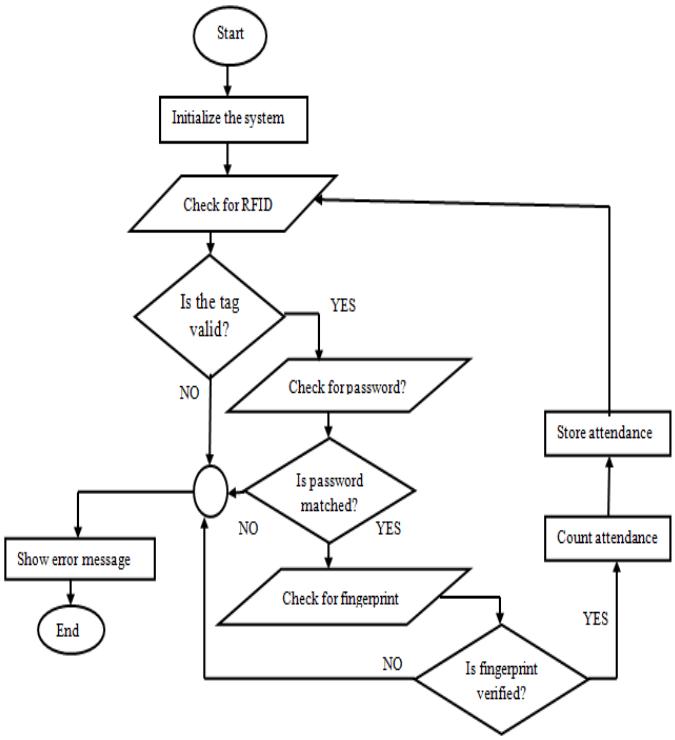


Figure 2. Flow chart for Identity verification

IV. SOFTWARE DEVELOPMENT

Visual Studio 2012 has been used to develop a windows application in C# language. In this application the instructor or the teacher must give his specific password to access the software. The students can't access this application due to password protection and to avoid any misuse of this software. After log-in procedure, the teacher can connect or disconnect the hardware device with the software. The course teacher

name and the course number are then showed on the application. The students are then allowed to give their access parameters to the hardware device for attendance. If the student ID is verified then the software saves the student ID, student name and percentage of attendance in a text document. A serial port object has been instantiated to receive data from hardware device. When the device connect button is pressed the specific serial COM port is opened and the textbox under this button shows device connectivity status. A data receive event handler has been used to receive data when serial port event occurs due to incoming data on serial receive buffer.

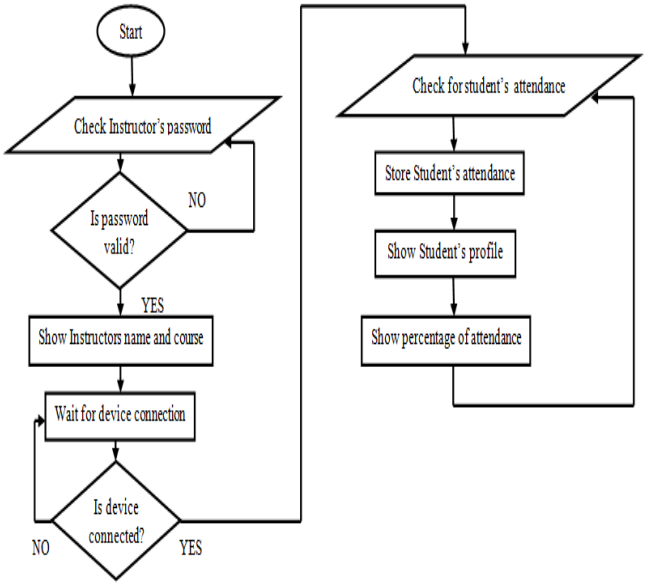


Figure 3. Flowchart of the desktop application

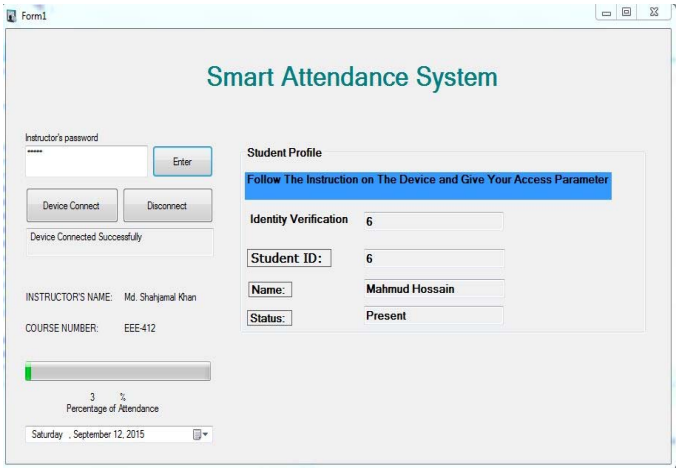


Figure 4. Desktop application

V. PERFORMANCE ANALYSIS

The performance evaluation of the system is done in different ways. The system is shown in the figure 5.

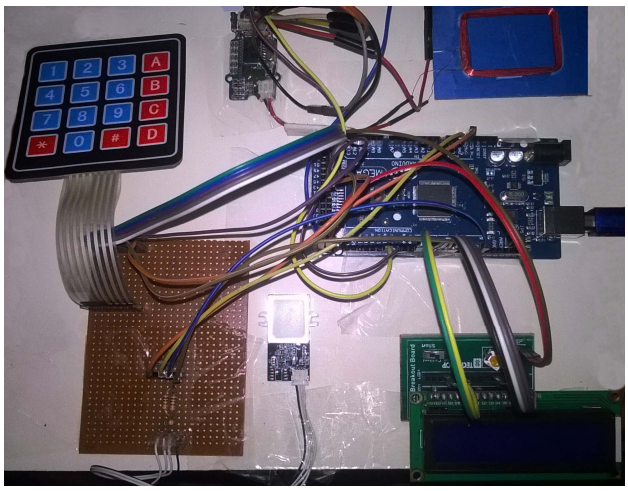


Figure 5. Smart attendance system

A. *Unencumbered Interfacing:*

Both the hardware and the software interfacing has been made user friendly, so the operator need not to be trained. Since modular design is adapted, interfacing the hardware is easier. This also provides advantage in replacement and maintenance of any subsystem. The hardware provides digital output to the processing unit which is more immune to external noise. The developed system performance shown in the figure 6

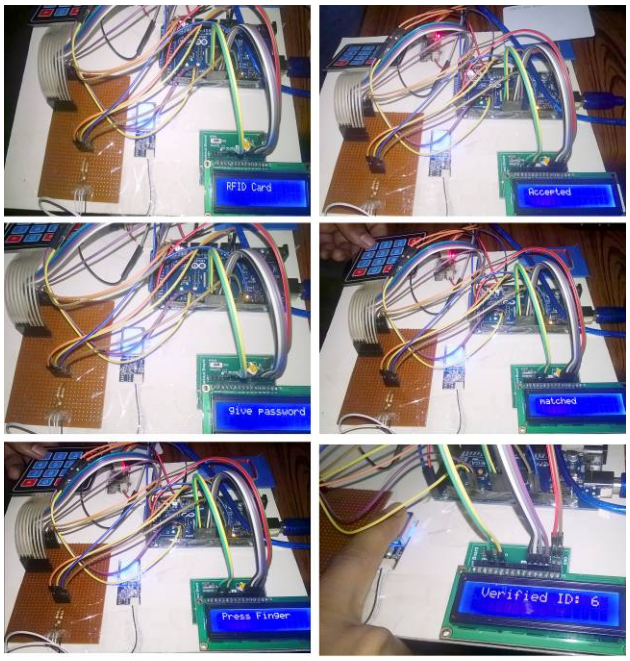


Figure 6. Operation of Attendance system

B. *Cost Analysis:*

Even though cost of fingerprint sensor and RFID module is moderate, the system works perfectly. As it is a prototype design and modular too, the costing for this prototype is considerably high. But if the whole system is developed and

integrated commercially then cost effectiveness of the system will be increased. If the Arduino Mega 2560 is replaced with low cost Atmel or PIC microcontroller chip then it will minimize cost considerably.

C. Fast Response Time

The response time of the system is moderately fast. If the finger is placed properly then typically it will take less than a second for identification. As the RFID data receive task is serial interrupt driven, the response for RFID identification is very fast. Though the keypad response time depends on user input, efficient use of the system keeps the response time remarkably less.

VI. Comparative Analysis

The following table shows the comparative analysis between the proposed system and the similar types of systems that were developed previously.

Table I : Comparative Analysis of Different System

Authors	Password	RFID	Finger print	Others
Shailendra et al. [1]	**Yes**	No	No	No
M. I. Moksin et al.[2]	No	No	No	**Barcode**
M.A. Abas et al.[3]	No	**Yes**	No	**Motion Detector**
A. Kassem et al.5]	**Yes**	No	No	No
T. J. Zhi et al.[7]	**Yes**	No	No	**NFC**
P. Subpratatsavee et al.[8]	No	No	No	**NFC and Image Verification**
N. I. Zainal et al.[11]	No	No	**Yes**	No
M. Basheer K. P. at el.[12]	No	No	**Yes**	No
L. Jian-po et al.[14]	No	No	**Yes**	**Wireless**
E. Peter U et al.[15]	**Yes**	No	No	**Barcode**
Proposed method	**Yes**	**Yes**	**Yes**	No

In the above mentioned table, it is seen that only one or two step authentication for attendance is used in most of the methods. But the proposed method shows three steps of authentication which make it more reliable.

VII. Future Work

In future, this work can be improved by incorporating IoT applications with the existing system. If IoT application is incorporated, then the attendance status can be seen from the mobile devices which are connected to the dedicated server. For IoT based attendance management system, GPRS & light weight protocol can be used. MQTT is a light weight data transmission protocol. Since small data packet can be used in this system, MQTT protocol will be perfect choice here.

VIII. Conclusion

The system described in this paper is for taking general purpose attendance. Though several methods of taking attendance was proposed previously, most of the systems did not integrate more than two sub systems which makes them vulnerable when proxy attendance is given by students on behalf of their absent friends. With minor modification such as face identification, this system can also be used for security purpose.

Acknowledgment

The authors would like to express their gratitude to the department of Electrical and Electronic Engineering of Chittagong University of Engineering and Technology for their continuous and generous help. The authors are also thankful to ASRRO (Andromeda Space and Robotic Research Organization) and its member for their constructive participation and encouragement for the successful accomplishment of this project.

References

[1] Shailendra, M. Singh, M. A. Khan, V. Singh, A. Palit, S. Wadar, "Attendance Management System" in *proc. of 2nd IEEE International Conference on Electronics and Communication Systems (ICECS),* 26-27 Feb. 2015,pp. 418 – 422.

[2] M. I. Moksin, N. M. Yasin, "The Implementation of Wireless Student Attendance System in an Examination Procedure" in *proc. of International Association of Computer Science and Information Technology - Spring Conference, 2009 (IACSITSC '09.),* 17-20 April 2009, pp. 174 – 177.

[3] M. A. Abas, T.B. Tuck, M. Dahlui, "Attendance Management System (AMS) with Fast Track Analysis"in *proc of International Conference on Computer, Control, Informatics and Its Applications (IC3INA),* 21-23 Oct. 2014, pp.35 – 40.

[4] S. M. Hassan, A. M. Anis, H. Rahman, J. S. Alam, S. I. Nabil, M. K. Rhaman, "Development of Electronic Voting Machine with the Inclusion of Near Field Communication ID Cards and Biometric Fingerprint Identifier" in *proc. of 17th IEEE International Conference on Computer and Information Technology (ICCIT),* 22-23 Dec. 2014, pp. 383 – 387.

[5] A. Kassem , M. Hamad , Z. Chalhoub, and S. EI Dahdaah, "An RFID Attendance and Monitoring System for University Applications", in *proc. of 2010 17th IEEE International Conference on Electronics, Circuits, and Systems (ICECS),* 12-15 Dec. 2010, pp.851 – 854.

[6] N. Arbain, N. F. Nordin, N. M. Isa, S. Saaidin, "LAS: Web-based Laboratory Attendance System by integrating RFID-ARDUINO Technology", in *proc. of IEEE International Conference on Electrical, Electronics and System Engineering (ICEESE),* 9-10 Dec. 2014, pp. 89 – 94.

[7] T. J. Zhi, Z. Ibrahim and H. Aris, "Effective and Efficient Attendance Tracking System Using Secret Code", in *proc. of IEEE International Conference on Information Technology and Multimedia (ICIMU),* 18-20 Nov. 2014, pp. 108 – 112.

[8] P. Subpratatsavee, W. Siripro, T. Promjun , W. Sriboon, "Attendance System using NFC Technology and Embedded Camera Device on Mobile Phone", in *proc. of IEEE International Conference on Information Science and Applications (ICISA)*, 6-9 May 2014, pp. 1 – 4.

[9] M. Othman, S. N. Ismail, H. Noradzan, "An Adaptation of the Web-based System Architecture in the Development of the Online Attendance System", in *proc.of IEEE International Conference on Open Systems (ICOS),* 21-24 Oct. 2012, pp. 1 – 6.

[10] M. Kassim, H. Mazlan, N. Zaini, M. K. Salleh, "Web-based Student Attendance System using RFID Technology", in *proc. of IEEE Control and System Graduate Research Colloquium (ICSGRC),* 16-17 July 2012, pp. 213 – 218.

[11] N. I. Zainal, K. A. Sidek, T. S. Gunawan, H. Mansor, M. Kartiwi, " Design and Development of Portable Classroom Attendance System Based on Arduino and Fingerprint Biometric", in *proc.of The IEEE 5th International Conference on Information and Communication Technology for The Muslim World (ICT4M),* 17-18 Nov. 2014, pp. 1 – 4.

[12] K. P. M. Basheer, C. V. Raghu, "Fingerprint Attendance System for classroom needs", in *proc. of Annual IEEE India Conference (INDICON),* 7-9 Dec. 2012, pp. 433 – 438.

[13] M.A. Meor Said, M.H. Misran, M.A. Othman, M.M. Ismail, H.A. Sulaiman, A. Salleh, N. Yusop "Biometric Attendance", in *proc. of IEEE International Symposium on Technology Management and Emerging Technologies (ISTMET)*, 27-29 May 2014, pp.258 – 263.

[14] L. J. po, Z. Xu-ning, L. Xue, Z. Zhi-ming, J. SUI, "Wireless Fingerprint Attendance System Based on ZigBee Technology", in *proc.of 2nd International Workshop on Intelligent Systems and Applications (ISA),* 22-23 May 2010, pp. 1 – 4.

[15] U. P. Eze, C. K. A J. Uzuegbu, L. Uzoechi, F.K . Opara, "Biometric-based Attendance System with Remote Real-time Monitoring for Tertiary Institutions in Developing Countries", in *proc. of IEEE International Conference on Emerging & Sustainable Technologies for Power & ICT in a Developing Society (NIGERCON),* 14-16 Nov. 2013, pp. 1 – 8.

[16] "GT511C1R datasheet",Optical Fingerprint recognition Embedded module, Taiwan.

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

Krill Herd Based Clustering Algorithm for Wireless Sensor Networks

Md. Shopon
Department of Computer Science and Engineering
University of Asia Pacific
Email: shopon.uap@gmail.com

Md. Akhtaruzzaman Adnan
Department of Computer Science and Engineering
University of Asia Pacific
Email: adnan.cse@uap-bd.edu

Md. Firoz Mridha
Department of Computer Science and Engineering
University of Asia Pacific
Email: firoz@uap-bd.edu

***Abstract*— Wireless sensor networks are principally categorized by insufficient energy resource. Naturally, communication between the nodes is the utmost energy consuming act that they perform. Hence, development of a well-organized clustering algorithm can play a vital part in enhancing the lifetime of network. Currently, nature inspired methodologies are very common in dealing with it. This work presents a centralized approach that deals with energy-awareness of wireless sensor networks using the Krill Herd algorithm. The performance of the suggested algorithm is assessed with famous clustering protocols. The simulation results show that suggested approach can maximize sensor network lifetime over other algorithms of the same category.**

Keywords*— *Krill Herd; Clustering; Optimization; Wireless sensor networks.

I. INTRODUCTION

Current developments in Micro-Electro-Mechanical Systems area brought some exciting news for wireless sensor networks [1]. It ensures us the development of low power and multifunctional wireless nodes which are able to communicate between short distances [2-3]. The sensors are furnished with some amount of control, communicating and sensing components. To build a wireless sensor network (WSN), some of the sensor nodes are put inside a network area and are used to monitor natural disaster [4-5], patients [6-7], environment [8-10], as well as in military target tracking [11-12], precision agriculture [13-14] etc.

WSNs are very tiny in structure and also resource constrained by processing performance, memory, and limited power supply [2]. Thus, in every situation, nodes depend on limited source of energy e.g. low powered batteries [15]. If once the sensors are installed and operating in the sensor network, changing the batteries usually is not a good option. Thus almost all current and related works in this particular arena are considering energy efficiency as their research topic [16-17].

Clustering reduces the amount of sensors that want to contact and communicate to the base station (BS) from a distance and also ensures the distribution of energy evenly amongst the sensors [1]. This paper develops a centralized cluster forming algorithm using Krill Herd (KH) to extend the network life span [18]. The simulation and evaluation of the proposed KH algorithm against other algorithms were performed on the basis of scalability, total lifespan of the network and also the volume of data directed to base station. The proposed algorithm was simulated against some established protocols like PSO, LEACH-C, and LEACH.

Rest of the research work is prepared as follows: the 2nd section discusses the literature review part. Section 3 summarizes radio and network models. Section 4 describes the projected configuration of clustering protocol using KH algorithm. In section 5, simulations of our protocol are shown, and section 6 concludes the work.

II. LITERATURE REVIEW

In literature, the algorithms that worked with cluster based approaches are very large in numbers. Among these, one of the most known and pioneering clustering protocol is Low Energy Adaptive Clustering Hierarchy (LEACH) [1]. A further improvement of the LEACH is LEACH-C [19], which constructs clusters in a centralized way, meaning all the calculations are done in BS. Very recently some researchers used nature-inspired algorithms in order to resolve the problems associated with clustering namely Ant Colony Optimization (ACO) [20-21], Particle Swarm optimization (PSO) [22-23], Cuckoo Search (CS) [24], Genetic Algorithm (GA) [16][25] etc. Krill Herd (KH) [18] is one of the latest inclusions in bio-inspired optimization algorithms. KH was proposed in 2012 by Amir Hossein Gandomi and Amir Hossein Alavi. The algorithm was inspired from the formation of various groups of marine animals which are non-random and under-dispersed. The name “krill” roots back to the word in Norwegian language, which stands for "small fry of fish”. Antarctic krills are one of the well-studied types of marine animal. Krill herds are accumulations with unparalleled orientation remaining on time scales of hours to days and space scales of tens to hundreds of meters. Among many, one of the qualities of this species is its capability to organize huge flocks.

In several literatures it was proved that KH algorithm performs better than the PSO and GA [26] . So the target of this research work is to find out whether KH performs well comparing to other algorithms in WSN. To the best of our

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

knowledge, this is the first work that implements KH algorithm in clustering in WSN. In this research, the clusters are created with the best possible set of nodes in each round. Later the data collected by them are passed on to the designated cluster head (CH). Therefore, the aim was to optimize the energy expenditure among the nodes on the basis of their residual energy. In turn, it will improve the network life span.

III. The System Model

A. Network Model

We are assuming the following configurations with our network model: the Base Station is fixed and is situated inside the network field. The nodes are static; always have some data to send to BS. All the nodes can function both as a cluster head and as a normal node. In order to reduce the amount of sent messages we used data aggregation.

B. Radio Energy Model

Our proposed algorithm implements the similar strategy for radio model which was used in [27]. In order to attain reasonable Signal-to-Noise-Ratio (SNR) in transferring l-bit data through distance d, we will use the following equation:

$$E_{TX}(l,d) = l.E_{elec} + l.\varepsilon_{elec}.d^2, \quad \text{if } d<d_0$$
$$= l.E_{elec} + l.\varepsilon_{elec}.d^4, \quad \text{if } d\geq d_0 \qquad (1)$$

Here E_{elec} stands for the amount of energy it loses to transmit 1 bit of data that is used to run the receiver circuit. And again, to receive *l* bit packet, the amount of energy dissipated by the radio will be calculated as follows:

$$E_{RX}(l) = l.E_{elec}$$

IV. Proposed Algorithm Description

A. Krill Herd Algorithm

Krill herd studies the herding behavior of the krill swarms considering the factors influencing the position of the krill individually as well as in the herd. The Lagrangian model [28] describes the functionality of Krill Herd algorithm. Fitness function of a krill is the combination of the highest density of the krill and the distance of food from the krill. The time dependent position of a krill is ascendant by the following three factors [26]:

- Movement induced by other krill individuals;
- Foraging activity; and
- Random diffusion

The fitness function of the algorithm in an n-dimensional space could be simplified as below:

$$\frac{ax}{dt}=N+F+D$$

Where *N* is Motion induced on *ith* krill individual due to the other krill individuals. *F* is Foraging motion and *D* is Random diffusion.

B. Cluster Setup Using Krill Herd Algorithm

In any wireless sensor network two types of communication scenario is observed. One is inter cluster and the second one is intra cluster communication [3]. This transmission may be through single-hop to the BS or through multi-hops among CHs headed for the BS. In this work we have used single hop technique. The main purpose of clustering is to increase intra cluster communication and select a proper cluster representative from all the nodes in each round. After gathering total information from the non-cluster head nodes, the cluster head will forward the data to the BS. In this approach the total amount of energy consumption decreases. But in this procedure if a single node is always acting as the cluster head of that cluster, that node will lose its energy level. So, it is required to assign a new node to be the Cluster head in every round. Krill herd will help making the decision of which node would be suitable for current round. The amount of energy and distance from non-cluster head or normal nodes are two criteria for nominating a fresh CH in a round.

C. Clustering Phases

Usually clustering protocols dwells in four foremost phases and in two stages. The major four phases are: i) Cluster head selection ii) Cluster formation iii) Data aggregation iv) Data communication. Two stages are i) Setup stage ii) Steady state stage. In the beginning of a single setup phase, sensors send information to the base station about their location and residual energy. Based on these data, BS measures the average energy. For one particular round only those sensors are selected as probable CH aspirant, who have greater energy than the average of that specific cluster in that round. This operation ensures that the nodes with a competent energy are designated as CHs for that round. Afterward, the BS runs KH algorithm to determine the fittest K number of CHs. In turn, it minimizes the cost function, as well-defined by :

$$f_1=max \sum d(,CHe,k)/ \quad CHe,k \vee$$

$K = 1,2,3$

$$f_2= \sum_{i=1}^{N} E(\ ) / \sum_{K=1}^{K} E(CHe,k)$$

$$\text{cost} = \beta \text{ x } f_1 + (1_+ \beta) \text{ x } f_2$$

Here f_1 function is the max average Euclidian distance among sensors and associated CHs. The $|C_{e,k}|$ is the number of sensors that fit into the cluster C_k of Krills. Ratio of the whole preliminary energy of all sensor nodes n_i, i=1,2,....,N in the network with all existing energy level of the CHs applicants of that particular round is well-defined by f_2. β is a user defined constant which is important to decide sub objectives role. Fitness function given here describes the objective of decreasing the remoteness of intra cluster amid CHs and their respective sensors, as quantified by f_1. As per the cost function mentioned earlier, a smaller value of f_1 and f_2 suggests a dense cluster construction with the utmost optimized set of sensors that have suitable level of energy to be the CH in that round [24].

If a WSN has *N* sensor nodes and *K* number of fixed clusters, the clustering for this sensor network may be identified as follows:

1. Initialize *K* krills to contain K arbitrarily designated CHs within the fit CH candidates [].

2. Assess cost function for each Krill *i*.
A) For every node n_i =1,2,....,*N*
X_{vector} (i) = random number (between the maxPosition and minPosition)
K_{vector} (i) = evaluate (X_{vector}(i))
B- K_{ib}=K; best fitness values for whole swarm so far
X_{ib}=X; best position for whole swarm so far
C- Find the krill that has the best fitness and its position

K_{gbest} = min(K)
X_{gbest}= return the position of K_{gbest}

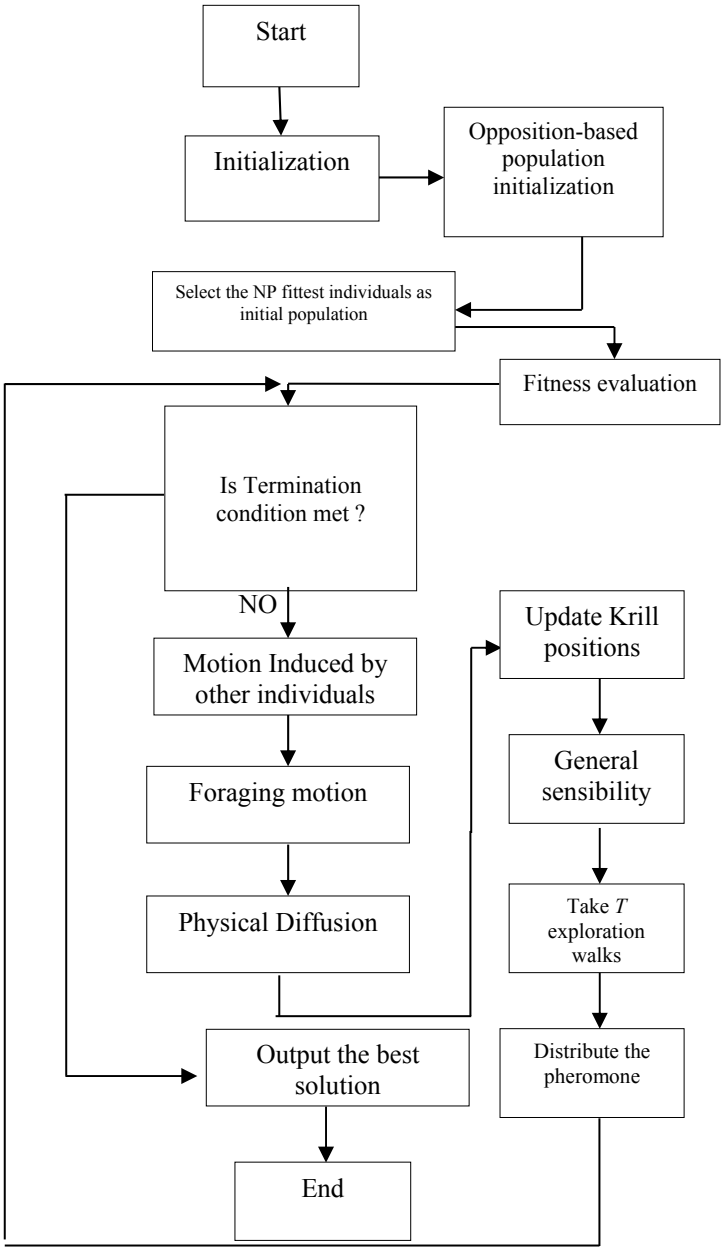


Fig 1: KH Algorithm flow chart

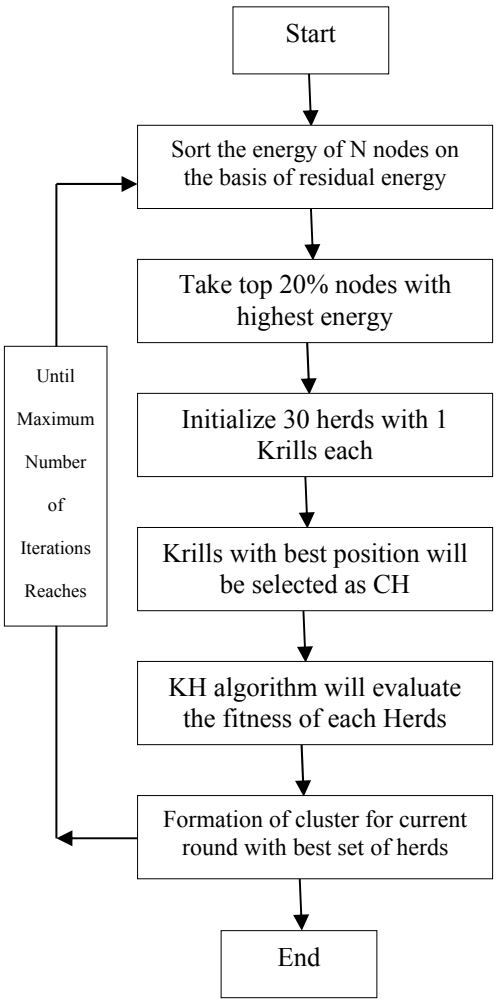


Fig 2: Selection of finest set of CHs using Krill Herd

3. Calculation of Motion:
- Motion urged by the presence of other individuals
- *D*, Foraging motion
- Physical diffusion

4. Find the best positioned krills.
5. Updating the positions of individual in the search space using the following equation.

dXi = delta_t*(N(i)+F(i)+D(i));
X(i)=X(i)+dXi;

6. Repeat steps 2 to 5 until the highest number of iteration is reached or any other ending criterion.

After getting the best set of clusters, cluster heads and their cluster members, the BS transmits the data that contains the CH-ID for each node back to all nodes. Figure 1 depicts the flow chart of Krill herd algorithm in selecting the best set of CHs. Figure 2 shows our proposed clustering algorithm and the whole procedure of clustering using KH algorithm in WSN.

V. Simulation and Analysis

This protocol was implemented in MATLAB R2013a. Our implementation of the simulation was in the setting of 100 nodes in an area of 500m x 500m. The BS is located in coordinates (250,250). It was proposed that 20% of the nodes will have 5J of initial energy. The other 80% will have 2J of initial energy. 5% of the total sensor nodes will be selected as cluster heads (K=5). The evaluation of our proposed clustering algorithm was done against three important protocols. First one is LEACH. LEACH is one of the pioneering clustering protocols. Second, LEACH-C, because we used a centralized clustering system which was first implemented in LEACH-C. Third is Particle Swarm Optimization (PSO), because PSO achieved the best optimization accuracy in clustering in several researches[25].

We simulated the network until all the nodes died. The data packet was set to 6500 bytes in length. The size of the control packet was 250 bytes. We used 30 Herds, P_a=0.25. Here P_a is the probability of krills that are discarded after every round. We set our constant β= 0.010. Table 1 shows the parameters we used in the simulation.

Table 1. Values of the Network Parameters

Parameters	Values
Number of nodes	100
Area size	500m X 500m
Coordinates of Base station	(250, 250)
Data size	6500 bytes
Control packet size	250 bytes
Number of clusters K	5% of total nodes

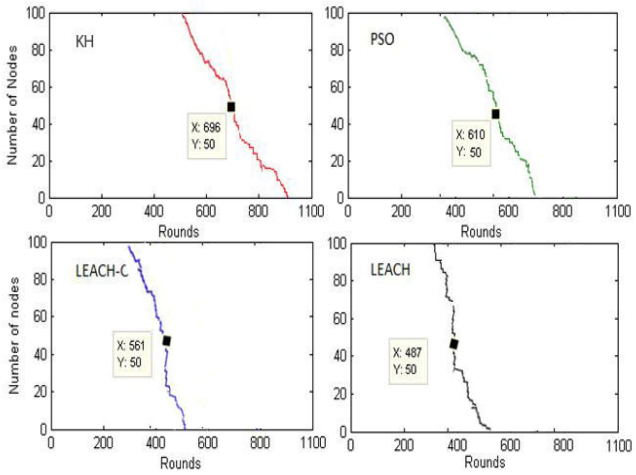


Fig 3: Lifetime of network after 50% node dies

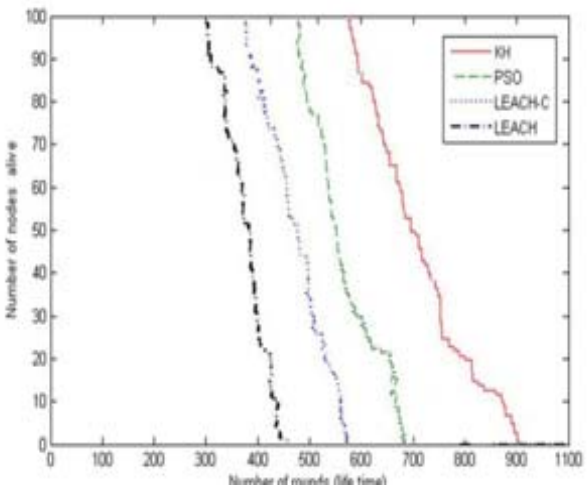


Fig 4: Lifetime of the network till all the nodes die

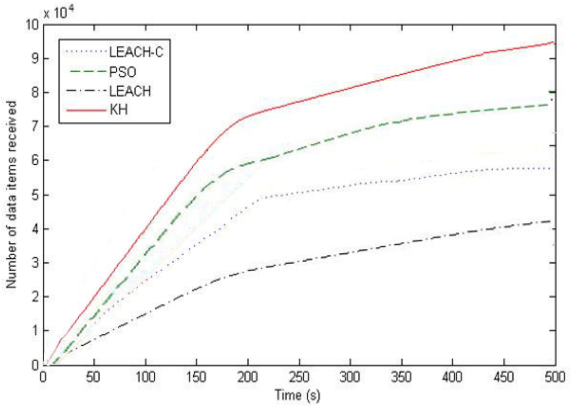


Fig 5: Total data received at the BS over time

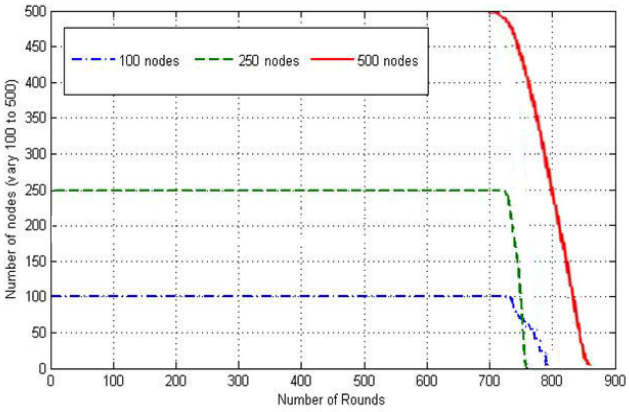


Fig 6: Scalability of the protocol

Figure 3 demonstrates the period after the loss of 50% of the total sensor nodes. Figure 4 shows total network life span, defined by number of sensors alive over rounds (here 100

nodes, 1100 rounds). Both the figures show that the clustering algorithm outperforms the overall network lifetime of PSO , LEACH-C and LEACH, by almost 9 %, 38 %, and 54 % respectively. Firstly, the reason is, with optimal CH distribution across the network and less amount of intra cluster distance, the proposed protocol gives better network lifetime. Thus, the energy consumption is condensed. Secondly, the adopted method produces decent adjustments where cost function values are acceptable to the WSN necessities.

Figure 5 shows the number of messages that are transmitted to the BS by all the techniques. KH increases the data delivery by factors of 17% over PSO, 53% over LEACH-C and 163 % over LEACH. KH algorithm can choose the node with high energy as CH by considering the minimum distance between the sensors and their respective CH and residual energy of the candidates of the Ch. Therefore, more packets are sent to the BS.

For the scalability evaluation, the KH clustering algorithm is simulated for a number of instances where the volume of sensor varies from 100, 200 to 500 nodes. This simulation illustrated the competence of the WSN scalability of our designed protocol. From Figure 6, it is noticeable that it produces reasonable results even when the number of nodes are augmented from 100 to 200 and 500 respectively.

VI. CONCLUSION AND FUTURE WORKS

In this paper, we used KH algorithm to define an energy-aware clustering algorithm for WSNs. Simulation results show that our projected algorithm achieved improved cluster formation by uniformly distributing the CHs all through the network area and thus maximizing the lifespan of the network than other popular algorithms like LEACH, LEACH-C, and PSO. Our future plan is to extend this algorithm to accomplish more improved results in more versatile scenarios.

References

[1] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, “Energy-Efficient Communication Protocol for Wireless Microsensor Networks,” in *33rd Hawaii International Conference on System Sciences*, 2000, pp. 1–10.

[2] W. C. Fang and T. H. Lin, “Low-power radio design for wireless smart sensor networks,” in *Intelligent Information Hiding and Multimedia Signal Processing 2006 IIHMSP06 International Conference on*, 2006, pp. 583–586.

[3] M. L. Florence and D. Swamydoss, “A SURVEY ON WIRELESS SENSOR NETWORK ARCHITECTURE , PROTOCOLS AND APPLICATIONS,” *J. Glob. Res. Comput. Sci.*, vol. 2, no. 2229–371x, 2011.

[4] M. Bahrepour, N. Meratnia, M. Poel, Z. Taghikhaki, and P. J. M. Havinga, “Distributed Event Detection in Wireless Sensor Networks for Disaster Management,” in *2010 International Conference on Intelligent Networking and Collaborative Systems*, 2010, vol. 2008, no. 11, pp. 507–512.

[5] G. Werner-Allen, K. Lorincz, M. Ruiz, O. Marcillo, J. Johnson, J. Lees, and M. Welsh, *Deploying a wireless sensor network on an active volcano*, vol. 10, no. 2. Published by the IEEE Computer Society, 2006, pp. 18–25.

[6] M. Patel and J. Wang, “Applications, challenges, and prospective in emerging body area networking technologies,” *Wirel. Commun. IEEE*, no. February, pp. 80–88, 2010.

[7] P. Honeine, F. Mourad, M. Kallas, H. Snoussi, H. Amoud, and C. Francis, *Wireless sensor networks in biomedical: Body area networks*, no. 1. IEEE, 2011, pp. 388–391.

[8] W. Wirawan, S. Rachman, I. Pratomo, and N. Mita, “Design of low cost wireless sensor networks-based environmental monitoring system for developing country,” *2008 14th AsiaPacific Conference on Communications*. pp. 1–5, 2008.

[9] J. M. Kahn, R. H. Katz, and K. S. J. Pister, “Emerging Challenges: Mobile Networking for Smart Dust,” *J. Commun. Networks*, vol. 2, no. 3, pp. 188–196, 2000.

[10] I. Vasilescu, K. Kotay, D. Rus, M. Dunbabin, and P. Corke, “Data collection, storage, and retrieval with an underwater sensor network,” *Proc. 3rd Int. Conf. Embed. networked Sens. Syst. SenSys 05*, vol. San Diego, p. 154, 2005.

[11] O. Ozdemir, R. N. R. Niu, and P. K. Varshney, “Channel Aware Particle Filtering for Tracking in Sensor Networks,” *2006 Fortieth Asilomar Conf. Signals Syst. Comput.*, no. 3, pp. 290–294, 2006.

[12] O. Ozdemir, R. Niu, and P. K. Varshney, *Tracking in Wireless Sensor Networks Using Particle Filtering: Physical Layer Considerations*, vol. 57, no. 5. 2009, pp. 1987–1999.

[13] Z. Zhang, “Investigation of wireless sensor networks for precision agriculture,” in *American Society of Agricultural and Biological Engineers*, 2004, vol. 300, no. 4, pp. 1157–1164.

[14] L. Ruiz-Garcia, L. Lunadei, P. Barreiro, and I. Robla, “A Review of Wireless Sensor Technologies and Applications in Agriculture and Food Industry: State of the Art and Current Trends,” *Sensors (Peterboroug*, vol. 9, no. 6, pp. 4728–4750, 2009.

[15] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, “Wireless sensor networks: a survey,” *Comput. Networks*, vol. 38, no. 4, pp. 393–422, Mar. 2002.

[16] M. A. Adnan, M. Abdur Razzaque, I. Ahmed, and I. F. Isnin, “Bio-mimic optimization strategies in wireless sensor networks: a survey.,” *Sensors (Basel).*, vol. 14, no. 1, pp. 299–345, 2013.

[17] X. Liu, “A Survey on Clustering Routing Protocols in Wireless Sensor Networks,” *Sensors*, vol. 12, pp. 11113–11153, 2012.

[18] A. H. Gandomi and A. H. Alavi, “Krill herd: A new bio-inspired optimization algorithm,” *Commun. Nonlinear Sci. Numer. Simul.*, vol. 17, no. 12, pp. 4831–4845, Dec. 2012.

[19] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, “An application-specific protocol architecture for wireless microsensor networks,” *IEEE Trans. Wirel. Commun.*, vol. 1, no. 4, pp. 660–670, 2002.

[20] A. Salehpour, B. Mirmobin, and A. Afzali-kusha, “An Energy Efficient Routing Protocol for Cluster-Based Wireless Sensor Networks Using Ant Colony Optimization,” in *IInternational Conference on Innovations in Information Technology*, 2008, pp. 455–459.

[21] M. Ziyadi, K. Yasami, and B. Abolhassani, “Adaptive Clustering for Energy Efficient Wireless Sensor Networks Based on Ant Colony Optimization,” in *2009 Seventh Annual Communication Networks and Services Research Conference*, 2009, pp. 330–334.

[22] T. Shih and S. Member, “Particle Swarm Optimization

Algorithm for Energy-Efficient Cluster-Based Sensor Networks," no. 7, pp. 1950–1958, 2006.

[23] N. M. A. Latiff, C. C. Tsimenidis, and B. S. Sharif, "Energy-aware clustering for wireless sensor networks using particle swarm optimization," in *IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications*, 2007, pp. 1–5.

[24] M. A. Adnan, M. A. Razzaque, M. A. Abedin, S. M. S. Reza, and M. R. Hussein, "A Novel Cuckoo Search Based Clustering Algorithm for Wireless Sensor Networks," in *Lecture Notes in Electrical Engineering*, 2016, vol. 362, pp. 621–634.

[25] M. A. Mehr, "Design and Implementation a New Energy Efficient Clustering Algorithm using Genetic Algorithm for Wireless Sensor Networks," *World Acad. Sci. Eng. Technol.*, vol. 53, no. 2, pp. 430–433, 2011.

[26] G. Wang, L. Guo, H. Wang, H. Duan, L. Liu, and J. Li, "Incorporating mutation scheme into krill herd algorithm for global numerical optimization," *Neural Comput. Appl.*, vol. 24, no. 3–4, pp. 853–871, 2014.

[27] V. Rodoplu and T. H. Meng, "Minimum energy mobile wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 3, no. 8. IEEE, pp. 1333–1344, 1999.

[28] G. G. Wang, A. H. Gandomi, and A. H. Alavi, "An effective krill herd algorithm with migration operator in biogeography-based optimization," *Appl. Math. Model.*, vol. 38, no. 9–10, pp. 2454–2462, 2014.

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

Automated Textile Defect Classification by Bayesian Classifier Based on Statistical Features

Md. Tarek Habib
Department of Computer Science and Engineering,
Jahangirnagar University
Dhaka, Bangladesh
md.tarekhabib@yahoo.com

Mohammad Shorif Uddin
Department of Computer Science and Engineering,
Jahangirnagar University
Dhaka, Bangladesh
shorifuddin@gmail.com

Shaon Bhatta Shuvo
Department of Computer Science and Engineering,
Daffodil International University
Dhaka, Bangladesh
shuvocste21@gmail.com

Farruk Ahmed
Department of Computer Science and Engineering,
Independent University, Bangladesh
farruk60@gmail.com

Abstract*—Textile inspection system, which carries a lot of importance in the production process of textile goods, has been part of a great deal of research for automating the process. Manual textile inspection is a lengthy, slow as well as erroneous job; therefore, automation of textile inspection is a demand of time. Machine vision based i.e. automated fabric inspection deals with two primary challenges, namely defect detection and defect classification. The quality and quantity of research done on defect detection and classification is still not up to the mark. Our focus is on detecting various defects in textile fabrics and classifying them. We extracted features using statistical techniques. Images of textile fabrics were used as sample. Inspecting the images, geometric and statistical features were found out. Through this paper, we bring in a suitable Bayesian classifier to classify the images into different classes of defective properties. Our approach has delivered acceptable accuracy compared to other works in the domain of textile defect detection and classification.

Keywords*—Textile defect, statistical features,defect detection, defect classification, Bayesian classifier, accuracy.

I. INTRODUCTION

Readymade garments items are the main source economic growth of Bangladesh. More specifically, textile industry is the backbone of Bangladeshi export sector. To gain maximum performance from this sector, the quality of products attracts the primary focus. As the competition is increasing day by day, there is no alternative of producing the best quality product to endure and dominate the global textile sector. Rapid as well as correct defect detection leads to minimizing the cost of improvement of quality. For detection of defects, the traditional approach is manual inspection that is lengthy and prone to human error. One of the best solutions of this problem could be checking the quality of the product in an automated way.

To work with the problem we have collected a sample dataset. Our dataset is comprised of 128 samples of fabrics where each samples consisting of 8 geometrical features along with 5 statistical features.

Features identification has been done using a statistical features extraction process. Various images of knitted fabrics have been converted into different predefined formats in order to extract the features. Once we identified various features from the given sample, Bayesian classifier has been deployed to those features in order to categorize the defects into various classes. Satisfactory results have been found throughout the work.

The arrangements for the remaining part of the paper are follows: Section II is used for Literature review. The methodology of our work has been discussed in Section III while, Implementations are discussed in Section IV. Section V has been used for experimental result and comparative analysis and finally in Section VI, we have discussed about the future prospect of our work.

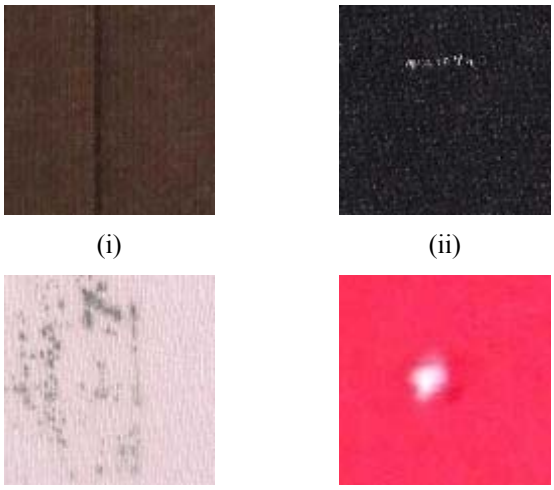
II. LITERATURE REVIEW

First of all, it is necessary to identify the various possible defects of fabric, i.e. in Textile defect detection and classification techniques, features from which the detection and classification is done play a pivotal role. Several feature sets have been used over the years; however, a suitable set is still debatable. Again, defect classes are also subject to research as different researchers used different features and classes. The features of various defects that we have chosen has been delimitated in Fig. 1.

Scientists have tried various automated artifact for the inspection of fabric [1]-[20]. Among all the approaches, only a few of them focused on identifying the classification rather than detects. It is observed that popular strategies for classification are done based on some popular machine learning schemes. Some activities [2]-[8],also concentrated on choosing of correct features [9]-[12].An approach[8]for classification of defect using Support Vector Machine (SVM) which was not preferable for real life implementation as the feature extraction process was very troublesome.

Classification of statistical conjecture was confined to binary classification [13], that is defective and without any defect [14]. It will not be possible to identify the classification of defect using the binary classification.

978-1-5090-5769-6/16/$31.00 ©2016 IEEE



(i) (ii) (iii) (iv)

Fig. 1. Various types of defects that occurr in knitted Textiles.

(i) Missing yarn, (ii) Color yarn, (iii) Spot, (iv) Hole.

Automation of fabric inspection is a real-time problem and needs a prompt suitable approach to deal with that, but the model-based approach [15] for clustering failed to meet the requirement.

Besides SVM and various model based clustering techniques, many works [6], [10] on classification have been done using Artificial Neural Networks (ANNs). In a remarkable work by M. T. Habib and M. Rokonuzzaman [16] on feature selection, ANN is used for defect classification but the size of the dataset was not very large and it was also done based on only geometric features. Various works [2], [9] uses Back-propagation learning approaches. Kumar [17] trained his network using Back-propagation algorithm and again trained separately using Least Mean Square Error simply known as LMS algorithm, but both of the cases it worked only for binary classification and cost was also very high. Though Saeidi *et al.* [2] trained their network in such a way so that it can handle multiclass problem but the application of the network was confined only to frequency domain and was not applicable for spatial domain. Kuo and Lee [9] also designed an ANN which had been implemented to handle multiclass problem, but the network failed to establish defect analysis and justification of feature in a broad category along with classification error. Another work [12] has been done to deal with multiclass problem which uses Learning Vector Quantization (LVQ) to train the network can be functional in both spatial as well as frequency domain but the problem was lack of defect analysis and no feature justification. The work on defect classification using ANN by Karayiannis *et al.* [6] used ANN for defect classification but feature justification for defect analysis failed to get the pass marks. Satisfactory results of defect classification using ANN has been found by Mitropulos *et al.* [1] but the dataset were too small to come into a conclusion. Islam *et al.* [4] and [7] also addressed multiclass problem using ANNs, application range has become limited as the sample along with the number of features size were too small.

It has been observed that SVM, model-based clustering, and various ANN techniques were not good enough to give the suitable result in classification of defective fabrics that is the main reason behind choosing Bayesian classifier for our work. So far various works [18]-[20] on defect classification used Bayesian classifier for their work, while Naïve Bayesian classifier has been used by Keerthika *et al.* [18] and Jeong *et al.* [19] also used Bayesian classifier for textile defect detection but was only applicable for binary classification as a result it could not handle the multiclass problem. A significant work [20] of defect classification based on Bayesian classifier has been done on geometrical features only. Since there has been no significant work of defect classification based on statistical features, the aim of our experiment is to accomplish the task of defect classification using only the discriminatory quality of statistical features.

III. Methodology

The pragmatic approach of our work consists of statistical defect detection as well as the classification. For identification of defects, we have deployed a statistical technique (thresholding technique) and we have used Bayesian classifier to classify different types of defects. The procedure begins with a supervised image of knitted textile goods, which subsequently transformed into a gray-level image. After that, the noises of the image need to be filtered to make it smooth. From the image a gray-level histogram is needed to form in order to calculate two threshold values by using histogram peach technique [21]. Segmentation of images takes place using those two threshold values that finally results in a binary image. Binary image comprises of defected object and defect free part a long with some noises. The noises are negligible compared to the minimum size of the defects that need to be detected. The minimum defect size that we want to detect is 3mm×1mm.Therefore, size-based threshold technique is used to remove noises. Different calculations on statistical features of various defects are needed to form a features vector. Finally, the classifications of various defects are done depending on the features vector using a previously trained Bayesian classifier.

A. Defect Types

We proposed four different types of defects based on most frequent appearing defects of knitted textile goods in Bangladesh. The classification of the defects are- color yarn, hole, missing yarn and spot. Fig. 1 shows these four distinct types of defects and which were described in details in [16].

B. An Appropriate Set of Statistical Features

For classifying defects, an appropriate set of statistical features have been selected. The set comprises of the following statistical features:

i) *Mean gray-level color intensity difference* ($\Delta\mu_{GS}$).

ii) *Mode gray-level Color Intensity difference* (ΔMo_{GS}).

iii) *Standard deviation of gray-level color intensity of pixels in defective regions* (σ_{GSDR}).

iv) *Coefficient of variation ofgray-level color intensity of pixels in defective regions* (CV_{GSDR}).

v) *Skewness of gray-level color intensity of pixels in defective regions* (γ_{GSDR}).

The details of all the five features can be found in [16].

C. Bayesian Classifier

Bayesian classifier is notable for classification and has been using for many years. Different unknown samples can be classified to various categories using Naïve Bayes classifier.

Let consider *M* and *N* are two events, therefore the Bayes theorem can be stated as:

$$P(M|N) = \frac{P(N|M)P(M)}{P(N)} \dots\dots\dots\dots\dots\dots\dots\dots (1)$$

Here,

- $P(N)$ represents as the evidence.
- $P(N|\ M)$ represents the class-conditional probability.
- $P(M)$ represents the prior probability for *M*.
- $P(M\ |\ N)$ represents the posterior probability for *M*.

Various advantages of using naive Bayesian are fast and computation is also easy, it is not impressionable to irrelevant data as it considers each attributes as independent element. Therefore, the probability of one attribute will not be affected by another attribute.

Again, the challenges of building a Bayesian classifier include bias, variance and noise of training data. The selection of distinguished features can reduce the training data noise while the effect of bias and variance is integrated within the system.

The methodical definition of Naïve Bayesian classifier can be demonstrated as:

$$C = arg \max_{k\in\{1,2,\dots,K\}} P(C_k) \prod_{i=1}^{n} P(x_i|C_k) \dots\dots (2)$$

Where,

- $P(C_k)$ = probability of class k
- $P(x_i|C_k)$ = probability of feature x_i given class k

As, the nature of attributes are continuous, the density estimation function can be stated considering the distribution as normal:

$$\varphi_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \dots\dots\dots\dots\dots\dots\dots (3)$$

Density function expresses the relative probability of a point, where, $\mu$and σ represent the mean and the standard deviation respectably. $\varphi_{\mu,\sigma}(x)$ is needed to calculate $P(x|C_k)$.

The model has been trained by calculating estimation of prior probability $P(C_k)$ and for every attribute A_i, for every attribute value v of A_i estimating $P(A_i = v|C_k)$.

Applying the probabilistic model for a given feature vector $(v_1, v_2, v_3, \dots, v_n)$, the class is picked that maximizes

$$P(C_k) \prod_{i=1}^{n} P(A_i = v_i|C_k) \dots\dots\dots\dots\dots\dots\dots\dots (4)$$

IV. Implementation

Our implementation included feeding the classifier with training samples to build the model. Since the classifier we built was solely on Bayesian theorem, the classifier needed a probability table described as "pTable". pTable is a three-dimensional array with rows representing classes, columns representing features and each cell consists of two levels. The first level contains the mean of the respective feature within that class and the second level contains the standard deviation of the same. The last column contains the prior probability of the class.

We fed the classifier with the sample data set. As mentioned earlier, our data set made up of 128 Textile samples along with various features, which have been represented using numerical values. An excel sheet has been used to represent those numerical values while the values has been adopted using image processing. Table I shows the distributions of the samples in different classes. Standard formulas have been used for finding means and standard deviations. The classifier needs to initialize and the values of the standard deviations along with mean are collected to get the classifier prepare to accomplish its job.

TABLE I. Samples and their distributions

Sample	Sample size	Training samples	Test samples
Horizontal Missing yarn	19	11	8
Vertical Missing yarn	19	10	9
Color Yarn	16	10	6
Spot	18	9	9
Hole	23	10	13
Defect free	33	20	13
Total	128	70	58

To make the classifier prepare to classify, samples are feed into the classifier one by one. The holdout method [22] is used in order to select the proportion of data reserved for training as well as testing. For every test sample of class classes, the posterior probability has been computed. After the calculation of posterior probability the class with the highest value represents each test sample. We performed the training of classifier with five distinguished classes, which are: horizontal missing yarn, vertical missing yarn, color yarn, spot, and hole. Each class was enumerated from 1, 2, 3, 4 and 5 respectively.

V. Experimental Result and Comparative Analysis

The results that we have achieved after implementation of our approach are promising enough. Table II describes the results obtained. Simple method of correctness ration has been used for accuracy calculation. It can be stated as:

$$accuracy = \frac{no.of\ correctly\ classified\ samples}{no.of tested\ samples} \times 100\%$$

TABLE II. DEFECT CLASSIFICATION PERFORMANCE

Class	Accuracy
Horizontal missing yarn	98.84%
Vertical missing yarn	99.05%
Color yarn	100%
Spot	99.45%
Hole	100.0%
Defect free	99.21%
Total	99.85%

To understand the merits of our work in textile defect classification, we need to delve into all results found. Since different researcher considered and classified classes in different way due to the variety of defects, as a result it has become difficult to compare between all the works. According to the review of Kumar [23] for automatic fabric defect detection more than 95% accuracy should be considered as industry standard. Considering this review the accuracy we have achieved has clearly crossed the standard.

The sample has been tested with various ratios between test dataset and training dataset. We have found poor accuracy while 20% of our sample dataset has been used for training purpose and the rest i.e. 80% of the sample data has been used for testing purpose. While increasing the percentage of training data to 25%, the accuracy has been improved but the improvement was very negligible. However, a drastic change in accuracy has been found after increasing the sample of training dataset to 33% to 50%. Fig. 2 demonstrates the accuracy curve rising along with the increase of training samples.

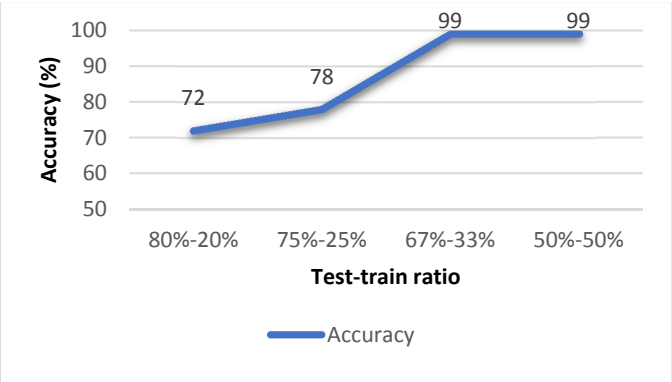


Fig. 2. Performance curve

VI. CONCLUSION AND FUTURE WORKS

We have exhibited an extensive pathway to classify textile defects in this paper. Since we completely relied on a dataset with features extracted with statistical techniques, the classification cannot be stated as final. We have considered thirteen features both of geometric and statistic characteristics. We are satisfied with the result of our work and can be considered well enough compared to other relevant works.

However, the accuracy can vary if we applied on a large dataset with similar features. Nevertheless, a lot of work remains to be accomplished. The data set could be enriched with more samples as well as with improvement in the quality of the images. We have a plan to work for reducing the amount of features to be used in machine vision based automated classification of textile defect detection.

REFERENCES

[1] R. Stojanovic, P. Mitropulos, C. Koulamas, Y. Karayiannis, S. Koubias, and G. Papadopoulos, "Real-time Vision Based System for Textile Fabric Inspection," *Real-Time Imaging*, vol. 7, no. 6, pp.507-518, 2001.

[2] R. Saeidi, M. Latifi, S. Najar, and A. Saeidi, "Computer Vision-aided Fabric Inspection System for On-circular Knitting Machine," *Textile Research Journal*, vol. 75, no. 6, pp. 492-497, 2005.

[3] W. Jasper, J. Joines, and J. Brenzovich, "Fabric Defect Detection Using a Genetic Algorithm Tuned Wavelet Filter," *Journal of the Textile Institute*, vol. 96, pp. 43-54, 2005.

[4] A. Islam, S. Akhter, and T. Mursalin, "Automated Txtile Defect Recognition System Using Computer Vision and Artificial Neural Networks," *Proceedings World Academy of Science, Engineering and Technology*, vol. 13, pp. 1-7, 2006.

[5] Y. Shu, Z. Tan, "Fabric Defects Automatic Detection Using Gabor Filters," in *Proceedings of the 5th World Congress on Intelligent Control and Automation*, pp. 3378-3380, June 2004, China.

[6] Y. Karayiannis, R. Stojanovic, P. Mitropoulos, C. Koulamas, T. Stouraitis, S. Koubias, and G. Papadopoulos, "Defect Detection and Classification on Web Textile Fabric Using Multiresolution Decomposition and Neural Networks," in *Proceedings of the 6th IEEE International Conference on Electronics, Circuits and Systems*, pp. 765-768, September 1999, Cyprus.

[7] A. Islam, S. Akhter, T. Mursalin, and M. Amin, "A Suitable Neural Network to Detect Textile Defects," *Neural Information Processing*, SpringerLink, vol. 4233, pp. 430-438, 2006.

[8] V. Murino, M. Bicego, and I. Rossi, "Statistical Classification of Raw Textile Defects," in *Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04),* pp. 311-314, 2004.

[9] C. Kuo, C. Lee, "A Back-propagation Neural Network for Recognizing Fabric Defects," *Textile Research Journal*, vol. 73, pp. 147-151, 2003.

[10] P. Mitropoulos, C. Koulamas, R. Stojanovic, S. Koubias, G. Papadopoulos, & G. Karayanis, "Real-time Vision System for Defect Detection and Neural Classification of Web Textile Fabric," *in Proceedings of SPIE Conference*, vol. 3652, pp.59-69, 1999, USA.

[11] H. A. Abou-Taleb, A. T. M. Sallam, "Online Fabric Defect Detection and Full Control in a Circular Knitting Machine,"*AUTEX Research Journal*, Vol. 8, No1, 2008.

[12] E. Shady, Y. Gowayed, M. Abouiiana, S. Youssef, and C. Pastore, "Detection and Classification of Defects in Knitted Fabric Structures," *Textile Research Journal*, vol. 76, pp. 295-300, 2006.

[13] F. Cohen, Z. Fan, and Z. Attali, "Automated Inspection of Textile Fabrics Using Textural Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, pp. 803-808, 1991.

[14] J. Campbell, A. Hashim, T. McGinnity, and T. Lunney, "Flaw Detection in Woven Textiles by Neural Network," in *Proceedings of the 5th Irish Neural Networks Conference*, pp. 92-99, September 1995, Ireland.

[15] J. Campbell, C. Fraley, D. Stanford, F. Murtagh, and A. Raftery, "Model-based Methods for Textile Fault Detection," *International Journal of Imaging Systems and Technology*, vol. 10, pp. 339-346, 1999.

[16] M. T. Habib, M. Rokonuzzaman, "Distinguishing Feature Selection for Textile Defect Classification Using Neural Network," *Journal of Multimedia,* vol. 6, no. 5, 2011.

[17] A. Kumar, "Neural Network Based Detection of Local Textile Defects," *Pattern Recognition*, vol. 36, pp. 1645-1659, 2003.

[18] G. Keerthika, D. Saravan Priya, "Feature Subset Evaluation and Classification using Naive Bayes Classifier", *Journal of Network Communications and Emerging Technologies (JNCET)*, vol. 1, issue 1, 2015.

[19] S. H. Jeong, H. T. Choi, S. R. Kim, J. Y. Jaong, and S. H. Kim, "Detecting Textile Defects with Computer Vision and Fuzzy Rule Generetion, part I: Defect classification by Image processing", *Textile Research Journal*, 2001.

[20] M. M. Mottalib, M. T. Habib, M. Rokonuzzaman, and F. Ahmed, "Fabric Defect Classification with Geometric Features Using Bayesian Classifier," in *Proceedings of the 3rd IEEE International Conference on Advances in Electrical Engineering*, pp. 137-140, December 2015, Dhaka, Bangladesh.

[21] D. Phillips, *Image Processing in C*. 2nd ed.,Kansas, USA: R & D Publications, 2000.

[22] P. N. Tan,M. Steinbach, and V. Kumar, *Introduction to Data Mining*. USA: Pearson Education Inc., 2006.

[23] A. Kumar, "Computer-Vision-Based Fabric Defect Detection: A Survey," *IEEE Transactions On Industrial Electronics*, vol. 55, no. 1, pp. 348-363, nuary 2008.

Maligant Kidney Tumor Ablation Using Electric Probe Heating

M. Tanseer Ali, Reazul Hasan Rasel, Mantasha Mubasshir, Salauddin Mahamood, Rafa Jarin
Department of EEE, Faculty of Engineering
American International University – Bangladesh (AIUB)
Dhaka, Bangladesh
tanseer@aiub.edu

S. M. Basitur Rashid
Department of Pathology
ENAM Medical College
Dhaka, Bangladesh
dr.basitroshid@gmail.com

***Abstract*— Thermal ablation technique applies to remove malignant tumors from an individual with minimal effect on the surrounding healthy tissue. It is one of the effective cancer treatment procedure. In this medical procedure, the malignant tissue is removed by heating it above 50°C. To do this it requires a local heat source, in which a small electric probe is inserted into the tumor tissue. Position, duration and power are some important factor to damage the tissue effectively. Less time cannot damage the affected tissue properly, on the other hand excess time heating will damage the healthy tissue. Finally, the result shows that 10-minute heating is the most effective time to remove 90% of the tumor cell while positioned at the center. In which the temperature of electric probe is 95^0 C and the tumor cell is destroyed damaging very small portion (maximum 10%) of healthy tissue.**

Keywords*— *Thermal Abolation Technique; Cancer Treatment; Electric probe heating;

I. INTRODUCTION

A lot of people suffer from malignant kidney tumors, mostly renal cell carcinoma. [1] To remove these tumors there are some specific treatments. But these treatments are complex and mostly cause more side effects e.g. damaging good cells, radiation effect. [2] According to the proposed ablation technique for malignant kidney tumor, we can avoid surgical procedures like excision or nephrectomy, depending upon the extent of invasion of tumor. [3] Research shows that in 1988-1992 kidney cancer rate was highest in France (16.1 in hundred thousand man and 7.3 in hundred thousand women per years and lowest in India (2.0 and 0.9 respectively). [4] It has shown in many research that high temperature is used to destroy harmful cancer cell and it is less harmful to human body than other techniques. [5]

Thermal ablation technique has become an alternative treatment to remove the tumor cells. This ablation technique is less harmful to healthy tissue and cost effective than other techniques. By inserting a four-armed electric probe through an electric current can run is accomplishes in this study. For this case bioheat equation models along with the temperature field in the tissue is used. The heat source which is resulting from the electric field is also known as resistive heating and The actual model comes from S. Tungjitkusolmun [6].

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

II. METHODOLOGY

In this case it has been considered that an electric probe is inserted into the tumor mass. The probe is made of a trocar (the main rod) which has four electrode arms. Trocar is electrically insulated but electrodes are not insulated. Electric current run through the probe which creates an electric field in the tissue. The electric field is strongest near the probe and this generates resistive heating, which dominates the probe's electrode arms. In this study thermal ablation technique using electric probe heating is observed by the absorption of temperature by the affected tissue. Temperature distribution is also analyzed and verified using damage tissue figures, temperature and damage tissue graph.

A. Simulation Model

The following model shown in Fig.1(a) and (b) demonstrates the block of kidney tissue with a tumor cell region and the electric probe with four electrodes is placed with in the vicinity.

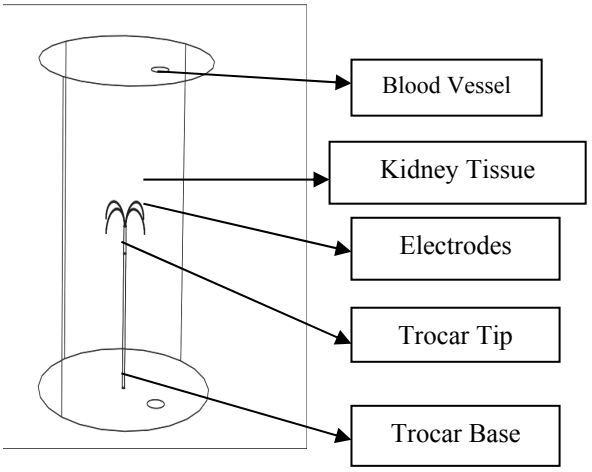


Fig. 1.(a) Kidney tissue and Electric probe Parts

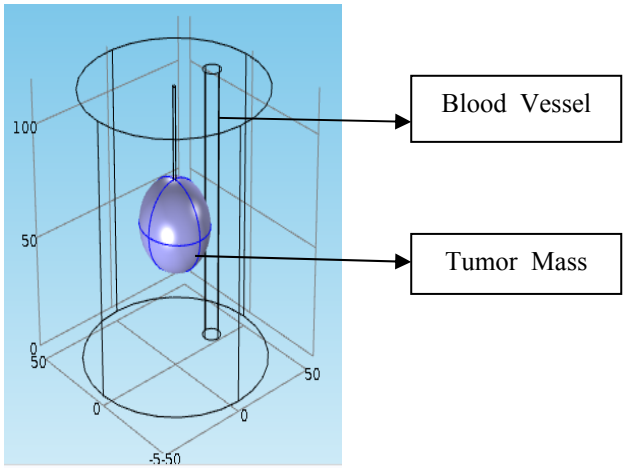


Fig. 1.(b) Tumor mass and Blood Vessel

To implement a transient analysis in this model, the Bioheat Transfer interface, the Electric Currents interface and a Electromagnetic Heat Source has been introduced. The standard temperature unit in COMSOL Multiphysics is kelvin (K) but in this model we use the Celsius temperature scale. The bioheat equation that transfer heat in the tissue is given as (1). [6]

$$\delta_{ts}\rho C\frac{\delta T}{\delta t}+\nabla.(-K\nabla T)=\rho_b C_b\omega_b(T_b-T)+Q_{met}+Q_{ext} \quad \ldots(1)$$

In (1) δ_{ts} is a time-scaling coefficient; ρ is the tissue density (kg/m^3); C is the tissue's specific heat (J/(kg·K)); and k is the thermal conductivity (W/(m·K)). On the right side of the equality, ρ_b gives the blood's density (kg/m^3); C_b is the blood's specific heat (J/(kg·K)); ω_b is its perfusion rate (1/s); T_b is the arterial blood temperature (K); and Q_{met} and Q_{ext} are the heat sources from metabolism and external heating, respectively (W/m^3).

From (1) if Q_{met} and T_b is considered as constant over time interval, and then the differential function of the temperature T can be solved for the for the given external spatial heating from the electric probe over certain period. Hence temperature effects from the heating probe can be measured from this equation. The specific heat and the thermal conductivity of the analyzed tissue can reveal how the heat is spread from the core temperature change of the electrodes.

B. *Material Properties*

For solving all the boundary values FEM (Finite Element Method) was used. Dielectric properties of Kidney tissue and Tumor were taken from the analysis [7][8], which are summarized in Table I.

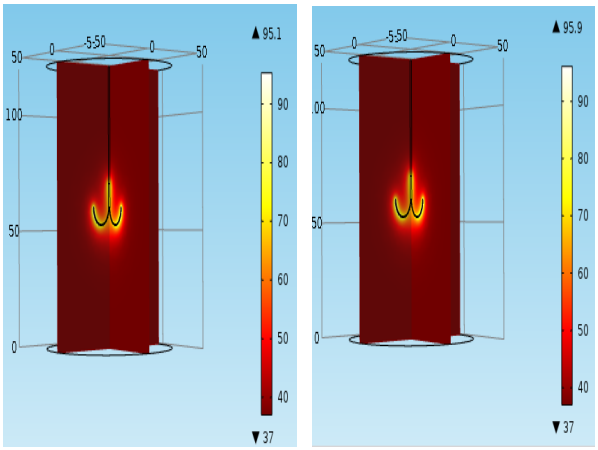
TEBLE I: DIELECTRIC PROPERTIES OF NORMAL KIDNEY TISSUE AND TUMOR

Properties	*Values*	*Units*
Heat capacity (constant pressure)	3540	J/(Kg·K)
Density	1079	Kg/m^3
Relative permittivity of Kidney	1	-
Electric conductivity of Kidney	0.333	S/m
Thermal conductivity of Kidney	0.50	W/(m·K)
Thermal conductivity of tumor	0.6	W/(m·K)
Electric conductivity of tumor	3	S/m
Relative permittivity of tumor	1	-

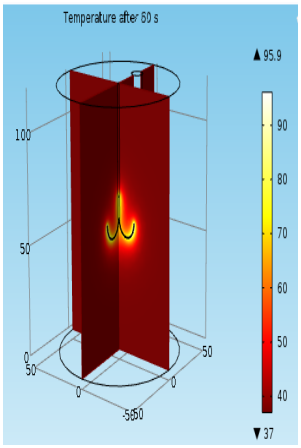
III. RESULTS INVESTIGATION

A. *Temperature of Electrodes and Surrownding Tissues:*

Fig. 2. (a), (b) and (c) show the simulated result of temperature of the electrodes and tissue after 5, 10 and 15 minutes. After 5-minute temperature increases from 37^0 C to maximum 95.1^{0}C and temperature spread in a small region. But after 5 and 10 minutes temperature increases to 95.9^{0}C which is enough to destroy the tumor cells. Although after 5 min the heat spreads around most of the tumor cell and allowing more time consecutively starts to damage the healthy cells.



(a)After 5 min (b) After 10 min



(c) After 15 min

Fig. 2. Temperature of Electrodes and Tissue

B. *Tempareture And Damaged Tissue Graph:*

Fig. 3(a), the temperature at one electrodes after 10 minutes demonstrates that temperature increases from 0^0C to 95^0C. After first 5 min the temperature raises about 94^0C and takes 10 minutes when the temperature is in the saturation region and the temperature reaches to 95^0C.

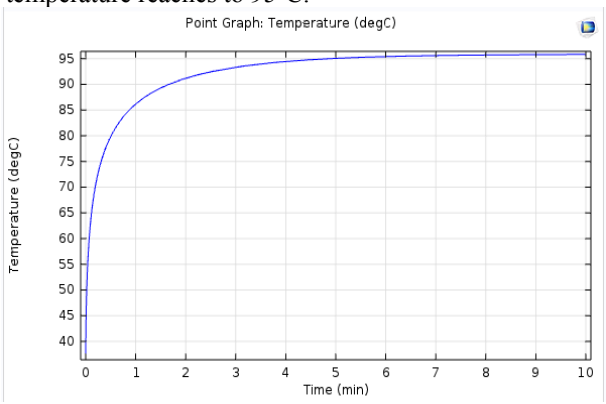


Fig. 3.(a) Temperature of One Electrode

Fig. 3(b) shows five cut points and damage tissue curve. Different color shows different cut point. The pink line shows the cut point outside the edge and blue is at the center of the malignant tissue. In this graph shows that with increasing time temperature increases along with tissue damage. At the edge after 10 minutes of heating about 30% cell is damaged. More heating time will increase the fraction of necrotic tissue but also starts to damage the healthy tissues.

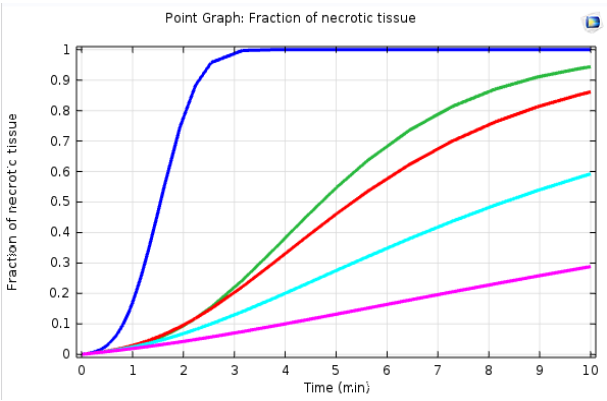


Fig. 3.(b) Fraction of Necrotic Tissue

C. *Electric Potential:*

Fig. 4 shows the electric potential of the electrodes. As the electric input power has been kept persistent over time, the electric potential for any given time reaches to highest value of 22V and lowest value of 0V.

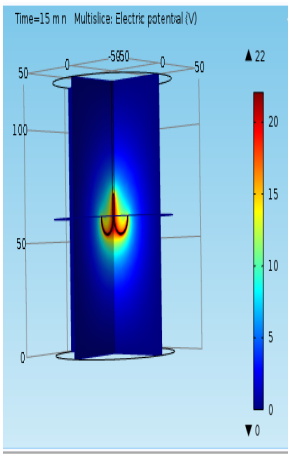
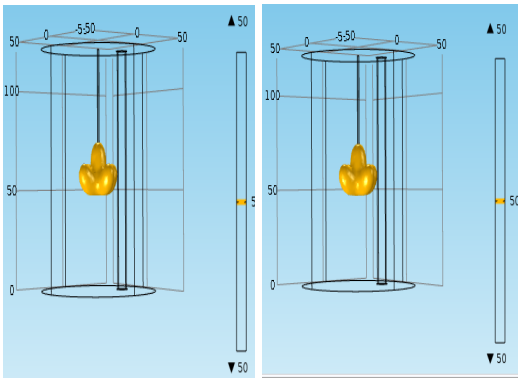


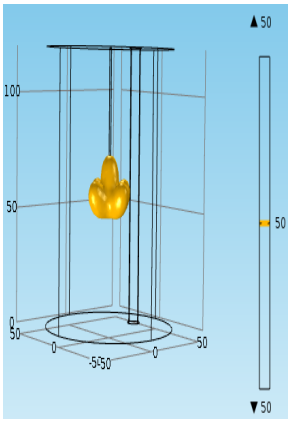
Fig. 4. Temperature of Electrodes and Tissue

D. *Isothermal Counter:*

Fig. 5.(a), (b) and (c) show the isothermal counter of the tumor mass after 5, 10 and 15 minutes accordingly. Iso-thermal surface is 50^0C and from the simulation it has been shown that after 5 minute it has covered a very small area. After 10 minutes, it covers whole of the tumor mass and eventually after 15 min the isothermal counter go past tumor cell, that means it started to damage the healthy tissues.



(a)After 5 min (b) After 10 min

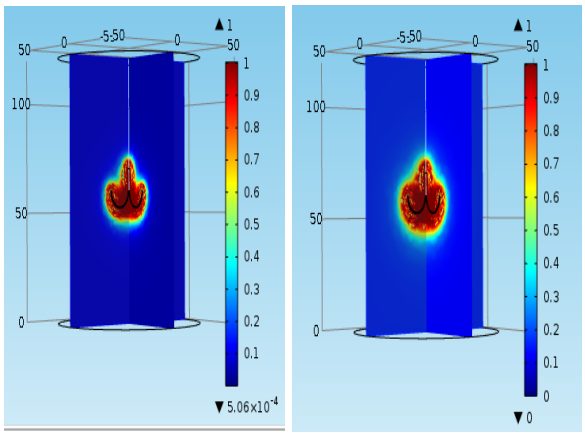


(c) After 15 min

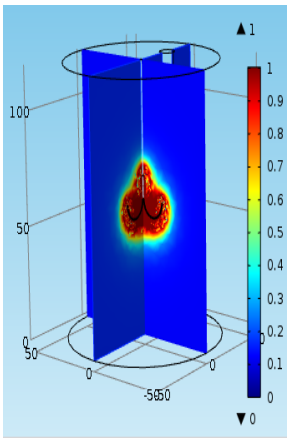
Fig. 5. Isothermal Counter Of Tumor Cell.

E. Damaged Tissue Cell:

Fig. 6.(a), (b) and (c) shows the damaged tissue after 5, 10 and 15 minutes accordingly. The red regions are 100% damaged tissue whereas green are identifying 50% damaged tissue region. Hence it can be observed that after 5 minutes, tissue has damaged only at the adjacent area to the electrodes. While after 10 minutes of persistent heating, it has damaged 80% area at the edge of the tumor cell. And then after 15 minutes it has damaged almost all the tumor cells, but starts to damage the healthy tissue.



(a)After 5 min (b) After 10 min



(c) After 15 min

Fig. 6. Damaged Tissue

IV. CONCLUTION

This work presents the analysis the heat transfer using the electric probe in a malignant renal mass. To accomplish this temperature distribution, electric potential of probes and tissue and their effect on kidney tissue and tumor is investigated in this paper. The observed results can be summarized as Table II.

TEBLE II: SUMMARY OF THE RESULTS

Time	*Temperature of electrodes and surrounding tissue (^{0}C)*	*Electric Potential (V)*	*Damaged tissue (%)*
5	95.1	22	80
10	95.9	22	100
15	95.9	22	Above 100

From the analysis, it shows that the damage of tumor cell is depend on the time how long the electricity pass and the temperature of electrodes. 95^{0}C is enough to destroy or remove the cancerous tumor cell from the healthy tissue and the effective time is 10 minutes. This method is more easy, cost effective and less risky than any other method such as microwave ablation [9]. To develop the proposed technique into a low cost effective treatment for malignant kidney tumor a practical implementation with tissue and tumor properties can be investigated for the future works.

REFERENCES

[1] Doss, J.D. and McCabe, C.W. "Method for localizing heating in tumor tissue" Google Patents, US Patent 4,016,886, 1977.

[2] Alexander R. zlotta, Thierry Wildschutz, Gilraviv, Marie-Odile Peny, Daniel van Gansbeke, Jean-Christophe Noel, and Claude C. schulman. Journal of Endourology. March 2009, 11(4): 251-258. doi:10.1089/end.1997.11.251.

[3] Eggers, P.E. and Shmulewitz, A. "Apparatus and method for characterization and treatment of tumors" Google Patents, US Patent 5,630,426,1997.

[4] Streib, G.H. and Schwindt, R.J."Probe station for low current, low voltage parametric measurements using multiple probes" Google Patents, US Patent 6,031,383, 2000.

[5] Mathew and Devesa. "Global increases in kidney cancer incidence, 1973–1992",April 2002 - Volume 11 - Issue 2.

[6] S. Tungjitkusolmun, S. Tyler Staelin, D. Haemmerich, J.Z. Tsai, H. Cao, J.G. Webster, F.T. Lee, Jr., D.M. Mahvi, and V.R. Vorperian, "Three-Dimensional Finite Element Analyses for Radio-Frequency Hepatic Tumor Ablation," IEEE Transactions on Biomedical Engineering, vol. 49, no. 1, 2002.

[7] S.Gabriel, R.W.Lau, and C. Gabriel," "The dielectric properties of biological tissues: II. Measurements in the frequency range 10 Hz to 20 GHz." , Phys. Med. Biol., vol. 41, no. 11, pp. 2251-2269,Nov.1996.

[8] R. L. Mcintosh and V. Anderson, "Erratum: 'A Comprehensive Tissue Properties Database Provided for The Thermal Assessment of A Human at Rest." , Biophys. Rev. Lett., vol. 08, no. 01n02, pp. 99-100, Jun. 2013.

[9] Liang,Ping et al,"Ultrasound Guided Percutaneous Microwave Ablation for Small Renal Cancer: Initial Experience".The Journal of Urology , Volume 180 , Issue 3 , 844 – 848.

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

Design & Simulation of Fuzzy Logic Based Speed Control of Electrical Vehicle with IPMSM taking Core loss into Account

Farzana Ahmed, Mohammad Abdul Mannan

Department of Electrical & Electronics Engineering
American International University- Bangladesh
Email: ahmedfarzana22@yahoo.com

***Abstract*—Because of advantages of Electrical Vehicles (EV), people are becoming more interested in using them rather than using mechanical differentials. In electrical vehicles different types of electrical machines such as Interior Permanent Magnet Synchronous Motor (IPMSM), Induction Motor etc. are used. The design of a controller is a challenging work, as the output of the motor has to match with vehicle input. So, far, most of the reported works have utilized proportional-integral (PI) controllers as the speed control. But, the disadvantage of PI controller are well known, as its design depends on the exact motor parameters & the performance is sensitive to system disturbances. Thus one of the main objective of this paper is to replace the conventional PI controller by an Fuzzy logic controller which is capable of handling highly non-linear IPMSM motor for high performance application in Electrical Vehicle taking core loss into account. The main advantages of FLC over the conventional controllers are that the design of FLC does not depend on machine parameters. The effectiveness of the proposed Fuzzy Logic controller based IPMSM taking core loss into account has been verified by using MATLAB Simulink Software. In simulation work different road conditions for EV are considered. After the simulation fuzzy logic controller is found to be robust for the speed control application of Electrical Vehicle with IPMSM taking core loss into account.**

Keywords—Electrical Vehicle; Interior Permanent Magnet Synchronous Motor; Fuzzy Logic Controller;Proportional Integral;

I. INTRODUCTION

In this modern era people prefer more automated and safe electrical vehicles because of its cleanest, most efficient, and most cost-effective form of transportation around the world. Electric vehicles use energy stored in its rechargeable controlling a vehicle in different conditions of road has become one of the biggest issues.For this reason, a need arises to build more efficient control, light weight, and compact electric propulsion systems, so as to maximize driving range per charge. Due to its high torque-current ratio, large power-weight ratio, high efficiency, high power factor, low noise and robustness, the permanent magnet synchronous motor (PMSM) is the preferred motor topology in today's automotive applications [1], [2].PMSM with interior or buried magnet in rotor known as IPMSM has higher power density compared with surface mounted magnet (SPMSM) according to industrial standard so far [3].

Permanent magnet synchronous motors (PMSM) have emerged as a very strong contender to replace induction motors used in electronically controlled variable speed applications in recent years. In most cases, PMSMs can provide superior performance in terms of increased efficiency and reduced noise. [2, 3].Depending on the placement of PM, PMSM are called either Surface Permanent Magnet Synchronous Motor (SPMSM) or Interior Permanent Magnet Synchronous Motor (IPMSM).IPMSM, having interior mounted permanent magnet rotor, is a good candidate for high-speed operation [4].

The challenge comes to researchers on how to control IPMSM efficiently due to its robustness and non-linearity characteristic. Initially fixed gain PI controllers were employed by researchers as speed controller of IPMSM drive system because of their simplicity [5, 6, 7]. However, it is very hard to determine suitable PI controller parameters in controlling complex non-linear, incompletely modeled or uncertain systems since the speed and the torque of the motor always change depends on terrain and traffic condition traveled by electric vehicle. Besides, this PI controller will exhibit poor transient response thus it's not suitable for this application due to complexity of this system[8].

Intelligent controllers offer many advantages as their design do not need the exact mathematical model of the system and theoretically they are capable of handling any non-linearity of arbitrary complexity [9]. Besides they exhibit excellent dynamic response. Among the various intelligent controllers, fuzzy logic controller (FLC) is the simplest and better in terms of response time, insensitivity to parameter and load variations [9], [10].

Fuzzy logic idea is similar to the human being's feeling and inference process. Unlike classical control strategy, which is a point-to-point control, fuzzy logic control is a range-to-point or range-to-range control[11].

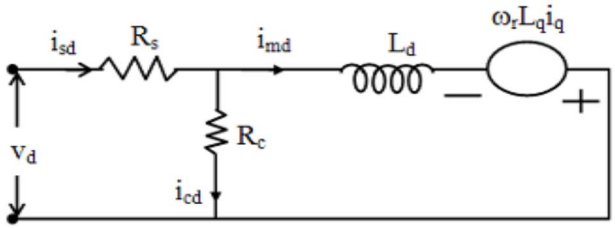
In this paper the FLC system is proposed to control of IPMSM as an application of electrical differential of an electrical vehicle. Overall fuzzy logic control system including IPMSM models taking core loss into account, Electrical Vehicle model & Fuzzy Logic Controller is simulated in MATLAB Simulink.

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

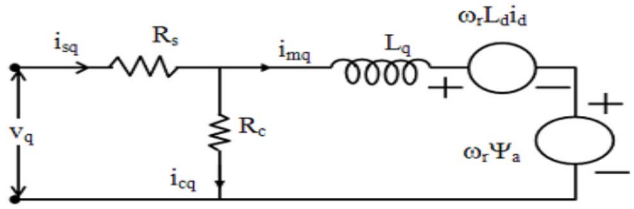
The output response is satisfactory compare with conventional controller output.

II. Dynamic & State Space Models of IPMSM

The dynamic equivalent circuit of an IPMSM taking core loss into account in synchronously rotating *d-q* axis reference frame [12,13] is shown in **Fig. 1.** Hysteresis loss and eddy current loss are the two components of core loss but for the sake of simplicity, the components of core loss are represented by a single quantity [12,14, 15] with a constant value of R_c as shown in **Fig. 1.**



(a). d-axis



(b). q-axis

Fig: 1: Equivalent Circuit of IPMSM model

According to **Fig. 1**, the *d-q* axis stator voltage of an IPMSM as follows:

$$V_d = R_s i_d + P\varphi_d - \omega_r \varphi_q \quad (1)$$

$$V_q = R_s i_q + P\varphi_d - \omega_r \varphi_q \quad (2)$$

where, $\rho = \left(\frac{d}{dt}\right), \quad R_{sc} = 1 + (R_s + R_c)$

The d-q axis flux linkages are:

$$\varphi_d = L_d i_{md} + \Psi_a \quad (3)$$

$$\varphi_q = L_q i_{mq} \quad (4)$$

The relation of current of IPMSM are given as:

$$i_{md} = i_d - i_{cd} \quad (5)$$

$$i_{mq} = i_q - i_{cq} \quad (6)$$

The mechanical dynamic of IPMSM can be expressed as:

$$T_e = T_L + J_m \rho_{\omega_r} B_m \omega_r \quad (7)$$

The electromagnetic torque is:

$$T_e = P_n[\Psi_n i_{mq} + (L_d - L_q) i_{mq} i_{md}] \quad (8)$$

The state equations of an IPMSM can be obtained by using the relation Eqs. (1-8)

$$p\omega_r = -\left(\frac{B_m}{J_m}\right)\omega_r + \left(\frac{P}{J_m}\right)(T_e - T_l) \quad (9)$$

$$pi_{md} = -\left(\frac{R_s}{R_{sc}L_d}\right) i_{md} + \left(\frac{L_q}{L_d}\right)\omega_r i_{mq} + \left(\frac{v_d}{R_{sc}L_d}\right) \quad (10)$$

$$pi_{mq} = -\left(\frac{R_s}{R_{sc}L_q}\right) i_{mq} + \left(\frac{L_d}{L_q}\right)\omega_r i_{md} - \left(\frac{\Psi_a}{L_q}\right)\omega_r + \left(\frac{v_q}{R_{sc}L_q}\right) \quad (11)$$

Eqs. (9) through (11) represent the state space model, having the states ω_r, i_{md}, i_{mq} of IPMSM. The input can be selected as v_d and v_q while the output can be selected as ω_r to control of speed of IPMSM .

III. Electrical Vehicle Mechanical Load's

The vehicle is considered as a load is characterized by different torques, which is mostly considered as resistive torques [12,13,16]. The vehicle torque is defined by the following relationship [12,17]:

The total resistive torque is defined as:

$$T_{res} = T_{roll} + T_{aero} + T_{slope} \quad (12)$$

The rolling force can be given as:

$$Troll = \mu MgRw \quad (13)$$

The aerodynamic Torque is:

$$Taero = (½)\, \rho Cx\, Sv2h\, Rw \quad (14)$$

The slope torque can be given as:

$$Tslope = Mg \cdot sin\beta\, Rw \quad (15)$$

IV. Electronic Differential Modeling

A simplified, trigonometry-based Ackerman steering model was used to obtain the Electronic Differential.

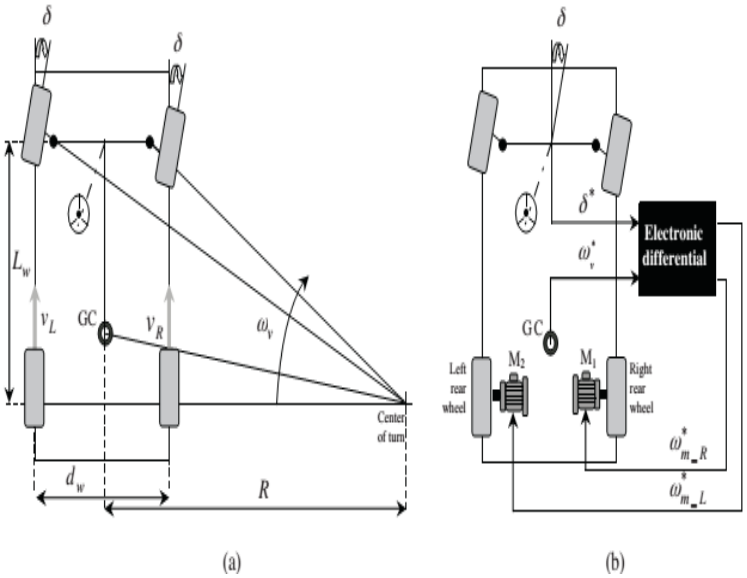


Fig. 2: (a) Structure of Electronic Differential. (b) Design Model for Vehicle Structure Driven in a Curve.

In Fig. 2 since the two rear wheels are directly driven by two separate motors, the speed of the wheel at the outer position of the curve will need to be greater than the speed of the inner wheel during curved steering (and vice versa); this helps the tire from losing traction in turns. Fig. 2(b) shows the vehicle structure describing a curve, where L_w represents the wheelbase, δ the steering angle, $d_w$the distance between the wheels of the same axle and ω_l and $\omega_r$the angular speeds of the left and right wheel drives, respectively.

$u_L$is the linear speed of the left wheel drive, $u_R$is the right wheel drive. The linear seed of each wheel drive is expressed as a function of the vehicle angular speed and the radius of the curve, according to Fig. 2.

$$u_L = \omega_v \left(R + \frac{d_\omega}{2}\right) \quad (16)$$

$$u_R = \omega_v \left(R - \frac{d_\omega}{2}\right) \quad (17)$$

The curve radius is related to the wheelbase and steering angle:

$$R = \frac{L_\omega}{tan\delta} \quad (18)$$

Substituting Eq. (18) into Eqs. (16) and (17), we obtain the angular speed in each wheel drive:

$$\omega_{rL} = \frac{L_\omega + ½\ d_\omega \tan\delta}{L_w}\omega_v \quad (19)$$

$$\omega_{rL} = \frac{L_\omega - ½\ d_\omega \tan\delta}{L_w}\omega_v \quad (20)$$

The difference between the angular speeds of the wheel drives is expressed by the relation:

$$\Delta\omega = \omega_{lR} - \omega_{rR} = \frac{(d_\omega \cdot \tan\delta)}{L_\omega}\omega_v \quad (21)$$

The direction of turn of an EV can be determined by the applied steering angle (δ) as follows:

$\delta > 0 = Turnright$
$\delta = 0 = Straightahead$
$\delta < 0 = Turnleft$

When the steering angle is applied the speed of inner wheel is reduced and the speed is increased of outer wheel. The driving wheel reference angular speeds are calculated by:

$$\omega^*_{lR} = \omega_v + \frac{\Delta_\omega}{2} \quad (22)$$

$$\omega^*_{rR} = \omega_v - \frac{\Delta_\omega}{2} \quad (23)$$

The speed references of the two motors are:

$$\omega^*_{lm} = kgear\omega^*_{lR} \quad (24)$$

$$\omega^*_{rm} = kgear\omega^*_{rR} \quad (25)$$

Where, $k_{gear}$is the gearbox ratio.

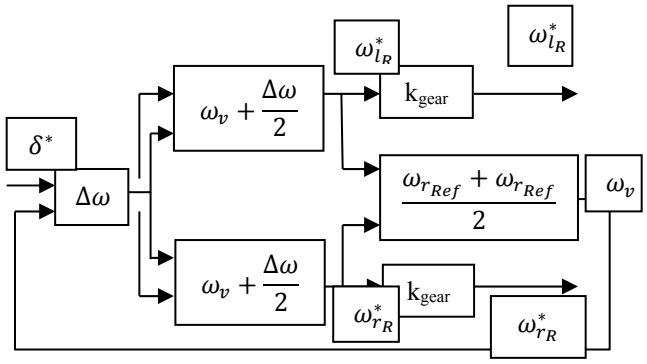


Fig. 3: Block Diagram to Show Use of the Electronic Differential

V. FUZZY LOGIC CONTROLLER

Fuzzy logic is widely used in machine control. The term "fuzzy" refers to the fact that the logic involved can deal with concepts that cannot be expressed as the "true" or "false" but rather as "partially true". Although alternative approaches such as genetic algorithms and neural networks can perform just as well as fuzzy logic in many cases, fuzzy logic has the advantage that the solution to the problem can be cast in terms that human operators can understand, so that their experience can be used in the design of the controller. This makes it easier to mechanize tasks that are already successfully performed by humans.

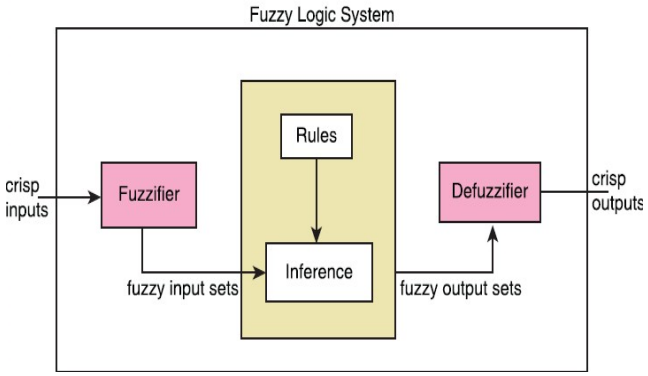


Fig. 4: *Simplified Block Diagram of Fuzzy logic Controller.*

Three essential parts of fuzzy logic controller are:

i. Fuzzifier:

The process of converting a numerical 'variable (real number) into a linguistic variable (fuzzy number) is fuzzification. In FLC, the input is always a crisp numerical value limited to the universe of the input variable and the output is fuzzy degree of membership in the qualifying linguistic set (always between 0 and 1).

To design the proposed FLC, the first assumption is that the difference of the actual speed from the desired or reference

speed, $e_\omega(k)$ and the difference of speed error, $\Delta e_\omega(k)$ are the only available controller input linguistic variables. The first difference of magnetizing q-axis current, Δi^*_{mq}, is considered as the output linguistic variable. For convenience, the inputs and output of the FLC are scaled with different coefficients K_ω, K_e and K_i respectively. The scaling factors K_ω, K_e and K_i are chosen for error of speed, change of error of speed and first difference of magnetizing q-axis current respectively as shown in **Fig. 5**. The triangular membership functions with overlap used for the input and output fuzzy sets are shown in **Fig. 6** in which the linguistic variables are represented by NB (Negative Big), NM (Negative Medium), NS (Negative Small), Z (Zero), PS (Positive Small), PM (Positive Medium) and PB (Positive Big). The grade of input membership functions can be obtained as follows.

$$\mu(x)=[w - 2|x-m|]/ w$$

where, $\mu(x)$ is the value of grade of membership, w is the width and m is the coordinate of the point at which the grade of membership is 1, x is the value of the input variable.

ii. Rule Base:

The fuzzy mapping of the input variables to the output is represented to fuzzy IF-THEN rules of the following form:

$$\text{If } \langle e^n_\omega \; is \; N\rangle \text{ and } \langle \Delta e^n_\omega \; is \; N\rangle \text{ Then } \langle \Delta i^n_{mq} \; is \; S\rangle$$

$$\text{If } \langle e^n_\omega \; is \; P\rangle \text{ and } \langle \Delta e^n_\omega \; is \; Z\rangle \text{ Then } \langle \Delta i^n_{mq} \; is \; B\rangle$$

The entire rule base is given in **Table 1.** There are total 49 rules to achive desired speed trajectory.

iii. Interface System:

From the rule base in **Table 1**, the inference engine provides fuzzy value of Δi^*_{mq} and then crisp numerical value of Δi^*_{mq} is obtained by using defuzzification procedure. The most popular method of inference and defuzzification is Mamdani's max-min (or sum-product) composition with center of gravity method [18]. In this work, Mamdani type fuzzy inference [18] is used. The center of gravity method [18] is used for defuzzification to obtain Δi^*_{mq}. The normalized output function is given as:

$$\Delta i^*_{mq}(n) = \sum_{i=1}^{N} \mu_i C_i / \sum_{i=1}^{N} \mu_i \qquad (26)$$

Where, N is the total number of rules, μ_i is the membership grade for i^{th} rule and C_i is the coordination corresponding to the maximum value of the respective consequent membership function $[C_i \in \{0.0, 0.9\}]$. After finding out Δi^*_{mq}, the actual desired first difference magnetizing current, Δi^*_{mq} can be found out by product of scaling factor as shown in **Fig. 5**

iv. Defuzzifier:

After the inference step, the overall result is a fuzzy value. This result should be defuzzified to obtain a final crisp output. This is the purpose of the defuzzifier component of a FLS. Defuzzification is performed according to the membership function of the output variable.

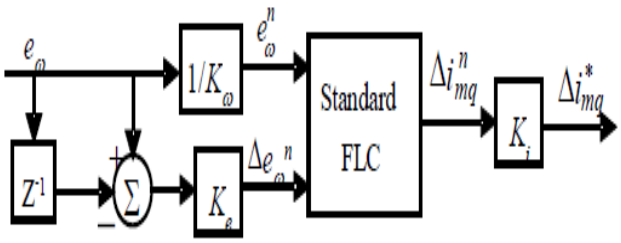


Fig. 5: Scheme structure of the proposed FLC.[18]

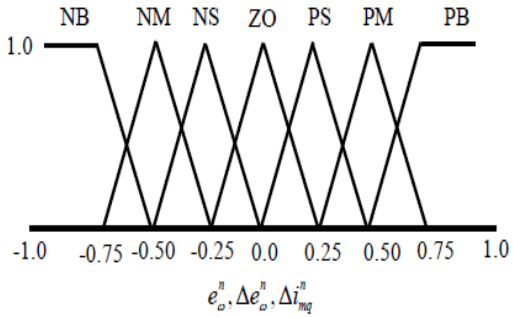


Fig. 6: Fuzzy sets and their corresponding membership function.[18]

Table:1. Fuzzy Rules

e^n_ω \ Δi^*_{mq}	Δe^n_ω NB	NM	NS	ZO	PS	PM	PB
NB	NB	NB	NM	NM	NS	NS	ZO
NM	NB	NM	NM	NS	NS	ZO	PS
NS	NM	NM	NS	NS	ZO	PS	PS
ZO	NM	NS	NS	ZO	PS	PS	PM
PS	NS	NS	ZO	PS	PS	PM	PM
PM	NS	ZO	PS	PS	PM	PM	PB
PB	ZO	PS	PS	PM	PM	PB	PB

The 49 rules shown in tabular form are written in the program according to the syntax provided by MATLAB. The document is saved with the extension '.fis'

VI. SIMULATION & RESULTS

Table 2& 3 shows the parameters of IPMSM which used in the simulation test. The simulation has been carried out using Matlab/ Simulink and Fuzzy Logic Toolbox. From the Graphs we can see when car will trun left then the right side speed will be increased & for left side it will be vice versa. When the car will go to up hills it will increase its torque & for down wards the torque will be decreased.

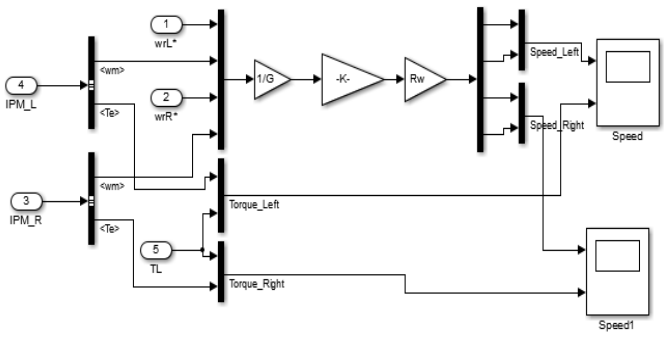


Fig: 6.a. IPMSM_ FLC Speed Measurement

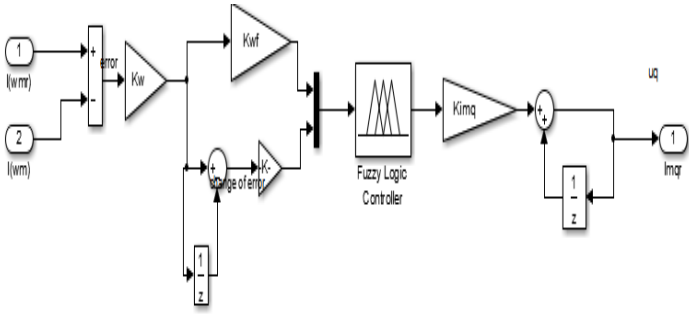


Fig: 6.b. Fuzzy Logic Speed Control

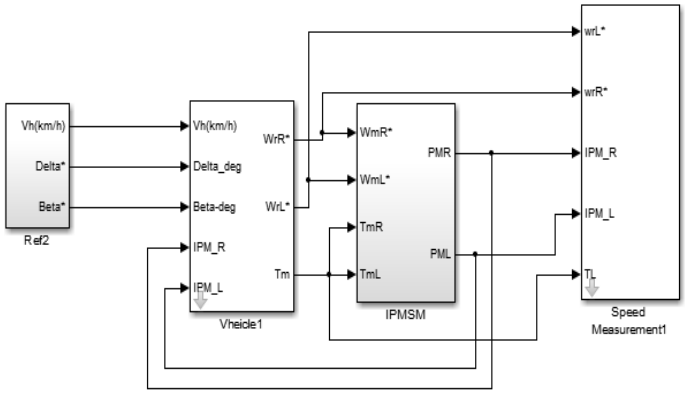


Fig: 7. Simulink Model of IPMSM Slope

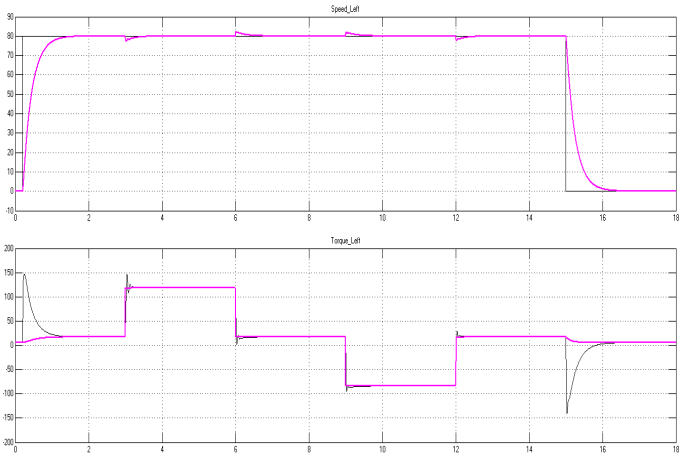


Fig: 7.a.Speed & Torque response of IPMSM_Left_FLC Slope

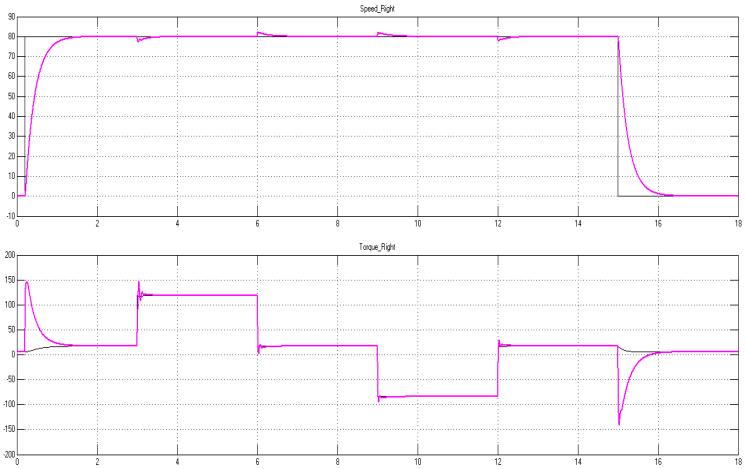


Fig: 7.b. Speed & Torque response of IPMSM_ Right_ FLC Slope

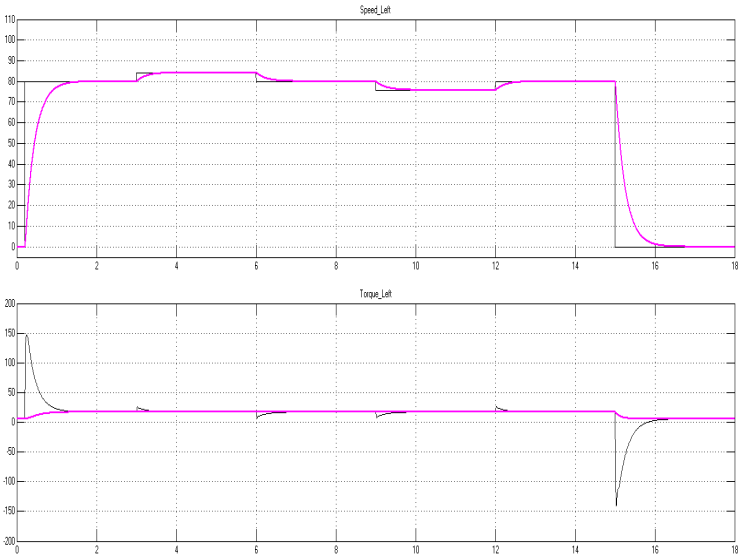


Fig: 7.c .Speed & Torque response of FLC_ Vehicle _Left

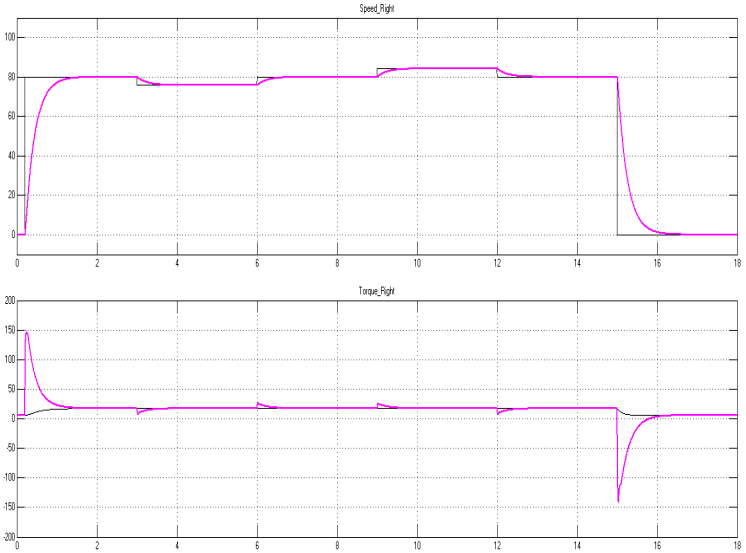


Fig:7.d.Speed & Torque response of FLC_ Vehicle _Right

Table 2: Rating and Parameters of the Vehicle

R_w	Wheel radius	0.32 m
H	Total transmission efficiency	93%
M	Vehicle mass	1300 Kg
C_d	Aerodynamic coefficient	0.46
A	Vehicle frontal area	2.6 m2
f_r	Vehicle friction coefficient	0.01
β	Grade angle of the road	Rad
L_w	Distance between two wheels and axes	2.5 m
d_w	Distance between the back and the front wheel	1.5

Table 3: Ratings and Parameters of the IPMSM

Stator resistance, R_s	5.8 Ω
Core loss resistance, R_c	100 Ω
Rated speed, Nm	1500 rpm
Rotor flux constant, λ_{af}	0.533 Wb
Motor moment of inertia, J	0.00529 kgm^2
Sampling time	0.00100
Number of pole pairs, P_n	2
Mechanical torque, T_m	40.77 Nm
d-axis stator inductance, L_d	44.8e-03
q-axis stator inductance, L_q	102.7e-03
V_{dc}	400 V

VII. CONCLUSION

This paper has discussed about the model of an electrical vehicle, including model of IPMSM taking core loss into account with a Fuzzy Logic controller. It also discussed about the three different road conditions and they are: For turning left, for turning right and for road with slope. The system was implemented in MATLAB Simulink. It overcomes the problem of over shoot & also comforts the designer as itis free from mathematical expressions. From the simulation studies, it is seen that the proposed controller is robust and effective equated to the other conventional controller.

References

[1] F. Gieras and M. Wang, Permanent Magnet Motor Technology, New York, NY: Marcel Dekker Inc., 2002.

[2] A. Consoli, G. Scarcella, G. Scelba, and A. Testa, "Steady-State and Transient Operation of IPMSMs under Maximum-Torque-per-Ampere Control," IEEE Trans. Ind. Appl., vol. 46, No. 1, pp. 121-129, JanJFeb. 2010.

[3] University of Paderborn. (2011). Control of Permanent Magnet Synchronous Motors for Automotive Applications. Retrieved June 7, 2013, from http://wwwlea.uni-paderborn.de/en/research/pmsm-control.html

[4] Y. Li and D. Gerling, "The Comparison of Control Strategies for the Interior PMSM Drive used in the Electric Vehicle," The 25th World Battery, Hybrid and Fuel Cell Electric Vehicle Symposium & Exhibition, Shenzhen, China, Nov. 5-9, 2010

[5] B.K. Bose, "A High Performance Inverter Fed Drive System of an Interior Permanent Magnet Synchronous Machine.‖ IEEE Trans. On Industry Applications, vol. 1A-24, no. 6, 1988, pp 987-997.

[6] S.R. Macminn, T.M. Jahns, "Control Techniques for Improved High Speed Performance of Interior PM Synchronous Motor Drives," IEEE Transactions on Industry Applications, Vol. 27, No. 5, Sept.-Oct. 1991, pp. 997 - 1004.

[7] T. M. Jahns, G. B. Kliman and T. W. Neumann, "Interior Permanent Magnet Synchronous Motors for Adjustable Speed Drives," IEEE Transactions on Industry Applications, Vol. 1A-22, No. 4, July-Aug.1986, pp. 738-747.

[8] F. J. Lin and Y. S. Lin, "A Robust PM Synchronous Motor Drive with Adaptive Uncertainty Observer," IEEE Transactions on Energy Conversion, 1999, vol. 14, no. 4, pp 989-995.

[9] M. N. Uddin, M. A. Abido and M. A. Rahman, "Real-Time Performance Evaluation of a Genetic Algorithm Based Fuzzy Logic Controller for IPM Motor Drives," IEEE Transactions on Industry Applications, Vol. 41, No. I, Jan. /Feb. 2005, pp. 246-252.

[10] M. N. Uddin, and M. A. Rahman, "High Speed Control of IPMSM Drives Using Improved Fuzzy Logic Algorithms," IEEE Trans. Ind. Appl., vol. 54, No. 1, pp. 190-199, Feb. 2007.

[11] "Modern Power Electronics & AC Drives" by Bimal k.Bose

[12] Mohammad Abdul Mannan, Toskiaki Murata and Junji Tamura, "Design and Simulation of Discrete Time Optimal Speed Control for an IPMSM Taking Core Loss into Account", The AIUB Journal of Science and Engineering (AJSE), Vol. 6, No. 1, August, 2007.

[13] Brahim Gasbaoui, Abdelfatah Nasri, "A Novel Multi-Drive Electric Vehicle System Control Based on Multi-Input Multi-Output PID Controller", SERBIAN JOURNAL OF ELECTRICAL ENGINEERING, Vol. 9, No.2, June, 2012

[14] Mohammad Ashraful Alam, Mohammad Abdul Mannan, "Design and Analysis of EV Controller using IPMSM Taking Core Loss into Account", Journal of Control & Instrumentation, Vol. 6, No. 2, 2015.

[15] Sumon Kumar Ghos, Mohammad Abdul Mannan, "Design and Analysis of Linear Quadratic Regulator for the application of EV Differential based on IPMSM", Journal of Control & Instrumentation, May 2016.

[16] Mohammad Ashraful Alam, Mohammad Abdul Mannan, Khondker Rezwanuzzaman, Mohammad Rashedul Amin, Khan Mohammad Shariful Islam, "Modeling and Performance Analysis of Electrical Vehicle with SPWM Inverter-fed IPMSM Taking Core Loss into Account", Journal of Control & Instrumentation, Vol. 7, No. 1, 2016.

[17] J. A.Haddoun, M.Benbouzid, D.Diallo, "Modeling, Analysis and Neural Network Control of An EV Electrical Differential", IEEE Transactions, Vol. 55, No. 6, pp.2286-2294, June, 2008.

[18] Mohammad Abdul Mannan, Asif Islam, Mohammad Nasir Uddin, Mohammad Kamrul Hassan, Toshiaki Murata, Junji Tamura, "Fuzzy-Logic Based Speed Control of Induction Motor Considering Core Loss into Account," *Intelligent Control and Automation*, 2012, 3, 229-235.

[19] Fuzzy Logic Based Efficiency Optimization and High Dynamic Performance of IPMSM Drive System in Both Transient and Steady-State Conditions. 0093-9994 (c) 2013 IEEE.

.

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

Algorithms Efficiency Measurement on Imbalanced Data using Geometric Mean and Cross Validation

Mustakim Al Helal, Mohammad Salman Haydar and Seraj Al Mahmud Mostafa
Department of Computer Science and Engineering, Faculty of Science and Information Technology
Daffodil International University, Dhaka-1207, Bangladesh
{mustakimsunny.cse, salman3045, seraj.mostafa}@diu.edu.bd

***Abstract*—The recent computing trend is producing tons of data every minutes where the amount of imbalanced data is quite high as far as real life data sets are concerned. In practical aspects of data mining, the imbalanced data set is prone to misguide a data mining model. However, data set needs pre-processing before mining. This work focuses on some practical data mining techniques and produces a valid evaluation process for imbalanced data set. A critical comparison of few well established algorithms are illustrated. Accuracy of few well known different algorithms such as, Decision Tree Classifier (DTC), Support Vector Machine (SVM), K-Nearest Neighbor (KNN) and Random Forest (RF) are also compared. The data was tested before and after Over-sampling using Synthetic Minority Over-sampling Technique (SMOTE) and then verified by using Geometric Mean (GM) and Cross Validation techniques. The results we achieved in this work demonstrates a critical comparison of some algorithms and most importantly performance measure that is valid for imbalanced data.**

***Index Terms*—Imbalanced data; SMOTE; Geometric Mean; Cross Validation; SVM; KNN; Decision Tree Classifier; Random Forest;**

I. Introduction

We, literally live in a world of data that became the most important asset to assist any computerized and non-computerized system. There may be raw data, unprocessed/processed data, meta data and such other data forms those need to be analyzed to extract the required information. Having a huge amount of data in these days is a privilege for different corporations such as, business, intelligence, hospitals, government and so on to extract knowledge on specific area. Mining of data therefore became an innate part of the data centric world to serve different purposes. Data mining is a specialized field of computer science where a huge amount of data set is considered and some hidden information from those data is revealed through specific techniques. But one challenge prevails over these data sets is being imbalanced. In fact, most real life data is imbalanced and noisy. Sometimes, data mining approaches might apparently look very promising having a great amount of accuracy which practically is not logical due to the imbalanced data set. So, researches have been carried out over the years on handling imbalanced data set and their various techniques for finding ways to balance inconsistent data sets. Intuitively, there are different ways to handle an imbalanced data set, however, the question is, which approach may give the appropriate solution? The solutions may come up by different parameter tuning and comparing algorithms before and after sampling the imbalanced data sets. The performance measures can tell us which way to follow for a fast convergence towards our goal. In this work, we focus on achieving a decision regarding an imbalanced data set. Our approach is a stochastic algorithm selection and tuning parameters. We focused to have an insight into an imbalanced data set that can raise issues with accuracy.

TABLE I
Imbalanced data and its Class distribution

Class	N	N[%]
unacc	1210	(70.023 %)
acc	384	(22.222 %)
good	69	(3.993 %)
v-good	65	(3.762 %)

Imbalanced Data Set: The chosen data set is significantly imbalanced which may form worst case scenarios can also be considered. Table I contains imbalanced data set with its different class distribution that is considered as our test case for this work. The data set containing car data provided by the UCI machine learning repository [1]. There are 1700+ data in the Car data set out of which 70% falls into **unaccepted** category which is quite significant. So we chose Over-sampling method to balance it. After we ran the process, a new set is produced with a balanced data which is illustrated in Table V. This balanced class values are obtained after the SMOTE algorithm is applied on the data set which is discussed in section IV.

II. Related Work

A good number of research have been carried out in this area. Learning from imbalanced data is always a difficult and also one of the frequent problems in data mining and machine learning problems. One related work is active learning from imbalanced data set by Ertekin S et al. [2] discussed on the concept of providing a more balanced class to a learner that tends to produce better results. Our approach is also similar to active learning and the objective is to provide user a better class value before applying state of the art algorithms. Re-sampling and hyper-parameter tuning is part of the process to provide the learning algorithm with much stable examples. Another work is related to our approach where feature selection was discussed. For an imbalanced data

978-1-5090-5769-6/16/$31.00

set, feature selection is not as straightforward as it is supposed to be. So, feature assessment is done by a threshold value by Chen, X [4] However, unlike our approach of evaluation of their prediction model is based on an area under ROC curve. Haibo and Edwardo [6] pointed on the failure of distributive characteristics of the data due to complex imbalanced data sets that leads to unfavorable accuracies across the classes of the data. Khoshgoftaar et al. discussed the inequality and degree of class in real-world data sets alomg with few applications, such as fraud detection, diagnosis of rare diseases [7] which becomes problematic due to imbalanced data. SMOTE algorithm [10] has been used for Over-sampling in various research work that emphasis on data balancing.

III. Methodologies

The main contribution of this paper is to critically analyze some well established classification algorithms on an imbalanced data set. We take a car data set that is relatively simple in terms of the attributes but is significantly imbalanced to one single class value. This is a multi-class classification problem having four different class values. However, the data in concentrated to one single value as much as 70%. Another challenge is that the data set does not have a big number of tuples. So, the data set is relatively week which is a best suit for the issues that we will try to solve. First, re-sampling was done to make it more of a fit into the feasible region. We applied four different algorithms for this data set to check the performance on imbalanced data. Decision tree classifier, Support Vector Machine (SVM), K-Nearest Neighbors (KNN) and Random Forest algorithm from Ensemble method were applied before and after the re-sampling. Each time a parameter tuning was performed which bring out the changes in performance dramatically. Geometric Mean (GM) technique was used on imbalanced data set for validation.

IV. Re-sampling effect on different algorithms

This research work is carried out in few simple steps to determine the actual results before and after sampling. As we see, the data set are distributed in four separate classes which are ***unacc***, ***acc***, ***good***, and ***v-good***, that gives a decision on the product whether it is **unaccepted**,**accepted**, in **very good** condition or in **good** condition. The main problem of this data set is having 70.023% imbalanced data that is the **unaccepted** class, which is much higher compared to the **accepted** class with only 22.222%. On the other hand the **good** and **v-good** class scored only 3.993% and 3.762% respectively. It shows an abnormality within classes which clearly picture an imbalanced data set. The aim was to find the score of the data before and after sampling from the imbalanced data and balanced data set. To follow the target, we divided the work in few steps which are described in the following sections.

Step 1: At the very first step, we pre-processed data using Sci-kit learn to apply different algorithms on the data set. The data set we have chosen has different class values both in string and numerical forms which is converted to numeric numbers. Table II shows the list of the numerical values ranging from 0 to 5 for separate attributes values in the data set. After this we scaled our feature values between 0 to 1 before feeding to our algorithms to maximize the performance of four different algorithms.

TABLE II
Numerical Representation of attributes value

Attributes value	Represented value
vhigh	3
high	2
med	1
low	0
small	0
big	2
5more	5
more	5
unacc	0
acc	1
good	2
vgood	3

TABLE III
Algorithms efficiency measurement from imbalanced and balanced data using Geometric Mean

Algorithms	Before Over-sampling (imbalanced data)	After Over-sampling (balanced data)
Decision tree classifier	0.911409	0.971213
SVM	0.919679	0.969262
KNN	0.924694	0.960892
Random Forest	0.924694	0.966673

Step 2: In this step, we split the imbalanced data set into Training and Test data on a random selection basis which are 70% and 30% respectively. Decision Tree Classifier, Support Vector Machine (SVM), K-Nearest Neighbor (KNN) and Random Forest were applied on training data one by one and then checked how accurately each model can predict on test or unknown data when the training data is imbalanced. The result is shown in the center column in the table III are the geometric mean (GM) scores which is discussed in detail in step four. GM score is calculated from contingency table of the nodes. True Positive, True Negative, False Positive and False Negative are used to calculate the GM score. GM is one of the most reliable techniques for calculating performance measure for imbalanced data set. We also used K-fold cross validation for performance measure which is illustrated in the table IV. From both the table III and IV it is clear that, after sampling K-fold cross validation slightly better scores are achieved overall for the algorithms which is exaggerated a bit. To some extent it is a biased result due to the imbalanced data. Another visible difference between GM and Cross validation is illustrated in

TABLE IV
ALGORITHMS EFFICIENCY MEASUREMENT FROM IMBALANCED AND BALANCED DATA USING CROSS VALIDATION

Algorithms	Before Over-sampling (imbalanced data)	After Over-sampling (balanced data)
Decision tree classifier	0.917203	0.983677
SVM	0.918399	0.962190
KNN	0.897539	0.974586
Random Forest	0.913146	0.974586

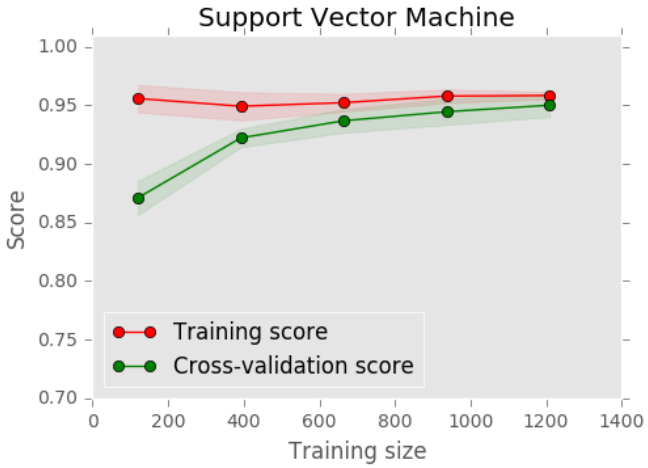


Fig. 1. Learning Curve For SVM before Over-sampling

KNN and Random Forest algorithm prior sampling. GM gives comparatively better score when the data is fully imbalanced and oversampling was not applied to the data set. Whereas, cross validation gives a low score on the imbalanced data particularly for these two algorithms.

In this step, we do an additional check on the trained model, whether or not it is suffering from over-fitting or under-fitting. It is also to be noted that, to add artificial data the model has to be in low bias. If two lines converge at a good score then it is likely to be a good model. If the two lines is not converging but they are going to converge at a good score then getting more data will help to converge. If training score is very higher than the cross validation score then this model is suffering from over-fitting problem then getting more data is likely to help us. We explicitly test the learning curve for SVM and RF algorithm since these two are more prone to over-fit the model due to high accuracy rate and we do this before and after Over-sampling.

We got the learning curves in figure 1 and 2 before Over-sampling. It is visible in 2 that, the training and cross validation curves are going to converge at a good score, that is, this model is not under-fitting which can be resolve by taking more data. Figure 1 shows that, two lines are converged at a good score, thus, this model also is not suffering from any generalization problem.

The result in figure 3 and 4 are the learning curves we

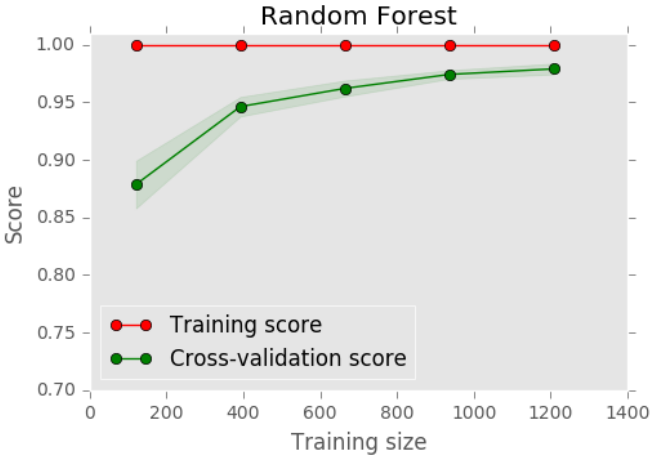


Fig. 2. Learning Curve For Random Forest before Over-sampling

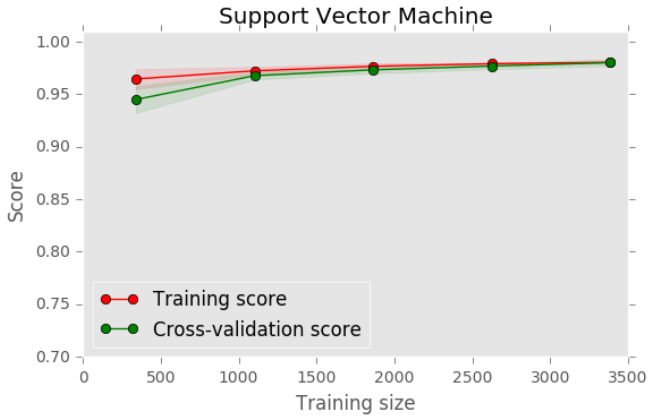


Fig. 3. Learning Curve For SVM after Over-sampling

got after Over-sampling. These curves clearly shows that, the models are not suffering from either any under-fitting or over-fitting problems. Also, the models point the increment of the training size which is more stable from which the effect of sampling can be understood.

Step 3: In the previous step we tested different models on imbalanced data set and compared the results. In this step we applied SMOTE Over-sampling technique to balance the data set. Table V shows the balanced data set which is clearly different than the imbalanced data set shown in Table I. After getting the balanced data we repeat the previous process of splitting training and test data and ran the four algorithms again. The right most column in table III shows the comparison to the previous result. We see a visible differences in accuracy of the algorithms between pre and post sampling of the data set. The comparison is pictured in figure 6.

In this step cross validation reapplied after SMOTE to compare the scores with the GM score which is shown in table III. Normal classification accuracy calculated as total number of correct prediction divided by total number of prediction but for imbalanced data it is not always a

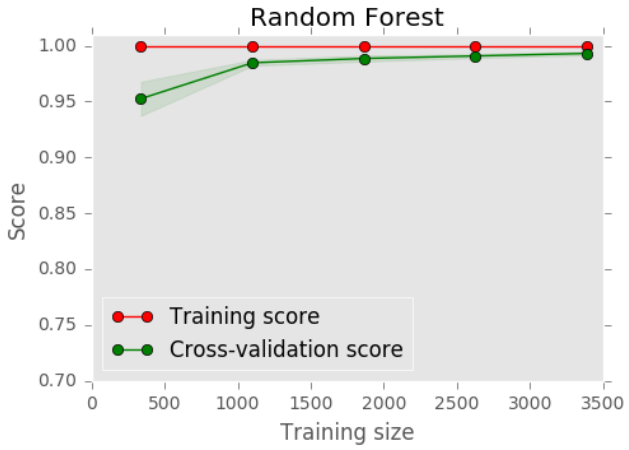


Fig. 4. Learning Curve For Random Forest after Over-sampling

TABLE V
BALANCED DATA AND ITS CLASS DISTRIBUTION

Class	N	N[%]
unacc	1210	(25 %)
acc	1210	(25 %)
good	1210	(25 %)
v-good	1210	(25 %)

good accuracy method. It often gives higher accuracy in imbalanced data set. In cross validation we used this normal accuracy method to calculate the accuracy in each of the 10 fold. The Cross validation scores are comparatively better than Geometric Mean which is shown in Table III and Table IV. Cross validation score is illustrated in figure 5 for all algorithm before and after Over-sampling.

Step 4: For an imbalanced data set the Geometric Mean (GM) is the most accepted and authenticated approach to evaluate performance of the model. So, in this step we apply Geometric Mean (GM) for each of the algorithms individually. The technique that we have applied to the data set is known as One Vs Rest. This is a multi-class classification problem where we have four different class values. So, One Vs Rest technique takes one single class value at a time and checks its validity against all other class values. Each time a confusion matrix is generated where we have True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) values. So, TP is summed up for each of the algorithms. Similarly, Tn, Fn and Fp are also calculated. The scenario is described below:

$$Tp = Tp(DTC) + Tp(KNN) + Tp(SVM) + Tp(RF)$$
$$Tn = Tn(DTC) + Tn(KNN) + Tn(SVM) + Tn(RF)$$
$$Fp = Fp(DTC) + Fp(KNN) + Fp(SVM) + Fp(RF)$$
$$Fn = Fn(DTC) + Fn(KNN) + Fn(SVM) + Fn(RF)$$

After this we need to calculate Specificity (Sp) and Sensitivity (Sn). The following formulas shows how the **Sp** and **Sn** are calculated:

$$Sp = \frac{Tn}{Tn+Fp}$$
$$Sn = \frac{Tp}{Tp+Fn}$$

And finally, the GM is calculated as:

$$GM = \sqrt{Sp * Sn}$$

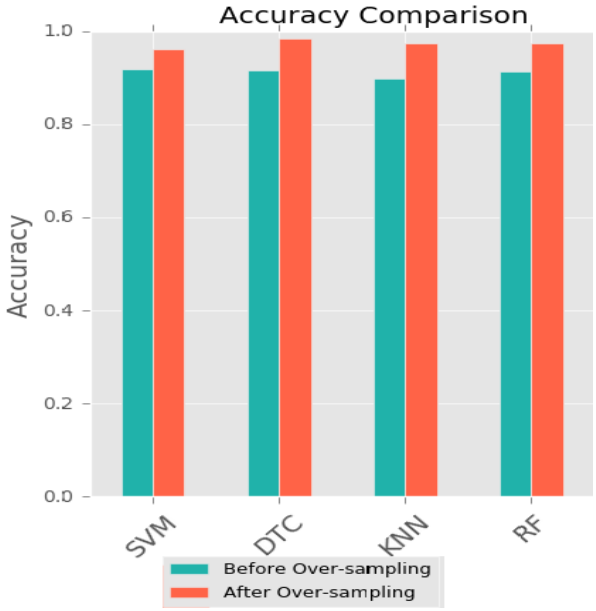


Fig. 5. Accuracy Comparison using cross validation on imbalanced and balanced data

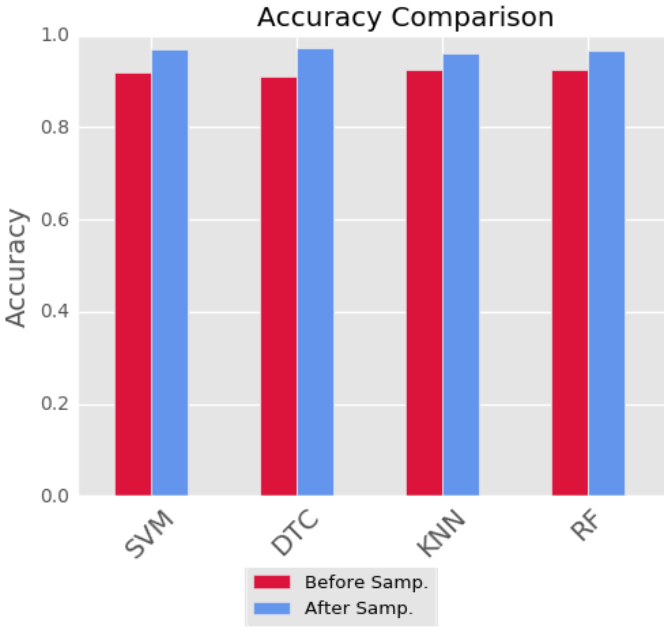


Fig. 6. Accuracy comparison using Geometric Mean on Imbalanced and Balanced data

V. Discussion and Future Work

This work is based on imbalanced data set where we have demonstrated a comparison of algorithms before and after re-sampling of data. Geometric mean played a vital role in authentication of the models for imbalanced data set, although, Cross validation gave us better accuracy compared to Geometric Mean on the same data set. Diverse experiments are needed in the future to figure out the impact of geometric mean in various real life applications. More noisy data can be used to investigate the outcome of GM. The work can be extended by feeding real time data set to the model for example, satellite data or traffic data those are likely to have more noise and also imbalance. Model over-fitting issue has to be considered with importance when experimenting more noisy data.

VI. Conclusion

Imbalanced data set is a serious issue in machine learning and data mining problems. We tried to mitigate the problem with some efficient techniques. Other focus of this work was to link up scattered ideas of solving imbalanced data problems. The accuracy we achieved is quite satisfactory but we cannot expect such results every times. So, reliable performance measurement is very important for such kind of problems. This work could be a good reference as extensive research work on balancing imbalanced data and the idea of integrating different learning algorithms.

References

[1] Lichman, M. (2013). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science (Accessed on 20-Sep-2016 at 10:20am @ GMT +6 standard).

[2] Ertekin, Seyda, et al. "Learning on the border: active learning in imbalanced data classification." Proceedings of the sixteenth ACM conference on Conference on information and knowledge management. ACM, 2007.

[3] Gu, Qiong, et al. "Data mining on imbalanced data sets." 2008 International Conference on Advanced Computer Theory and Engineering. IEEE, 2008.

[4] Chen, Xue-wen, and Michael Wasikowski. "Fast: a roc-based feature selection metric for small samples and imbalanced data classification problems." Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2008.

[5] Lobo, Jorge M., Alberto JimnezValverde, and Raimundo Real. "AUC: a misleading measure of the performance of predictive distribution models." Global ecology and Biogeography 17.2 (2008): 145-151.

[6] He, Haibo, and Edwardo A. Garcia. "Learning from imbalanced data." IEEE Transactions on knowledge and data engineering 21.9 (2009): 1263-1284.

[7] Khoshgoftaar, Taghi M., Moiz Golawala, and Jason Van Hulse. "An empirical study of learning from imbalanced data using random forest." 19th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2007). Vol. 2. IEEE, 2007.

[8] Branco, Paula, Luis Torgo, and Rita Ribeiro. "A survey of predictive modelling under imbalanced distributions." arXiv preprint arXiv:1505.01658 (2015).

[9] Hsu, Chih-Wei, Chih-Chung Chang, and Chih-Jen Lin. "A practical guide to support vector classification." (2003): 1-16.

[10] Chawla, Nitesh V., et al. "SMOTE: synthetic minority over-sampling technique." Journal of artificial intelligence research 16 (2002): 321-357.

[11] Akbani, Rehan, Stephen Kwek, and Nathalie Japkowicz. "Applying support vector machines to imbalanced datasets." European conference on machine learning. Springer Berlin Heidelberg, 2004.

[12] Kotsiantis, Sotiris, Dimitris Kanellopoulos, and Panayiotis Pintelas. "Handling imbalanced datasets: A review." GESTS International Transactions on Computer Science and Engineering 30.1 (2006): 25-36.

[13] Sokolova, Marina, and Guy Lapalme. "A systematic analysis of performance measures for classification tasks."Information Processing Management 45.4 (2009): 427-437.

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

Line Profile Based Fingerprint Matching

Hafsa Moontari Ali, Sonia Corraya
Department of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
{hafsa.moontari, sonia.corraya}@bracu.ac.bd

***Abstract*—Traditional fingerprint matching algorithms primarily focus on minutiae points on fingertip surface. In this paper, a novel approach is proposed for fingerprint matching that is based on ridge and valley characteristics of fingerprints. At first, the input fingerprint image is normalized and the registration point of that particular fingerprint is detected. Then a line profile is generated centering on that reference point. The distances between the reference point and ridges and the count of intersection points of line profile and ridges are stored in database. This process is repeated after every 15 degree angle to 345 degree in clock-wise direction and for orientation angle, the distances are stored sequentially. For matching intersection point count number along with the sequence of distance values are compared with the stored values. This new method can detect fingerprint from any orientation angle. Experimental result shows 90.87% accuracy of the proposed method.**

Keywords—Biometrics; Fingerprint match; Registration point; Ridge line; Line profile

I. INTRODUCTION

Fingerprint identification is a major physiological biometrics modality. Biometrics is an automated method to identify a person based on physiological or behavioral characteristics. Physical biometrics measures the innate physical characteristics of individual. Examples of physical biometrics include Fingerprint, Face, Hand geometry, Iris scan, Vascular scan, DNA etc. Out of these physical signs, fingerprint is the oldest and commonly used recognized biometric trait. Factors believed to be accountable for its vast application include individuality, uniqueness, permanence and reliability [1]. Biometric fingerprints are most widely used person identification tool. Every fingerprint has some unique local and global features on fingertip that are used by fingerprint verification methods. Local features are referred to as minutiae points and most of the fingerprint matching techniques are focused on minutiae points. One major shortcoming of minutiae based techniques is that if the minutiae points of a fingerprint are exposed, that person's identity can easily be stolen since these techniques only uses the features extracted from minutiae points [2]. For example, any minutiae based system would detect both Fig 2(a) and Fig. 2(b) as same fingerprint. We propose a novel line profile based fingerprint matching technique that uses the global level properties such as ridge and valley structures.

In this paper, we calculated the distances sequentially between the registration point and the ridges that the line profile intersects along with the total intersection points for any particular line profile. The distances are calculated from 0 degree to 345 degree incrementing the angle by 15 in clock-wise direction.

Rest of the paper is organized as follows: Related works are given in section II. Section III explains about most popular and commonly used fingerprint features. Section IV explains the proposed model and section V represents the experimental result with discussion. Finally, conclusion is drawn in section VI.

II. RELATED WORKS

The process of fingerprint identification using a ridge skin layout that is used to describe biometric characteristics is referred to as dactyloscopy. Minutiae based, correlation based and hybrid are three broad categories in Automatic Fingerprint Identification Systems. Minutiae based fingerprint identification system focuses on local level analysis of fingerprint features. On the other hand, global level analysis is frequently used for fingerprint classification and indexing problems. Fingerprint classification is an important indexing mechanism in the overall identification criteria since an accurate classification technique minimizes the fingerprint matching time in a large database. A Gabor filter based approach is explained in [3] to classify fingerprints without using singular point exactly. The core point in fingerprint plays an important role in classification and matching, especially in minutiae based techniques. Several algorithms have been proposed to detect the core point. Pankaj Mohindru, Govind Sharma and Pooja proposed a minutiae extraction method in [4] where the fingerprint image pixels are normalized using Fuzzy Inference System (FIS) and false minutiae points are deducted by calculating Euclidean distance between the bifurcation and termination points. Minutiae selection plays a vital role in minutiae based fingerprint detection techniques. In [5], seven minutiae selection methods are stated and shown that how the selection of minutiae points affects the performance of on-card-comparison algorithm. At first the ideal case, i.e. all minutiae points are considered and then six different test cases of minutiae selection under truncation based methods are represented. In order to identify partial fingerprints, all the features of fingerprint images are considered in [6]. Here, the fingerprints are compared pixel-wise by calculating their correlation coefficients. To develop minutiae-only algorithms with light architecture, another concept Minutiae Cylinder Code (MCC) was emerged which associates a local structure to each minutiae point. In [7], a 3D geometric hash table is proposed to store binary minutiae cylinder codes where the

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

access keys are the descriptors of the global geometric configuration. Monika and M. Kumar proposed a novel fingerprint matching technique [8] which generates authentic result irrespective of rotation or replacement. In this technique, the fingerprints are matched globally and if they are matched only then it is validated by matching the texture by Local Binary Pattern (LBP). In [9] to mitigate the complexity of determining the best correspondence among fingerprint images, fixed length representation of fingerprints is developed which generates exact alignment between the features of corresponding fingerprints and leads to high-speed matching with minimal computational effort. This method is also invariant to affine transformation. Image segmentation plays another key role in AFIS. A novel technique of automatic segmentation of fingerprint ROI is presented in [10] which employs divide and conquer approach using force field and heuristics. To solve the challenging task of matching a particular fingerprint against a large scale fingerprint database, a novel distributed approach called DFRS is proposed in [11]. It uses an innovative framework integrating key technologies for data compression and space reduction. Another method to increase the accuracy and robustness of fingerprint verification system is proposed in [12] where the spectrum features are extracted from fingerprint images. Lexicographic ordering is used to transform the 2D fingerprint image into 1D by scanning and dividing the image into square blocks. Finally the features are verified using support vector machine (SVM).

III. FINGERPRINT FEATURES

Fingerprints mainly exhibit local level and global level features. The local level analysis of fingerprint provides information about individuals through the locations of ridge-discontinuities, known as minutiae points [13]. Termination point and bifurcation point are most common types of minutiae points. Termination points are formed when a ridge ends abruptly and bifurcation points are created when a ridge line forks into Y-shaped junction. Contrary to local level analysis, global level analysis of fingerprints extract singular regions like loop, delta, whorl which were formed as a result of various types of pattern configuration of ridges and valleys. Fig. 1 shows the common fingerprint features.

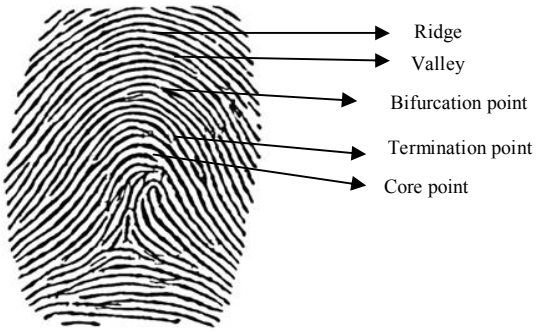


Fig. 1. Features of Fingerprint image

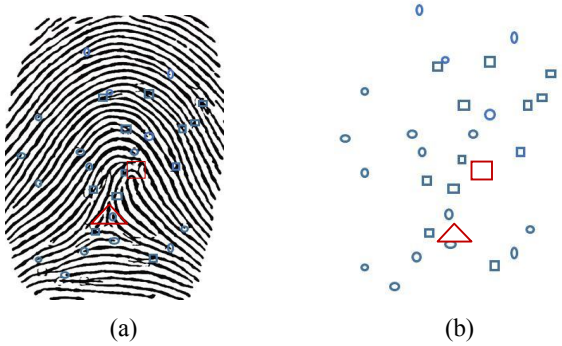


Fig. 2. (a)Fingerprint image depicting minutiae features and core points (b) only minutiae features and core point

IV. PROPOSED MODEL

The proposed line-profile based fingerprint matching system has four steps as shown in Fig. 3.

A. Input fingerprint image

Input gray fingerprint image is scanned by computer. Resolution of the input image is 500 dpi and size is 320 x 480 pixels.

B. Pre-processing: Image enhancement

1) De-noising by Median filter and binarization:

For image enhancement, at first Median filter was used inspired from the fingerprint image de-noising filter results of [14]. Next, for converting the Median filtered image, an adaptive thresholding was used for converting the grayscale image into binary. In this binarization, each image pixel is assigned new value (0 or 1) according to the mean intensity of a local neighbourhood (13 X 13) [15].

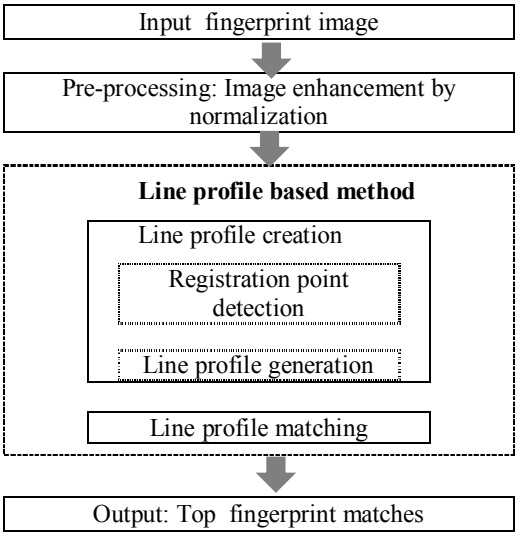


Fig. 3. Steps of the proposed line profile based fingerprint matching model

For limited enhancement result by the median filter and binarization based technique, we finally applied normalization; a direct gray scale enhancement method as gray scale based techniques performs better than techniques which have binarization and thinning at earlier steps [15].

2) Image Normalization:

Normalization or contrast stretching changes the range of pixel intensity values. It reduces the difference in gray level values along the ridges and valleys without affecting the original structure. The Equation of linear normalization of a grayscale image is as followed [16]

$$IN = (I - Min) + newMin \quad (1)$$

where,

I = old / original image

IN= new image

(Min,Max) = range of intensity values in original image.

$(newMin, newMax)$ = range of intensity values in new image.

C. Line profile based methodology

1) Line profile Creation:

a) Registration point detection

For global ridge structure and fingerprint alignment, core point based systems are not sufficient. If core point exists, it is detected as registration point but for fingerprints having no core points, delta is the registration point. If neither core nor delta is found, a point with high curvature regions or Low Coherence strength is detected as the registration point [17].

For double core point fingerprint images, core point having the maximum distance from boundary is selected as the reference point. While matching both core points are considered initially and only one is selected based on the matching score progress.

b) Line profile generation

Taking the detected registration point as reference or centre, line profile is created for storing and matching purpose.

Length 3.6285 mm to 4.5357 mm of line profile is decided, as registration points usually are placed at min 1.81425 mm to 2.26785 mm distance from the fingerprint image boundary as per our observation. Again distances among ridge intersection points which are far from registration points might vary for different fingerprints of same finger. For this we focused only the registration point area as ROI (Region Of Interest) and ignored far points. 15 degree difference in orientation is selected after observing intersection point count changing trend at all 360 degree orientations. Fig. 4 shows a line profile placed on reference point at starting position (0 degree angle). For this 1st angle, values are shown in TABLE 1.

TABLE 1: LINE PROFILE VALUES AT 1ST ORIENTATION ANGLE

Orientation (clock-wise) angle (Degree)	*1st*	*2nd*	*3rd*		*22nd*	*23rd*
	0	15	30	…	330	345
Total number of intersection points	N1=27	N2	N3	…	N22	N23
Sequence of distances among intersection points (nanometre)	2001314.35, 3023870.02, 2112100.63, 2773189.1, 2345468.29, 3123333.33, 1000581.37, 4430032.58, 1564537.03, 2564430.01, 8100002.14, 2423974.5, 557823, 1227610.84, 3789345.65, 11000051.3, 3435972.6, 1900034.10, 2053509.30, 2971540.59, 2212134.54, 1767781.3, 3519842.30, 931304.87, 3000018.90, 1245789.13, 3000231.53					

* reference point is included in intersection point count

2) Line Profile Matching Pseudo Code:

For cost optimization, complete line profile of the test fingerprint image is not created beforehand for matching like other traditional fingerprint matching systems. Pseudo code for matching fingerprint of the proposed novel system is given below.

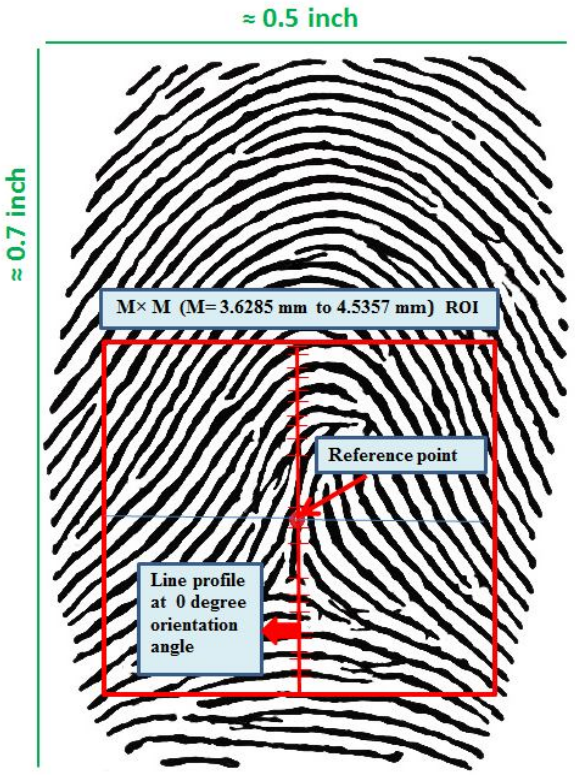


Fig. 4. Line profile is placed on reference point at starting position (0 degree angle)

Line Profile Matching Pseudo Code:

```
Consider the registration point as centre.
initialization:
starting_angle= 0;
orientation_angle_increment_degree= 1;
// sequence of the total number of intersection points //at
different orientation angel.
matching_sequence_length =0;
iteration_limit=15
for ( count = starting_angle: clock-
wise_orientation_angle_increment_degree: starting_angle
+ iteration_limit)
{
   calculate matching_sequence_length;
    if (matching_sequence_length ==2)
   {
   set  orientation_angle_increment_degree=15
   set iteration_limit=345
   }
}
return matching_sequence_length;
```

D. *Output: Top fingerprint matches*

For faster and optimized search, an initial database search condition is applied. While matching, if core is detected as the registration point of the test image, then the proposed method only search for matching fingerprints in those images which have core as their registration points. Same search condition is applied for other cases (delta and other registration point). Along with line profile structure of any fingerprint, this registration point information is also stored in the database like other associated personal information.

The proposed system not only detects the best matched fingerprint but also outputs the top M(M is the number of top matching results user wants to see) matching results. This feature is important for suspecting criminals in security systems. As an example, for M=3, output result format is shown in TABLE II.

For attendance or access based system M is to be set at value 1. A matching threshold value can be set as well for result acceptance or rejection. As example, for cases like below where total number of matching finger print found >1, 3rd field or property of the line profile ,'Sequence of distances among intersection points' is then considered and matched for obtaining a single output result. Example is shown in TABLE III.

TABLE II. OUTPUT EXAMPLE FOR M=3

M	Similarity or Matching score (%)	Total no of matching finger print found
1	99.93	1
2	90.01	1
3	87	2

TABLE III. OUTPUT EXAMPLE FOR M=1

M	Similarity or Matching score (%)	Total no of matching finger print found
1	95	2

V. EXPERIMENTAL RESULT

A. *Experimental environment*

Fingerprints used to evaluate the performance of the proposed method are selected carefully. 50 fingerprints are collected from NIST-4 database [18]. From each of the five classes (whorl, right loop, left loop, arch and tented arch) 10 images are collected from the NIST-4 database. Another 10 images are real scanned image with the same resolution (512X 512). These 10 images are different fingerprints of same 2 fingers. So, a total of 60 fingerprints are used to evaluate the proposed method.

This new method is implemented and tested in MATLAB R2014a environment by using a computer of Intel Core i5, 2.6 GHz with 4 GB RAM in Windows 8 platform.

While evaluating, for fingerprint matching 3 is used as the value of M (number of top match result). And for fingerprint detection, with M=1, threshold for matching acceptance is set to 91.03% for both fields (Total number of intersection points and Sequence of distances among intersection points)

B. *Performance evaluation:*

Results obtained in above configuration and settings are given in TABLE IV and in Fig 5. Fig. 5 shows the accuracy for different fingerprint types. According to the experimental result with experimental data, the overall accuracy of the proposed system is 97.04%.

TABLE IV: ACCURACY FOR DIFFERENT FINGERPRINT IMAGE

Sl	Fingerprint Type	Accuracy (%)
1	whorl	99
2	right loop	96.05
3	left loop	95.09
4	arch	96.03
5	tented arch	96.09
6	fingerprint of same finger	100
	Average	97.04

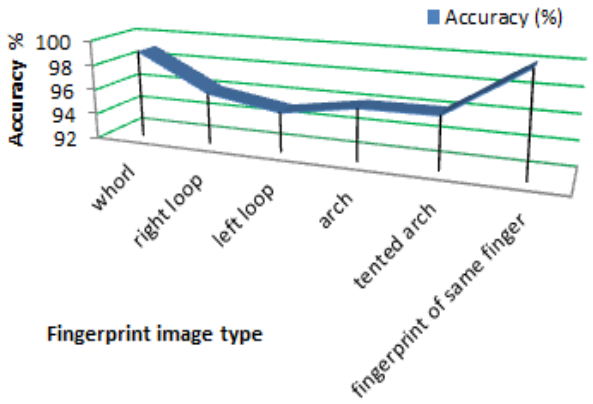


Fig. 5. Graph of different type of fingerprint accuracy

VI. CONCLUSION

A novel approach for fingerprint matching is proposed in this paper focusing on ridge patterns of fingerprint image. Our approach is reliable compared to traditional minutiae point based system as most of the hacking threats are also focused on minutiae points. The new method of this paper can be used as a matching validation technique with the existing fingerprint matching approaches for improving accuracy. In future machine learning can be applied in this method to enhance the overall performance.

REFERENCES

[1] Mohammed S. Khalil, Mohamad Dzulkifli, Muhammad Khurram Khan, and Qais Al-Nuzaili, "Fingerprint verification using statistical descriptors", Digital Signal Processing, vol. 20, No. 4, pp.1264-1273, July 2010.

[2] Davide Maltoni, Dario Maio, Anil Jain, and Salil Prabhakar, Handbook of fingerprint recognition, Second edition, Springer Science & Business Media, 2009

[3] Dhruv Batra, Girish Singhal, and Santanu Chaudhury. "Gabor Filter Based Fingerprint Classification Using Support Vector Machines" Proceedings of the IEEE INDICON 2004, pp 256-261, IEEE, 2004

[4] Pankaj Mohindru, Govind Sharma, and Pooja, "Fingerprint Minutiae Extraction using Fuzzy Logic" , International Journal of Computer Application, vol.101, no.10, pp. 24-26, September 2014

[5] B. Vibert, C. Charrier, J.-M. Lebars, C. Rosenberger, "Comparative Study of Minutiae Selection Algorithms for ISO Fingerprint Templates", in SPIE Electronic Imaging. International Society for Optics and Photonics, United States, pp. 94090C-94090C, February 2015

[6] O. Zanganeh, B. Srinivasan , and N. Bhattacharjee, "Partial Fingerprint Matching through Region-Based Similarity", Digital lmage Computing: Techniques and Applications (DlCTA), 2014 International Conference on. IEEE, Wollongong, NSW, pp. 1-8, November 2014.

[7] Y. Wang, L. Wang , Y. m. Cheung , P. C. Yuen, "Fingerprint Geometric Hashing Based on Binary Minutiae Cylinder Codes", 22nd International Conference on Pattern Recognition, Stockholm, pp. 690-695, August 2014

[8] Monika, and M. Kumar, "A novel fingerprint minutiae matching using LBP", 3rd International Conference on Reliability, Infocom Technologies and Optimization, Noida, pp. 1-4, October 2014.

[9] Akhil Vij, and Anoop Namboodiri, "Learning Minutiae Neighborhoods : A New Binary Representation for Matching Fingerprints", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Ohio, pp. 64-69, 2014

[10] Kamlesh Tiwari, and Phalguni Gupta, "An efficient technique for automatic segmentation of fingerprint ROI from digital slap image", Neurocomputing, vol. 151, no.3, pp. 1163-1170, March 2015

[11] Bing Li, Zhen Huang, Jinbang Chen, Yifan Yuan, and YuxingPeng, "DFRS: A Large-Scale Distributed Fingerprint Recognition System Based on Redis", 22nd International Conference in MultiMedia Modeling, USA, pp. 138-149, Springer International Publishing, 2016

[12] H. Kasban, "Fingerprints verification based on their spectrum", Neurocomputing, vol.171, pp. 910-920, 2016

[13] Daniel Peralta, MikelGalar, Isaac Triguero, Daniel Paternain, Salvador García, EdurneBarrenechea, José M. Benítez, HumbertoBustince, and Francisco Herrera, "A survey on fingerprint minutiae-based local matching for verification and identification: Taxonomy and experimental evaluation", Information Sciences, vol. 315, no. C, pp.67-87, September 2015.

[14] Dr.E.Chandra, K.Kanagalakshmi, "Noise Elimination in Fingerprint Image Using Median Filter", International Journal of Advanced Networking and Applications, vol. 02, Issue: 06, pp. 950-955 , 2011

[15] Shlomo Greenberg, Mayer Aladjem, Daniel Kogan and Itshak Dimitrov, "Fingerprint Image Enhancement using Filtering Techniques", IEEE proceedings of 15th International Conference on Pattern Recognition, vol 4, pp.1-18, 2000

[16] Neeraj Bhargava, Anchal Kumawat, Ritu Bhargava, "Fingerprint Matching of Normalized Image based on Euclidean Distance", International Journal of Computer Applications ,vol 120, no.24, June 2015

[17] H B Kekre, V A Bharadi, "Fingerprint's Core Point Detection Using Orientation Field", IEEE proceedings of International Conference on Advances in Computing, Control, & Telecommunication Technologies, 2009.

[18] C.I. Watson and C.L. Wilson. National Institute of Standards and Technology, March 1992. [https://www.nist.gov/srd/nist-special-database-4] (accessed on 29/10/2016)

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

Discovering Knowledge regarding Academic Profile of Students Pursuing Graduate Studies in World's Top Universities

Shibbir Ahmed[*], Abu Sayed Md. Latiful Hoque[*], Mahamudul Hasan[‡], Rahnuma Tasmin[¥], Deen Md. Abdullah[€], Anika Tabassum[*]

Department of Computer Science and Engineering
[*]Bangladesh University of Engineering and Technology
[‡]University of Dhaka, [¥]Military Institute of Science and Technology, [€]Primeasia University
Dhaka, Bangladesh
{shibbirahmedtanvin, munna09bd, rabi.mist, annisizzler}@gmail.com

***Abstract*— Each year a large number of students apply for graduate studies into well-known universities all around the world. Students trying to pursue higher study in a better institution strengthen their academic profile by obtaining better CGPA, high scores in standardized tests like GRE, TOEFL, IELTS, noteworthy publication records. Some of the applicants become successful by getting admission in desired university with financial support. Discovering knowledge from those successful applicants' entire academic profile by utilizing it for appropriate mining algorithm will be beneficial to the prospective graduate applicants. In this paper, we present a technique to discover knowledge from successful graduate applicant's entire academic records who got opportunity to study abroad with funding. We have preprocessed the dataset rigorously after gathering all relevant information from successful applicants. For Association rule mining, we have mentioned a technique to discretize the records of CGPA, Standardized test scores, research and job experience etc. Then we have applied Predictive Apriori Association rule algorithm and categorized interested association rules based on the predictive accuracy which are obtained from the dataset. Finally we extract knowledge which is very useful for the students who want to apply for graduate studies into suitable universities considering their entire academic profile.**

Keywords— Data Mining; Knowledge Discovery; Association Rule Mining; Predictive Apriori Algorithm; Higher Study Abroad.

I. INTRODUCTION

After completing undergraduate studies, many students want to pursue graduate study in a better institution outside their homeland. Those students try to build up their academic profile to make them competitive and successful candidates for graduate studies in their preferable academic institutions. Along with the undergraduate CGPA, standardized test scores such as GRE (Graduate Record Examination), TOEFL (Test of English as a Foreign Language), IELTS(International English Language Testing System) as well as research and job experience play significant role to make an applicant competent for graduate studies with financial support in a renowned university. Before applying to renowned graduate schools an applicant should be aware of the facts such as, which category of students get admission with funding? What are the factors of applicants' academic profile really influences admission with funding? How current applicants should match their profile with previous successful applicants' profile in context of getting admission into a specific graduate university? Concepts and techniques of data mining [1, 2] are very much useful to discover such types of hidden knowledge from successful applicants' academic profile. From Bangladesh, each year noteworthy amount of students try to apply for graduate studies in renowned universities in USA, Australia, Canada, Japan, Germany, Singapore, Malaysia etc. Some of them get admission with full funding. A non-profit organization collects those students' entire academic data [3] for betterment of prospective graduate applicants of Bangladesh. Discovering knowledge from successful applicant's academic records using appropriate data mining and machine learning algorithm [4] will be crucial for forthcoming graduate applicants which is the main objective of this research study.

In this paper, we have developed a technique to discover knowledge using Predictive Association Rule Mining from academic profile of students who have already got admission into their desired academic institution with funding. Here, we have preprocessed the dataset extensively after gathering all relevant information from successful applicants. For Association rule mining, we have manipulated a technique to discretize the records of applicant's CGPA, Standardized test scores, research and job experience. After applying Predictive Apriori Association rule mining algorithm, we have categorized interested association rules based on the predictive accuracy which is a metric combining support and confidence. Finally we extract knowledge from those significant rules which may be very useful for the students who have profound desire to get graduate admission into a renowned institution with best possible financial support.

II. BACKGROUND AND MOTIVATION

Pursuing graduate study in a renowned university in a renowned country is a cherished dream of many students. Those students prepare themselves from the very beginning of their undergraduate career. Getting inspired from elder ones and gathering knowledge from various sources, students build

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

up their academic profile since undergraduate studies. After completing undergraduate studies, students try to obtain better scores in standardized test scores such as in GRE, TOEFL, IELTS etc. Along with these, students try to enrich publication records and gathering research and job experience. But the most crucial step is to apply to appropriate graduate studies after selecting appropriate institutions matching applicants' entire academic profile. For this, appropriate knowledge is required which can be obtained from successful applicants of previous sessions who build up their academic profile and eventually get admission into desired academic institution. Here, only statistical analysis and personal observation may not be sufficient in this regard because of the nature of the data. That is why we are motivated to do research and apply data mining techniques to discover knowledge regarding academic profile of students pursuing higher study abroad with a view to assisting prospective future graduate applicants to apply for graduate studies in appropriate graduate universities.

III. Related Work

Educational data mining (EDM) has been deployed to explore data originating in an educational context. There is a survey [5] which illustrates the most relevant studies carried out in this field. Mining educational data to analyze students' performance from the extracted knowledge has been presented in recent studies [6, 7, 8]. Besides these, there is also a book chapter [9] where educational data mining methods, it's ongoing trends and it's linkage with learning analytics community are presented. A latest survey [10] of educational data mining based on analysis of recent works can be visualized as well. Mining significant association rules from educational data using critical relative support approach has been proposed recently in a study [11]. There is a significant research study [12] which describes the measure of interestingness for association rules in educational data. Mining techniques of rare rules are demonstrated in a research [13] clearly. Although there is a study showing the drawbacks and solutions of applying Association Rule Mining in learning management systems [14], the mined association rules reveal various factors like student's interest, curriculum design; teaching and learning outcomes has been successfully illustrated in significant research studies [15,16].

A technique of preprocessing academic data before Mining Association Rule [17] with synthetic dataset has been proposed recently for checking the suitability of the system with the real institutional dataset. A significant survey [19] on Association Rule Mining algorithm has been proposed where it is shown that the quality of small rule sets generated by Apriori can be improved by Predictive Apriori algorithm applied on UCI Repository of machine learning datasets. Another important study [20] has compared the performance of Apriori and Predictive Apriori based on the interesting measures using Weka. A recent study [21] has done noteworthy comparison between Apriori and Predictive Apriori algorithm in which Predictive Apriori has been found better and faster than Apriori algorithm in mining association rules from a dataset containing crimes data concerning women taken from UCI repository. Knowledge discovery regarding educational dataset of a university using Association Rule mining has been presented in our recent study [18] where Apriori algorithm is applied to extract knowledge of students' personal statistics, academic performance as well as impact of courses and curriculum. However, Predictive Apriori performs better than Apriori and higher study abroad dataset requires different algorithms for preprocessing. Thus, we are encouraged to improve the technique and discover hidden knowledge regarding academic profile of students pursuing higher study abroad using Predictive Apriori Association Rule Mining algorithm.

IV. Methodologies

A. Data Analysis

Several personal and academic information of a particular student is stored in the universal table created from the information found in the successful applicant's records [3]. Those are collected for data preprocessing and data analysis.

TABLE I. Selected Data from Universal Database

Academic Data	Undergrad Department
	Undergrad University
	Undergraduate CGPA
	Research Area
	Research Experience
	Job Experience
	Publication Records
Standardized Test Score Data	GRE (Verbal, Quantitative, Analytical Writing, Total Score)
	TOEFL/ IELTS (Reading, Listening, Speaking, Writing, Total Score)
Personal Data	Gender

The personal and academic data as well as Standardized test scores data stated in the above Table I are considered. As we have experimented with the successful applicants' data who pursue higher study abroad, we have analyzed all records of them such as, Outgoing University, Outgoing Country, Intended Semester, Admission In (M.Sc. or PhD). We have considered these graduate admission records stated in the above Table II for knowledge discovery regarding academic profile of successful graduate applicants.

TABLE II. Selected Graduate Admission Records from Universal Database

Graduate Admission Records	Outgoing University
	Outgoing Country
	Outgoing State in Country
	Intended Semester
	Admission In (M.Sc. or PhD).
	Funding
	Application Step

B. Preprocessing for Mining Higher Study Abroad Database

First of all, all-inclusive customized universal database is created in which records of all successful applicants from Table I and Table II are utilized. In this universal table redundant records and incomplete or null value records are omitted from the existing database for the suitability of applying mining algorithm. So, records such as name and other personal information have been omitted in the customized universal table shown in Table III. But this dataset is yet not ready for

applying association rule mining algorithm. For this we have to discretize the continuous attributes such as CGPA, test scores etc.

TABLE III. A SMALL PORTION OF CUSTOMIZED UNIVERSAL DATABASE

Gen-der	Outgoing Country	Outgoing University	Depar tmnt	Research Interest	GRE Score	TOEFL/IELTS Score	Publications	Funding	…
Male	USA	UC Berkley	EEE	Signal Processing	330	115	10	TA	
Female	USA	UIUC	CSE	Data Mining	315	90	2	RA	
…	…	…	…	…	…	…	…	…	

The customized universal database of Table III has been transformed into an equivalent transformation table by transforming the continuous valued attribute as discrete valued attribute representing some knowledge for the suitability of implementing Predictive Apriori algorithm of Association Rule Mining. As for example, CGPA, GRE, TOEFL, IELTS score are continuous attributes and it has been transformed into five classifications as Very High, High, Medium, Low and Very Low for better analysis. One algorithm has been used for discretizing all continuous scores of GRE, TOEFL and IELTS. Another algorithm has been used for transforming undergraduate CGPA into those five classifications. For transforming the CGPA of latest academic level, Algorithm 1 has been developed to populate the transformed table in such a way that there is no continuous CGPA value in an entry.

Algoithm 1: CGPA_Discretization( )

Input: all acquired *CGPA* of a student in undergraduate level of study in the universal table
Output: *transformed_ CGPA* for the Transformation Table
for i**=1** to | $Student_{list}$ |
  if (*CGPA>= 3.75 && CGPA<=4.00)*
    transformed_CGPA= 'VeryHigh'
  else if (*CGPA>= 3.50 && CGPA<3.75)*
    transformed_CGPA = 'High'
  else if (*CGPA>= 3.25 && CGPA<3.50)*
    transformed_CGPA = 'Medium'
  else if (*CGPA>= 3.00 && CGPA<3.25)*
    transformed_CGPA = 'Low'
  else if (*CGPA<3.00)*
    transformed_CGPA = 'VeryLow'
end for

Similarly the scores of all standardized test scores in customized universal table are also transformed by an algorithm named as 'Test Score Discretization' in Algorithm 2. As the real data set contains test scores of GRE, TOEFL, IELTS with various range we have considered another variable discretization format table and transformed the continuous value of various test scores to these five classified definitions.

To construct the entire transformed table as given in Table VII, we have used the customized universal table and stated discretized format of Table IV, Table V and Table VI.

Algoithm 2: TestScore_Discretization( )

Input: *scores of GRE(verbal, quant, awa), TOEFL, IELTS (reading,listening,speaking,writing)* from Universal Table of $Student_{list}$
Output: discrete *level* of *scores* for the Transformation Table
for i**=1** to | $Student_{list}$ |
  if (*score*>=*90%*)
    level = "*VeryHigh*"
  else if (*score*>=80% && *score*<90%)
    level = "High"
  else if (*score*>=65% && *score*<80%)
    level = "Medium"
  else if(*score*>=50% && *marks*<65%)
    level = "low"
  else if(*score<50%*)
    level = "VeryLow"
end for

As there are standardized test scores of GRE, TOEFL and IELTS with different range of scores, it requires different discretization format tables for all these different types of scores based on real dataset. For example, Discretization format table for GRE scores (in Table IV), for TOEFL (in Table V), and for IELTS (in Table VI) are illustrated below.

TABLE IV. DISCRETIZATION FORMAT TABLE FOR GRE SCORES

Classified Name	Range of Scores (S)			
	Verbal	Quant	AWA	Total
VeryHigh	165≤S<170	165≤S<170	4.5<S≤6.0	320≤S≤340
High	160≤S<165	160≤S<165	3.5<S≤4.5	310≤S<320
Medium	150≤S<160	150≤S<160	3≤S≤3.5	300≤S<310
VeryLow	140≤S<150	140≤S<150	2≤S≤2.5	290≤S<300
Low	130≤S<140	130≤S<140	S<2	260≤S<290

TABLE V. DISCRETIZATION FORMAT TABLE FOR TOEFL SCORES

Classified Name	Range of Scores (S)	
	Reading/Listening/Speaking/Writing	Total
VeryHigh	27≤ S ≤30	108≤S≤120
High	24≤ S < 27	95≤S<108
Medium	18 ≤ S < 24	80≤S<95
VeryLow	10 ≤ S < 18	60≤S<80
Low	*S <10*	S<60

TABLE VI. DISCRETIZATION FORMAT TABLE FOR IELTS SCORES

Classified Name	Range of Scores (S)	
	Reading/Listening/Speaking/Writing	Total
VeryHigh	S > 7.5	S > 7.5
High	7.0 ≤ S ≤ 7.5	7.0 ≤ S ≤ 7.5
Medium	6.0 ≤ S ≤ 6.5	6.0 ≤ S ≤ 6.5
VeryLow	5 ≤ S ≤ 5.5	5 ≤ S ≤ 5.5
Low	S <5	S <5

We have considered Job and research experience in terms of years and publication records as five classified values (0: VeryLow, 0-2:Low, 3-4:Medium, 5-6:High; >6:very High) based on numeric value given in the real dataset of academic profile of successful graduate applicants. All other attributes in the customized universal are discrete and suitable for Association Rule mining.

TABLE VII. TRANSFORMED TABLE FROM CUSTOMIZED UNIVERSAL TABLE

Gender	Outgoing Country	Outgoing University	Department	ResearchArea	CGPA	GRE_Score	TOEFL/ IELTS_Score	Funding	……
Male	USA	UC Berkley	EEE	Signal Processing	VeryHigh	VeryHigh	VeryHigh	FullWaiver_TA	
Female	USA	UIUC	CSE	Data Mining	Medium	High	Medium	PartialWaiver_RA	
….	….	….	….	….	….	….	….	….	

V. Experimental Platform

In this experiment, data up to the last three years including different sessions such as 'Fall', 'Spring' are considered. There are more than a thousand records of successful applicants who are now pursuing higher study abroad with financial support. We have categorized relevant academic and standardized test records as well as job and research experience and built into a customized universal table structure of about four hundreds successful applicants ignoring missing values as well as null values. Finally we transformed it into a transformed table structure which contains only discretized attributes for applying Predictive Apriori Association Rule mining algorithm.

After preprocessing step, a customized transformed table of about 392 students of different universities of Bangladesh has been obtained who have already got opportunity to study MSc or PhD in the renowned universities abroad. We have manipulated the transformation table containing all continuous data like CGPA, test scores and transformed into five discrete values. Then, we have used Weka Explorer [22], a popular suite of machine learning software written in Java. We have applied Predictive Apriori algorithm to the customized transformation table with predefined minimum support and confidence using Weka to generate interesting Association Rules. Choosing predictive accuracy precisely and selecting important association rules from huge number of generated rules are very important in this regard.

VI. Experimental Results and Analysis

A. Influence of Gender

The influence of gender has been found in the overall academic profile of successful applicants. We have selected just few rules related to gender of successful applicants with higher predictive accuracy. There are multiple factors that affect admission with funding and academic records. The result of Fig. 1 points out that the male students have very high predictive accuracy of getting funding as RA with 1 year research experience. Another rule indicates that getting PhD admission with full funding as TA among female students is better than the male students.

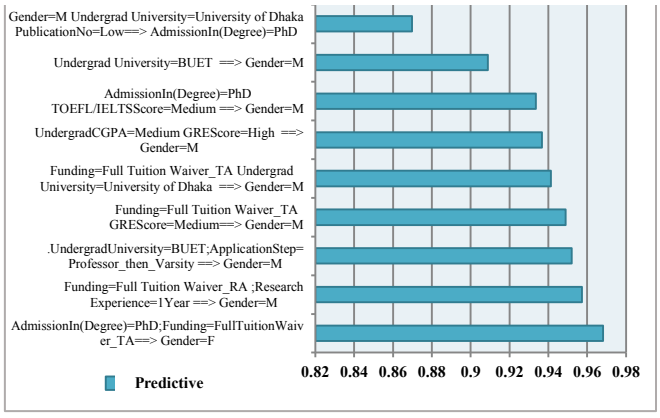


Fig. 1. Influence of Gender

B. Influence of Undergraduate Department & University

In Fig. 2, there are several interesting association rules with high predictive accuracy. Rule no. 1 indicates that if admission opportunity is obtained by a Civil Engineer, then he usually achieve funding in previous years. Again another implies that if a BUET graduate tries to apply for PhD with funding, he or she should apply to a university after contacting with professor.

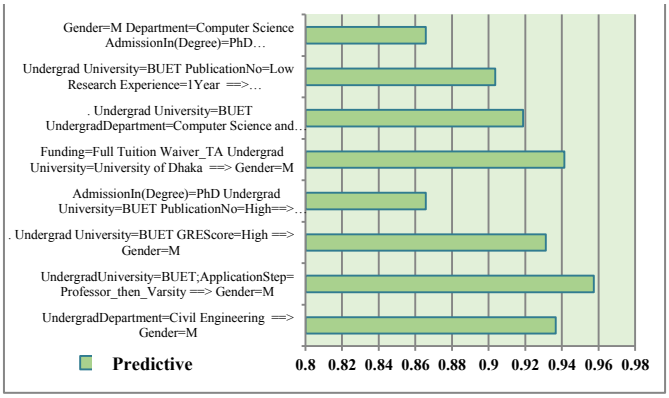


Fig. 2. Influence of Undergraduate University & Department

C. Influence of Research Area

The analyzed Association Rules show that research area influences funding as RA. In rule number 6, we find that if anyone has research interest with Data Mining he or she may get funding with RA as the rule has higher predictive accuracy than research area of bioinformatics. Again according to rule no. 3, if anyone does not have fixed research interest, he or she should focus on getting high GRE scores for funding opportunity according to the rule stated in Fig. 3.

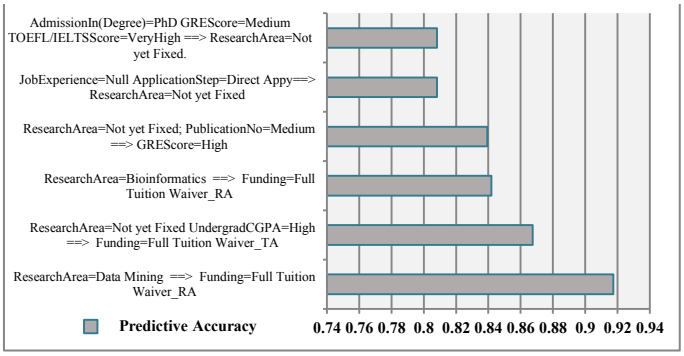


Fig. 3. Influence of Research Area

D. Influence of Undergraduate CGPA

We have selected few interesting association rules based on higher predictive accuracy presented in the Fig. 4. For example a rule indicates that if any student has very high CGPA and high GRE Score, he or she should contact professor before applying to a university. Again another significant rule indicates that the number of male students with undergrad CGPA high is relatively large in the last three years dataset of successful applicants.

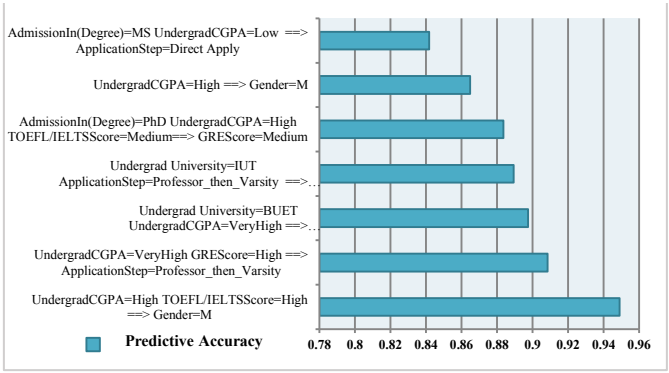


Fig. 4. Influence of Undergraduate CGPA

E. *Influence of GRE Score*

The students who get medium range of GRE scores generally gets high score in quant but low in verbal. By analyzing the rules illustrated in the Fig. 5, we have observed such kind of fact. It is discovered that with a higher predictive accuracy that students with high GRE scores should direct apply to university and funding opportunity will be relatively higher in this kind of cases. Again, if a student applies for PhD and gets full funding as RA; he or she usually gets very high GRE score as a significant rule of below figure indicates.

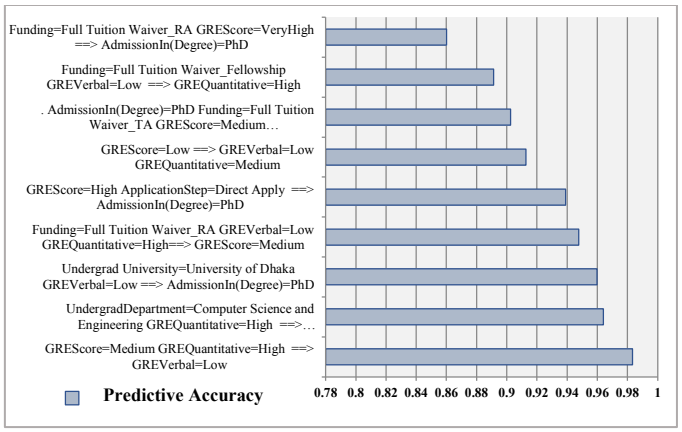


Fig. 5. Influence of GRE Score

F. *Influence of TOEFL/IELTS Score*

Several significant rules can be selected to indicate the influence of TOEFL/IELTS score based on higher predictive accuracy. One rule in the Fig. 6 indicates that if a student with low CGPA gets funding then his or her TOEFL score is very high. Again, another significant rule indicates that if a student with low CGPA directly applies to a university abroad, he or she gets full funding because of high TOEFL/IELTS speaking scores. Another important rule indicates that if a very low CGPA holder applies after getting high speaking and listening score in TOEFL/IELTS; he or she gets admission in PhD with funding.

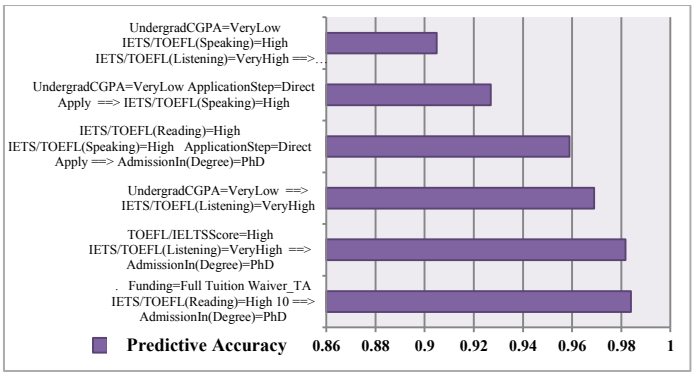


Fig. 6. Influence of TOEFL/IELTS Score

G. *Influence of Research Experience & Publication Record*

After analyzing the generated Association Rules (in Fig. 7) we observed various impacts of research experience and publication records of a successful applicant. According to the fifth rule if an applicant has research experience more than 2 to 4 years and applies for PhD admission, the probability of getting full tuition waiver with RA is much higher. Again another rule implies that, if an applicant has very high publication records and he or she applies after contacting with professor, he or she usually gets PhD admission with funding.

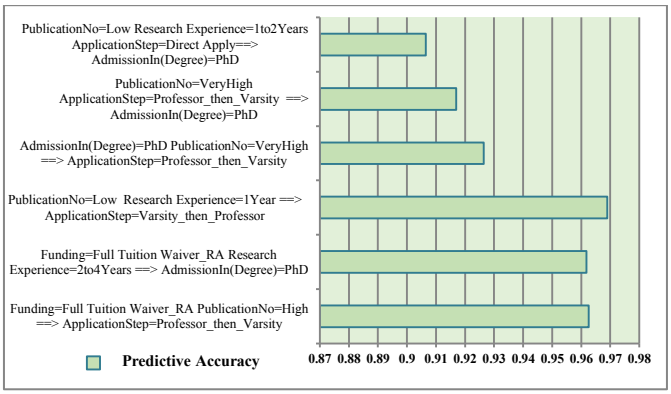


Fig. 7. Influence of Research Experience & Publication Record

H. *Influence of Job Experience*

From the generated association rules based on predictive accuracy we have analyzed influence of job experience of the applicant and stated few significant rules in Fig. 8. For example, according to the last rule in the Fig. 8, if an applicant has job experience of more than 4 years and contacts with professor before applying; he or she may probably get admission in PhD program. Again, second rule implies that if a fresher applies for PhD program with no job experience and he or she directly applies to university without contacting with any professor, the probability of getting full tuition waiver with RA is higher according to the generated predictive accuracy 0.90342.

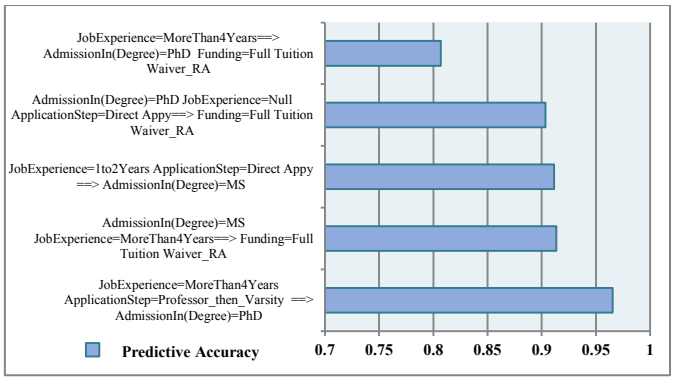


Fig. 8. Influence of Job Experience

VII. CONCLUSIONS

Discovering knowledge from entire academic profile of successful graduate applicants and applying it properly for providing or achieving guidelines will be very much beneficial for the prospective graduate applicants. Here, we have considered the data of three years which are voluntarily given by the successful applicants, but there many other students who get admission with funding yet to record in the list. We can apply the same technique to extract knowledge from all the students pursuing higher study abroad in last decade. Again, we have considered the students of Bangladesh going abroad for graduate studies. We can modify the technique so that it can be applicable to the analysis of applicants of other countries like Indonesia, Malaysia, China etc. In this research, we have preprocessed the dataset efficiently after gathering all relevant information from successful applicants before applying the data mining algorithm. For Association rule mining, we have manipulated a technique to discretize the records of applicant's continuous attributes like CGPA, Standardized test scores of GRE, TOEFL and IELTS, research and job experience etc. Then we have applied Predictive Apriori Association rule mining algorithm to generate association rules. We have categorized interested association rules based on higher predictive accuracy. Finally we extract knowledge from those significant rules which may be very helpful for the students who have profound desire to obtain graduate admission into a renowned institution abroad with financial support.

ACKNOWLEDGMENT

The authors would like to gratefully acknowledge collaborators of Higher Study Abroad Bangladesh chapter and all the graduate students staying abroad for their generous assistance to get the necessary records for this research study.

REFERENCES

[1] J. Han , M. Kamber, and J. Pei, "Data Mining: Concepts and Techniques", 3rd ed. San Francisco:Morgan Kaufmann Publishers, 2011.
[2] Aggarwal, Charu C. "*Data mining: the textbook"*. Springer, 2015.
[3] (2015, Oct.) All bangladeshis abroad Universal Database [Online]. Available:http://www.higherstudyabroad.com/database/all-bangladeshis-abroad/
[4] Charu C. Aggarwal, "Data Mining: The Textbook", 1st ed., Springer International Publishing Switzerland, 2015.
[5] Romero, C., & Ventura, S. "Educational data mining: a review of the state of the art". IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40(6), 601-618, 2010.
[6] Baradwaj, B. K., & Pal, S. "Mining educational data to analyze students' performance". arXiv preprint arXiv:1201.3417, 2012.
[7] El-Halees, A. "Mining students data to analyze e-Learning behavior: A Case Study". Department of Computer Science, Islamic University of Gaza PO Box, 108, 2009.
[8] Romero, C., Ventura, S., & García, E. "Data mining in course management systems: Moodle case study and tutorial". Computers & Education, 51(1), 368-384, 2008.
[9] Baker, R. S., & Inventado, P. S. "Educational data mining and learning analytics". In Learning analytics, pp. 61-75, Springer New York, 2014.
[10] Peña-Ayala, A. "Educational data mining: A survey and a data mining-based analysis of recent works". Expert systems with applications, 41(4), 1432-1462, 2014.
[11] Abdullah, Z., Herawan, T., Ahmad, N., & Deris, M. M. "Mining significant association rules from educational data using critical relative support approach". Procedia-Social and Behavioral Sciences, 28, 97-101, 2011.
[12] Merceron, A., & Yacef, K. "Interestingness measures for association rules in educational data". In Educational Data Mining, 2008.
[13] Romero, C., Romero, J. R., Luna, J. M., & Ventura, S. "Mining rare association rules from e-learning data". In Educational Data Mining, 2010.
[14] E. Garcia, C. Romero, S. Ventura, T. Calders, "Drawbacks and solutions of applying association rule mining in learning management systems," In Proceedings of the International Workshop on Applying Data in e-Learning, 2007.
[15] Kumar, V., & Chadha, A. "Mining association rules in student's assessment data". International Journal of Computer Science Issues, 9(5), 211-216, 2012.
[16] Parack, S., Zahid, Z., & Merchant, F. "Application of data mining in educational databases for predicting academic trends and patterns". In Technology Enhanced Education (ICTEE), 2012 IEEE International Conference on (pp. 1-4). IEEE, 2012.
[17] A.S.M. L. Hoque, R. Paul, S. Ahmed, "Preprocessing of Academic Data for Mining Association Rule," In Proceedings of the Workshop on Advances in Data Management: Applications and Algorithms(WADM), 2013.
[18] Ahmed, S., Paul, R., & Hoque, A. S. M. L. "Knowledge discovery from academic data using Association Rule Mining". In Computer and Information Technology (ICCIT), 2014 17th International Conference on (pp. 314-319). IEEE, 2014.
[19] Stefan Mutter, Mark Hall and Eibe Frank. "Using Classification to Evaluate the Output of Confidence-Based Association Rule Mining". In the Proceedings of 17th Australian Joint Conference on Artificial Intelligence, Cairns, Australia, December 4-6, 2004. DOI: 10.1007/978-3-540-30549-1_47.
[20] Sunita B. Aher and Lobo L.M.R.J. "A Comparative Study of Association Rule Algorithms for Course Recommender System in E-learning". International Journal of Computer Applications (0975 – 8887), Volume 39– No.1, February 2012.
[21] Divya Bansal and Lekha Bhambhu. "Execution of APRIORI Algorithm of Data Mining Directed Towards Tumultuous Crimes Concerning Women". International Journal of Advanced Research in Computer Science and Software Engineering, ISSN: 2277 128X, Volume 3, Issue 9, September 2013.
[22] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I. H. Witten, "The WEKA Data Mining Software: An Update." *SIGKDD Explorations*, 11(1), 2009.

Sequence Classification: A Regression Based Generalization of Two-stage Clustering

Nusrat Jahan Farin*†, Nafees Mansoor†, Sifat Momen†, Iftekharul Mobin† and Nabeel Mohammed†
*Computer Science and Engineering Department
Jahangirnagar University, Savar, Dhaka
*†Computer Science and Engineering Department
University of Liberal Arts Bangladesh (ULAB), Dhanmondi, Dhaka
Email: nusratfarin89@gmail.com, nafees.mansoor@ulab.edu.bd, sifat.momen@ulab.edu.bd,
iftekharul.mobin@ulab.edu.bd, and nabeel.mohammed@ulab.edu.bd

***Abstract*—Two-stage clustering based approaches has been used for sequence classification. However, certain parameters of these process are either hand picked or found through exhaustive searches. In this paper we propose a simple regression based approach, derived from parameter values found in a previous study, which generalizes the method to find values of three different parameters through proposed equations. We tested the applicability of our method on eight UCR sequence datasets and found our method to be comparable with hand picked approaches and better than single clustering based approaches.**

Keywords - sliding window size; cluster number; time-series data; kmeans++

I. Introduction

Identification and retrieval of appropriate features from time-series data is a challenging [1], and at the same time a rewarding task. These classes of data typically inherent sequence-based partitions, which are not always known apriori. Clustering-based techniques have been successfully applied to this problem domain, mainly due to the structure-discovery mechanism built into such grouping techniques [2].

The approach and results of [3] forms the basis of this study, where a two stage clustering approach is successfully applied in a classification task on some parts of the popular UCR time series data set [4]. Their approach mainly consists of four steps with three manually chosen parameters.

At the first step, the method divides the data into subsequences of a manually selected length sl. There exists several techniques to divide the subsequences [5]. These subsequences are then clustered into k_1 clusters and each cluster center is used as a bin center to create a Bag-of-Words (BoW) feature vector from the subsequences. These BoW feature vectors are then further grouped into k_2 clusters. A classification scheme, which is further detailed in section IV, is used to utilize these k_2 clusters to predict the class of time-series data.

In [3], the values of sl, k_1 and k_2 are all selected through some non-specified methods. It appears that the values reported, the ones which found to give the best performance on each individual data set. This, however, is not an approach that can be repeated under practical constraints.

This paper reports on the finding of some initial work on prediction of the parameter values from information known a priori. Particularly the contribution of this work is a natural extension of [3], where the parameter values found to be useful are used to fit linear models where the dependent variables are the parameters and the independent variables are the number of classes and sequence length.

The results indicate that the proposed models, though extracted from only three data points, are effective and can match the performance of the exhaustively found parameter values and consistently outperforms the single clustering approach.

The rest of the paper is organized as follows. The following section describes the method and evaluation strategy of [3] in details. Section III describes the method used to derive the proposed linear models and how they have used in experiments. In section IV the results are described and compared with the results obtained in [3]. The paper is concluded with the future works in section VI.

II. Existing Methods

Time series data is sequential data, where the ordering is temporal. There are several popular data sets, which have been widely used in research projects. The UCR time-series is a collection of multiple data sets [6]. It has been used as a benchmark collection of data in multiple studies. This study reports results using 8 data sets from the UCR time-series collection, the details of which can be found in Table I.

In [3], the time-series data is divided into some subsequences. The set of the subsequences is $S = \{S_{nl} | n = 1...., X, l = 1, ..., N\}$, where X is the number of time-series data, n is the number of sequences, l is the length of subsequeces and N is the number of length.

Each sequence is encoded as a vector and aggregated into a data matrix which is then used in the first of clustering stage. The used clustering algorithm is Kmeans++ [7]. The study reported in [3] uses brute-force technique to determine the number of clusters rather than any particular methodology.

As it is possible for the original sequences to be of variable length, a fixed-length encoding scheme is necessary for feature comparison. The cluster centers obtained from the first clustering stage are treated as 'words' in a Bag-of-Words (BoW) encoding scheme. Briefly, each subsequence is mapped to one of the cluster centers by finding the cluster center that is closest

978-1-5090-5769-6/16/$31.00 ©2016 IEEEs

to the subsequence vector in terms of cosine length. For cluster center, the number of subsequences mapped to it is counted, from which it is possible to calculate a frequency histogram, which is the BoW vector.

In [3], numeric vector N is created, such that $V_n = \{x_1, x_2, x_3, ...x_k\}$ where n = 1, 2, 3, ..., N. After the creation of the vector, the numeric value is used for the second cluster. In the second cluster kmeans++ algorithm is used once again [3]. The numeric vector is used for the clustering, where the cluster number is determined randomly for different datasets. In [3], clear and generalized formulas for determining the first and second cluster numbers are absent. Moreover, discussion on the sliding window size is also not presented.

However, the classification performance and the generated cluster quality are checked in [3]. In [3], two methods named as the Pseudo F measure [8] and the Davies-Bouldin (DB) index [9] is compared for measuring the performance. Authors in [3] presented that the high Pseudo F measure indicated the good quality for the clustering, where as the DB index indicated high quality for low measures. Thus the method in [3] performs better than the comparative methods of DB index for three datasets, namely Synthetic, Beef and OliveOil.

In the systems of two-stage clustering process for time-series data, four datasets called Beef, Synthetic, Coffee and OliveOil are used. In [3], kmean++ algorithm is used for two stage cluster, where cluster number and the sliding window of sequence length are predetermined. Authors in [3] use random value for experiments and the best results are explained. Rational explanation on use the particular cluster number and sliding window size are not provided in [3]. Thus, All of the numbers may confuse the users. To avoid the confusion of the user, some generalized formulas are proposed. Thus, using these formulas the sliding window size, number of first and second cluster can be determined.

Two-stage clustering process requires to determine the sliding window size, cluster number etc. The existing systems determine these things arbitrarily for different database. In this paper, we propose a new process to determine the sliding window size, first cluster number and second cluster number for performing the second stage cluster. In the first clustering process the featured data of the time-series data are extracted. Then the featured data is used for the second cluster. Before executing the second cluster, the result of first cluster is converted to numerical vector. The numeric vector is used in the second clustering. Finally, featured data of the time-series data are obtained through the second clustering procedure. The paper also compares the result with the existing methods [3].

A. Splitting up The Time-Series Data

Tomoharu Nakashima uses arbitrary number for different database for sliding window, first cluster number and second cluster number [3]. The series of subsequences are used for the first clustering described below.

B. Initial Clustering Procedure

In the first clustering procedure the extracted subsequences are used to find out the features of time-series data. KMeans++ algorithm is used to cluster the extracted subsequences. KMeans [10] algorithm gives the center of each cluster where kmeans++ the cluster number to be predefined. In these methods, to determine the cluster number formula 3 is used in this paper. Cosine distance [11] of each data from the cluster center is measured by cosine similarity [12] (equation 2).

$$Similarity(A, B) = \frac{\sum_{i=1}^{n} A_i.B_i}{\sqrt{\sum_{i=1}^{n}(A_i)^2}\sqrt{\sum_{i=1}^{n}(B_i)^2}} \quad (1)$$

$$Distance_{cosine} = 1 - Similarity(A, B). \quad (2)$$

Cosine gives the similarities between the subsequences. All values of the subsequences are assumed as positive where the maximum value of the similarity is 1 and the minimum is 0. The first cluster result shows the featured data of the time-series [13] data. These data is converted into numeric values. kmeans++ Procedure for clustering is given in Algorithm 1 [7]. The conversion procedure is described in the next section.

Algorithm 1 KMeans++ algorithm

1: Choose an initial center randomly from the subsequences.
2: Compute the vector containing the square distances between all points in the dataset.
3: Choose a second center from subsequences randomly using the probability distribution
4: Recompute the distance vector until K_f cluster center determined.
5: Choose a successive center using the average, simple arithmetic mean of all element of subsequences
6: Recompute the distance vector until convergence.

C. Time-Series Data to Numeric Vector

Time-series data in the database are converted into numeric vectors. Numeric vector is implemented from the result of first cluster. This process determines each and every data point of the cluster and determine the repetition frequentness of the cluster in introductory time-series data. We convert the time-series data to numeric vector using the existing method. These methods are described in [3]. Detailed procedure is presented in Algorithm 2:

D. Second Clustering

The kmeans++ algorithm is used for the second cluster. The proposed methods for the second-stage clustering is presented at the next section.

The overview of [3] is presented in section III.

III. Proposed Method

A. Data Analysis

[3] reports results of their proposed two stage clustering system using different values for the three parameters discussed above. The method used to determine the parameter values are not discussed. Assuming the parameter values were chosen to achieve good classification rates, it is reasonable to attempt to find a pattern in the values. As [3] reported results on four

Algorithm 2 Time-series data to numeric vector

1: Let num = 1 and len = 1.
2: Set V_n = (x_1,....., x_n)
3: **if** C_{nl} the subsequences Sub_{nl} belongs to by Identifying its nearest cluster center **then**
4: **return** Increase the C_{nl}-th of V_n by 1.
5: **end if**
6: **if** len = S **then**
7: **return** num = num + 1 and len = 1
8: **else**
9: len = len + 1
10: **end if**
11: **if** n > N **then**
12: **return** terminate the process.
13: **else**
14: go to step 2.
15: **end if**

different datasets, for each parameter it is possible to obtain four values. Although this is a small number, it is sufficient to fit a straight line through the points.

Figure 2 and Figure 3 shows the number of clusters in first-stage clustering and second-stage clustering respectively plotted against the number of classes.

Figure 4 plots the chosen window size(s) vs the sequence length of the data set.

Each of the figures also shows the best fit straight line. The main focus of this study revolves around these fitted lines, as their equations are treated as general method of determining the values of the three parameters. These equations are shown in Equations 3, 4, 5

$$K_f = Round(-17 * N_{classes} + 130) \tag{3}$$

Where k_f = First Cluster Number and $N_{classes}$ = Number of Classes

$$K_s = Round(4.34 * N_{classes} + 1.5) \tag{4}$$

Where k_s = Second Cluster Number and $N_{classes}$ = Number of Classes

$$S_w = Round(0.5 * S_l + 10) \tag{5}$$

Where S_w = Sliding Window Size and S_l = Sequence length of time series data

IV. Computational Experiment

In this section the performance of the proposed method is investigated. Experiments are done on 8 datasets from the UCR collection. These datasets are available through the UCR Time-series classification/clustering page [4], [6]. All details of the datasets, used in this experiment, is given in Table I.

In this experiment two types of comparison is made. The first comparison is between the existing proposed methods for the two-stage clustering [3] and the proposed method of the paper. And the second comparison is between the proposed methods and the first clustering process. All the comparison of the result is discussed in section V.

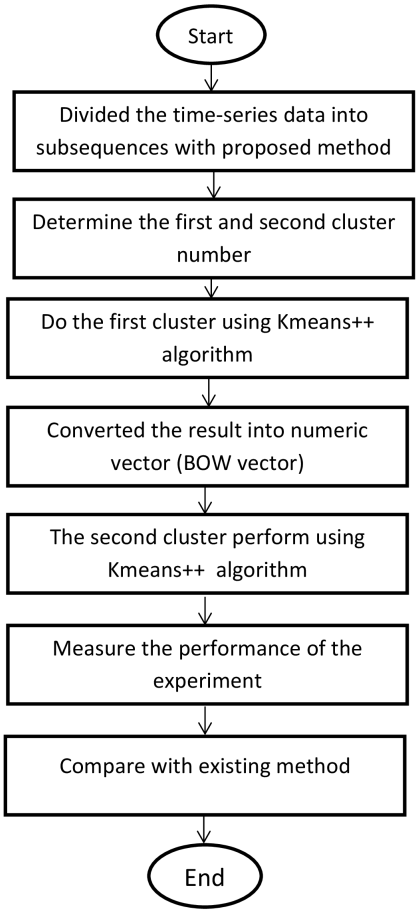


Fig. 1: System overview of the proposed method

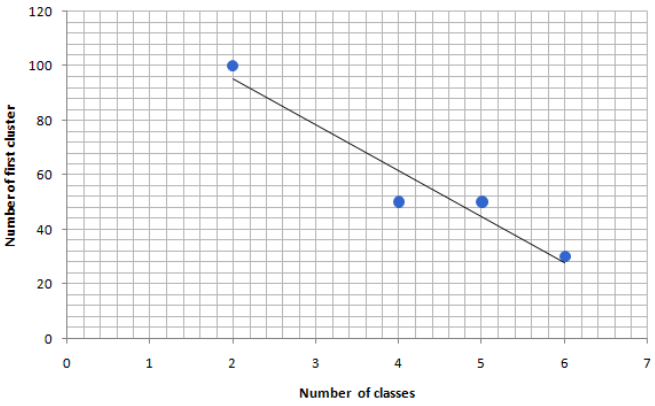


Fig. 2: Number of classes vs number of clusters in first-stage clustering

V. Results and Discussion

The results obtained using the proposed parameters estimation methods are compared with the results reported in [3].

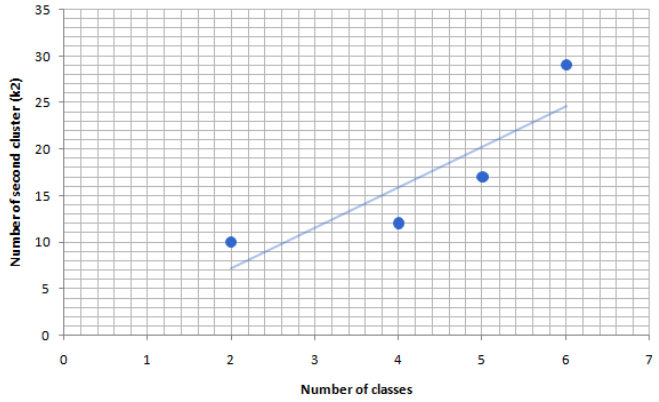


Fig. 3: Number of classes vs number of clusters in second-stage clustering

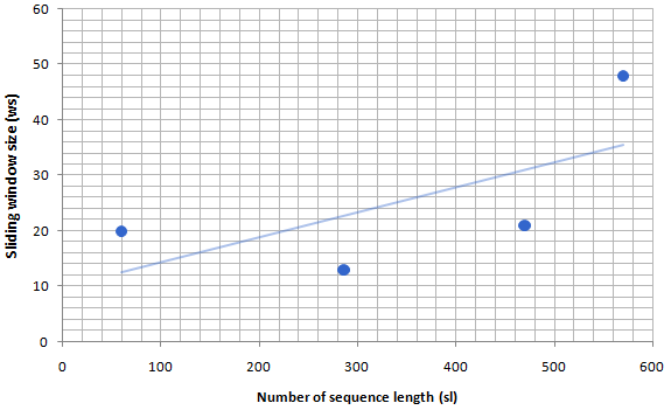


Fig. 4: Sequence length vs window size

The comparison is presented in Table II, which shows that ther results obtained in this study are comparable in our accuracy with those reported in [3]. In some of situation terms of the proposed method give better classification accuracy [3]. Figure 5 shows a bar-chart for visual comparison of the results.

In Table III, the comparison between the proposed methods and the single clustering methods for the proposed formula is given. In most of the cases the proposed methods give the better result than the single clustering process. In Figure 6 the overview of the comparison is given.

In Table III the classification performance of eight bench-

TABLE I: THE EIGHT BENCHMARK DATSETS USED IN THE EXPERIMENT

Datasets	classes	training seq.	test seq.	seq. length
Coffee	2	28	28	286
OliveOil	4	30	30	570
synthetic	6	300	3	60
Beef	5	30	30	470
BeetleFly	2	20	20	512
BirdChicken	2	20	20	512
Car	4	60	60	577
Earthquakes	2	139	322	512

TABLE II: COMPARISON PERFORMANCE WITH EXISTING METHOD [3]

Datasets	Comparison Method	Proposed Method
Coffee	0.954±0.032	0.964±0.03
OliveOil	0.850±0.043	0.8±0.03
synthetic	0.977±0.007	0.977±0.005
Beef	0.603±0.057	0.47±0.05

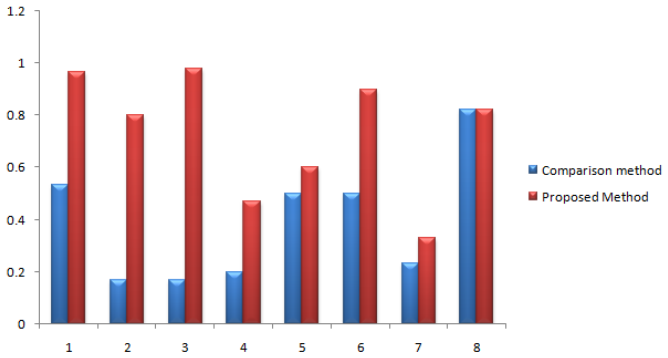


Fig. 5: Comparison between the existing method of single clustering and proposed method

mark datasets are given. The average classification rate is given here over ten trials along with standard deviation. The existing system [3] shows the four datasets classification performance. The results of the experiment is compared with the existing results in Table II.

VI. CONCLUSION

Two-stage clustering based approach to a time-series data classification task is proposed in [3]. The process involves setting/choosing values of three different parameters. The original study does not report on the method used to choose the parameters. The contribution of this paper is in proposing three equations which can be used as a generalized method of choosing the parameter values based on known data. Experiments were conducted using the proposed parameter estimation equations and the classification accuracy obtained from these experiments were comparable. And most of the time results produced by proposed method, are better than those reported by [3]. The comparison was done on the four datasets used in the previous study. Results were reported on a

TABLE III: CLASSIFICATION PERFORMANCE ON BENCHMARK DATASETS WITH SINGLE-STAGE CLUSTERING

Datasets	Comparison Method	Proposed Method
Coffee	0.535	0.964±0.03
OliveOil	0.167	0.8±0.03
synthetic	0.167	0.977±0.005
Beef	0.2	0.47±0.05
BeetleFly	0.5	0.6±0.001
BirdChicken	0.5	0.9±0.01
Car	0.233	0.33±0.03
Earthquakes	0.82	0.82±0.057

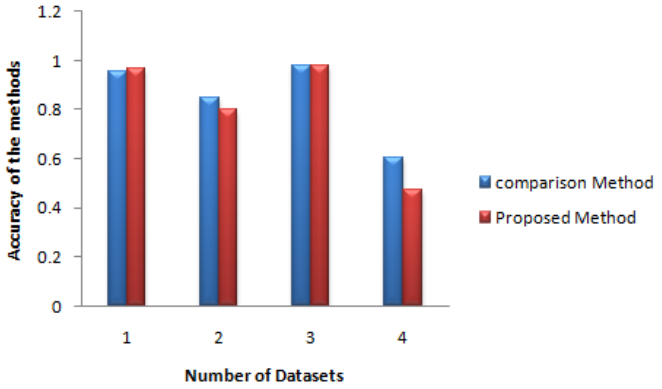


Fig. 6: Comparison between the existing method of two-stage clustering and proposed method

further four data sets which demonstrate the proposed method is consistently superior to single-clustering-based approach.

REFERENCES

[1] Granger, Clive WJ Some properties of time series data and their use in econometric model specification. In *Journal of econometrics*, pages 121–130, Vol. 16, Num. 01, Elsevier. 1981.

[2] TO SUBSPACE CLUSTERING HAVE. A number of approaches to subspace clustering have been proposed in the past two decades. 2011.

[3] Youhei Kuroda Md. Atiqur Rahman Ahad Tomoharu Nakashima, Gerald Schaefer. Performance evalution of a two-stage clustering technique for time-series data. ICIEV 2016, Dhaka University, 2016.

[4] Jiaping Zhao and Laurent Itti. Classifying time series using local descriptors with hybrid sampling. *IEEE Transactions on Knowledge and Data Engineering*, 28(3):623–637, 2016.

[5] Seyedjamal Zolhavarieh, Saeed Aghabozorgi, and Ying Wah Teh. A review of subsequence time series clustering. *The Scientific World Journal*, 2014, 2014.

[6] Bing Hu Nurjahan Begum Anthony Bagnall Abdullah Mueen Yanping Chen, Eamonn Keogh and Gustavo Batista. The ucr time series classification archive. www.cs.ucr.edu/~eamonn/time_series_data/, 2015 (accessed July, 2016).

[7] David Arthur and Sergei Vassilvitskii. k-means++: The advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 1027–1035. Society for Industrial and Applied Mathematics, 2007.

[8] Tadeusz Caliński and Jerzy Harabasz. A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 3(1):1–27, 1974.

[9] David L Davies and Donald W Bouldin. A cluster separation measure. *IEEE transactions on pattern analysis and machine intelligence*, (2):224–227, 1979.

[10] Sugato Basu, Arindam Banerjee, and Raymond Mooney. Semi-supervised clustering by seeding. In *In Proceedings of 19th International Conference on Machine Learning (ICML-2002*. Citeseer, 2002.

[11] Hui Ding, Goce Trajcevski, Peter Scheuermann, Xiaoyue Wang, and Eamonn Keogh. Querying and mining of time series data: experimental comparison of representations and distance measures. *Proceedings of the VLDB Endowment*, 1(2):1542–1552, 2008.

[12] Jun Ye. Cosine similarity measures for intuitionistic fuzzy sets and their applications. *Mathematical and Computer Modelling*, 53(1):91–97, 2011.

[13] Rohit J Kate. Using dynamic time warping distances as features for improved time series classification. *Data Mining and Knowledge Discovery*, 30(2):283–312, 2016.

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

Generation and Verification of Digital Signature with Two Factor Authentication

Narayan Ranjan Chakraborty[1], Muhammad Taifur Rahman[1], MD. Ekhlasur Rahman[1], Mohammad Shorif Uddin[2]

[1]Department of Computer Science and Engineering, Daffodil International University
[2]Department of Computer Science and Engineering, Jahangirnagar University

***Abstract*—This paper is intended to provide a cloud-based digital signature platform with biometric authentication and establishes an enhanced security solution in the field of cryptography. We proposed a new schema of digital signature where signature generation and verification is done on the cloud environment. Like Public Key Infrastructure each user owned a pair of key private key and public key. A signer has to confirm his biometric identification (vein pattern) and a one-time key verification to access the private key for signing a document. This one-time key (OTK) is only shared between the signer and the receiver additionally they use it for a single document signing or verification. Before the verification of a signature or a document, a receiver also needs to prove his identity using his vein pattern. To complete the verification process, he/she must provide the public key of the signer and the shared one-time key to the system. Thus our system gives a confidential interaction between the signer and the receiver of the document.**

***Index Terms*—Cryptography; Digital Signature; Finger Vein; MD5; One-time Password.**

I. Introduction

In todays world e-commerce, e-government, e-business is omitting their conventional documents into e-document. But information tempering, signature forgery of alteration has been increasing remarkably. As a result, from business organization to the ordinary user who signs a document digitally can be a victim of digital crime. To assure the security of digital signature,most widely accepted scheme is Public Key Infrastructure (PKI) which is relied on two pair of keys: private key and public key. [8] Although PKI is widely accepted method, it has some drawbacks . [14] A public key algorithm (e.g. RSA) is used to produce this pair of keys. [15] Frauds target this private key of clients or Certificate Authority (CA) to pretend as a signer and using this key can easily sign the documents. So the main concern of the PKI based security system is to protect the keys. Generally, private keys are stored on a device using a password or PIN which can be guessed easily. Another approach to protect private keys is to store them on a smart card which needs to be purchased from a trusted source and needed to be carried with for signing. If for any reason the cardholder loses his card then he could end up facing some unexpected issues. Somehow someone else gets that card and wishes to use the private key of a signer and signs a message, then it is impossible to detect that the signer is not the actual cardholder and he did not sign the message. Furthermore, CA may use old keys to issue new certificates for the clients [7].

II. Literature Review

Xiaodong Liu, Quan Miao and Daxing Li [1], presented A New Special Identity Based Signature Scheme where a signer needs to provide his/her fingerprint in the Key Generation Center (KGC) during registration. After that finger print is converted into a public key string and issued a corresponding private key by the KGC. To claim this private key, it is mandatory for a signer to confirm his identification using his fingerprint. During the verification, receiver used the public key (came from the singer) to reconstruct the fingerprint. After that the singer is recommended to provide his finger print on the spot to match with the reconstructed finger print. Although this system identifies the signer using his biometric, the signer is forced to the verification process.

Ahmed B. Elmadani [2], proposed a system that generates unique keys (Ks, Kr) for the singer and the receiver from their extracted face image to digitally sign a message. Both unique keys are combined together to encrypt and decrypt the calculated fingerprint of the document. Singer sends the fingerprint of the document (which is already encrypted by combined key, Ksr), the document itself and his Signing Key. These three parameters are encrypted with the receivers key. Finally, the receiver uses his/her key to decrypt the received encrypted message and generates a new fingerprint of the document to match with the previously received document fingerprint although the system gives an error rate 1.65.

Hitachi, Ltd [3], Developed a secure digital signature technology where finger vein is used as a secret key instead of using a smart card or a password for user authentication. The scheme is strong as the Water Signature claimed by Hitachi Ltd.

Sambangi Eswara Rao and S. Ravi Kumar [4], used iris as a biometric and proposed a solution to generate private key from the biometric template (512 byte iris code template) and used one-way hash function to provide protection of the biometric template. This Paper presented two systems using RSA and DSA. Additionally, they suggested not to transmit the biometric template over the internet for authentication and eliminated the key management issues of PKI by avoiding the private key storage. Finally, authors also discussed the problem about the correctness of all bits of the template.

A.M. Al-Khouri and J. Bal [5], suggested to use three-factor authentication for digital identification. Authors described a

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

trio technology that integrated PKI with smart card and biometrics for precise identification. The Main advantage of this proposal is that nobody can use the smart card but the cardholder.

Wojciech Kinastowski [6], proposed a protocol for data exchange and message signing in the cloud environment. According to the protocol public and private key pair is stored on the cloud where the private key is encrypted with a password that is only known to the signer. To access this private key a one-time password is generated using a hardware security module (HSM) and send it to the signers cell phone. The protocol provides an excellent usability (eliminate dedicated devices like smart card, card reader etc.) and cross platform requirement in the cloud communication, however it does not provide sufficient verification to detect the precise identification of the signer. If the Singer lost his cell then anyone can access the private key using his OTP.

III. Proposed System

A. System Architecture

In our proposed system, signature generation and verification process are done in the cloud platform. As all we know in PKI, each user has a pair of keys: private key and public key. Private keys are stored on a central secure server. Our main target is the protection of this private key and provide a maximum possible security level, to achieve this goal we will use a hardware security module (HSM) to the application server.

To access this private key and sign a document signer has to face two-factor authentication i.e. something you are - Biometric identification and something you have One-Time Key (OTK) - a randomly generated code. So the system allows a user to sign if his biometric vein image template is matched with the previously stored template in the database and it generates a key (OTK) then send it to the signers mobile. After that the signer return it back to the application server to complete the signature operation. This key is used for a single session only and known between the signer and the receiver to provide confidential interaction between them. The scenario of signature generation is represented in fig. 1

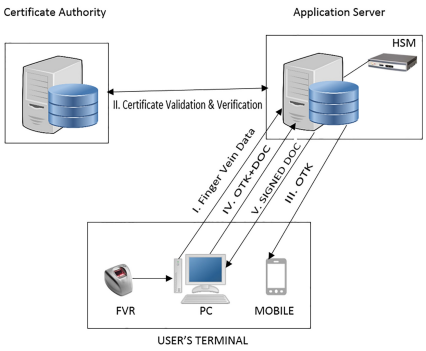


Fig. 1. Generation of Digital Signature

Signer sends his finger vein data using a vein reader to make an authentication request to the application server from the user terminal. The server manages the users certificates and validates them with the help of the certificate authority. The server returns the signed document to the user when the signature process is done.

During signature verification, receiver also needs to provide his vein data to prove his authenticity. This time receiver just provides the OTK and public key to the server to complete the verification process. Then server displays the verification information of the document to the receiver. This story is exposed in fig. 2

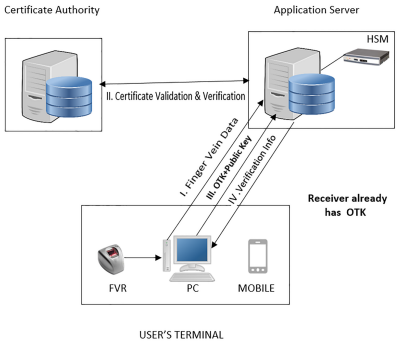


Fig. 2. Verification of Digital Signature

Forum of European Supervisory Authorities has been considered positively the concept of implementing digital signature remotely or the server-side. [16] Additionally, some trusted provider like Ascertia and DocuSign has been introduced cloud based digital signature system which is performed in the server-side environment and authorize the user with SMS-tokens. [17], [18]

B. System Algorithm

In this section we will discuss about the functions and algorithms which is used in the system.

One time key: We have used `random_int` function which use Mersenne Twisters algorithm that generate cryptographically secure pseudo random integer.

```
function one_time_key($min, $max){
    $otk = random_int( $min, $max );
    Return $otk;
}
```

Public key: We have used md5 (which takes signed document contents) algorithm which is also encrypted by using modified base64 algorithm that makes documents more secure than before.

```
function generate_public_key($document_content){
    $unique_random_string=md5($document_content);
    $base64_string=base64_encode($unique_random_string);
    $modified_$base64_string=str_replace(  +  , . , $base64_string);
    $public_key = substr($modified_base64_string,0, $length);
    Return $public_key;
}
```

C. Biometric Selection

PKI does not identify the actual signer. If we use biometric instead of PKI, it removes dedicated devices and identify the actual signer. [1] Authors suggest to use some biometric like fingerprint, face, hand geometry, voice, iris and retina to verify the signers identity. We choose vein pattern as a biometric, it overcomes the various frauds associated with other biometric. [9], [10], [11]
Finger vein biometric authentication technology detects an individual using the pattern of vein inside of a finger. Its false acceptance rate is one in a million, false reject rate is 1:10,000 and failure rate is extremely low. [12], [13] Vein pattern is usually blood vessels which carry blood, so the only live human body is workable to the authentication process. There is deoxyhemoglobin in our blood which consumes infrared lights, vein pattern then looks like black/dark outlines. With the help of infrared lights and a special camera image of vein pattern is captured. This image is transformed into the template and compared with the saved template during the authentication. We let this job to the vein reader such as Hitachi M2SYS reader, just put your finger and get authenticated.

IV. User panel implementation

To complete the signing and the verification process user must go through some system predefined procedures which are much easier than the existing system. We illustrated those procedures below:

A. Sign Up

To complete the registration process user needs to use his finger vein data only. Now the question is how system will collect the remaining information. We are assuming there is a public database maintained by the government contains the basic information (user name, address etc) about their citizens. This prevents the fake information of any user who will use our system. So the system will try to find his data in the database and if any match found, it will collect his information and save in the system. Finally as shown in fig. 3 user have to provide his current email address to complete the registration.

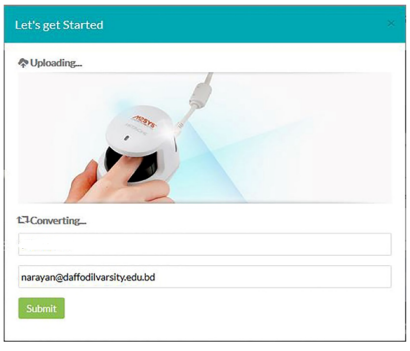


Fig. 3. User signup

B. Login

During login into the system user needs to provide his vein data only which is easier than providing password or PIN. If the vein data match with our system database, then the system will allow user to sign or verify the document.

Fig. 4. User login

C. User authentication and document signing

Before sign a document user uploads it to the system. Scenario is exposed in fig. 5.

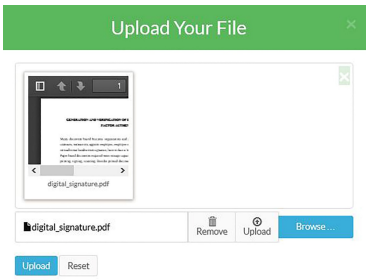


Fig. 5. File upload page

When the signer is intended to sign a document the system will generate a One-Time Key (OTK) and send it to his email. Using an authentication form system ask the OTK from the signer as a proof of authenticity. From fig. 6 it is shown that signer specifies the email address of the receiver where he will get the same OTK.

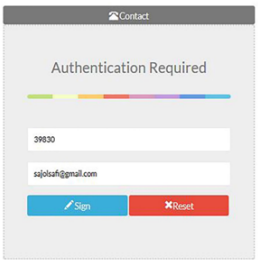


Fig. 6. User authentication

Thus the OTK is only shared between the sender and the receiver and system will provide a confidential interaction. Finally during signature generation, send a copy of OTK to the receivers cell phone (e-mail) for document verification. Finally

signed document is returned to the signer using browser default download option for his further uses as displayed in fig 7.

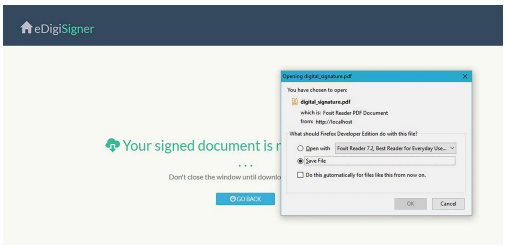


Fig. 7. Downloading signed document

After downloading we will get the signed document as displayed in fig. 8. Signature is drawn on the right corner of the bottom of the document. User can click on the signature to see the signature properties.

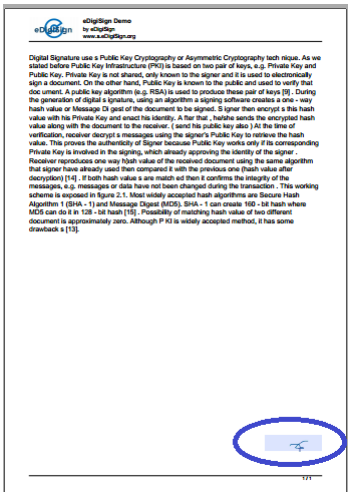


Fig. 8. Signed document

D. Document verification procedure

During document verification, receiver login to the system. After that system will ask the shared OTK and the public key of the signer. As all we know public key is available to all but OTK is only known between sender and receiver. Fig. 9 shows the verification process where first field contains OTK and second one contains public key of the signer.

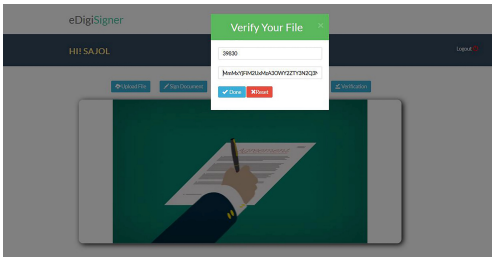


Fig. 9. Verification page

As displayed in fig. 10 receiver also may find the public key from the signature property of the signer.

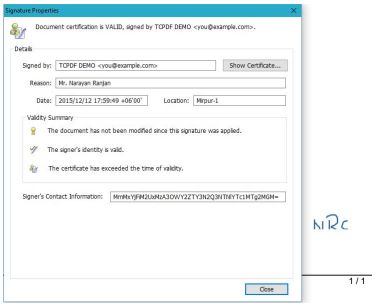


Fig. 10. Public key of signer

If provided information of OTK and public key is valid, the system will display the verification information about the document and the signer as shown in fig. 11.

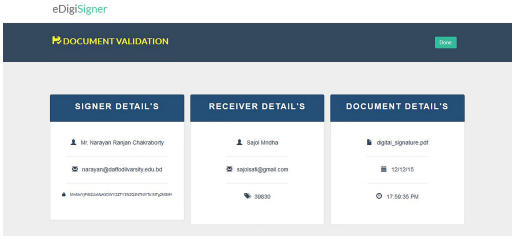


Fig. 11. Document validation page

V. Conclusion

If a smartcard or a PIN is used to identify an individual then anyone can involve in signature generation or verification. Moreover there are a number of websites are still ongoing that provide fake ID including US, Canada, Bangladesh etc. Our main approach is to establish an improved digital signature scheme that satisfy the current business needs and fulfill the user satisfaction as well as obtain a scope for the future expansion within the organizational constraints. Although biometric technology is quite expensive, it can provide increased level of security in the field of cryptography and digital signature.

VI. Limitations

Some restrictions of our system are listed below, 1) Since our system provides a cloud based digital signature platform, internet connection is required to adopt all of the services. 2) Required additional hardware such as a vein reader to detect vein pattern images. But this could be resolved if we could use smart phones to detect vein. 3) The Biometric vein technology is quite expensive

VII. Future work

Our system provides a confidential interaction between a signer and a receiver, only specified receiver can validate the document or message. If we send encrypted document to the

receiver and share a Key (such as OTK) between them to decrypt the message then the data confidentiality also can be preserved. Additionally nowadays many mobile devices are coming up with the fingerprint image detection feature, in near future they will eventually meet the finger vein detection and our proposed system can be implemented with the mobile device integration.

REFERENCES

[1] Xiaodong Liu, Quan Miao and Daxing Li, A New Special Biometric Identity Based Signature Scheme, International Journal of Security and its Applications, vol. 2, no. 1, Jan. 2008.
[2] Ahmed B. Elmadani, Trusted Document Signing based on use of biometric (Face) keys, International Journal of Cyber-Security and Digital Forensics, vol. 4, no. 1, pp. 289-296, 2012.
[3] Successful development of biometric digital signature technology, available at http://www.hitachi.com/New/cnews/130218.htmlLast accessed on 15-07-2016 at 7:09pm
[4] Sambangi Eswara Rao and S.Ravi Kumar, Novel Biometric Digital Signature System for Electronic Commerce Applications Using Java, International Journal & Magazine of Engineering, Technology, Management and Research, vol. 1, no. 10, pp. 287-293, Oct. 2014.
[5] A.M. Al-Khouri and J. Bal, Digital Identities and Promise of the Technology Trio: PKI, Smart Cards, and Biometrics, Journal of Computer Science, vol. 3, no. 5, pp. 361-367, 2007.
[6] Wojciech Kinastowski, Digital Signature as a Cloud-based Service, The Fourth International Conference on Cloud Computing, GRIDs, and Virtualization, 2013.
[7] AyshaAlbarqi, EtharAlzaid, Fatimah Al Ghamdi, SomayaAsiri and JayaprakashKar, Public Key Infrastructure: A Survey, Journal of Information Security, vol. 6, pp. 31-37, Jan. 2015.
[8] Rachana C.R., The Role of Digital Signatures in Digital Information Management, International Monthly Refereed Journal of Research in Management & Technology, vol. 2, pp. 103-109, Mar. 2013.
[9] Arulalan.V, Balamurugan.G and Premanand.V, A Survey on Biometric Recognition Techniques, International Journal of Advanced Research in Computer and Communication Engineering, vol. 3, pp. 5708-5711, Feb. 2014.
[10] Rupinder Saini and NarinderRana, Comparison of Various Biometric Methods, International Journal of Advances in Science and Technology, vol. 2, pp. 24-30, Mar. 2014.
[11] Dr. Rajinder Singh and Shakti Kumar, Comparison of Various Biometric Methods, International Journal of Emerging Technologies in Computational and Applied Science, pp. 256-261, Feb. 2014.
[12] Learn about Barclays brings finger vein biometrics to internet banking, available at http://www.wired.co.uk/news/archive/2014-09/05/barclays-finger-scanner. Last accessed on 17-07-2016 at 11:47pm.
[13] Learn about M2-FingerVein Non-invasive finger vein reader available at http://www.m2sys.com/finger-vein-reader/. Last accessed on 18-07-2016 at 03:00pm
[14] Carl Ellison and Bruce Schneier (2000), Ten Risks of PKI: What Youre not Being Told about Public Key Infrastructure, Computer security journal, vol. xiv, pp. 1-8, 2000.
[15] Behrouz A. Forouzan, Data Communications and Networking, 4th ed., New York: McGraw-Hill, 2007.
[16] Forum of European Supervisory Authorities for Electronic Signatures (FESA), Public Statement on Server Based Signature Services, available at http://www.fesa.eu/public-documents/PublicStatementServerBasedSignatureServices-20051027.pdf.
[17] Learn about Cloud Signing - Multiple Signing in Options available at http://www.ascertia.com/Solutions/ByTechnology/cloud-signing
[18] Secure CoSign Digital Signature Use via One-Time-Password (OTP) Authentication available at http://www.arx.com/files/documents/cosign-digital-signatures-and-otp.pdf

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

An Efficient Approach of Identifying Tourist by Call Detail Record Analysis

Ratul Sikder and Md. Jamal Uddin
Department of Computer Science and Engineering
Bangabandhu Sheikh Mujibur Rahman Science
and Technology University, Bangladesh
(ratulsikder121, jamal.bsmrstu)@gmail.com

Sajal Halder
Department of Computer Science and Engineering
Jagannath University, Bangladesh
sajal@cse.jnu.ac.bd

***Abstract*—Tourist identification with a lower effort can be highly demandable for the tourism department and related organizations of a country. Nowadays, most of the people including tourists use cell phones to keep communication resulting in corresponding data for every transaction (call, SMS, MMS, mobile data) named call detail record. This kind of data is usually collected and stored by telecom operators mainly for billing reasons. Call detail record (CDR) can be mined to get the approximate location of cellular mobile phones. This article describes a framework that identifies tourists among total population by analyzing cellular phone location data from call detail record. The framework also includes an efficient yet effective data scan method from huge CDR database of the total cell phone users.**

***Index Terms*—Tourist Identification; CDR Analysis; Mobile Tracking; Call Detail Record**

I. Introduction

Call Detail Record(CDR) is the relative information about the corresponding cellular transaction and the transactions are SMS, MMS, phone calls, data communications. A CDR database contains cellular phones important information including the approximate location of the source and destination of every cellular transaction [1]. When a cell phone user tries to communicate with the cell phone then a corresponding CDR is created and stored in CDR database [2]. There is various information related to cellular network activity stored in CDR. CDR databases are the massive source of data repository for various research fields. CDR database can be analyzed with numerous data mining techniques to discover hidden knowledge about human mobility, behavior, usage pattern etc [3].

Almost every country maintains a tourism department to help tourist, increase economy related to tourism, give support to non-government tourism organizations and companies etc. Many of these functions directly or indirectly depend on the successful identification of tourists. Tracking is a good way of identifying tourists based on the general behaviors of them. We can track cell phone users from location information of CDR data. As CDR is a passive source of data, gathered by telecom operator, no direct cell phone users' intervention is required [4]. Though the location information from CDR is not highly accurate but it will not affect our tourist detection framework because we use some threshold to identify tourists. Threshold values can be altered to see the difference.

Global Positioning System (GPS), on the other hand, can serve the purpose of cell phones location tracking more accurately [5]. There are many smartphones that have A-GPS(Assisted GPS) device built in. These A-GPS devices can get the current location, almost as like as GPS devices. Using A-GPS enabled mobile device, we can continuously track users' trajectory behavior. But there are some notable limitations of using A-GPS as a mobile tracking device. Firstly, a lot of mobile phones do not have A-GPS device. Secondly, mobile A-GPS device consumes a lot of power which is not acceptable for mobile phone [6]. Thirdly, the A-GPS device is kept off in maximum A-GPS enabled mobile devices by the users. So, it is not a wise decision to track cell phone users and find tourists among them by A-GPS tracking.

Our main goal is to detect tourists among all cell phone users efficiently by using CDR location data. In this paper, we study the problem of CDR database as a source of cellular phones location data. And we study the general behavior of tourist's trajectory with respect to some threshold values. Then we will address an efficient CDR database scan method from our previous study to reduce continuous scan. Then we develop a framework to detect tourists on the basis of the general tourist behavior. This framework will further identify tourist by analyzing CDR location data. Our main contributions can be summarized as follows:

- We first introduce a CDR database scan method to reduce continuous scan and high resource use, specially designed for tourist detection. This method overcomes the limitations of CDR location tracking for this case.
- Then we propose a method to detect tourists among the total population by analyzing the CDR location data of cellular phones.

The remaining sections of the paper are organized as follows. In section II describes related work on tourist identification by using CDR data. Section III introduces our proposed method for identifying tourists by CDR analysis efficiently. Implementation of our proposed technique has been described in section IV. Experimental analysis has been shown in section

978-1-5090-5769-6/16/$31.00

V. Application of our system, discussion as well as the future research scopes and limitations are shown in section VI.

II. Related Work

Call detail record can be analyzed to find hidden relationships of human mobility in various sectors[3] [4]. Urban and country dynamics like population deployment, popular areas, the connection among different geographical zones etc. can be mined from mobile phone location data in call detail record and some effective relationships can be found [7]. In this research paper [7] the authors have successfully used CDR location data for their trajectory purposes. Some researchers also worked on CDR database to find unusual crowd events [8]. CDR location data can be used in various data mining techniques. In research paper [9] traffic zones have been divided using K-means clustering algorithm. Real-time road traffic monitoring is also possible by using latest cellular networking technologies but these are costly [10]. So, it is clear that CDR location data can be an interesting data repository for numerous mobility analysis.

There are some works on tourist behavior analysis by using CDR data. In research paper [11], tourist's spatial and temporal behavior is unveiled in urban places. The researchers used GPS technology for their work. In [12], tourists' movements and flow patterns are characterized. This type of tracking is also done by GPS technology. To understand tourists' behavior on a large scale some frameworks have proposed [5] which track the tourists by GPS device.

There are some highly related works to us on tourists. But almost all of them are based on GPS/A-GPS tracking. In section I, we already studied the problems of GPS/A-GPS tracking. On the other hand, some work on CDR database [3] [4] can mine general human mobility. Some CDR based works also tried to characterize tourists' behavior. But there are a few works on tourist identification by passive location data.

Our proposed methods for tourist identification use CDR as passive location data. We will get rid of the problems of using GPS/A-GPS tracking as we will use CDR location data. The problem of using CDR data is that it produces and stored only when there is a cellular transaction like call, SMS, MMS [2]. This is one of the main problems of CDR data. So, with CDR location data, continuous as well as instant tracking is not possible. But an analysis on tourists' behavior [13] shows that tourists use their cell phone on a tour. A big percentage use cell phones frequently on a tour, on the other hand, there is also a big number who do not want to use the cell phone on a tour. But all of them use the cell phone at least a minimum level for important communication. We designed our framework to overcome this limitation.

III. Proposed Method

Our work on efficient tourist detection can be divided into two parts. They are:

- Efficient scan of CDR database.
- Tourist detection by using CDR location data.

TABLE I
Sample of CDR Attributes

Attribute	Content
network_plmn	The international operator code of the physical network
network_msc	An identifier for the Mobile Service Center servicing the call
network_bts	An identifier of the Base Transceiver Station
network_lattitude	The expected or real customer latitude at the beginning of the call
network_longitude	The expected or real customer longitude at the beginning of the call

A. Efficient scan of CDR database

Efficient scan of CDR database means nothing but a time interval scan of the database for cellular phones location. In our work, we mainly target the tourists who do not travel very short distance. And usually, maximum tourist need to travel some non-short distance to reach their destination. There is some information regarding the location in a CDR data. We have used an online CDR generator[14] to generate CDR dataset and analyze its attributes. Table I shows some attributes related to location information of CDR data.

We will scan the CDR database with a time interval and we will try to get the last available location data of each user. This interval should be something that wont decrease the probability of detecting tourist gradually as well as will not give high resource pressure on CDR database. Figure 1 depicts a 4 hour time interval CDR database scanning.

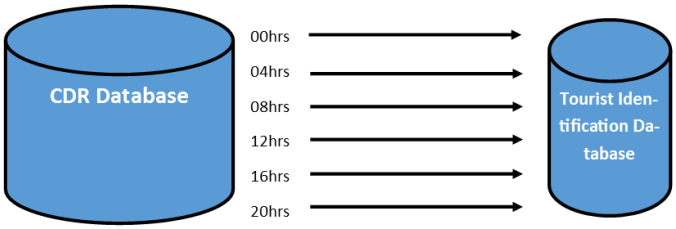


Fig. 1. Time interval scan of CDR database

B. Tourist identification by using CDR location data

1) Preliminaries: Tourist identification means to detect tourist based on their location data analysis. We will propose a set of rules to identify tourists among the total population. This will involve some general knowledge on tourists behavior. Efficient scanning of CDR database will give us a table of cell phone users approximate latitude and longitude. We will then develop appropriate algorithms based on our set of rules. These algorithms will be implemented on the data like table II. In turn, the algorithms will return a probability of a cell phone user being tourist or not.

2) Procedure: We have discussed scanning and gathering CDR location data from CDR database from the previous section. We will collect the network latitude and network longitude with the corresponding unique identity for every scanning phase. Then we will insert these into our tourist

TABLE II
TOURIST IDENTIFICATION TABLE

cellular_id	Ti	loc_00	loc_04	loc_08	loc_12	loc_16	loc_20
5017	82	23.005639 89.830058	23.005639 89.830058	22.963011 89.815795	22.832525 89.525229	22.832525 89.525229	22.855734 89.540696
5018	85	23.153573 89.924615	23.153573 89.924615	23.425575 90.166921	23.721791 90.482932	23.343036 91.142701	22.510990 91.799645
5019	73	23.960504 90.183234	23.960504 90.183234	23.960504 90.183234	24.116404 91.119009	24.922957 91.840274	24.922957 91.840274
5020	8	24.768432 90.369329	24.768432 90.369329	24.768432 90.369329	24.738136 90.394406	24.769502 90.367718	24.769502 90.367718

detection table which will remind the last few location histories. An example of our tourist identification database is given below:

Here, $cellular\_id$ is the derived identity from the actual cellular number. loc_00 means location information collected at 00 hrs of the day. loc_04 means location at 4 hrs and so on for loc_08, loc_12, loc_16, loc_20. As our CDR database scan interval is 4 hours, so we took sample after every 4 hours of a day starting from 12am. Here, T_i is an important integer variable, stands for tourist indicator. This tourist indicator variable Ti shows the probability of a cell phone user to be a tourist for that period. The value of Ti ranges between 0-100. The lower value means there is a lower chance of the users being tourist and the higher value means the opposite. So the summary is:

- Call Detail Records(CDR) of all cell phone users will be scanned on a regular basis(ex. 8am, 12pm, 4pm) for the current location of the subscribers.
- There will be a set of rules R_t which will responsible for triggering the tourist indicator T_i.
- Tourist indicator T_i is a variable, range between 0-100 for each cell phone user. T_i=100 means all rules indicate that the subscriber of the phone is tourist. T_i=0 means the opposite.

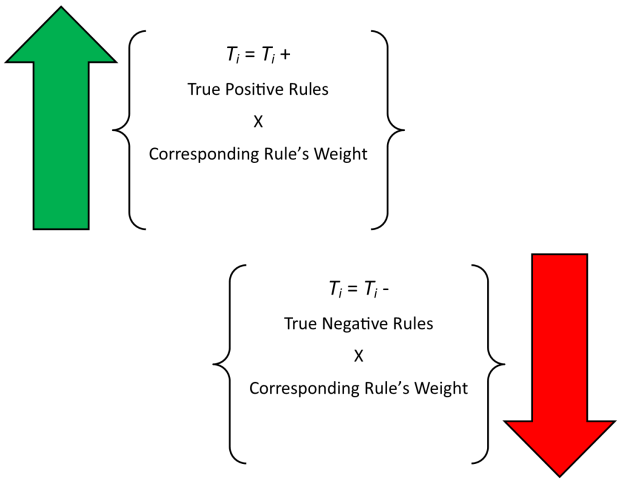


Fig. 2. T_i manipulation concept

So, the set of rules R_t plays an important role in the tourist detection. The algorithm depends hugely on these rules. Moreover, the success or failure to detect tourists depends on these rules. The concept of change in T_i is shown in figure 2.

3) Set of Rules: Rt: We divided the set of rules into two subsets. One subset is about the rules for which a cell phone user can be identified as a tourist. And the other subset includes the rules for which a cell phone user can be identified as not-tourist.

$NegativeRules(NegativeRt)$: A cell phone user should not be considered as tourist if

1) The phone's location is not changing.
2) The phone's location is changing among some neighbors small areas.
3) The phone's location is changing periodically/semi-periodically among some regions.
4) The average duration of staying some neighbors region for a long time.(ex. more than 7 days)
5) The phone is not belonging to a defined tourist place.

$PositiveRules(PositiveRt)$: A cell phone user should be considered as tourist if

1) The phone's location is changing rapidly towards some predefined tourist place.
2) The phone is not belonging to its home locations.
3) The phone is belonging to a defined tourist place.

The changes in T_i is depicted in the flowchart in figure 3. Note that, these set of rules is the key to success in our work and these rules might not enough for some condition. So, improving, modifying and adding these rules is a routine and future work.

IV. IMPLEMENTATION

A. Efficient Scan of CDR Database

To track each user's location, we need to scan the whole CDR database continuously. But continuous scanning on a large database is non-efficient. But if we scan the database after some time interval, then we can reduce this problem to a great extent. Usually, tourists travel time is greater than 8-12 hours. So, tourist detection with our methods does not require continuous scanning. A good time interval, such as 4-6 hours is enough for our tourist detection method. In this paper, we use 4-hour time interval among CDR database scanning. There is a problem regarding CDR data. This data is generated only when there is a cellular transaction (call, SMS, MMS, data connection etc.) by the mobile device. So, when we scan the CDR database, we may not find the current location

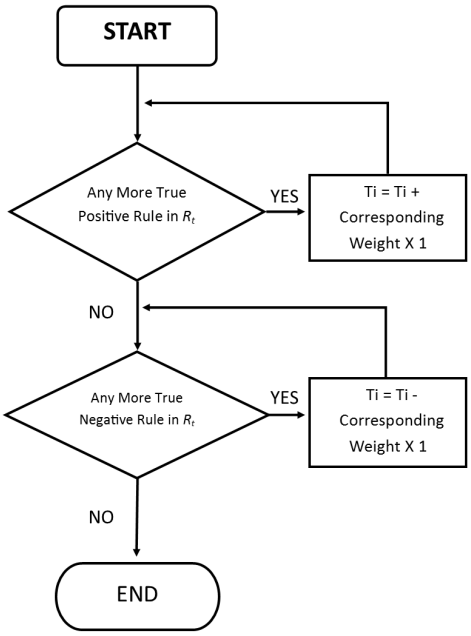


Fig. 3. Flowchart for T_i on the basis of R_t

data of the users. Because there is a good probability that the cell phone users did not receive or make any types of cellular transaction for a long time. We will try to address this problem in our tourist detection method. The probability of a tourist of not making or receiving any cellular transaction during the total travel period is very low [13]. So, we can easily expect at least some cellular transaction during the travel period. Our algorithm for tourist detection can detect tourists with such minimum available CDR location data. So finally, our efficient scan of CDR database means scanning the whole CDR database after a time interval and fetch location data with respect to corresponding cellular phone identity. The CDR data, on the other hand, is very private. So, for maintaining a high level of security and privacy, we will collect the location data against a derived attribute from the actual cellular phone identity. We will not use the real phone number, rather we will use some non-reversible methods to derive another corresponding users identity and then using those.

B. Tourist Identification by Using CDR Location Data

1) Preliminaries: Now, we have the set of rules R_t to maintain tourist indicator variable T_i. With respect to R_t, we have to develop an efficient algorithm to handle the tourist indicator variable T_i efficiently. This means to increase T_i when one or some rules from $PositiveRt$ and decrease when one or some rules from $NegativeRt$ becomes true. Each rule will have a separate section in the total algorithm. Moreover, each rule should have its own weight because some rules are important decision maker than other in the field of tourist detection. The algorithm should also aware of the fact of overrating or underrating. And finally, the value of T_i for each cell phone user should be between 0-100. The inputs of the algorithm are previous T_i value and the available location values corresponding to a cellular ID. The input location data is given to the algorithm in sorted order, beginning from the current time-stamp to earlier timestamps. Here, loc0 is the most recent location data and $loc1$, $loc2$, $loc3$ etc. are gradually older. $locH$ is the calculated home location of the cell phone user. The output is the new T_i value.

Algorithm 1 Tourist Identification

1: **procedure** TOURISTINDICATOR (T_i, $locH$, $loc0$, $loc1$, $loc2$, $loc3$, $loc4$, $loc5$)
2: Assign $FALSE$ to all the **booleans** corresponding to every rule
3: **if** Distance ($loc0, loc1$) < $MinTravelDistance$ **then** NegativeRt1 = $TRUE$
4: **end if**
5: **if** Distance (Region($loc0$), Region($loc1$)) < $MaxNeighbourDistance$ **then** NegativeRt2 = $TRUE$
6: **end if**
7: **if** Region($loc0$) equals to (Region($loc1$) OR Region($loc2$) OR...OR Region($loc5$)) **then** NegativeRt3 = $TRUE$
8: **end if**
9: **if** Region($loc0$) does not belong to T_z **then** NegativeRt5 = $TRUE$
10: **end if**
11: **if** (Direction($loc0, loc5$) indicates one or more T_z and Distance($loc0, loc5$) > $MinTravelDistance$ **then** PositiveRt1 = $TRUE$
12: **end if**
13: **if** (Region($loc0$) is not Region($locH$)) **then** PositiveRt2 = $TRUE$
14: **end if**
15: **if** Region($loc0$) belongs to T_z **then** PositiveRt3 = $TRUE$
16: **end if**
17: **for** `<each TRUE rule in PositiveRt>` **do**
18: `<calculate weighted values and add with the` T_i `value>`
19: **end for**
20: **for** `<each TRUE rule in NegativeRt>` **do**
21: `<calculate weighted values and subtract from the` T_i `value>`
22: **end for**
23: **if** T_i < 0 **then** T_i = 0
24: **end if**
25: **if** T_i > 100 **then** T_i = 100
26: **end if**
27: Return T_i
28: **end procedure**

2) Algorithms: Algorithm 1 is the main algorithm to maintain T_i. Distance($location, location$) is a simple distance measuring algorithm from two geometric coordinate values. Region($location$) is a function that returns the area code of the corresponding location. All area symbolize by corresponding

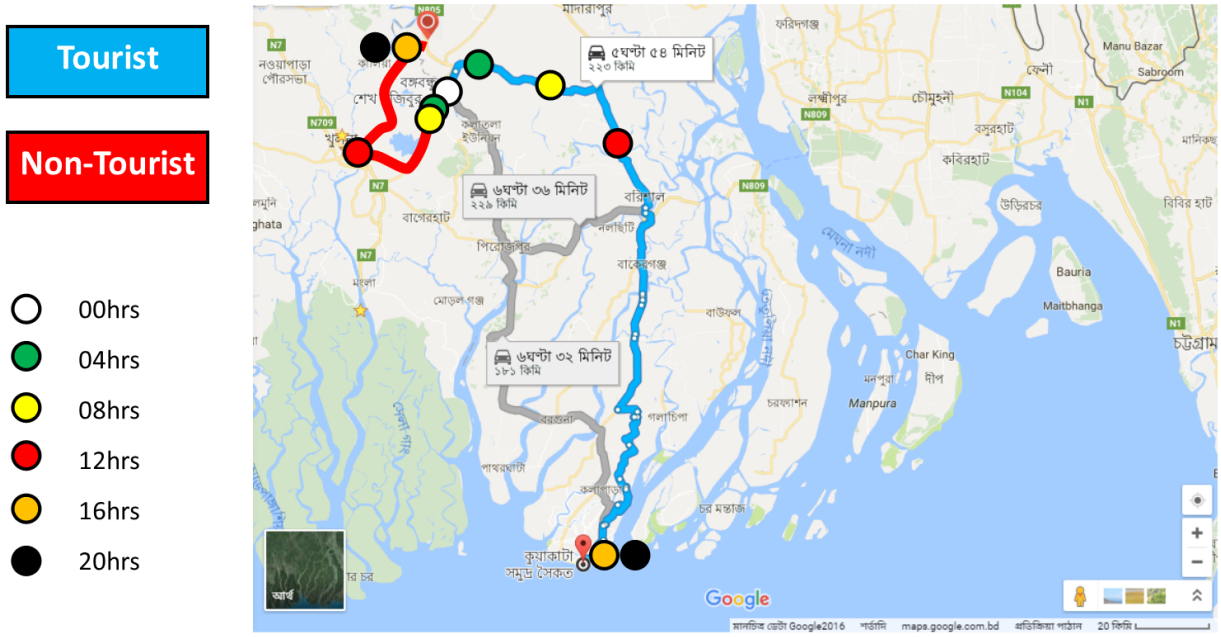


Fig. 4. Trajectories of tourist and non-tourist. The blue path is traveled by a tourist and the red path is traveled by a non-tourist. Various colored small circles indicate the location data scanned at the corresponding time of a day. Google Map is used to depict this. ©2015 Google Inc, used with permission. Google and the Google logo are registered trademarks of Google Inc.

area code forms the set A_z. Each area code must have a defined coordinate value. T_z is the set of tourist zones area code. T_z is a subset of A_z. Direction($location0, location1$) is a complex algorithm that returns a short list of T_z which is in the running direction from $location1$ to $location0$. The initial values of all T_i are zero. The home location $locH$ is calculated over a long period for measuring cell phone users home location.

V. Experimental Analysis

Figure 4 shows two trajectory paths, starting from the white small circle. The blue path is imagined as a tourist traveling path and the red path is imagined as a non-tourist traveling path. The end of the blue path is a predefined tourist zone and the end of the red path is not. The small various colored circles are the location data scanned from CDR database at the corresponding time of a day. For testing the effectiveness of our algorithms, especially the set of rules R_t, we will now analysis these trajectory data and see whether our algorithm can detect tourist or not.

From figure 4 we see that, both locations of 00hrs(12am) is same. At 04hrs(4am), we see that the blue colored traveler travels comparatively long distance than the red one. So, our $MinTravelDistance$ threshold will deny red colored traveler as tourist and opposite for blue one. The very next transaction will also perform similarly. The next orange colored circle will increase T_i for both of them but for previous values, the aggregate value of T_i will remain high. The next yellow colored transaction will further trigger other rules from R_t. For example, now the blue colored traveler travels a high distance and belonging to a defined tourist zone. But red colored traveler does not belong to a tourist zone. So, negative rules 2 and 5 will decrease R_t a lot of red ones and positive rules 1, 2 and 3 will increase R_t massively of blue ones. Now we can say that our algorithm worked successfully for detecting a cell phone user as tourist or not. It should be noted that the weight of some rules are much higher than others. For example, negative rule 5, positive rules 2, 3 have higher weight compared to others.

VI. Discussion and Conclusion

Our proposed framework can identify tourists by using CDR data, a passive source of cell phones location information. Our method to scan huge CDR database is also efficient in terms of resource usage without sacrificing the quality of our main goal. Our work can be useful to them who are directly or indirectly related to the tourism sector. Government tourism department can utilize our methods to enable SMS based system to help tourists, spread important information about the corresponding tourist zones among tourists, give emergency informations to them etc. Moreover, security agencies like tourist police, coast guard etc. can use this method to identify tourists and take proper steps to give them security. This work can be also useful to non-government tourist organizations and companies.

Scanning data from CDR database is not allowed in normal phenomenon. But our proposed method will create a non-reversible identity for cell phone users. So, direct intervention of tourism department of the government can solve the security issue. The set of rules to detect tourist is sufficient but these can be improved a lot to detect tourist more efficiently. Other algorithms stated in our work are already available in open source form on the internet. These algorithms can be further

specialized for this work to get improved result and/or minimize resource usage. It is very important to declare the weights of each rule correctly. This can be further a research scope. Our methods are proposed for detecting long traveled tourists. So, the threshold values should be something with respect to this phenomenon. Finally, the tourism department of a country can utilize our CDR based method to identify tourists. Successful identification may lead to further achievements like assessing important variables regarding tourism. Moreover, tourist detection history can be a big source of data for further research. We will try to improve our detection algorithms in future. We would be glad if someone finds this work on tourism helpful.

ACKNOWLEDGMENT

The authors are grateful to the anonymous reviewers for their comments that improved the quality of our paper. This research was supported by the research fund of Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Bangladesh. Md. Jamal Uddin is the corresponding author.

REFERENCES

[1] J. Steenbruggen, E. Tranos, and P. Nijkamp, "Data from mobile phone operators: A tool for smarter cities?" *Telecommunications Policy*, vol. 39, no. 3, pp. 335–346, 2015.

[2] R. A. Becker, R. Caceres, K. Hanson, J. M. Loh, S. Urbanek, A. Varshavsky, and C. Volinsky, "A tale of one city: Using cellular network data for urban planning," *IEEE Pervasive Computing*, vol. 10, no. 4, pp. 18–26, 2011.

[3] R. Becker, R. Cáceres, K. Hanson, S. Isaacman, J. M. Loh, M. Martonosi, J. Rowland, S. Urbanek, A. Varshavsky, and C. Volinsky, "Human mobility characterization from cellular network data," *Communications of the ACM*, vol. 56, no. 1, pp. 74–82, 2013.

[4] Y. Xu, S.-L. Shaw, Z. Zhao, L. Yin, Z. Fang, and Q. Li, "Understanding aggregate human mobility patterns using passive mobile phone location data: a home-based approach," *Transportation*, vol. 42, no. 4, pp. 625–646, 2015.

[5] S. Phithakkitnukoon, T. Horanont, A. Witayangkurn, R. Siri, Y. Sekimoto, and R. Shibasaki, "Understanding tourist behavior using large-scale mobile sensing approach: A case study of mobile phone users in japan," *Pervasive and Mobile Computing*, vol. 18, pp. 18–39, 2015.

[6] A. Carroll and G. Heiser, "An analysis of power consumption in a smartphone." in *USENIX annual technical conference*, vol. 14. Boston, MA, 2010.

[7] R. Trasarti, A.-M. Olteanu-Raimond, M. Nanni, T. Couronné, B. Furletti, F. Giannotti, Z. Smoreda, and C. Ziemlicki, "Discovering urban and country dynamics from mobile phone data with spatial correlation patterns," *Telecommunications Policy*, vol. 39, no. 3, pp. 347–362, 2015.

[8] Y. Dong, F. Pinelli, Y. Gkoufas, Z. Nabi, F. Calabrese, and N. V. Chawla, "Inferring unusual crowd events from mobile phone call detail records," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2015, pp. 474–492.

[9] H. Dong, M. Wu, X. Ding, L. Chu, L. Jia, Y. Qin, and X. Zhou, "Traffic zone division based on big data from mobile phone base stations," *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 278–291, 2015.

[10] A. Janecek, D. Valerio, K. A. Hummel, F. Ricciato, and H. Hlavacs, "The cellular network as a sensor: From mobile phone data to real-time road traffic monitoring," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 5, pp. 2551–2572, 2015.

[11] L. Kellner and R. Egger, "Tracking tourist spatial-temporal behavior in urban places, a methodological overview and gps case study," in *Information and Communication Technologies in Tourism 2016*. Springer, 2016, pp. 481–494.

[12] M. A. M. Toha and H. N. Ismail, "A heritage tourism and tourist flow pattern: A perspective on traditional versus modern technologies in tracking the tourists," *International Journal of Built Environment and Sustainability*, vol. 2, no. 2, 2015.

[13] J. E. Dickinson, J. F. Hibbert, and V. Filimonau, "Mobile technology and the tourist experience:(dis) connection at the campsite," *Tourism Management*, vol. 57, pp. 193–201, 2016.

[14] Administrator. Call detail record generator. [Online]. Available: http://www.gedis-studio.com/online-call-detail-records-cdr-generator.html

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

Enhancing Performance of Naïve Bayes in Text Classification by Introducing an Extra Weight using less Number of Training Examples

Shahnaj Parvin Shathi[1], Md. Delowar Hossain[1], Md. Nadim[1], Sayed Golam Rasul Riayadh[1], Tangina Sultana[2*]
Department of Computer Science and Engineering, Department of Electronics and Communication Engineering
Hajee Mohammad Danesh Science and Technology University, Bangladesh.
E-mail: {shahnajshathi14, delowar.cit, mnadims.cse, riaydh06, dristy.043}@gmail.com

***Abstract*— This paper presents an effective and efficient method for classifying text documents in order to deliver feasible information retrieval using naïve bayes algorithm. Today lots of algorithms have earned good score in the field of information retrieval, Naïve Bayes is one of them. In this paper, a Weight Matrix is introduced during training text documents which is combination of term frequency (TF) and inverse class frequency(ICF) and later this weighted term is powered by a significant number and added with the posteriori value during the prediction time of Naïve Bayes (NB) algorithm to establish a better and efficient performance of the classification task. Here the precedence base element TF results an additional weight for each term (word) of the text. On the other hand, ICF gives each common word a low score. Finally the combinational term 'Weight Matrix' gives an extra weight and balances weight where necessary. As a result, improve the performance accuracy of the NB classifier. Experimental results show that NB with Weight Matrix rarely demotes accuracy compared to standard Naïve Bayes, instead of enhancing accuracy dramatically.**

Keywords—*Feature selection; ICF; Naïve Bayes*; *Text categorization;* TF*;Weight Matrix*

I. INTRODUCTION

Text is the most common vehicle for the formal exchange of information. Nowadays a huge amount of information is being associated with the web technology and the internet. As the information of internet and improvement of digital articles are increasing day by day, people urgently need effective and efficient tool to automatically classify the information, retrieve essential information from unstructured text and assign them in a predefined category. To collect necessary information from these raw texts, significant features have to be extracted from raw texts. The motive to classify a given data instance into pre-defined set of categories is known as "Text Categorization". It includes query, colander and store the huge amount of resources.

NB is a probability based supervised learning algorithm. It is famous for its conceptual and computational simplicity. In our paper, we study how the accuracy of the standard NB can be increased significantly for English text documents. A term namely 'Weight Matrix' is introduced for the accuracy enhancement procedure. It needs to be noted that minimum text documents are used for training. NLTK [1] standard library is used for preprocessing raw texts. Preprocessing training documents includes some sequential steps: tokenization, stop words removal, digit removal, punctuation removal, converting all words in lowercase presentation and stemming. After preprocessing a well-structured text is found and feature extraction is performed on this structured text.

Finally, separate learned classes are constructed from this structured training data. In our experiment, we apply Standard NB and our proposed Weighted NB successively for training and testing. We collect dataset from bbc [2] and among the huge dataset we choose 100 documents randomly for the research work. Then divide the chosen documents equally, half is used for training and half is used for testing. Considering total chosen data sets, 2-fold validation is done to measure the performance accuracy and we get around 83% accuracy from Weighted NB,considering minimum text documents for training. For 2-fold validation check we divide the total chosen documents into two groups, 1st half is used for training and 2nd half is used for testing and vice versa.

In this paper section II describes the related works in text classification field. The proposed system is stated in section III. Section IV evaluates the detailed methodology of the system. On the other hand, experimental setup is assigned in section V and finally conclusions and future plans are imputed in section VI.

II. RELATED WORKS

Traditionally two major methods are used for Text Categorization (TC): one of them is rule based approach and another is machine learning approach. In rule based method classification rules are generated by experts. This method is accurate but not cost effective. On the other side, in machine learning approach the grouping methods are created automatically with the help of some statistical algorithm. This approach is cost effective and a new domain of system is easy to construct. So this automated technique is most famous which classify texts into their predefined categories based on the contents they have [3, 4]. A lots of training algorithms for TC have been developed in past few years, these are: probability based algorithm Naïve Bayesian method, Decision tree learning algorithms [5], K-Nearest Neighbors [6], Support Vector Machine [7], etc.

Among those automated classification approaches, NB is one of the most commonly used algorithm. It calculates the

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

probability value of the contents belonging to a particular class. In [8], A. S. Patil & B. Pawar used Simple NB for automatic text categorization and got approximately 80% accuracy. Deep feature weighting (DFW) for Naïve Bayes was introduced in [9] where DFW estimates the conditional probabilities of NB by deeply computing feature weighted frequencies from training data. Furthermore, various smoothing methods are mentioned in [10] for increasing NB learning and applied it for the classification of short text. Although a number of approaches have been proposed, they are not faultless and still needs improvements.

III. Proposed System

We propose an enhanced system that will improve the performance efficiency of naïve Bayes algorithm in the text classification task. Here we introduce a term namely 'Weight Matrix' which is combination of TF and ICF and later this Weight Matrix value is powered by a significant number that is retrieved from any new instance. We used ICF instead of IDF(Inverse Document Frequency) because in Centroid-based text classification ICF is more efficient than IDF and improves classification accuracy[11]. The term TF specifies the importance of a word to a document in a collection or corpus. It works as a weighting factor for the system. On the other side, the ICF defines if a word is common or rare across the class. It works as a 'balancing factor', which is used to balance the values of the Weight Matrix. In testing, when we take a new instance to determine its class, the posteriori probability value is calculated for each class. After that the introduced powered Weight Matrix value is added to the posteriori probability value for each class. The maximum weighted posteriori probability value will be the result (class) of the prediction process.

IV. Methodology

NB is a multinomial supervised learning method which is based on probabilistic measurement applying Bayes theorem. For a supervised learning setting, probabilistic model supported NB classifiers can be trained very efficiently. In NB two types of probability functions are calculated for training, one is prior probability and another is conditional probability. This calculated probability values gained in training are then used in the testing process where we predict the probability for a new instance to be in a particular class.

Probability of a individual class P(C_n) = Number of documents for a individual class / Total number of documents. Here, P(C_n) is the prior probability.Likelihood function of an individual given class is calculated by (1), this is also called conditional probability.

$$P(W_n|C_n) = [count(W_n|C_n) + 1] / [count(C_n) + |V_n|] \quad (1)$$

Here, count($W_n|C_n$) is the count of each individual word according to that class(C_n); count(C_n) is the count of all words in the original class ;V_n is the number of vocabulary.

The frequencies of each isolated informative words of each particular class are f_1 , f_2 ,……, f_n .Now, Posteriori probability for each class,

$$P(C_n|W_n) = \underset{C_n \in C}{\operatorname{argmax}} P(C_n) * P(W_n|C_n)^{f_n} \quad (2)$$

According to Naïve Bayes the new instance will be classified to that category for which value of the posteriori probability will maximum.

Our idea is to power the weight value of each individual matched word by the frequency of a new instance and then add the powered Weight Matrix value with the posteriori probability of Naïve Bayes to improve the efficiency of the classification task. This Weight Matrix is the combination of term frequency $TF(t, d_1)$ and inverse class frequency ICF_{tc}. $TF(t, d_1)$ adds high discriminating power to specific documents whereas ICF_{tc} offers a less pronounced relevance signals. If the total number of classes is N and number of classes with term t_c in it is N_c , then ICF_{t_c} is mapped as follows,

$$ICF_{t_c} = \begin{cases} \ln \frac{N}{N_c+1}, & if N_b = 0 \\ \ln \frac{N}{N_c}, & if N_b > 0 \end{cases} \quad (3)$$

If we have n documents $(d_1, d_2, d_3, \ldots\ldots, d_n)$ then TF of those documents,

$$TF(t, d_1) = \frac{Number\ of\ times\ term\ t\ appears\ in\ document\ d_1}{Total\ number\ of\ terms\ in\ d_1}$$

$$TF(t, d_2) = \frac{Number\ of\ times\ term\ t\ appears\ in\ document\ d_2}{Total\ number\ of\ terms\ in\ d_2}$$

and so on. If a class have n number of documents then the TF for each class is

$$TF(t, d_c) = TF(t, d_1) + TF(t, d_2) + \ldots..+TF(t, d_n) \quad (4)$$

From (3), and (4), we get the Weight Matrix for each class calculated as below:

$W_{t,d_c} = TF(t, d_c) * ICF_{t_c}$. After this completion of training stage when a new text instance is taken for testing then the Weight Matrix values are powered and updated for each class with respect to the matched feature frequency of the new text instance. So, the powered weight(W) which is calculated with respect to each class is,

$W = (W_{t,d_c})^f$, here f is the frequency of matched words in the new instance. So the final probability value P(x) for each class of a new text instance x is calculated by (5) is,

$$P(x) = \operatorname{argmax} [P(C_n|W_n) + W] \quad (5)$$

That's why, for the Weighted NB the new instance will be classified to that category at which the value of P(x) will maximum.

V. Experimental Setup

In this section, we discuss the entire experimental setup of our research. To start with we need collection of English text documents which is pre-classified into different categories. We choose bbc news articles as a standard for our experiment, then the documents are preprocessed (tokenization, stop words removal, digit removal, punctuation removal, converting all words in lowercase presentation and stemming) and finally the dataset is subjected to training and testing the classifier. We discussed the steps briefly as given bellow:

A. Creation of Dataset

Our data set is the collection of documents with their pre-defined categories mentioned in Table I. The dataset consisted of 2225 documents with five different categories- Sports, Business, Entertainment, politics & technology.

TABLE I. Total Dataset

Topics	No. of documents
Sports	511
Business	510
Entertainment	386
Politics	417
Technology	401

Among this large dataset we choose 100 documents randomly and half of the documents are used for training and the rest is used for testing.

B. Preprocessing Text Documents

Before performing feature extraction preprocessing is performed to ignore the unimportant and irrelevant words, reducing dimension and to turn the text in a standard format.

1) *Tokenization:* Break a group of text into words, phrases, or other meaningful elements which is called tokens. Here we separate each word by whitespace.
2) *Stop words removal:* Python library NLTK [1] supports a list of stopwords which don't aggregate useful or relevant information for the classification task. That's why unimportant words are removed.
3) *Digit removal:* Any kind of digits are removed by using there Unicode representation.
4) *Punctuation removal:* Remove punctuation marks as well as special symbols(', ", :, ;, [ ], { }, ( ), _, -, /, \,|, ?, <>, #, & etc.)
5) *Lower case conversion:* Convert all the words into lowercase format. For example, the term 'bowler' and 'Bowler' have the same impact for a class but they will consider as separate words because the case difference. For this reason, convert all words into lowercase is necessary.
6) *Stemming:* Represent the root form of different words.

C. Traning & Testing

NB gives best performance with the increase of training examples but if less training examples are taken it will give poor performance. Hence we try to establish a system offering better performance even if using less text documents for training. After preprocessing required probability matrix and Weight Matrix are calculated in the training stage.

TABLE II, describes the Weight Matrix calculation for some elements in 'sports' class. f_n denotes the frequency of each term and MC (Matched Classes) specifies the number of classes with term t_c in it. Here the 1 & 8 no. term have high TF but as they are common for maximum number of classes so ICF decreases significantly and the resulting weight is downcast. On the other side 3,6,7 no. terms have almost average TF as well as high ICF because they matched with less number of classes, as a result their Weight Matrix value increases more. So, it is the ICF, which have a great effort in balancing the Weight Matrix value of each term for each particular class.

TABLE II. Weight Matrix Calculation for Some Elements in Sports Class

No.	Term	f_n	TF(t,d_c)	MC	ICF_{tc}	Weight Matrix, $TF(t, d_c).ICF_{tc}$
1.	Play	7	0.004688546	3	0.5108256	0.002395029
2.	Pitch	3	0.002009377	1	1.6094379	0.003233967
3.	Club	4	0.002679169	1	1.6094379	0.004311956
4.	Bowler	2	0.001339584	0	1.6094379	0.002155978
5.	Toss	1	0.000669792	0	1.6094379	0.001077989
6.	Bowl	6	0.004018754	0	1.6094379	0.006467935
7.	wicket	3	0.002009377	0	1.6094379	0.003233967
8.	Ever	5	0.003348961	4	0.2231436	0.000747299

Our task is to add this Weight Matrix values after being powered by a significant number, with the posteriori probability ($P(C_n|W_n)$), which is calculated for each class when considering a new instance. This new matrix increases probability values for each class but for a particular class this value increases more and we classify the new text instance as a member of that class. After adding this new matrix value, the efficiency of this classifier increases significantly and somewhere gives a better result where NB fails. The posteriori probability value is changeable and will be more efficient with the increase of training examples. Let C is the set of all categories. The algorithms for calculating 'Weight Matrix' and how this matrix is used in testing are as given below:

WEIGHT MATRIX CALCULATION TRAINING

1.For each category c_i in C:
  Let w_j is the term for each category c_i
  For each w_j in c_i:
Let W(c_{ij}) is the Weight Matrix for each class c_i.
Set W(c_{ij}) := TF(c_{ij}) * ICF(c_{ij})
  where, TF(c_{ij}) is the term frequency of term w_j in class c_i,
  and ICF(c_{ij}) is the inverse class frequency of term w_j in class c_i.

WEIGHTED NB TESTING:

1.Given a test document D.
2.Let V be the vocabulary of all words($w_1, w_{2,\ldots\ldots w_n}$) in D and f_n is the number of times w_n occurs in D. Let, maximum countable frequency, $f_{mc} = 4$ *

*$f_{mc} = 4$, because it is observed that significant words don't occur more than four times in a test document. f_{mc} is changeable based on the size and type of texts.

3.Set W(c_i, D) :=0

where, W(c_i, D) is the weight of each class $c_i$for D

4.Let, f_{w_j} is the frequency of term w_j in D.

For each category c_i in C:

For each w_j€ V:

If f_{w_j}>f_{mc}, then:

Set W(c_i, D) := W(c_i, D) + W(c_{ij})

Else:

Repeat for k=1 to f_{w_j}:

Set W(c_i, D) := W(c_i, D) + W(c_{ij})

here, W(c_{ij}) is calculated during training for each category.

5.Return the category:

argmax P(c_i| D)= P(c_i) $\prod_{i=1}^{n} P(w_i|\, c_i)$ + W(c_i, D)

w_i is the word occurring in D at i^{th} position

A practical example of testing a new instance is shown below. We take this new instance from Entertainment class.

Standard NB gives the following estimated posteriori probability values-

Sports: 0.0025326876897012133
Business: 0.002712519455458494
Entertainment: 0.003078122415535175
Politics: 0.0021951416864381106
Technology: 0.0022546516651089587

Now, Weight Calculation for new instance with respect to each class,
for Sports class-0.0546242601911833267
for Business class-0.024711140579482021
for Entertainment class-0.065033581976851565
for Politics class-0.0189186822764950014
for Technology class-0.0238229514302171683

After adding the weights with the posteriori value of Standard NB we get the following estimated values for Weighted NB-

Sports: 0.05715694788088454
Business: 0.027423660034940515
Entertainment: 0.06811170439238674
Politics: 0.021113823962933112
Technology: 0.026077603095326127

From the above analysis it can be concluded that after adding the Weight Matrix value the estimated value is increased as a result accuracy is also increased.

D. Performance Measure

We chose 100 documents randomly from bbc dataset. In our experiment we employed 2-fold validation, where 1st fifty documents are chosen for training and 2nd fifties are for testing and vice versa. Confusion matrix is used to describe the performance of Standard NB and Weighted NB. TABLE III & TABLE IV shows the Confusion matrix for Standard NB and Weighted NB when the 1st half of 100 chosen documents is used for training and 2nd half is used for testing (group I). Now, in the 2nd half documents, which is used for testing, there are 50 documents- 10 Sports, 10 Business, 10 Entertainment, 10 Politics and 10 Technology. In TABLE III, of the 10 sports documents, Standard NB predicted that four are sports, four are politics and two are technology related text; of the 10 Business documents, Standard NB predicted that four are Business, four are politics and two are technology related documents and so on.

TABLE III. Confusion matrix for Standard NB (group I)

		Predicted class				
Actual class		Sports	Business	Entertainment	Politics	Technology
	Sports	4	0	0	4	2
	Business	0	4	0	4	2
	Entertainment	0	0	4	1	5
	Politics	0	1	0	5	4
	Technology	0	1	0	2	7

In TABLE IV, of the 10 sports documents, Weighted NB predicted that nine are sports, one is Entertainment related documents; of the 10 Business documents Weighted NB predicted that nine are Business, one is technology related documents and so on.

TABLE IV. Confusion matrix for Weighted NB (group I)

		Predicted class				
Actual class		Sports	Business	Entertainment	Politics	Technology
	Sports	9	0	1	0	0
	Business	0	9	0	0	1
	Entertainment	0	0	8	2	0
	Politics	0	1	0	9	0
	Technology	0	2	0	1	7

By comparing TABLE III and IV (group I), it can be said that Weighted NB can do better prediction than Standard NB.

TABLE V & TABLE VI shows the Confusion matrix for Standard NB and Weighted NB When the 2nd half of 100 chosen documents is used for training and 1st half is utilized for testing (group II). In the 1st half documents there are-10 Sports, 10 Business, 10 Entertainment, 10 Politics and 10 Technology. In TABLE V, of the 10 sports documents, Standard NB predicted that eight sports, two Politics related documents; of the 10 Business documents, Standard NB predicted that five Business, four Politics, one Technology related text and so on.

TABLE III. Confusion matrix for Standard NB (group II)

		Predicted class				
Actual class		Sports	Business	Entertainment	Politics	Technology
	Sports	8	0	0	2	0
	Business	0	5	0	4	1
	Entertainment	0	0	8	2	0
	Politics	1	0	0	9	0
	Technology	0	0	0	4	6

In TABLE VI, of the 10 sports documents, Weighted NB predicted all the ten documents as sports; of the 10 Business documents, Weighted NB predicted that eight Business, one Politics, one Technology related text and so on.

TABLE IVI. Confusion matrix for Weighted NB (group II)

		Predicted class				
Actual class		Sports	Business	Entertainment	Politics	Technology
	Sports	10	0	0	0	0
	Business	0	8	0	1	1
	Entertainment	0	1	7	0	2
	Politics	0	0	0	10	0
	Technology	0	0	0	2	8

By comparing TABLE V and VI (group II), it can be said that Weighted NB can do better prediction than Standard NB.

Most commonly performance measuring tools are used to verify the accuracy of our classification approach. They are- precision, recall, and F-measure [12] and calculations are done with the help of the Confusion matrices shown before.

If A = relevant documents, B = retrieved documents, then-

Precision= $\frac{|A \cap B|}{|B|}$, Recall = $\frac{|A \cap B|}{|A|}$

and F-measure = 2*$\frac{Precision*recall}{precision+recall}$

TABLE VII & TABLE VIII shows the precision and recall calculation for Standard NB and Weighted NB when the 1st half of 100 documents is used for training and 2nd half is used for testing (groupI). Using the three measures for group I, we observe the average F-measure of Standard NB is 0.49 and average F-measure of Weighted NB is 0.80.

TABLE V .Categorization Accuracy for Standard NB (group I)

Category	Precision	Recall	F-measure
Sport	1.00	0.40	0.57
Business	0.57	0.40	0.47
Entertainment	1.00	0.40	0.57
Politics	0.31	0.50	0.38
Technology	0.35	0.70	0.46
Average	0.65	0.48	0.49

TABLE VIII. Categorization Accuracy for Weighted NB (group I)

Category	Precision	Recall	F-measure
Sport	1.00	0.90	0.95
Business	0.69	0.90	0.78
Entertainment	0.89	0.80	0.84
Politics	0.60	0.90	0.72
Technology	0.70	0.70	0.70
Average	0.65	0.48	0.80

TABLE IX & TABLE X shows the precision and recall calculation for Standard NB and Weighted NB when the 2nd half of 100 chosen documents is used for training and 1st half is utilized for testing (group II). Using the three measures for group II, we observe the average F-measure of Standard NB is 0.72 and average F-measure of Weighted NB is 0.86.

TABLE IX. Categorization Accuracy for Standard NB (group II)

Category	Precision	Recall	F-measure
Sport	0.89	0.80	0.84
Business	1.00	0.40	0.57
Entertainment	1.00	0.80	0.88
Politics	0.43	0.90	0.58
Technology	0.86	0.60	0.71
Average	0.83	0.70	0.72

TABLE X. Categorization Accuracy for Weighted NB(group II)

Category	Precision	Recall	F-measure
Sport	1.00	1.00	1.00
Business	0.89	0.80	0.84
Entertainment	1.00	0.70	0.82
Politics	0.77	1.00	0.86
Technology	0.73	0.80	0.76
Average	0.88	0.86	0.86

After measuring the average F-measure for NB and Weighted NB for group I and group II, it is observed that Weighted NB achieved the highest average accuracy(83%), followed by Standard NB with average accuracy of 61% when the training examples is minimum. Here, we use only 50 documents for training and Weighted NB gives such a significant performance with this minimum number of training set. That concludes, if the training set in increased than Weighted NB will give much better performance. So, it is proved that with the Weight Matrix, NB gives better performance even if we use little training examples.

Fig. 1 is flow chart showing the accuracy of Standard NB and Weighted NB. The average F-measure of Standard NB and Weighted NB for group I are 0.49 and 0.80. Similarly, for group II, they are 0.72 and 0.86. Finally, we get the average values of Standard and Weighted NB are 0.61 and 0.83 respectively, considering group I and II.From this result it can be concluded that Weighted NB is much better than Standard NB.

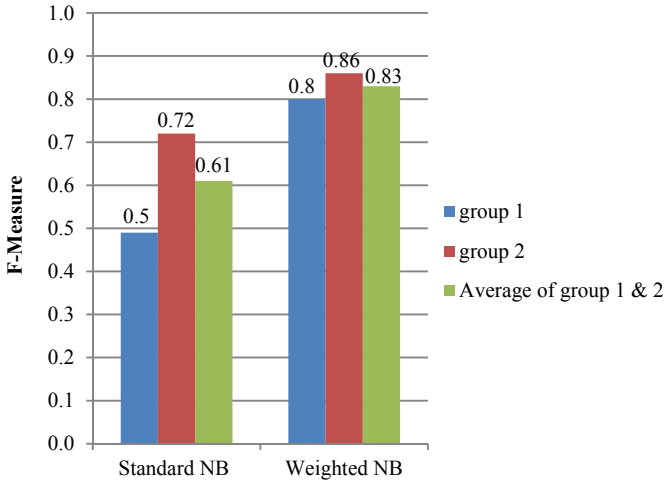


Fig. 1. Average Classification accuracy for Standard NB and Weighted NB.

VI. Conclusions And Future Plans

The Weighted NB approach proposed in this paperutilizes the most effective features set of text documents namely, 'Weight Matrix' which is later added with the posteriori value for the classification task. It is proved that this Weighted NB gives significantly better performance than standard NB, even if it uses minimum number of training examples. Experimental results show that Weighted NB approach for the classification of text documents in five categories achieves average 83% accuracy when we consider 50 documents for training and other 50 documents for testing among 2225 documents. Here Weighted NB gives much better performance with the little number of training sets whereas normal NB gives average 61% accuracy. After comparing the accuracy terms it is clear that Weighted NB is better than Standard NB considering bbc dataset.

However, in this experiment we only use Weight Matrix to increase the accuracy of NB. Other terms also can be introduced to increase the accuracy and to remove unwanted errors. Weighted NB gives poor performance in some cases. In future we are willing to find out the error for which it gives such performance.

Acknowledgement

This paper owes a lot for the suggestions and constructive criticism of Md. Ahsan Habib, Assistant Professor, Department of Computer Science & Engineering, HSTU, Dinajpur.

References

[1] E. Loper, E. Klein, and S. Bird, "Preprocessing Raw Text" in Natural Language Processing With Python, 1st ed. O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472, 10 July *2009*, ch. 3, pp. 79-123.

[2] D. Greene and P. Cunningham. "Practical Solutions to the Problem of Diagonal Dominance in Kernel Document Clustering", Proc. ICML 2006.

[3] F. Sebastiani, "Machine Learning in Automated Text Categorization," ACM Computing Surveys, vol. 34, No. 1, March 2002, pp. 1-47.

[4] J. Grimmer and B. M. Stewart, "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts," Political Analysis, January 2013, pp. 1–31, doi:10.1093/pan/mps028.

[5] D. D. Lewis and M. Ringuette (1994), " *A Comparison of Two Learning Algorithms for Text Categorization*," In Proc. of Third Annual Symposium on Document Analysis and Information Retrieval, pp. 81-93.

[6] G. D. Guo, H. Wang, D. Bell, Y. X. Bi *and* K.Greer, "Using kNN model for automatic text categorization," Soft Computing, 10*(5), pp.* 423- 430, 2006.

[7] C. Cortes and V. Vapnik, "Support-vector networks," Machine learning, vol. 20, pp. 273-297, 1995.

[8] A. S. Patil, B.V. Pawar, "Automated Classification of Web Sites using Naive Bayesian Algorithm," Proceedings of the International MultiConference of Engineers and Computer Scientists, Hong Kong 2012, Vol. I, 14-16.

[9] L. Jiang, C. Li, S. Wang, and L. Zhang, "Deep feature weighting for naïve Bayes and its application to text classification,"Engineering Applications of Artificial Intelligence, vol. 52, June 2016, pp. 26-39.

[10] Q. Yuan, G. Cong, and N. M. Thalmann, "Enhancing Naive Bayes with Various Smoothing Methods for Short Text Classification,"
WWW 2012 Companion, April 16–20, 2012, Lyon, France, ACM 978-1-4503-1230-1/12/04.

[11] V. Lertnattee and T. Theeramunkongt, " Analysis of Inverse Class Frequency in Centroid-based Text Classification," International Symposium on Communication and Information Technologies 2004 (ISCIT 2014), pp. 1171-1176, Sapporo, Japan, October 26-29, 2004.

[12] David M W Powers, "Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation," School of Informatics and Engineering, Flinders University of South Australia, Technical Report SIE-07-001, December 2007.

A Framework for Event Anomaly Detection in Cognitive Radio Based Smart Community

S. M. Nadim Uddin*, Nafees Mansoor†, Musfiqur Rahman‡, Nabeel Mohammed§ and Sazzad Hossain¶
*‡ Department of Electronics and Telecommunication Engineering, University of Liberal Arts Bangladesh.
†§¶ Department of Computer Science and Engineering,University of Liberal Arts Bangladesh.
Email:*sm.uddin.ete@ulab.edu.bd, †nafees.mansoor@ulab.edu.bd ‡musfiqur.rahman.ete@ulab.edu.bd§nabeel.mohammed@ulab.edu.bd
¶sazzad.hossain@ulab.edu.bd

***Abstract*—With the advancement of technology, a surge of research interest in cognitive radio based networks in smart communities has been mounting. It is anticipated that CR-enabled networks will play a vigorous role in the enrichment of communication efficiency in neighborhood sensor area network. This paper presents a framework for Cognitive Radio based event anomaly detection mechanism. A skeleton for intelligent learning, detection and decision mechanism for Local Controller Unit and a Primary Controller Unit is also proposed and discussed in the model. The proposed model has four distinct layers namely sensors, routers, Local Controller Unit and Primary Controller Unit. A scheme for emergency situation detection and notification has been proposed. This paper also introduces a cluster formation scheme for better accuracy in data transmission among different hierarchical layers. The network module of the proposed model is later simulated and validated for some important performance communication metrics.**

***Keywords*—Cognitive radio, Neighborhood area network, WSN, Anomaly detection.**

I. Introduction

With the development of the communication technology and microelectronic devices, a surge of interest in designing smart networks based on cognitive radio mechanism for smart communities has been increasing among academics as well as researchers. Neighborhood Sensor Area Network (NSAN) is a variation of the Neighborhood Area Network (NAN) in smart community and consists of the communications of smart router units and sensors with a back-end local control center. Though adopting wireless mesh topology can be a suitable solution for efficient communication, to overcome the network congestion problem and the growing demand of radio spectrum, proper utilization of the radio spectrum is essential[1]. Cognitive radio enables dynamic spectrum allocation technique to utilize radio spectrum efficiently [1][2][3].

However, data delivery in sensor networks is faulty and unpredictable [4], which lead to data anomaly. Anomaly refers to the problem of recognizing patterns in data that do not conform to expected behavior [5]. Anomaly detection in sensor networks poses a set of unique challenges. A sensor network comprises of sensors that collect different types of data, such as binary, discrete, continuous etc. Failures in wireless sensor networks can occur for various reasons due to fragility, depletion of batteries or destruction by an external event. In addition, nodes may capture and communicate incorrect readings because of environmental influence on their sensing components. Moreover, links in any ad hoc wireless networks are failure-prone [6], causing network partitions and dynamic changes in network topology. Additionally, the anomaly detection techniques need to be light- weight and in a distributed fashion due to resource constraints [5].

For the past few years, different anomaly detection approaches have been proposed to improve communication in wireless networks. Most of these techniques use the machine learning and statistical approaches for anomaly detection [7]. Moreover, these techniques can be categorized into three groups, namely unsupervised clustering, semi-supervised classification and supervised classification. Comparing with the other two groups, clustering approaches exhibit higher reliability in communication [8]. However, communication networks for smart grids also require to consider additional parameters, such as, application requirements, link capacity, traffic settings, cost, scalability, etc. [9][10]. It is also observed that very little considerations have been made for anomaly detection in cognitive radio oriented smart community. Thus, anomaly detection in this area remains to be at the infant stage.

In this paper, a framework for cognitive radio based anomaly detection mechanism in smart communities has been proposed. A conceptual architecture with a wireless hybrid topology has been proposed to integrate sensing, computation and decision-making for enabling efficient detection of anomaly. In the proposed model, two distinct types of anomaly detection mechanism using Exponentially Weighted Moving Average and Anomaly factor based on Brier's Score are proposed. The remainder of this paper is organized as follows. Section II demonstrates the abstract and deployment view of the proposed framework. In Section III, simulations results are presented and discussed. The paper ends with Section IV, where this section presents conclusion and future works.

II. Proposed Framework

A. System architecture

The proposed model consists of five levels of hierarchy namely sensors, routers, local controller unit (LCU), Database layer and primary controller unit (PCU). The first layer of the architecture is composed of hundreds of sensors situated in the street areas are grouped into clusters, sending their data

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

directly to routers. Sensors are assumed to be dispersed in a 2 dimensional space and are quasi-stationary. Sensors transmit at the same fixed power levels, which is dependent on the transmission distance. It is assumed that the energy consumption among nodes is not uniform. The routers, the second layer of the architecture, are responsible for transmitting the received data to the LCU. The third layer is LCU, a test and control unit that is responsible for applying control over the routers and sensors. Routers and the LCUs are connected via wireless mesh network in a multi-hop manner. The forth layer is the Database layer that will store the aggregated data received from layers below, and feed the PCU with collected data. PCU and the LCUs keep databases of unusual behaviors of different devices at different levels of the hierarchy. The databases are used to predict damages or determine emergency situation levels.

In the proposed model, each node at a level of the hierarchy sends control message to the node of upper level or among themselves. The control message is used for connection set up, cluster formation, message delivery completion, instantaneous data fetch command, etc. Information messages transmitted and received among nodes are defined as sensed data from the terminal sensor devices, combined data from Nodes to routers, combined data from routers to LCU and Filtered and combined data from LCUs to PCU. A Primary Controller Unit (PCU) is proposed to be placed in the network for data storage, processing and decision making. It is the top layer and applies control over the entire neighborhood area. It accesses sensing units directly in emergency situations, or indirectly through the DB layer in normal conditions.

Data packet, in the model, follows a basic structure consisting of sender node id, information, destination id, hop count, timer, sequence number etc. The routing path for source to destination is determined by the greedy forwarding algorithm.

B. Network model

Each grid is proposed to be hexagonal in shape and is divided into six sub-grids. The proposed environment is a mesh network in which the LCU is located at the center of the grid to make LCU-router and LCU-LCU communication easier. According to the proposed model, one LCU can serve six sub-grids and is efficient for installation costs and maintenance. The topology ensures equidistant LCUs in a community. The connection between the router and the terminal sensor nodes is set by a control message sent by the router. Upon reception of the message by the sensor node, an acknowledgement message is sent to the router and connection is established. This process continues until all nodes of different clusters are connected to the router. The router collects the data from the cluster of nodes, verifies and uploads the data to the LCU. Unable to connect, the router waits for a threshold time and tries again. After trying for a threshold number of attempt, the router takes a log of the nodes along with the associated data and sends notification to the LCU and waits for instructions. The connection establishment mechanism between LCU and router is similar to the connection between the sensors and router. LCU receives data from the router and analyses data of that router. If the data complies with the database, LCU keeps the data for threshold time and then rejects it. If the data varies from the threshold, LCU finds out the problem node or the cluster. Appropriate commands from the database are selected and sent it to routers as a counter-measure. If appropriate commands are not found in LCU's database for a specific event, a control message along with the data is sent to PCU for analysis and counter-measure. PCU receives information data from the LCUs in a community. Depending on the variance of the data, it sends command to LCU or put the system into a warning state. Figure 1 illustrates the hierarchical data processing of the proposed model.

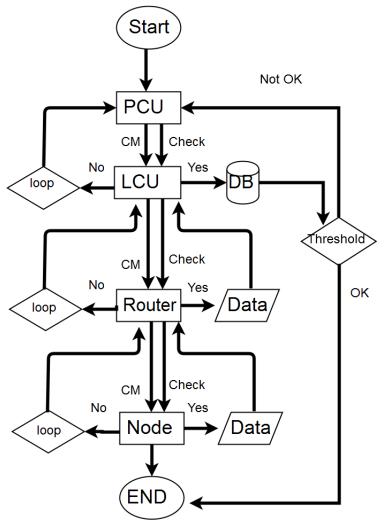


Fig. 1: Flow chart of data movement in the proposed network.

C. Detection mechanism

In the sensor receiver, the returning signal comprises the wanted data and the unwanted niose. The router determines whether point anomaly has occurred in nodes i.e. an individual data instance that can be considered as anomalous with respect to the rest of data. To determine point anomaly, router calculates the Exponentially Weighted Moving Average (EWMA) of each nodes which acts as the baseline or the reference model for each node. Let $\Theta \leq 1$ denote a constant and ζ_t is the mean of data value at time t, the Exponentially Weighted Moving Average (EWMA) [11] of a node Γ_t is,

$$\Gamma_t = \Theta\zeta + (1 - \Theta)\Gamma_{t-1} \tag{1}$$

The EWMA value changes with time as it depends on the previous value. However, if the data value of a node crosses EWMA by a threshold value, the router recognizes the event as important and sends command to LCU for further instruction. Figure 2 illustrates a basic detection of point anomaly of a node.

LCU maintains network through techniques that operate in a semi- supervised mode and it is assumed that the training data has been labeled for instances for only the normal class.

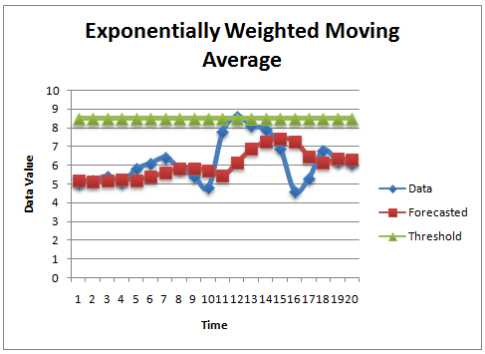


Fig. 2: Detection of point anomaly.

The LCU checks if a collection of related data instances from routers is anomalous with respect to the entire data set. If so, it marks it as a collective anomaly and checks for problem node or router in divide and conquer manner i.e. going through each division of the router. On each division, based on [12], LCU uses hyper-ellipsoids for k dimensional inputs for calculating Radial Basis Function for the Multi Layer Parametric approach in case of data classification. .

Let w_{jk} be the vector of weights of the node from the LCU, $\Phi(.)$ be the activation function and $||x-\mu_j||$ be the Euclidean distance, then the Response Function of system, τ_k can be defined as,

$$\tau_k = \sum_{j=1}^{m} w_{jk} \Phi \left(\frac{\|x-\mu_j\|}{\sigma_j} \right) \tag{2}$$

where m $\in N$ is the number of nodes where N is the set of natural numbers, $\mu_j(x)$ is the centroid of the network and σ_j is the smoothing factors which may vary depending on the distance.

The activation function which shapes the network system is selected to be Gaussian Radial Basis Function (RBF). Let d be the dimension of data from an input and $\sigma_j^2(x)$ be the variance, then the RBF, Φ can be defined as,

$$\Phi(t|x) = \frac{1}{[2\pi]^{\frac{d}{2}} \sigma_j^k(x)} e^{-\frac{\|t-\mu(x)\|^2}{2\sigma_j^2(x)}} \tag{3}$$

For generality, it is assumed that the smoothing factors σ_j are equal for all nodes. Based on validation, LCU assigns an anomaly factor for each router through which anomaly can be detected. Inspired from Brier's Score, the factor δ can be expressed as,

$$\delta = \frac{1}{N} \sum_{t=1}^{N} (\alpha_t - \beta_t)^2 \tag{4}$$

Where α_t is the category of the anomaly, β_t is the outcome of the observation at time t and N is the total observation. Depending on the variance of the output, LCU decides commands from the database or sends data to PCU. During warning state, an emergency protocol command from PCU is carried out to routers via LCU to monitor specific regions. In such case, a new cluster of nodes is formed to focus on a specific region and the nodes can bypass normal procedures and directly send data to other routers and LCUs for accuracy and effective data transmission. In emergency situations, each node in every level is capable of sending some emergency messages to other nodes. This type of message includes instantaneous information of node, connection hierarchy bypass message, power messages which include emergency shutdown, emergency network buildup, system restoration messages etc. PCU retrieves data from that region more frequently and calculates the data variations. Depending on the variation of data, the PCU turns off the state, puts the system on halt or sends alarm to house owners and local authority and stores the data in database for future decisions.

D. Cluster formation

The proposed clustering mechanism is inspired from the clustering scheme for cognitive radio ad-hoc network in [3]. In the existing clustering mechanism, cluster-head selection is based upon a parameter called cluster-head determining factor (CHDF) where CHDF of a node is calculated over number of common channels and number of neighboring nodes.

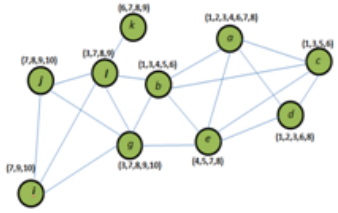


Fig. 3: Connectivity graph of a CRN with the accessible channels' sets in the brackets.

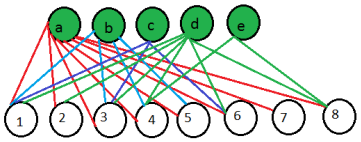


Fig. 4: Bipartite graph constructed by node.

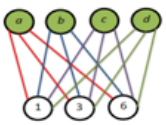


Fig. 5: Maximum edge biclique graph of node.

The proposed cluster formation stage starts once nodes in the network finish the neighbor discovery process. Next, the nodes share accessible channel lists (ACLs) C_i and neighbors list N_i among 1-hop neighbors (where i = 1, 2, 3 n). The

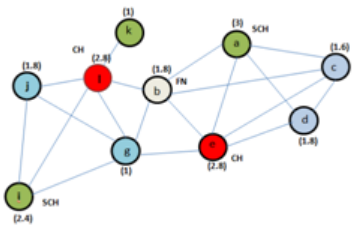


Fig. 6: Cluster-head, Secondary Cluster-head and Cluster Member selection.

proposed clustering scheme is defined as a maximum edge biclique problem. Based on neighbor list N_i and accessible channels list C_i, each CR_i constructs an undirected bipartite graph G_i (A_i, B_i, E_i). Here, A_i = $CR_i \cup N_i$, and B_i= C_i. An edge (x, y) exists between vertices x $\in A_i$ and y $\in B_i$ if y $\in C_i$, i.e., channel y is in the channel list of CR_i. From the bipartite graph, each node in the network constructs its own maximum edge biclique graph. From the maximum edge biclique graph, node determines new C_i and N_i values. The proposed clustering scheme aims to allocate maximum number of free common channels per cluster with suitable amount of member nodes. A parameter called Cluster Head Determination Factor (CHDF) is used to select cluster heads. Every CR calculates CHDF based on equation (5).

$$CHDF = \sqrt[C_i]{{C_i}^{N_i}}; i = 1, 2, 3, \ldots \quad (5)$$

Where, C_i is number of free common channels and N_i is the number of neighboring nodes of CR_i. A node declares itself as cluster head if its own CHDF value is higher than all its neighbors. Once the CHDF value of a node CR_i is lesser than any of its neighbor, CR_i joins the neighboring node that has the highest value as cluster member (CM). After the cluster formation, CH selects SCH from the CMs based on the CHDF value. The SCH takes charge of the cluster if current CH moves out, which shrinks the possibility of re-clustering.

III. Simulation and Results

A. Simulation setup

To simulate and analyze performances namely throughput, delay and routing overhead of the proposed network model for the proposed network model, discrete-event simulator NS2 has been used and the performance analysis are conducted using PERL scripts.

A simulation area of 10000 m^3 is considered for the simulation purpose. Drop-tail method is used for the queuing purpose. IEEE 802.11 is considered as the MAC type and TCP is considered for Transport Layer. Maximum packet queuing delay is considered as 50 μs. In the simulation, AODV is used as the routing protocol while data traffic is generated with Constant Bit Rate (CBR) with packet size is set to 512 bytes. Varied packet rate ranging from 100 packets/sec to 800 packets/sec is considered to evaluate the performance of the network for different traffic load. The number of nodes is considered to be 100. Initial energy for all nodes is considered to be 100 Joules. The simulations are run for 240s each.

B. Evaluation

This section of the paper discusses the simulation results of the proposed model in terms of throughput, energy consumption, delay and overhead.

1) Performance based on Throughput: In this paper, throughput is defined as the number of successfully received data packets at the destination node in a unit time and it is represented in Kbps.

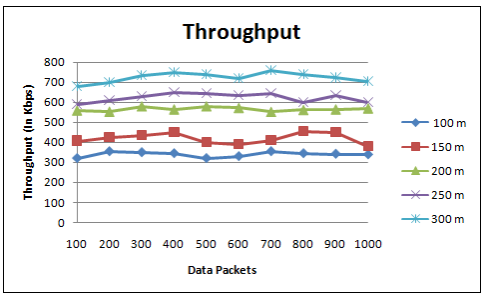


Fig. 7: Throughput for the proposed network.

Figure 7 shows that with an increasing traffic load, throughput for all scenarios increases. However, throughput is higher in the network with the longer radio transmission range, than the other radio transmission ranges for all different data flow rates. This is because; a network with longer ranged transmission finds lesser number of hops to transmit packets to the destination. Thus, with decreasing number of hops, number of links throughout the network, probability of link failure and rate of packet retransmission reduce significantly. Therefore, with higher transmission range, throughput of the network increases for all different traffic loads.

2) Performance based on Packet Delivery Ratio: In this paper, the packet transmission delay is defined as the average time required for transferring data packets from the source node to the destination node. Packet delivery ratio is expressed in percentage.

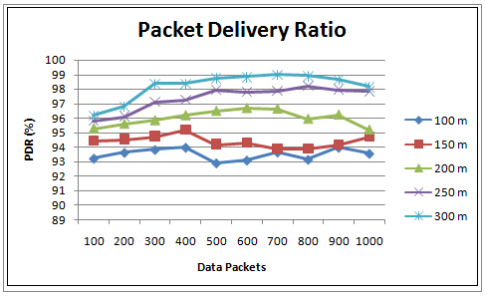


Fig. 8: Packet delivery ratio for the proposed network.

From the figure, it is observed that the packet transmission delay increases with increasing data flow rate in both scenarios because when traffic load is increased, more data packets need to be transmitted from the source to the destination node. As

a result, the intermediate nodes are required to process more packets which eventually increases individual data processing sessions among the nodes. Thus, when higher number of packets propagates, source node and the intermediate nodes need longer time to forward the packets to the next hop, which increases the cumulative packet transmission delay. Moreover, lesser number of intermediate nodes is engaged to forward the data in long transmission ranged network, which results lesser data processing sessions. Therefore, a network with long transmission ranged radios results lesser packet transmission delay than that of a network with short transmission ranged radios.

3) Performance based on Overhead Ratio: The network overhead is defined as the total transmitted control packets over total received data packet at the destination node.

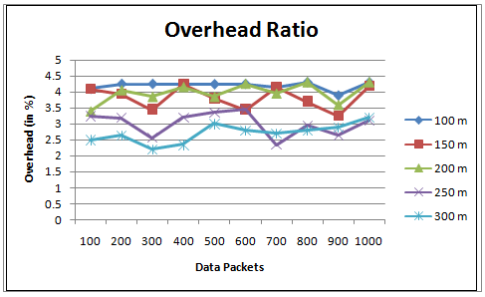


Fig. 9: Overhead ratio for the proposed network.

From the figure, it is seen that the overhead ratio increases with increasing data flow rate in both scenarios. Increase in data rate means nodes need to process more data packets individually which eventually results in reduction of a node's efficiency for packet forwarding. It is also observed from Fig. 7 that the overhead ratio is lesser in a network with longer radio transmission ranged network compared to the network with shorter radio transmission ranged network for all different traffic loads. That is because, in a shorter radio transmission ranged network, more nodes will have to process increased traffic load and as a result, retransmission of data packets due to exceeding node's capacity to process data packets will increase which eventually increases number of transmitted control packets.

IV. Conclusion

In this paper, a framework for cognitive radio based event anomaly detection mechanism in neighborhood sensor area network has been proposed. A concept of adapting exponentially weighted moving average for point anomaly and anomaly factor for cumulative anomaly has been outlined and a clustering scheme for emergency situations has been proposed. The next research project will involve designing a robust algorithm for labeling anomalies and system optimization as well situation handling mechanism.

References

[1] W. Saad, Z. Han, A. Hjorungnes, D. Niyato, and E. Hossain, "Coalition formation games for distributed cooperation among roadside units in vehicular networks," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 1, pp. 48–60, 2011.

[2] N. Mansoor, A. M. Islam, M. Zareei, S. Baharun, T. Wakabayashi, and S. Komaki, "Cognitive radio ad-hoc network architectures: a survey," *Wireless Personal Communications*, vol. 81, no. 3, pp. 1117–1142, 2015.

[3] N. Mansoor, A. M. Islam, M. Zareei, S. Baharun, and S. Komaki, "Spectrum aware cluster-based architecture for cognitive radio ad-hoc networks," in *Advances in Electrical Engineering (ICAEE), 2013 International Conference on*. IEEE, 2013, pp. 181–185.

[4] J. Zhao and R. Govindan, "Understanding packet delivery performance in dense wireless sensor networks," in *Proceedings of the 1st international conference on Embedded networked sensor systems*. ACM, 2003, pp. 1–13.

[5] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM computing surveys (CSUR)*, vol. 41, no. 3, p. 15, 2009.

[6] A. Woo, T. Tong, and D. Culler, "Taming the underlying challenges of reliable multihop routing in sensor networks," in *Proceedings of the 1st international conference on Embedded networked sensor systems*. ACM, 2003, pp. 14–27.

[7] M. Markou and S. Singh, "Novelty detection: a reviewpart 1: statistical approaches," *Signal processing*, vol. 83, no. 12, pp. 2481–2497, 2003.

[8] A. L. Dos Santos, E. P. Duarte Jr, and G. M. Keeni, "Reliable distributed network management by replication," *Journal of Network and Systems Management*, vol. 12, no. 2, pp. 191–213, 2004.

[9] Z. Zhu, S. Lambotharan, W. H. Chin, and Z. Fan, "Overview of demand management in smart grid and enabling wireless communication technologies," *IEEE Wireless Communications*, vol. 19, no. 3, pp. 48–56, 2012.

[10] Y. Dong, Z. Cai, M. Yu, and M. Sturer, "Modeling and simulation of the communication networks in smart grid," in *Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on*. IEEE, 2011, pp. 2658–2663.

[11] J. M. Lucas and M. S. Saccucci, "Exponentially weighted moving average control schemes: properties and enhancements," *Technometrics*, vol. 32, no. 1, pp. 1–12, 1990.

[12] T. Brotherton, T. Johnson, and G. Chadderdon, "Classification and novelty detection using linear models and a class dependent-elliptical basis function neural network," in *Neural Networks Proceedings, 1998. IEEE World Congress on Computational Intelligence. The 1998 IEEE International Joint Conference on*, vol. 2. IEEE, 1998, pp. 876–879.

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

PSO-NF based Vertical Handoff Decision for Ubiquitous Heterogeneous Wireless Network(UHWN)

Nurjahan*, Saoreen Rahman†, Tanusree Sharma ‡, S M Reza§, M M Rahman¶ and M S Kaiser‖
Institute of Information Technology, Jahangirnagar University, Savar, Dhaka-1342, Bangladesh*‡§¶
Bangladesh University of Business and Technology (BUBT), Dhaka-1216, Bangladesh†
Anglia Ruskin IT Research Institute, Anglia Ruskin University, Chelmsford, UK‖
Email:{nurjahannipa*, sadiajanim‡, smrezaiit§, mrrajuiit¶}@gmail.com, saoreen@bubt.edu.bd†, m-shamim.kaiser@anglia.ac.uk‖

***Abstract*—In this paper, Hybrid of Particle Swarm Optimization (PSO) and Neuro-Fuzzy (NF) has been proposed for improving Vertical Hand-off (VHO) decision for the Ubiquitous Heterogeneous Wireless Network (UHWN). The mobile users have made VHO decision based on data-rate, dwelling time and service type, residue energy and network connection time. To reduce the computational complexity, the pre-decision of VHO is made based on preference of user and number of non-occupied channel. The main aim is the improvement of Quality of Service (QoS) of link and reduction of hand-off frequency. This method is able to reduce call blocking probability, hand-off latency and hand-off call dropping probability, improve the throughput and also balance the load of radio access technologies. The execution of the proposed system has been performed using NS 2 simulator. The numerical results of the proposed algorithm outperforms the state-of-the-art method.**

***Keywords*—PSO; Neuro-Fuzzy; RSSI; VHO; QoS**

I. Introduction

Heterogeneous Wireless Network (HWN) is a gathering of different types of wireless devices. And these devices are always accessible and available at whatever time at any place. Wireless communication devices consists of various pervasive devices such as mobiles, GPS units, cordless telephones, wireless computer parts, ZigBee technology and satellite television etc. When these wireless devices are able to omnipresent high speed data access at any time and at any place is called Ubiquitous heterogeneous wireless network (UHWN). For anytime and anywhere connectivity in UHWN, integration of various network access technologies is obligatory to support high speed data rate, mobility without call interruption, reducing call dropping probability.

Cellular telecommunication system is divided into various small cells. Small cells are good for frequency reuse and coverage. So, when users move from one place to another, UE becomes connected to various cells of various technologies. To continue the connection, hand-off occurs. Otherwise, call becomes dropped and user becomes dissatisfied.

Hand-off confers to the process of the transition of a data connection or active call in a non-interrupted way. The transitions occurs from one base station to another base station which may be geographically apart. Generally, two types of hand-offs are widely used. One type of them horizontal hand-offs, is used in a homogeneous wireless network. Hand-off triggers within the same network. When mobile terminal moves within the same wireless network technologies, horizontal hand-off focuses on reducing numbers of calls failed. Horizontal hand-off handling capability exists in today's network technologies. In another hand, Hand-off triggering among different access technologies is called vertical hand-off (VHO). Here, the mobile terminal moves into different networks where different access technologies are used. In order to maintain higher throughput within different collocated networks, VHO is indispensable.

So, it becomes a vital issue to enhance the overall performance of seamless connectivity in a HWN. To correctly take the decision, balance the load, maintain energy efficiency and QoS, efficiently handle resource utilization, provide security, reduce cost, VHO algorithm should be carefully maintained. Complexity becomes higher to consider all the factors. The best candidate network selection is the main complexity from the pervasive networks based on the numerous criteria. Choosing best candidate network depends on various performance metrics such as ubiquitous mobility with coverage, delaying in the time of handover, handover frequency , handover shortage probability, throughput and combinations of these parameters.

Considering these issues, a hybrid PSO-NF based VHO decision algorithm has been introduced. For tuning the parameters of a fuzzy inference system (FIS), NF uses neural network. NF controller parameters are: data rate (DR), network connection time (NCT), service type (ST), Energy consumption (EC) and dwelling time (DT) . To train the NF system, hybrid back propagation method and least mean square algorithm is used. On the other hand, PSO is used to select initial parameters intelligently. It increases convergence speed and makes a decision in a short period time. This method is able to reduce call dropping probability, maintain quality of service, balance load, lessen hand-off call blocking probability and handle the energy more efficiently. Initial step is used to avoid ping pong effect. Initial step decision depends on two parameters: user preference and number of free channel. This also avoids algorithm complexity.

Remainder paper is well organized as below: section III represents PSO-FNN based VHO algorithm, section IV represents the Simulation and Analysis and section V finally draws

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

the conclusion.

II. Related Works

The choice of VHO algorithm is random. Many of VHO algorithm can give disputable results. To get indisputable results, it is important to choose the method very cautiously. Various approaches has been proposed for improving for VHO decision [1]–[3], [6]–[19], [21], [22], [24]. The decision parameters can be received signal strength (RSS), bandwidth, network connection time, handover latency, power consumption, monnetary cost, security, bit error rate (BER) and signal to interference ratio (SIR), user preferences and many more [4], [5]. In another hand, dynamic programming or artificial intelligece, for example, neural nework(NN) or fuzzy logic, pattern recognition, or genetic algorithm (GA) is considered more effective in case of vertical hand-off decision prediction [2], [6], [10], [12], [15], [19], [21], [22]. As fuzzy logic (FL) systems and NN classifiers are non-linear and capable of generalization. And GA has better uncertainty and imprecision tolerance, good learning ability and adaptability. But problem occur when the networks needs to scalable. For the initial parameter and weight optimization in expert systems, PSO is used effectively [14]–[16], [18].

RSS based VHO algorithm increases handover delay, as it takes sample RSS points and make average of it [3], [9]. Problem also occurs when close RSSI among candidate network technologies are found. In that case, number of hand-offs becomes higher and throughput becomes lower. Packet delay and throughput becomes lower [9]. VHO based on SINR provides better throughput because it directly related with signal to interference plus noise ratio [8]. Considering only bandwidth, the handover occurrence becomes higher due to shadowing and fading. Problems with data rate, energy efficiency and handover failure probability exists. Also call dropping probability becomes higher [8]. To keep low the number of handover occurrence, to get higher throughput and take user preferences into consideration, cost function based VHO is introduced [11], [13]. But problem occurs in the calculation of security and interference levels and network resource wastage [11], [13].

Fuzzy logic and ANN are more suitable than traditional techniques because of their cognitive uncertainty, learning mechanism, adaptation capability, fault tolerance capability, parallelism and generalization capability. It is impossible to draw out rules from NN. Integration of special information about problem is ambiguous into the NN to make easy learning [6]. User Preferences, energy efficiency are not considered in the paper. So, quality of service are not ensured [21].

There need a trade-off of handover delay time and system load in fuzzy based system [2]. Bad network condition and variation of user requirement needs variable weights but three input parameter weight is fixed [4]. Load balancing, users preferences and service quality are not taken into consideration [12]. Load conditions, allocation of resources, energy efficiency, higher data rate acquisition is unreasonable in [10]. However, in fuzzy logic based VHO algorithm, knowledge acquisition is difficult. Expert knowledge is needed to set rules. And another limitation is fuzzy logic works worst with higher number of inputs.

Balancing load, HCBP is successfully performed in [12], [15], [19]. Bandwidth and received signal strength are used as the parameters of FNN controller. And PSO is used for initial parameter learning. But User preferences, Data rate and energy efficiency are not taken into consideration and consent for lower computational cost is illogical [15].Genetic Algorithm is used to set the beginning weights and threshold for neural network [12], [19]. User preferences, Energy efficiency is not taken into consideration [12]. In [19], genetic algorithm and fuzzy algorithm correctly determines whether a handoff is necessary or not. User Preferences, Load balancing are not considered in the paper. Cost function is used to select best network for handover and particle swarm optimization for optimizing weight in real time is proposed in [14]. Energy are not handled efficiently, also the user preferences are not reflected. Grey model along with fuzzy inference rules to predict the received signal is proposed in [17]. Then the prediction is fine-tuned by PSO which optimizes better than self-tuning algorithm. But energy efficiency, user preference and call blocking probability are not taken into consideration [17]. Do not consider energy efficiency, data rate and user preferences. Handling data fusion is also a major issue [16]. Another dynamic method is proposed in [18]. But this method is not energy efficient.

III. PSO-NF based VHO Algorithm

A HWN environment are designed consisting of LTE(4G), UMTS(3G), GSM(2G), WLAN and picocell. An UE receives signal from these collocated networks considers many factors to connect to a specific network. The hand-off scenario based on our proposed PSO-NF based VHO algorithm is shown in 1.

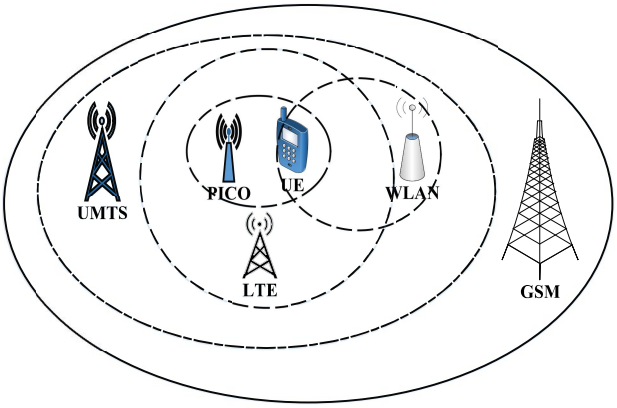


Fig. 1. Vertical Handoff Scenario for the proposed system

Small base station, Picocell, often provides maximum signal strength in a small coverage area. Where signals of other various radio technologies don't reach in a blocked area such as in a shopping mall,train station, stock exchange,office or in an indoor building, called a dead zone, gets coverage from picocell and helps the users.

A. Flowchart of Vertical Handoff Decision Algorithm

In this paper, A HWN environment consists of UMTS, LTE, GSM, WLAN and picocell networks is considered. Main

three parts of this paper is : intial screen-out step, PSO-NF controller and finally performing hand-off decision. First, User Entity (UE) predicts network according to RSSI.

Then initial screen-out method selects some candidate networks and removes some networks according to user preference and channel utilization. The main flow of our proposed method is shown in 2.

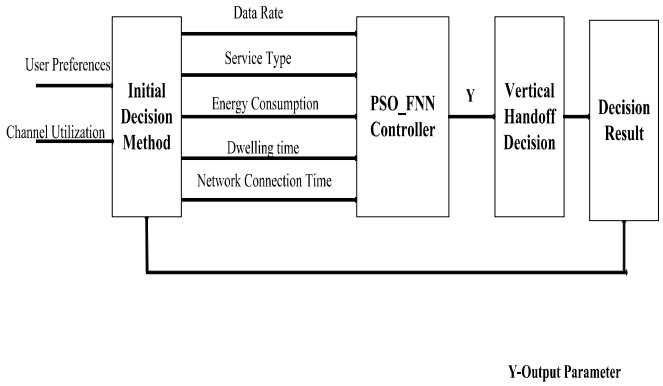


Fig. 2. Flowchart of Vertical Hand-off Decision Algorithm

The parameters data rate, network connection time, service type, energy consumption and available energy and the time a device spends in a cell without moving are feed into hybrid controller. HO decision finally is performed based on output parameter Y. PSO is used with the FNN controller to optimize the weight of the it.

B. Initial Screen-out Method

According to RSSI, user entity discovers various radio access technologies first. Then, some networks are removed according to the number of non-occupied channel and priority of networks. If user entity belong to the network that is preferred by users, then no hand-off occurs. But if current network is not user preference list, then hand-off is occurred to the network which are preferred by any users with low mobility nature of user entity. In case of no user preference and low mobility, hand-off decision is performed by the NF controller. But with user preference or without user preference, if any entity travels back and forth in a high speed, no hand-off will occur.

For avoiding ping-pong effect, initial decision process is performed before hand-off decision process and ensure the user satisfaction. If no hand-off is needed, then the flow directly go to the decision step. This reduces hand-off decision complexity in large extent. So, this is enable to guarantee the QoS and to cut-down the number of candidate networks primarily. The user entity with busy channel and no user priority, is removed from the candidate networks.

C. NF Vertical Hand-off Controller

FL and NN are two different tools in constructing intelligent systems. NN mainly deals with raw data where as fuzzy logic are good at reasoning. With the help of parallel calculation and also learning capabilities, fuzzy systems along with ANN can perform human-like knowledge representation. Combining the advantages of two intelligent systems, fuzzy neural network (FNN) is used to acquire excellent performance. FNN still suffers convergence problem and it is necessary to adjust the weights more intelligently. PSO, a good optimization method is used to set initial parameters. And for increasing the convergence speed.

1) Structure of NF: Fuzzy logic has the ability to represent human decision making capability in IF-THEN forms. Crisp input are fuzzified into membership function which values lie between 0 and 1. Then through output membership function, de-fuzzification is performed. Here a five layer NF is considered to implement the proposed system. It comprises the input, input membership, fuzzy rule, output membership and finally output. For training NF, back propagation along with Least Mean Squares(LMS) algorithms are used [23]. The proposed structure for NF is shown in 3 [20].

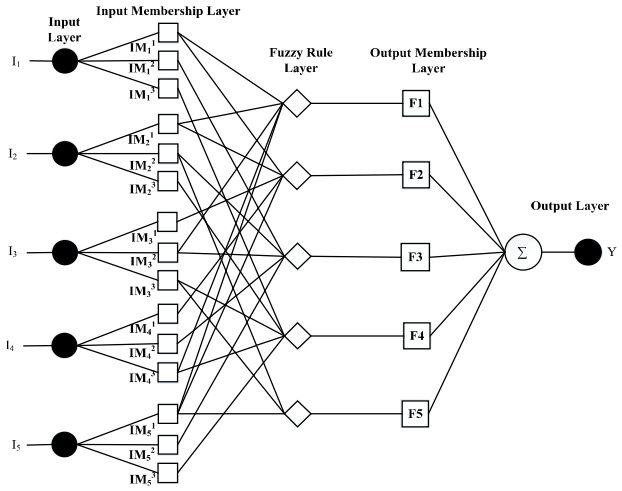


Fig. 3. Structure of NF system. I: Input, IM: Input Membership Function, Y: Output.

NF structure considers input as I_c, output as Y.

$$I_c = \begin{bmatrix} I_1 \\ I_2 \\ I_3 \\ I_4 \\ I_5 \end{bmatrix} = \begin{bmatrix} DR \\ NCT \\ ST \\ EC \\ DT \end{bmatrix}$$

Where, DR = Data Rate, NCT = Network Connection Time, ST = Service Type, DT = Dwelling Time, EC=Energy Consumption.

For membership function, we adopt Gussian function. Input membership layer is responsible for transforming the crisp input into linguistic level. The output of this layer can be calculated as:

$$Y_{cd}^2(I_c) = exp(-\Sigma_{c=1}^{r}[(I_c - m_{cd})^2/s_{cd}^2)]) \quad (1)$$

$d = 1, 2, 3, ..., x$. Here, x is the number of neurons in membership layer. m_{cd} and s_{cd} are the average and standard deviation of Gaussian function. Fuzzification process receives crisp input and determines the degree of the fuzzy set. Here,

fuzzy set (L, M, H), represents Low, Medium and High which is applicable for all input variables from input layer.

The membership functions are calculated from fuzzy rules. The linguistic variable in the input are $IM_c^1 = Low$, $IM_c^2 = Medium$, $IM_c^3 = High$. The output of each rule, F_c, is calculated as:

$$F_c = \beta_{c1}I_1 + \beta_{c2}I_2 + \beta_{c3}I_3 + \beta_{c4}I_4 + \beta_{c5}I_5 + \beta_{c6} \quad (2)$$

where β_{cd} is the parameters of input d.

Final output of the FNN controller is calculated by:

$$Y = \Sigma_{c=1}^{L}[\frac{\Pi_{d=1}^{L} IM_c^d(I_c)(F_c)}{\Sigma_{c=1}^{R}(\Pi_{d=1}^{l} IM_c^d(I_c)}] \quad (3)$$

Here, L is linguistic variables number and R is the rules number.

2) Particle Swarm Optimization: Initial parameters of NF is an important matter. This parameter and threshold is optimized by PSO.

PSO belongs to swarm intelligence. This technique performs stochastic optimization . Particles velocity and positions is continuously updated in the search space. According to the fitness value of each particle, personal best and global best value is also updated.

The change of velocity is followed by the mathematical equation given below:

$$s_{mn}(e+1) = ws_{mn}(e) + a_1 * r_{1n}[pbest_{mn}(e) - p_{mn}(e)] + a_2 * r_{2n}(e)[gbest_n(e) - p_{mn}(e)] \quad (4)$$

And The change of any particles position is expressed as

$$p_{mn}(e+1) = p_{mn}(e) + s_{mn}(e+1) \quad (5)$$

Where,
$p_m(e)$ presents current location of the mth particle at time e.
$p_m(e+1)$ represents location of mth particle at time $e+1$.
$s_m(e+1)$ represents velocity of mth particle at time $e+1$.
$s_{mn}(e)$ is the inertia component.
$a_1 * r_{1n}(e)[pbest_{mn}(e) - p_{mn}(e)]$ is the cognitive component.
$a_2 * r_{2n}(e)[gbest_{mn}(e) - p_{mn}(e)]$ is the social component.
The a_1 and a_2 is positive cognitive coefficients whose value close to 2 and affects the step size of particle toward its pbest and gbest.

Step by step of PSO:

- First task is to initialize the position and velocities of a group of particles. This can be set arbitrarily. Initialize the speed $s_{mn}(0) = 0$.
- Particles fitness value is calculated by the equation (6).

$$f = C|r_m - p_m| \quad (6)$$

  where, C is the positive constraint; r_m is the realistic value. p_m is the expected value.
- $pbest$ is the best value in the observation of personal m and $gbest$ is the global best. These values are initially determined.
- The positions and velocities are updated continuously according to (4) and (5). Calculate the fitness value of each. Update fitness value of current position of personal and global comparing with initial.
- Checking the stopping condition, declare new gbest value is the optimal best value if terminated. if not, then repeat the calculation.
- Lastly, Initial parameters of NF is declared from the co-ordinates of gbest.

Algorithm 1: Proposed Vertical Hand-off Decision Algorithm

Input: $user_{pref}$, $channel_{utilization}$
begin
if ($current_{network} \neq user_{pref}$)
begin
if ($user_{pref}$ == no && $channel_{utilization}$ == low)
begin
Input I_c from available networks, N.
Train FNN by hybrid back propagation and least mean square algorithm.
For (Available FNN Layer)
Calculate output for each layer using (1)-(2)-(3).
Return output Y.
End For
Select network, $N \leftarrow Y : Y \in NFC$.
Perform hanfoff for network N.
end
elseif ($user_{pref}$ == yes && $channel_{utilization}$ == low)
Perform hanfoff for network, $N \leftarrow user_{pref}$.
else
no handaff.
end
else
no handoff.
end

D. VHO Decision

The hand-off decision is decided from the value of Y of NF controller. If the value Y is between 0 and 0.20, handoff will be performed to UMTS, If the value of Y is between 0.20 and 0.40, hand-off will be performed to GSM, If the value of Y is between 0.40 and 0.60, hand-off will be performed to Pico cell, if the value of Y is between 0.60 and 0.80, hand-off will be performed to LTE and If the value of Y is between 0.80 and 1.00, hand-off will be performed to WLAN. Markov chain process is able to calculate probability of transition from one state to another. Probability is calculated as:

$$p_{i,j} = p(s_{t+1} = i | s_t = j) \quad (7)$$

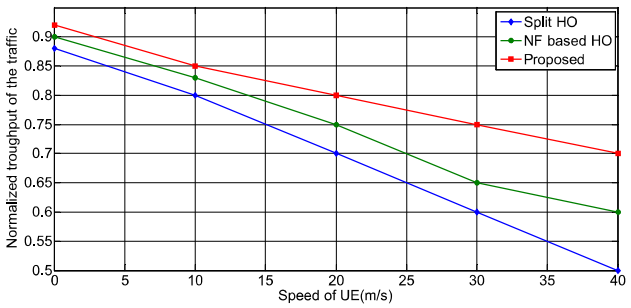


Fig. 4. Normalized throughput of traffic during handoff

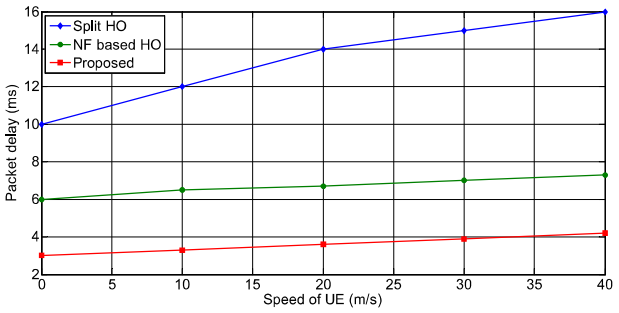


Fig. 5. Packet delay as a function of speed of UE

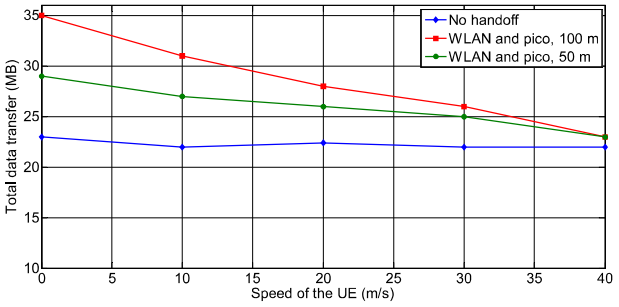


Fig. 6. Effect of UE speed on aggregated data rate

Here $i, j \epsilon s$ and s is the state space. GSM, picocell, LTE, UMTS and WLAN forms the state space for this case.

IV. Simulation and Results

This section includes the simulation setup and results. The simulation has been done in NS2 considering mobility of UEs. In the model we have considered LTE (4G), UMTS(3G), GSM(2G), WLAN and Pico networks.

A. Simulation Parameters

The simulation parameters are ordered in the Table 1.

TABLE I. Simple table

Parameters	Value
UMTS cell range (m)	1000
WLAN 802.11 cell range (m)	50-100
Pico cell range (m)	0-30
LTE cell range (m)	800
GSM cell range (m)	1500
Propagation model	Rayleigh
Antenna	Omni-directional
AP transmit power	0.02 to 0.06mW
Application Traffic for UE	Cluster based routing
Simulation duration (seconds)	500
Data rate	200Kb, 1Mbps, 4Mbps
Applications type	H.261, MPEG-4, HDTV
Packet Size (byte)	1000
WLAN 802.11 cell range (m)	512, 800, 1024

B. Simulation Result

When a UE resides in pico or WLAN coverage, the average RSSI is higher and better throughput can be achieved. Here we have compared our results with the split HO [26] and NF based HO [25]. Figure 4 shows the normalized throughput of traffic during handoff. It has been found that the normalized throughput of the proposed VHO algorithm outperform compared to NF based HO and split HO techniques.

Figure 5 shows the effect of UE's speed on the packet delay. It has been found that the packet delay is higher for the split HO compared to NF based HO and proposed VHO. The packet delay is less for the proposed HO algorithm. It is due to the fact that the proposed algorithm converged very fast compared to its counter parts.

Figure 6 illustrated the affects of UE's speed on total data rate. In this case, our proposed approach outperforms. It is because our approach reduces the HO frequencies compared to NF based HO and Split HO algorithms.

V. Conclusion

Firstly, network discovery is performed based on received signal strength. To avoid complexity, user preference and channel utilization is used as initial decision step. This also increases user satisfaction and avoid unnecessary hand-off. Then NF based VHO algorithm for the ubiquitous heterogeneous wireless network has been proposed to choose best network accurately. The parameters of VHO decision method based on service type, data-rate, time of staying in a cell without transition, network connection time and residue energy. The evaluation of the proposed method has been performed using NS 2 simulator. The result proves that the proposed method outperform the state-of-the-art method. This work can be extended by comparing the algorithm with other machine learning approaches and statistical methods.

References

[1] E. S. Navarro, Y. Lin and V. W. S. Wong, *An MDP-Based Vertical Handoff Decision Algorithm for Heterogeneous Wireless Networks*, IEEE Transactions on Vehicular Technology, vol. 57, no. 2, pp. 1243-1254, Mar. 2008, ISSN: 1558-2612, doi: 10.1109/WCNC.2016.7564804.

[2] X. Liu, J. Ling-ge and H. Chen, *A Novel Fuzzy Logic Vertical Handoff Algorithm with Aid of Differential Prediction and Pre-Decision Method*, Communications, 2007. ICC '07, IEEE International Conference on, Glasgow, pp. 5665-5670, June 24-28, 2007, ISBN: 1-4244-0353-7, doi: 10.1109/ICC.2007.939.

[3] X. Yan, N. Mani and Y. A. Sekercioglu, *A Traveling Distance Prediction Based Method to Minimize Unnecessary Handovers from Cellular Networks to WLAN*, in IEEE Communications Letters, vol. 12, no. 1, pp. 14-16, Jan. 16, 2008, ISSN: 1089-7798, doi: 10.1109/LCOMM.2008.071430.

[4] X. Yan, Y. A. Sekercioglu and S. Narayanan, *A survey of vertical handover decision algorithms in Fourth Generation heterogeneous wireless networks*, Computer Networks, vol. 54, no. 11, pp. 1848-1863, 2010.

[5] M. Adnan, H. Zen and A. Othman, *A Survey of Vertical Handover Decision Algorithms in Fourth Generation Heterogeneous Wireless Networks*, Asian Journal of Information Technology, vol. 13, no. 4, pp. 247-251, 2014, ISSN:1682-3915.

[6] N. Nasser, S. Guizani and E. Al-Masri, *Middleware vertical handoff manager: a neural network-based solution*, IEEE International Conference on Communications (ICC-07), Glasgow, Scotland, pp. 5671-5676, June 24-28, 2007, ISBN: 1-4244-0353-7, doi: 10.1109/ICC.2007.940.

[7] Z. Na, Y. Cui, Y. Xu and L. Chen,*A Novel Mobility Prediction Algorithm Based on LSVR for Heterogeneous Wireless Networks*, International Journal of Future Generation Communication and Networking, vol. 5, no. 3, Sep. 2012.

[8] K. Yang, I. Gondal, B. Qiu, and L. Dooley, *Combined SINR Based Vertical Handoff Algorithm for the Next Generation Heterogeneous Wireless Networks*, in the Proc. of IEEE Global Telecommunication Conference, pp. 4483-4487, Nov. 26-30,2007, ISBN: 978-1-4244-1042-2, doi: 10.1109/GLOCOM.2007.852.

[9] S. Mohanty and I. F. Akyildiz , *A Cross Layer (2 + 3) Handover Management Protocol for the Next Generation Wireless Systems*, IEEE Transaction on Mobile Computing", Vol. 5, pp. 1347-1360, Aug. 28, 2006, ISSN: 1536-1233, doi: 10.1109/TMC.2006.142.

[10] J. Hou and D. C. O'Brien, *Vertical handover-decision-making algorithm using fuzzy logic for the integrated Radio-and-OW system*, in IEEE Transactions on Wireless Communications, vol. 5, no. 1, pp. 176-185, Jan-2006.

[11] A. Hasswa, N. Nasser and H. Hassanein, *Tramcar: A Context-Aware Cross-Layer Architecture for Next Generation Heterogeneous Wireless Networks*, IEEE International Conference on, Istanbul, 2006, pp. 240-245, doi: 10.1109/ICC.2006.254734.

[12] W. Zhang, *Handover decision using fuzzy MADM in heterogeneous networks*, 2004 IEEE Wireless Communications and Networking Conference (IEEE Cat. No.04TH8733), 2004, pp. 653-658 Vol. 2, doi: 10.1109/WCNC.2004.1311263

[13] Y. Zhang, Z. Zheng and L. Chen, *A Cost-based Vertical Handoff with Combination Prediction of SINR in Heterogeneous Wireless Networks*, Journal of Theorotical and Applied Information Technology, vol. 49, no. 1, 2013.

[14] K. Ahuja, B. Singh and R. Khanna, *Particle swarm optimization based network selection in heterogeneous wireless environment*, Optik - International Journal for Light and Electron Optics, vol. 125, no. 1, pp. 214-219, 2014, ISSN 0030-4026.

[15] W. Nan, S. Wenxiao, F. Shaoshuai and L. Shuxiang, *PSO-FNN-Based Vertical Handoff Decision Algorithm in Heterogeneous Wireless Networks*, Procedia Environmental Sciences, vol. 11, pp. 55-62, 2011, ISSN 1878-0296, doi:10.1016/j.proenv.2011.12.010.

[16] Y. Yee, S. Tan, H. Lim and S. Chien, *Application of Particle Swarm Optimizer on Load Distribution for Hybrid Network Selection Scheme in Heterogeneous Wireless Networks*, ISRN Communications and Networking, vol. 2012, Article ID 340720, 7 pages, 2012. doi:10.5402/2012/340720

[17] S. Venkatachalaiah, R. Harris, and J. Murphy, *Improving Handoff in Wireless Networks using Grey and Particle Swarm Optimisation*, CCCT, vol. 5, pp. 368-373, 2004.

[18] S. Goudarzi, W. Hassan, M. Anisi, S. Soleymani and P. Shabanzadeh, *A Novel Model on Curve Fitting and Particle Swarm Optimization for Vertical Handover in Heterogeneous Wireless Networks*, Mathematical Problems in Engineering, vol. 2015, pp. 1-16, 2015, doi:10.1155/2015/620658.

[19] A. Calhan and C. Ceken, *An Optimum Vertical Handoff Decision Algorithm Based on Adaptive Fuzzy Logic and Genetic Algorithm*, Wireless Pers Commun, vol. 64, no. 4, pp. 647-664, 2010, doi:10.1007/s11277-010-0210-6.

[20] M. S. Kaiser, M. H. Chaudary, Raza Ali Shah and K. M. Ahmed, *Neuro-Fuzzy (NF) based relay selection and resource allocation for cooperative networks*, ECTI-CON2010: The 2010 ECTI International Confernce on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, Chaing Mai, 2010, pp. 244-248, ISBN: 978-1-4244-5607-9, .

[21] A. Calhan and C. Ceken, *Artificial Neural Network Based Vertical Handoff Algorithm for Reducing Handoff Latency*, Wireless Pers Commun, vol. 71, no. 4, pp. 2399-2415, 2012, doi:10.1007/s11277-012-0944-4.

[22] P. Haoliang, S. Wenxiao, L. Shuxiang and X. Chuanjun, *A GA-FNN based vertical handoff algorithm for heterogeneous wireless networks*, International Conference on Computer Science and Automation Engineering (CSAE), IEEE, Zhangjiajie, China, pp. 37-40, 2012, doi: 10.1109/CSAE.2012.6272723.

[23] M. S. Kaiser, Z. I. Chowdhury, S, Al Mamun, A. Hussain, and M. Mahmud, *A Neuro-Fuzzy Control System Based on Feature Extraction of Surface Electromyogram Signal for Solar-Powered Wheelchair*, Cognitive Computation, Mar. 24, 2016, doi:10.1007/s12559-016-9398-4.

[24] A. M. Vegni, G. Tamea, T. Inzerilli, R. Cusani, *A combined vertical handover decision metric for QoS enhancement in next generation network*, Proceeding WIMOB '09 IEEE International Conference on Wireless and Mobile Computing, Networking and Communications, Pages 233-238, 2009, doi: 10.1109/WiMob.2009.47.

[25] A. Calhan and C. Ceken, *An Adaptive Neuro-Fuzzy Based Vertical Handoff Decision Algorithm for Wireless Heterogeneous Networks*, 21st Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, Instanbul, 2010, pp. 2271-2276, doi: 10.1109/PIMRC.2010.5671693.

[26] C. Singhal, and S. Kumar, and S. De, and N. Panwar, and R. Tonde, and P. De, *Class-Based Shared Resource Allocation for Cell-Edge Users in OFDMA Network*, IEEE Transactions on Mobile Computing, pp. 48-60, Jan. 2014, doi: 10.1109/TMC.2012.210.

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

A Comprehensive Approach towards User-Based Collaborative Filtering Recommender System

Mahamudul Hasan†, Shibbir Ahmed*, Md. Ariful Islam Malik†, and Shabbir Ahmed†
Department of Computer Science and Engineering
†University of Dhaka, *Bangladesh University of Engineering and Technology, Dhaka, Bangladesh
Email: {munna09bd, shibbirahmedtanvin, malikariful}@gmail.com, shabbir@cse.du.ac.bd

***Abstract*—Recommender system refers to an information system that predicts the intuition of user observing behavior of all the users. Collaborative filtering is an approach in recommender system which produces recommendations based on similarity as well as knowledge of users' relationships to items. In this paper, we combine some traditional similarity metrics to find three types of similar users which are super similar, super dissimilar and average similar. We also introduce a new similarity metric which is used in case of average similar user pairs effectively. Finally we evaluate the proposed method for recommendation by experimenting with real data of Movielens as well as Epinions. Thus we can conclude that our proposed similarity metric paves the way to take a comprehensive approach towards user-based collaborative filtering recommender system and performs better than other traditional similarity metrics.**

***Index Terms*—Machine Learning; Recommender System; Information Retrieval; User-Based Collaborative Filtering.**

I. Introduction

The usage of World Wide Web is growing at an exponential rate. In order to meet users preference the the complexity of website is increasing. Today the huge amount of information in online make people puzzled on selection of some criteria. However the behavior of users can be tracked by the web browsing. Thus, Recommender systems can be used to categorize those information and recommend some sorts of acts after further processing. Such acts include preferable videos in *Youtube*,books in *Amazon*, friend finder in *Facebook*.

A recommender system is an information system that is used to predict the best items or products to the users according to their past behavior and possibly using other kinds of data [1]. By using that rating recommender system produces a rating model, then by using that rating model recommendation has been made to the users for further items according to the user preferences. To make predictions about a user's interests recommender system has to learn a user models. Thus, collaborative filtering select items based on the similarities between the preferences of different users which is discussed in [2].

The novelty of this work lies in designing a collaborative filtering recommender system model for the purpose of recommendation by using *super similar* and *super dissimilar* terminology and for *average similar* term using our proposed similarity metric. We take all popular similarity metrics for computing the *support matrix*. For a user, the *support matrix* provides us a decision of super similarity, average similarity or super dissimilarity among all other users. Here, we have considered the users preference that changes over time, as well as we have taken the confidence among users into account. The results that we have shown in performance evaluation chapter, show that our proposed recommender system along with this similarity model achieved better performances compared to state-of-the-art similarity metrics in terms of *average mean absolute error*, *coverage*, *precision* and *recall*.

II. Motivation

Capturing user preferences is a gigantic task. Simply asking the users what they want is too intrusive and prone to error, yet monitoring behavior unobtrusively and then finding meaningful patterns is difficult and computationally time consuming. Capturing accurate user preferences is however, an essential task if the information systems of tomorrow are to respond dynamically to the changing needs of their users. The motivation that we discuss here also takes a different stance to the general context of information filtering. As above, a vast amount of available content is assumed to already exist; the motivation of filtering is used to heightened user activity (which often translates to increased revenue for web-based businesses). Building tools like recommender systems, that offer personalised views of a web site's content to visiting users, encourages people to actively engage with the site: two thirds of the movies rented by Netflix.com were recommended, Google news recommendations result in 38% more clickthroughs, and 35% of the product sales on Amazon.com were recommended items [3]. Thus, we are trying to create a new similarity metrics considering this fact with the concept of confidence based similarity, and empirically evaluates the benefits associated with this approach compared to some traditional approach.

III. Related Work

Over the last few years, research into recommender systems has evolved: the particular target scenarios that have been explored have mirrored changes to the way people use the Internet. In the early 1990s, the first filtering system, Tapestry, was developed at the Xerox Palo Alto Research Center [4]. Similar concepts were later applied to *Usenet* news by the *GroupLens* research project, which extended previous work

978-1-5090-5769-6/16/$31.00

by applying the same principles to the Internet discussion forum, which had become too big for any single user to manage [5]. The initial success that recommender systems experienced is reflected in the surge of e-commerce businesses that implement them; Schafer et al. review and describe a number of mainstream examples [6], [7]. The cited sites, like *Amazon* and *CDNow*, implement recommenders to build customer loyalty, increase profits, and boost item-cross selling. The technology behind the web can be characterized as an information system composed of agents [8]. In the recent years, online recommender systems have begun to provide a technological proxy for this social recommendation process [9]. Recommender systems (RS) [10] are used in a variety of applications. Examples are web stores, online communities, and music players. Products can be based on the top overall sellers on a site, on the demographics of the consumers, or on an analysis of the past buying behavior of the consumers as a prediction for future buying behavior.

It is important to note that recently CF technologies with content based filtering technologies are usually integrated to provide powerful hybrid filtering solutions. Successful research has been done in projects like GroupLens [11], [12], Ringo [13], Video Recommander citeHSRF95 and MovieLens [14]. A variety of memory-based CF systems have been developed [15]. The Pearson correlation co efficient was used in GroupLens [12]. The Ringo project [13] focuses on testing different similarity metrics, including correlation and mean squared difference. Breese et al. [16] propose the use of vector similarity, based on the vector cosine measure often used in information retrieval systems. Lately, researchers have introduced dimensionality reduction techniques to address data sparsity [17], but as pointed out in [18], [19]. The resulting model of model-based CF systems is usually very small, fast and essentially as accurate as memory-based methods [16]. Combining separate recommendation approach implements content-based and collaborative techniques separately and combines their predictions [20] [21], [22]. A Recent study regarding user similarity computation for Collaborative Filtering Using Dynamic Implicit Trust [23] is also noteworthy.

Among the recent works, advances in collaborative filtering [24], user-based collaborative filtering for tourist attraction recommendations [25], an evolutionary approach for combining results of recommender systems techniques based on collaborative filtering [26] and an recommendation algorithm based on weighted Slope one algorithm and user-based collaborative filtering [27] are really concise and praiseworthy.

IV. Methodologies

In this research, we present our proposed support matrix that has been made from standard traditional similarity metrics. Through support matrix for a user we can cluster the group of users those who are super similar, average similar and super dissimilar. And then using our proposed similarity metrics we have calculated the rating for an item to an individual user. Then we select the top-N items for recommendation to a user.

A. Comparative Analysis of Traditional Similarity Metrics

Some flaw of Traditional Similarity Metrics [28] has been stated below:

- Low similarity regardless of the similar ratings by two users.
- High similarity regardless of the difference between the two users ratings.
- Ignoring the proportion of common ratings will lead low accuracy.
- Discarding the absolute value of rating will become difficult to distinguish different users.
- The combination of Jaccard and MSD can make up for the partial shortages of Jaccard and MSD.
- By taking into account all of the problems, no solution exist currently.

B. Motivation Towards Support Count

We have already analyzed the drawbacks of the traditional similarity measures and the improved variants. In most of the recommender systems, most users only rate a small number of items. In order to improve the recommended accuracy, we propose a support matrix model. For our cases we have taken only N=8 traditional similarity model into account so it provide us much more accurate and precious recommendation compared to the state of the art works.

C. Computing Support Matrix

In our experimentation we have used *8* different similarity metrics; PCC, SPCC, CPCC, ACOS, COS, JMSD, MSD and Jaccard [28]. There has been varieties of literature on these measures reporting their individual performances.

Another important fact regarding all of those individual similarity metric is that some provide good performance for some items or users but for rest of them might not follow this fact. Again some other work well in some other situation. So in total each of the individual performance have been combined to our *support matrix*. The algorithm of the support matrix also has been given in 1.

D. Similarity Analysis

From the algorithm 1, we get a support count between every possible users. Since we use 8 similarity metrics, so if we can find support count 8, that means all of the similarity metrics provide the decision of super similarity. Again, if we can find support count 0, that means non of the similarity provide similarity value greater than their median value, which implies the super dissimilarity. So, By analyzing this effect we can classify it as super similarity, average similarity, and super dissimilarity. For the case of super similar user pairs, we calculate the rating by setting the highest similarity between the users. Accordingly, for the super dissimilar user pairs we use a negative similarity. Finally, for the average similar user pairs we use the proposed confidence based similarity for the purpose of better computation.

Algorithm 1 Support Matrix Algorithm

INPUT: Train Data Sets.
OUTPUT: User-User Support Matrix.
for all $x \in U$ **do**
  for all $y \in U$ **do**
    $SupportMatrix_{(x,y)} \leftarrow 0$
  end for
end for
for all $SimilarityMetric \in TraditionalSimilarityMetrics$ **do**
  for all $x \in U$ **do**
    for all $y \in U$ **do**
      $sim_{(x,y)} \leftarrow CalculateSimilarity_{(x,y)}$
    end for
  end for
  $median \leftarrow CalculateMedian(Similarity_{(x,y)})$
  for all $x \in U$ **do**
    for all $y \in U$ **do**
      if $sim_{(x,y)} >= median$ **then** ▷ Similarity above Median
        $SupportMatrix_{(x,y)} \leftarrow SupportMatrix_{(x,y)} + 1$
      end if
    end for
  end for
end for

E. Confidence Based Similarity Computation

In this paper, we have proposed a confidence based similarity. We have already known to the confidence between two users. Confidence expresses the reliability of the affiliation between the users based on the number of co−rated items and influenced when the amount of co−rated items are changed. Confidence specify that a user is more reliable if his/her additional co−rated items are available in the system and the degree of reliability of an affiliation is influenced by the changes amount of co−rated items between the users. The confidence between the users is determined as the target user U_a is confident on recommender U_b as follows:

$$Confidence_{(a,b)} = \frac{I_a \cap I_b}{I_b} \tag{1}$$

Where $I_a \cap I_b$ represent the amount common items rated by both user and I_b denotes the total amount of items that recommender U_b rates in the system.

Now from other perspective, The confidence between the users is determined as the target user U_b is confident on recommender U_a as follows:

$$Confidence_{(b,a)} = \frac{I_a \cap I_b}{I_b} \tag{2}$$

Now, if we closely analyze the case is that, user U_a is confident on recommender U_b as well as user U_b is confident on recommender U_a, we can combine the confidence among them. So, the proposed similarity method that we named as Confidence Confidence Similarity is :

$$CCS_{(a,b)} = Confidence_{(a,b)} * Confidence_{(b,a)} \tag{3}$$

$$CCS_{(a,b)} = \frac{(I_a \cap I_b) * (I_a \cap I_b)}{I_a * I_b} \tag{4}$$

But one important fact is that , In some case some of the users tend to rate low, even they like the items very much. That is not taken into account in our Confidence Confidence Similarity method. So we have proposed our *Preference Network* below.

F. Preference Network

The Confidence Confidence Similarity *(CCS)* only considers the proportion of common rating, and does not consider the absolute value of rating. This leads to the difficulty of distinguishing between the users. Again, Mean Squared Difference ($MSD^{resnick's}$) only calculates the average difference between both users at current time, but ignores the proportion of common ratings. This may lead to the low accuracy. We have known the above mentioned similarity and theirs problem. Now we make a preference network by combining the time based *MSD* similarity and our the *CCS* similarity by the following formula 5. The combination of *CCS* and time based *MSD* can make up for the partial shortages of *CCS* and *MSD*. The algorithm of the preference network has been given in Algorithm 2.

Algorithm 2 Preference Network Algorithm

INPUT: Train Data Sets.
OUTPUT: User-User Similarity of Preference Network.
for all $x \in U$ **do** ▷ Calculate MSD
  for all $y \in U$ **do**
    $MSD^{resnick's}_{(x,y)} \leftarrow CalculateMSD(x,y)$
  end for
end for
for all $x \in U$ **do** ▷ Calculate CCS
  for all $y \in U$ **do**
    for all $i \in I$ **do**
      if $r_{x,i} \in TrainData_r$ **then** ▷ Calculate Time Difference
        $Numerator \leftarrow Numerator + 1$
      end if
    end for
    $ConfidenceSimilarity_{(x,y))} \leftarrow Numerator^2/(i_x \times i_y)$
  end for
end for
for all $x \in U$ **do** ▷ Calculate Preferred Network
  for all $y \in U$ **do**
    $PreferredNetwork_{(x,y))} \leftarrow ConfidenceSimilarity_{(x,y))} \times MSD^{resnick's}_{(x,y)}$
  end for
end for

$$Preference_{(a,b)} = MSD^{resnick's}_{(a,b)} * CCS_{a,b} \quad (5)$$

G. Proposed Prediction Method

For prediction purpose in the classical collaborative filtering prediction , with similarity defined in our own way. To calculate the predicted rating p^i_x for user x of an item i, the following Deviation From Mean (DFM) as aggregation approach is used in traditional system:

$$p^i_x = \bar{r_x} + \frac{\sum_{n \in k_x}[sim(x,n) \times (r^i_n - \bar{r_n})]}{\sum_{n \in k_x} sim(x,n)} \quad (6)$$

If we separate the deviated term (offset) it looks like :
For the user $x \in U$ and item $i \in I$ the offset

$$offset_{u,i} = \frac{\sum_{n \in k_x}[sim(x,n) \times (r^i_n - \bar{r_n})]}{\sum_{n \in k_x} sim(x,n)} \quad (7)$$

Now we introduce three offset term, For Super Similar Users.

$$SSO_{x,i} = \frac{\sum_{n \in k_x}[superSimilarity_{(x,n)} \times (r^i_n - \bar{r_n})]}{\sum_{n \in k_x} superSimilarity_{(x,n)}} \quad (8)$$

For Average Similar Users.

$$ASO_{x,i} = \frac{\sum_{n \in k_x}[Preference_{(x,n)} \times (r^i_n - \bar{r_n})]}{\sum_{n \in k_x} Preference_{(x,n)}} \quad (9)$$

For Super Dissimilar Users.

$$DSO_{x,i} = \frac{\sum_{n \in k_x}[superDissimilarity_{(x,n)} \times (r^i_n - \bar{r_n})]}{\sum_{n \in k_x} superDissimilarity_{(x,n)}} \quad (10)$$

But for our use, Using this function we propose an Algorithm 3 for prediction purpose which provide more accurate prediction.

Algorithm 3 Prediction algorithm , for each of the item for a user $r_{x,i}$

INPUT: Train Data Sets, Test Data Sets.
OUTPUT: Predicted Rating.

function CALCULATESIMILARITY(x, n)
  for all $x \in U$ **do**
    for all $n \in U$ **do**
      $S_{x,n} \leftarrow$CALCULATESUPPORTMATRIX(x, n)
      $sim_{x,n} \leftarrow$CALCULATEPREFSIMILARITY(x, n)
    end for
  end for
end function

for all $r_{x,i} \in TestData$ **do**
  for all $n \in U$ **do**
    if $r_{n,i} \in T_r$ **then**
      if $S_{x,n} >= Max$ **then** ▷ Here, Max = 5, For All Super Similar Users
        $offset_{x,i} \leftarrow offset_{x,i} + SSO_{x,i}$
      end if
      if $S_{x,n} = Avg$ or $S_{x,n} = Min$ **then** ▷ Here, Avg = 4, Min = 3, For All Average Similar Users
        $offset_{x,i} \leftarrow offset_{x,i} + ASO_{x,i}$
      end if
      if $S_{x,n} < Min$ **then** ▷ Here, Min = 3, For All Super Dissimilar Users
        $offset_{x,i} \leftarrow offset_{x,i} + DSO_{x,i}$
      end if
    end if
  end for
  $r_{x,i} \leftarrow \bar{r_x} + offset_{x,i}$
end for

V. EXPERIMENTATION

A. Experimental Setup

For our research purpose we actually focus on offline analysis.We use two datasets, Movielens and Epinions for the experimental purpose.

1) Movielens Dataset: The two data sets of MovieLens (http://www.movielens.umn.edu) and Epinions are used in our experiments. The first MovieLens dataset is called ML−100K, there are 100,000 ratings with 943 persons and 1682 movies. Another is ML−1M, it includes 6040 users and 3952 movies with 1,000,209 ratings. In both data sets, each person has rated at least 20 movies. The user profile includes age, sex, and profession. The movie includes 19 types. The density of the user−item matrix is 6.3% in ML−100K and 4.1% in ML−1M. For our experimentation we use ML−1M.

2) Epinions Dataset: Epinions data set is collected from (http://www.epinions.com). Epinions founded in 1999 is a product and shop review site where users can review items (such as movies, books, and software) and users can also assign items numeric ratings in the range 15. The Epinions data set consists of 49,289 users who have rated a total of 139,738 different items at least once. There are 40,163 users who have rated at least one item. The sparseness of the data set is hence more than 99.99%.

TABLE I
MAIN PARAMETERS OF THE DATABASE USED IN THE EXPERIMENTS

Datasets	Users	Items	Rating Range	Time
Movielens	6040	3952	1-5	In second
Epinions	49289	139738	1-5	In second

B. Experimental Evaluation

1) Identify Metrics: Throughout the years, recommender systems and other information filtering systems have been measured using a variety of metrics. The most commonly used metrics include *mean absolute error*, *precision*, *recall*

and *coverage*. Each metric has its strengths and weaknesses with respect to the task that is identified for the system.

Mean absolute error (MAE) belongs to a group of statistical accuracy metrics that compares the estimated ratings against the actual ratings. More specifically, MAE measures the average absolute deviation between a predicted rating and the user's true rating.

$$MAE = \frac{\sum_{i=1}^{n_a} |r_{a,i} - p_{a,i}|}{n_a} \tag{11}$$

Coverage is the percentage of the total data that were accurately predicted. For our cases, the more we predict an item with higher accuracy the more we increase the coverage of recommendation. The precision is the proportion of recommended items that the testing users actually liked in the testing set. This measure is also as high as possible for good performance. The precision is computed as follows:

$$Precision = \frac{n}{topN} \tag{12}$$

The recall score is the average proportion of items from testing set that appear among of the ranked list from the training set. This measure should be as high as possible for good performance. Assuming M_T is the number of items which are in the testing set and liked by the active user, n is the amount of items which the testing user likes and appears in the recommended list. Hence, the recall is computed as follows:

$$Recall = \frac{n}{M_T} \tag{13}$$

If we round the the calculated rating for the prediction purpose then we can predict an item closely to the users test rating that are given by the users.By rounding we got more accurate result than previous non rounding information.

2) Performance Evaluation: Here we show the results obtained using the databases specified in Table I. Figs. 1 and 2 show respectively the results obtained with Movielens and Epinions. As may be seen in these figures, the results obtained for all the quality measures (MAE, coverage, precision and recall) with our proposed method are better than that the ones obtained with the traditional metrics. For each of the user we have a set of users those who are similar too closely. By using the support matrix of all traditional similarity measures we compute the super similarity score. We have shown that MAE values we obtain for setting different values for super dissimilar users keeping the similarity value for super similar users constant. After selecting the super similar and super dissimilar user pairs the remaining part of the users are kept as average similar user.User preferences have been changed over time. So, by considering this fact we calculate our similarity metrics.

Graphic (i) in Fig.1 implies that the MAE error obtained for Movielens when applying Pearson Correlation Coefficient(PCC), Constraint Pearson Correlation Coefficient(CPCC), Cosine Similarity (COS),Jaccard Mean Squared Deviation (JMSD) and our proposed method named as Preference Similarity(Pref) . The our method leads to fewer errors, particularly in the case of when we use 90 percent of data as a training set. Since, at this case for training purpose we got enough data to train properly. Graphic (ii) informs about the coverage obtained. As may be seen, our method can improve the coverage from any other metrics. In relation to any traditional metric used. This fact must be emphasized since the new latter similarity metrics which are usually proposed improve the MAE while resulting in a worse coverage. Graphics 1c and 1d inform respectively about the precision and recall. These quality measures are improved for any value used in the number of recommendations, N. Consequently, the GA method improves not only the accuracy and the coverage, but also provides better recommendations.

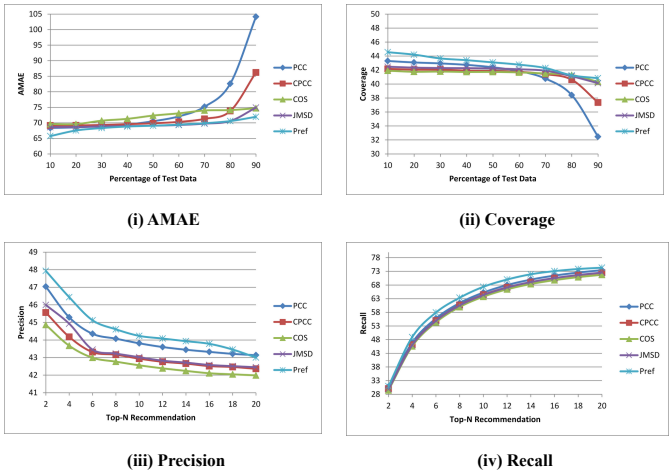


Fig. 1. Result using Movielens Dataset

3) Result Using Movielens DataSet: Graphic 1(i) informs about the MAE error obtained for Movielens for different metrics. Graphic 1(ii) implies about the coverage obtained. Our method can improve the coverage from any other metrics. In relation to any traditional metric used, this fact must be emphasized since the new latter similarity metrics which are usually proposed improve the MAE while resulting in a worse coverage. Graphics 1(iii) and 1(iv) indicates respectively about the precision and recall. These quality measures are improved for any value used in the number of recommendations, N. Consequently, the GA method improves not only the accuracy and the coverage, but also provides better recommendations.

4) Result Using Epinions DataSet: Graphic 2(i) in Fig. 2 informs about the MAE error obtained for Epinions dataset when applying Pearson Correlation Coefficient(PCC), Constraint Pearson Correlation Coefficient(CPCC), Cosine Similarity (COS),Jaccard Mean Squared Deviation (JMSD) and our proposed method named as Preference Similarity(Pref). Graph 2(ii) implies the coverage value. But most importantly the fact is that in case for precision and recall, our method improves its results in both case. This fact must be emphasized since

the new latter similarity metrics which are usually proposed improve the MAE while resulting in a worse precision and recall. In order to recommend an item to a user we always recommend the top most 10 or 20 item that he/she would be like most.For that case our method successfully predict the item above another traditional metrics.

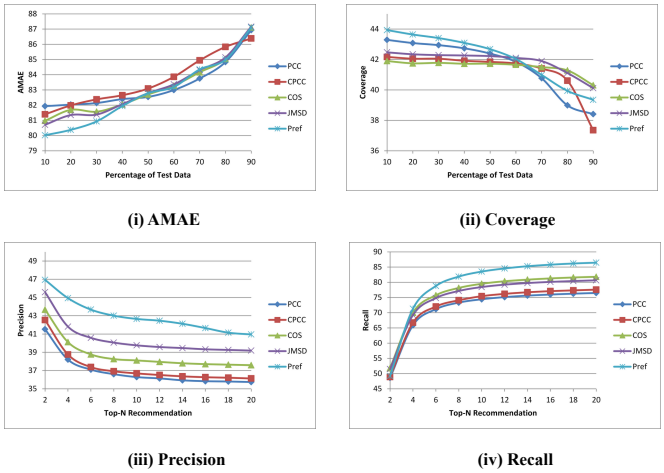


Fig. 2. Result Using Epinions DataSet

So, we are being able to show that our performance is quite good enough for the most used Movielens dataset . But for the case of Epinions we show that although our prediction not too accurate for the low rating (this might be $\in$ [1,3]) but for the *topN* recommendation it can being able to predict better than the another traditional similarity metrics.

VI. Conclusions

This paper has presented a new similarity along with the concept of super similarity, average similarity and super dissimilarity. At the time of prediction for a user, firstly we cluster from all user those who are super similar, average similar or super dissimilar ,Secondly by using our own similarity metrics along with the stated approach we predict the rating. After predicting all item we choose the top-N recommendation for the users that he/she might like that item or not. We have improved our performance over traditional similarity metrics to some acceptable amount.

References

[1] B. Shapira, F. Ricci, P. B. Kantor, and L. Rokach, "Recommender systems handbook." 2011.
[2] C. C. Aggarwal, *Recommender Systems: The Textbook*. Springer, 2016.
[3] O. Celma and P. Lamere, *Music Recommendation Tutorial*. Presented at the 8th International Conference on Music Information Retrieval, September 2007.
[4] D. Goldberg, D. Nichols, B. Oki, and D. Terry, "Using collaborative filtering to weave an information tapestry," *Communications of the ACM*, pp. 35:61–70, 1992.
[5] J. Konstan, B. Miller, D. Maltz, J. Herlocker, L. Gordon, and J. Riedl, "Grouplens: Applying collaborative filtering to usenet news," *Communications of the ACM*, pp. 40:77–87, 1997.
[6] J. Schafer, J. Konstan, and J. Riedl, "Recommender systems in e-commerce," *ACM Electronic Commerce*, pp. 158–166, 1999.
[7] J. Schafer, J. Konstan, , and J. Riedl, "E-commerce recommendation applications," *The Knowledge Engineering Review*, pp. 5:115–153, 2001.
[8] "Architecture of the world wide web, volume one." World Wide Web Consortium (W3C), December 2004.
[9] K. Swearingen and S. Rashmi, "Interaction design for recommender systems," *Designing Interactive Systems*, 2002, citeseer.ist.psu.edu/swearingen02interaction.html.
[10] P. Resnick and H. R. Varian, "Recommender systems, volume 40," 1997, http://doi.acm.org/10.1145/245108.245121.
[11] J. A. Konstan, B. N. Miller, D. Maltz, J. L. Herlocker, L. R. Gordon, and J. Riedl, "Grouplens: Applying collaborative filtering to usenet news, volume 40," *ACM Press, citeseer.ist.psu.edu/konstan97grouplens.html*, pp. 77–87, 1997.
[12] P. Resnick, N. Iacovou, M. Suchak, P. Bergstorm, and J. Riedl, "Grouplens: An open architecture for collaborative filtering of netnews," *Proceedings of ACM Conference on Computer Supported Cooperative Work*, 1994, citeseer.ist.psu.edu/resnick94grouplens.html.
[13] U. Shardanand and P. Maes, "Social information filtering: algorithms for automating word of mouth," *ACM Press/Addison-Wesley Publishing Co.*, pp. 210–217, 1995, http://doi.acm.org/10.1145/223904.223931.
[14] B. J. Dahlen, J. A. Konstan, J. L. Herlocker, N. Good, A. Borchers, and J. Riedl, "Jump-starting movielens: User benefits of starting a collaborative filtering system with."
[15] D. Pennock, E. Horvitz, S. Lawrence, and C. L. Giles, "Collaborative filtering by personality diagnosis: A hybrid memory- and model-based approach," *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence, UAI*, 2000, citeseer.ist.psu.edu/pennock00collaborative.html.
[16] J. S. Breese, D. Heckerman, and C. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering," 1998, citeseer.ist.psu.edu/breese98empirical.html.
[17] J. Wang, A. P. de Vries, and M. J. T. Reinders, "Unifying user-based and item-based collaborative filtering approaches by similarity fusion," *SIGIR 06: Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, 2006, http://doi.acm.org/10.1145/1148170.1148257.
[18] Z. Huang, H. Chen, and D. Zeng, "Applying associative niques to alleviate the sparsity problem in collaborative filtering." 2004, http://doi.acm.org/10.1145/963770.963775.
[19] G. R. Xue, C. Lin, Q. Yang, W. Xi, H. J. Zeng, and Y. Yu, "Scalable collaborative filtering using cluster-based smoothing." 2005, http://doi.acm.org/10.1145/1076034.1076056.
[20] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE Transactions on Knowledge and Data Engineering*, 2005, http://dx.doi.org/10.1109/TKDE.2005.99.
[21] M. J. Pazzani, "A framework for collaborative, content-based and demographic filtering," *Artificial Intelligence Review*, 1999, citeseer.ist.psu.edu/pazzani99framework.html.
[22] M. Claypool, A. Gokhale, T. Miranda, P. Murnikov, D. Netes, and M. Sartin, "Combining content-based and collaborative filters in an online newspaper," 1999, citeseer.ist.psu.edu/article/claypool99combining.html.
[23] F. Roy, S. M. Sarwar, and M. Hasan, "User similarity computation for collaborative filtering using dynamic implicit trust," in *International Conference on Analysis of Images, Social Networks and Texts*. Springer, 2015, pp. 224–235.
[24] Y. Koren and R. Bell, "Advances in collaborative filtering," in *Recommender Systems Handbook*. Springer, 2015, pp. 77–118.
[25] Z. Jia, Y. Yang, W. Gao, and X. Chen, "User-based collaborative filtering for tourist attraction recommendations," in *Computational Intelligence & Communication Technology (CICT), 2015 IEEE International Conference on*. IEEE, 2015, pp. 22–25.
[26] E. Q. da Silva, C. G. Camilo-Junior, L. M. L. Pascoal, and T. C. Rosa, "An evolutionary approach for combining results of recommender systems techniques based on collaborative filtering," *Expert Systems with Applications*, vol. 53, pp. 204–218, 2016.
[27] P. Wang, Q. Qian, Z. Shang, and J. Li, "An recommendation algorithm based on weighted slope one algorithm and user-based collaborative filtering," in *Control and Decision Conference (CCDC), 2016 Chinese*. IEEE, 2016, pp. 2431–2434.
[28] A. M. H. T. X. Z. Haifeng Liu, Zheng Hu, "A new user similarity model to improve the accuracy of collaborative filtering," *Knowledge-based systems*, vol. 56, pp. 156–166, 2014.

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

Comparison of Cyclostationary and Energy Detection in Cognitive Radio Network

Risala Tasin Khan
Institute of Information Technology
Jahangirnagar University
Savar, Bangladesh
risala@juniv.edu

Md. Imdadul Islam
Computer Science and Engineering
Jahangirnagar University
Savar, Bangladesh
imdad@juniv.edu

Shakila Zaman
Institute of Information Technology
Jahangirnagar University
Savar, Bangladesh
shakila.ju@gmail.com

M. R. Amin
Electronics and Communication Engineering
East West University
Dhaka, Bangladesh
ramin@ewubd.edu

***Abstract*—In cognitive radio network (CRN) three types of spectrum sensing techniques are widely used: matched filter detection, cyclostationary detection, and energy detection. Matched filter detection technique provides maximum SNR at receiving end but the detector need to be matched with the input signal i.e. impulse response of the system need to be made as the delayed version of mirror image of input signal. If the modulation scheme and window function of baseband signal of primary user (PU) is not known to a secondary user (SU) then it is difficult to use matched filter detection. In this paper we only consider cyclostationary detection and energy detection of their simplicity compared to matched filter detection. Main focus of the paper is to measure the probability of successful access and corresponding delay of a SU in CRN. We found that energy detection technique is better compared to cyclostationary detection.**

***Index Terms*—Mean delay; probability of false alarm; SNR; matched filter detection and Winner filter detection.**

I. INTRODUCTION

Recently, cognitive radio network (CRN) is getting huge attention due to its nature of accessing the spectrum opportunistically which is considered as the solution of the underutilized spectrum access problem. In CRN the secondary user (SU) senses the spectrum to find a hole before trying to transmit anything [1].Appropriate sensing time of information is essential to make sure whether a user is present or not in CRN. Different types of detection algorithms are used to sense the presence of PU and SU to reduce the spectrum conflict of PU and SU. The accuracy of sensing is call requirement to accurate sensing which provides successful access operation of CRN. For known condition, energy detection (symbolically expressed as ED) technique of [2-4] is the simplest way of sensing. Actually the threshold value of received signal energy is the key parameter of ED. If the energy of received signal is greater than the threshold then the detector assumes the presence of a user otherwise the physical channel is free. A cognitive user is permitted to transmit signal if the spectrum and time slot of the corresponding physical channel is found free. If the sensing time of a detector is kept larger the detection accuracy becomes higher but speed of the network is deteriorated. Therefore, several studies have been done regarding optimal spectrum sensing time [5-7]. No matter how precisely spectrum sensing is done or whatever technique is used for spectrum sensing, there is always a probability of detecting the presence/absence of PU falsely which affects the spectrum sensing time.In [7], the ROC curves are shown to present the false alarm measurement against the spectrum sensing time. The spectrum handoff may be performed for the condition where multiple SUs trying to access in the same spectrum band. When several SUs tries to use the spectrum at the same time, they struggle to get access opportunity. One of the simplest technique of channel access is the combination of RTS (Request To Send) and CTS (Clear To Send) used in Internet and wireless LAN under IEEE 802.11. This scheme is proposed for channel access by SUs in [8-9]. In this scheme, a RTS packet is send by a node to an access point (AP).Access point is responsible to provide a clear signal for accessing the required channel. After getting the RTS packet, AP checks whether another node is also sending an RTS packet. If there is no available RTS packet without the first one, the AP announce a CTS packet for the all node and broadcast it to get the acquire channel for a specific node.

In [10], authors provides the way to minimize the access delay of a user when many SUs try to access the traffic channel at the same time. In Cyclostationary detection method, cycle frequency domain profile is used in the existence of low signal to noise ratio (SNR) and interference, which helps to increase spectral efficiency [11-12]. To detect the spectrum of primary user, the most essential requirement for cyclostationary detection is the information of at least one cycle frequency of the PU's signal [13].

The rest of the paper is organized as: section II provides statistical model to evaluate the mean delay of access, section III provides analytical and simulation results based on study of section II and section IV concludes the paper and provides the way of further improvement of the work.

II. SYSTEM MODEL

Let the test signal, $T = \frac{1}{N}\sum_{n=0}^{N-1}|x(n)|^2$ composed of N samples and $N = \tau F_S$.

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

Where τ is the sensing period on a traffic/control channel and F_s is the sampling frequency of modulated wave. Now the decision levels of two hypothesis mode,

$$L = \begin{cases} T < \xi; & H_0 \\ T \geq \xi; & H_1 \end{cases} \quad (1)$$

; where ξ is the threshold value of test statistics T.
The probability of false alarm and detection [14-15],

$$P_F = Q\left[\left(\frac{\xi}{\sigma_R^2} - 1\right)\sqrt{\tau F_S}\right] \quad (2)$$

$$P_D = Q\left[\left(\frac{\xi}{\sigma_R^2} - 1 - \rho\right)\sqrt{\frac{\tau F_S}{1+2\rho}}\right] \quad (3)$$

Where σ_R^2 is the variance of received signal, $\rho = \sigma_R^2 / \sigma_n^2$ called SNR and σ_n^2 is the variance of noise. The function $Q(y)$ is defined as,

$$Q(y) = \frac{1}{\sqrt{2\pi}} \int_y^\infty e^{-u^2/2} du = \frac{1}{2} erfc\left(\frac{y}{\sqrt{2}}\right) \quad (4)$$

Let m users are in competition to access a traffic channel. Probability of success a particular SU is $(1\text{-}p)^{m-1}p$. Therefore the probability of success of anyone is, $m(1\text{-}p)^{m-1}p$. Here p is the probability that a particular SU sends RTS to get access on a traffic channel.
Therefore the probability of failure of all SUs is,

$$P_{f,m} = 1\text{-}\, m(1\text{-}p)^{m-1}p. \quad (5)$$

If there are total M users in the network then the probability of m SUs are in competition is,

$$P_{a,m} = \binom{M}{m} P_f^{M-m}(\tau)\left\{1 - P_f^{M-m}(\tau)\right\}^m \quad (6)$$

; where M-m SUs are not in competition due to getting false alarm.
The probability of successful access of a SU in an access period where the SU is in competition with m other SUs,

$P_{s,m}$ = *success in first attempt* $\vee$ *failure in* 1^{st} *TS but success in* 2^{nd} *TS* $\vee$ *failure in* 1^{st} *two TSs but success in* 3^{rd} *TS* $\vee$ ... ... $\vee$ *failure in* 1^{st} K-1 *consecutive TSs but success in Kth TS*

$$= (1\text{-}p)^{m-1}p + (1\text{-}p)^{m-1}p\, P_{f,m} + (1\text{-}p)^{m-1}p\, P_{f,m}^2 + \ldots \ldots \ldots + (1\text{-}p)^{m-1}p\, P_{f,m}^{K-1}$$

$$= (1\text{-}p)^{m-1}p \sum_{k=0}^{K-1} P_{f,m}^{K-1}$$

$$= (1\text{-}p)^{m-1}p \frac{1 - P_{f,m}^K}{1 - P_{f,m}} \quad (7)$$

Now the probability of success of a SU on an access period is,

$$P_S = \sum_{k=0}^{K-1} P_{s,m} P_{a,m} \quad (8)$$

The expected value of delay when m SUs are in competition,

$$\bar{D}_{S,m} = \sum_{k=0}^{K-1} k t_a P_{f,m}^k (1-p)^{m-1} p \quad (9)$$

; where t_a is the length of TS in access period.
The overall expected delay,

$$\bar{D}_w = \sum_{m=0}^{M} \bar{D}_{S,m} P_{a,m} + \tau \quad (10)$$

In the one-order cyclostationary detection technique, the probability of false alarm,

$$P_{F,S} = 1 - \exp\left(-\frac{\xi}{2\sigma_R^2}(2\tau F_S + 1)\right) \quad (11)$$

The two-stage cyclostationary detection technique the probability of false alarm,

$$P_{F,S2} = P_{F,S} P_F \quad (12)$$

The right side of (12) will be found from (2) and (11).

III. RESULTS

Taking typical value of parameters: F_s = 6MHz, τ = 4ms, K = 100, t_a = 0.64ms and p = 1/3 of [10] the profile of expected delay $\bar{D}_w$ and probability of success of a SU P_S on an access period against the number of SUs is shown in fig.1 and 2 respectively taking the probability of false alarm as a parameter. For the increasing number of users the mean delay is also increased since the more SU participate in the sizing traffic channel. Again $\bar{D}_w$ decreases with large false alarm probability (here we consider high value of P_F as 0.14, 0.22, 0.28 and 0.33 to observe the impact of false alarm) since more SUs abstain from participation of seizing traffic channel at higher P_F. The profile of probability of success P_S of SU is revered to $\bar{D}_w$; where P_S decreases with increase in the number of users M or decrease in P_F.

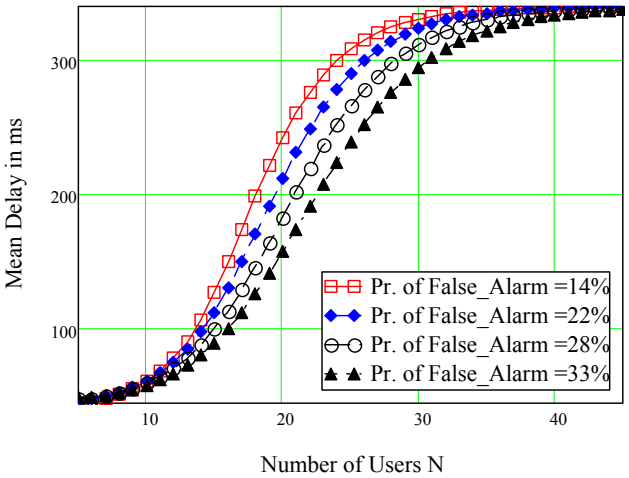


Fig. 1The variation of expected delay against the number of SUs on a network (p = 1/3)

The profile of $\bar{D}_w$ is again shown in fig.3 for p = 1/10 i.e. the probability of a SU to be active is now reduced. Now the network can accommodate more users for same access delay is visualized from fig.3. The impact of received SNR in dB is shown in fig.4. The PF decreases with increase in received SNR hence the same profile of found in fig.4 like fig.3.

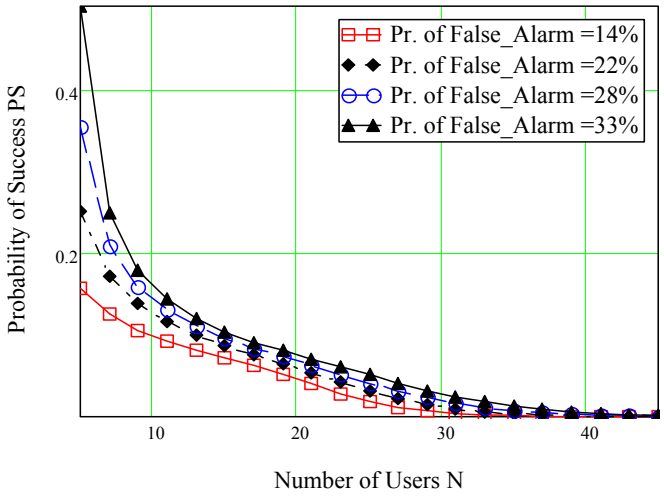


Fig. 2The variation of probability of success of a SU on an access period against the number of SUs

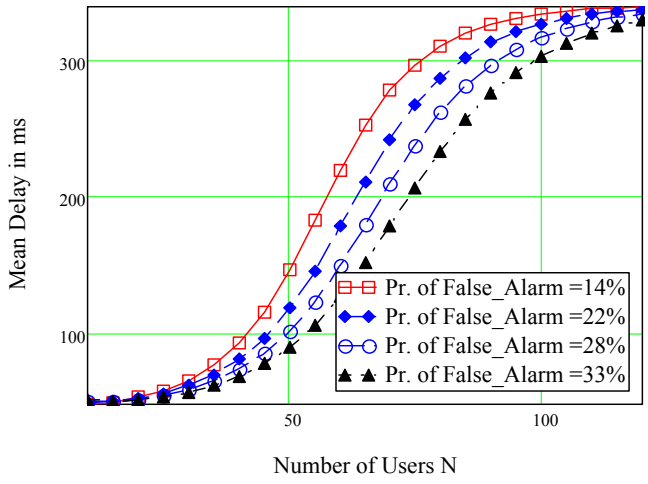


Fig. 3The variation of expected delay against the number of SUs on a network (p = 1/10)

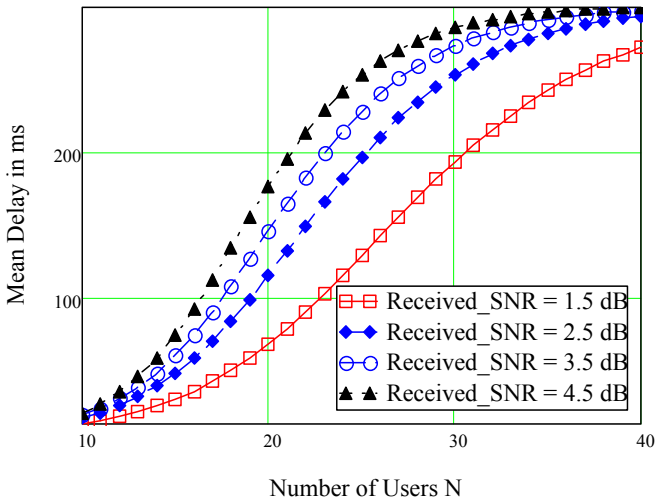


Fig. 4 The variation of expected delay against the number of SUs taking received SNR as a parameter

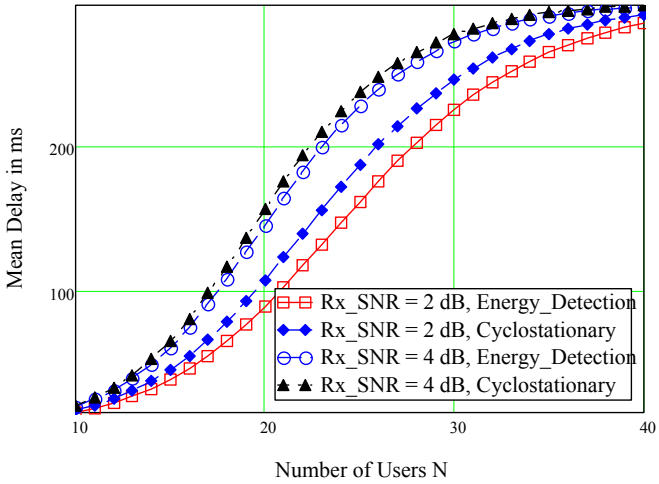


Fig.5 Comparison of energy detection and 2-stage cyclostationary detection

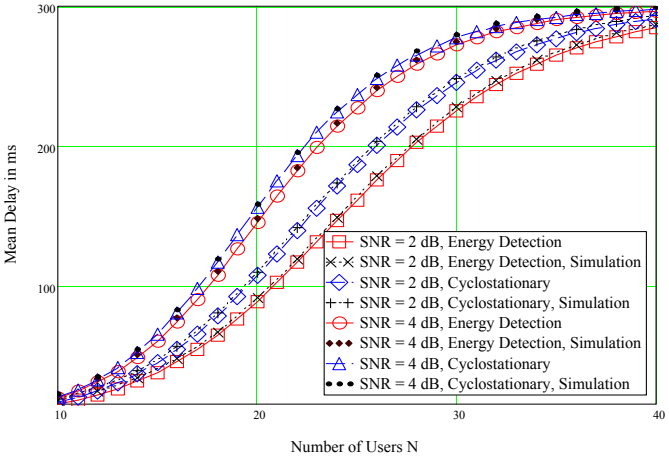


Fig.6 Comparison of analytical and simulation results of energy detection and 2-stage cyclostationary detection

Finally fig.5 compared energy detection and cyclostationary detection technique. The two stage cyclostationary detection technique provides lower P_F; more SUs are in completion therefore the delay of access will be higher for cyclostationary detection. Two pairs of curves are drawn for received SNR (symbolically used in graph as R_x_SNR) of 2dB and 4dB; where the delay of 4dB case is higher for the same reason mentioned previously.

Finally the analytical results of fig.5 are verified with simulation, shown in fig.6 where the results show +95% confidence level.

IV. Conclusions

In this paper the impact of the number of participant SUs on mean delay of access and the probability of success of SUs are analysed taking the false alarm probability as a parameter. Finally the performance of energy detection and cyclostationary detection are compared analytically and verified with simulation. We excluded the matched filter detection since it requires distinct reference signal (time/phase synchronization and modulation scheme) otherwise the performance will be heavily deteriorated. Finally, the size of cells is reduced in 3G and 4G mobile hence it becomes very difficult to maintain cell boundaries. In this case the received SINR becomes weighted squared sum of SINR from

surroundings therefore the Wiener filter detection technique of [16] can be applied for such network to compare with fig.5.

REFERENCES

[1] J. Mitola, 'Cognitive Radio: an integrated agent architecture for software defined radio,' Doctor of technology thesis, Royal Inst. Technol. (KTH), Stockhom, Sweden, 2000.

[2] Risala T. Khan, Tanzilah Noor Shabnam, Md. Imdadul Islam and M. R. Amin, 'Enhancement of Performance of Cognitive Radio Network with Incorporation of MRC Scheme at Secondary Receiver,'*IACSIT International Journal of Engineering and Technology*, Vol. 4, No. 4, pp. 495-499 ,August 2012.

[3] Risala T. Khan, Md. Imdadul Islam and M. R. Amin, 'Traffic analysis of Cognitive radio Network under the concept of Medium Access Model,'*Journal of Information Processing System*, Vol. 10, No. 04, pp. 602-617, Dec. 2014

[4] JJ Lehtomaki, R Vuohtoniemi, K Umebayashi, 'On the measurement of duty cycle and channel occupancy rate,' IEEE J. Sel. Areas Commun, 31(11), pp. 2555-2565, 2013.

[5] S. Stotas and A. Nallanathan, 'Optimal sensing time and power allocation in multiband cognitive radio networks,'*IEEE Trans. Commun.,* vol. 59, no. 1, pp. 226-235, Jan. 2001

[6] Y. C. Liang, Y. Zong, E. C. Y. Peh, and A. T. Hoang "Sensing-throughput tradeoff for cognitive radio networks," *IEEE Trans. Wireless Commun*., vol. 7, no. 4, pp. 1326-1337, Apr. 2008

[7] D. J. Lee and M. S. Jang, 'Optimal spectrum sensing time considering spectrum handoff due to false alarm in cognitive radio networks,'*IEEE Commun, Lett.,* vol. 13, no. 12, pp. 899-901, Dec. 2009

[8] J. Jia, Q. Zhang and X. Shen, 'HC-MAC: a hardware-constrained cognitive MAC for efficient spectrum management,'*IEEE J. Sel. Areas. Commun.*, vol. 26, no. 1, pp. 106-117, Jan. 2008.

[9] Q. Zhao, L. Tong, A. Swami and Y. Chen, 'Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: a POMDP framework,'*IEEE J. Sel. Areas, Commun*, vol. 25, no. 3, pp. 589-600, Apr. 2007.

[10] Dong-Jun Lee and Hyuk Wu, 'Spectrum Sensing Time Minimizing Access Delay of IEEE 802.11-Like MAC in Cognitive Radio Networks,' *IEEE Communications Letters*, Vol. 15, No. 11, pp. 1249-1251, Nov. 2011.

[11] DeepaBhargavi, Chandra R. Murth, 'Performance Comparison of Energy, Matched-Filter and Cyclostationarity-Based Spectrum Sensing,'*Signal Processing Advances in Wireless Communications*, pp. 1-5, 20-23 June 2010.

[12] Artem Tkachenko, Danijela Cabric, and Robert W. Brodersen, 'Cyclostationary Feature Detector Experiments using Reconfigurable BEE2,' *2007 2nd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks,* pp.216 – 219, 17-20 April 2007.

[13] JarmoLunden, Visa Koivunen, AnuHuttunen, and H.Vincent Poor, 'CollaborativeCyclostationary Spectrum Sensing for Cognitive Radio Systems,'*IEEE Transactions on Signal Processing,* Vol.57, no.11, pp. 4182 – 4195, 12 June 2009.

[14] Shokri-Ghadikolaei, H., Abdi, Y., Nasiri-Kenari, 'Analytical and learning-based spectrum sensing time optimization in cognitive radio systems,' *IET Communications*, 7(5): 480-489, 2013.

[15] Hong Du, Shuang Fu and Tingyi Shang, 'A Spectrum Handoff Scheme by Considering the Interference Probability in Cognitive Radio,' International *Journal of Future Generation Communication and Networking*, Vol. 9, No. 3, pp. 69-78, 2016.

[16] M. Mohammad Karimi, B. Mahboobi and M. Ardebilipour, 'Optimal spectrum sensing in fast fading Rayleigh channel for cognitive radio,' *IEEE Communications Letters*, vol.15, no.10, pp.1032-1034, Oct'2011.

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

Noise Aware Level Based Routing Protocol for Underwater Sensor Networks

Arnisha Akhter, Md. Ashraf Uddin
Department of Computer Science and Engineering
Jagannath University
Dhaka- 1100, Bangladesh
Email: arnishaakhter@yahoo.com,
ashraf@cse.jnu.ac.bd

Md. Anwarul Islam Abir, Md. Manowarul Islam
Department of Computer Science and Engineering
Jagannath University
Dhaka- 1100, Bangladesh
Email: anwarulislamabir@gmail.com,
mano.cse.du@gmail.com

Abstract— **In marine science, underwater sensor networking (UWSN) is the most important and interesting arena. A novel noise aware Level Based Routing Protocol (LBRP) is proposed in this paper to contribute in this field. The LBRP is essentially a level-based routing protocol. The sensor nodes need not to store any state information and only a small fraction of the total nodes are involved in routing. It also promises the best use of energy consumption and ensures more reliable packet transmission in comparison with others. To guide a packet from source to destination, LBRP calculates the noise, account residual energy, the level in which the node is situated and distance from the forwarding node. The forwarding node is selected calculating a formula proposed in this paper that uses the described terms. Simulation has been experienced both in noisy and noiseless environment and executed to endorse better performance of the proposed protocol. The LBRP aims to provide a robust, scalable and energy efficient under water routing protocol.**

Keywords— Underwater Sensor; Sensor Node; Fitness; Lifetime; Noise; Routing.

I. INTRODUCTION

With the provision of remote real-time wireless data access, UWSN enables real time monitoring of selected ocean areas. A number of issues need to be addressed while using sensor networks as an effective technology for underwater systems [1]. Autonomous and individual sensor nodes perform data collection operations as well as store and forwarding operations to route the data that has been collected to a central node. Limited battery storage and computation, low bandwidth and high error rates are the main challenges of deploying an underwater network. In addition, noisy environments, node failure, scalability to a large number of sensor nodes, survivability in harsh environment are common in underwater sensing and routing. Water current drives the nodes away in the speed of 3-6 km/h in average [2]. Underwater sensor nodes use acoustic modem with propagation speed 1500 m/s [3]. The proposed protocol is devised by cogitating upon limited battery and limited bandwidth. This protocol provides shorter end-to-end delay, evades control packet to guide data packet to the destination entirely this hoards up huge amount of total energy and considers the noise which is calculated for finding the better fitness of the nodes. Key contributions of the proposed routing protocol are listed as follows:

- Level based architecture
- Considers noise level in link
- Initiates partial path routing technique based on noise and the fitness of the nodes

This section describes the potentials and contributions of this paper. The works related to the theme of this paper are described in the second section. The following section three proposes the Noise Aware Level Based Routing Protocol and Multicasting Technique using fitness for UWSNs. Section four shows the simulation and performance analysis of the proposed protocol. The paper is concluded by section five along with future research directions.

II. RELATED WORKS

It is challenging to design a robust, scalable and energy-efficient underwater routing protocol because of limited life of battery. Moreover, most of the research works concerning UWSN are done on the issues related to physical layer. A new arena of the network layer is routing techniques. Providing an efficient routing protocol is the demand of time today.

Vector-based forwarding (VBF) [3] guides the packet from the source to the destination only considering the nodes that are within the range R of the vector. Thus the energy of the network is saved. Nodes are mobile and high communication time in dense network is needed for multiple nodes act as relay nodes. To reduce the high communication time and to handle node mobility, VBF routing protocol has been modified and HH-VBF (Hop by hop VBF) [4] routing protocol is proposed. Another work regarding energy analysis is a routing protocol for underwater wireless sensor networks [5] which is based on analyzing the total energy consumption in underwater acoustic sensor networks. Different functioning principles (relaying, direct transmission and clustering) are proposed and the total energy consumption for each case is analyzed.

FBR (Focused Beam Routing) [6] is a scalable routing technique for multi-hop ad-hoc networks, based on location information and it is suitable for networks containing both

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

static and mobile nodes. Link Expiration Time-Aware [LETA] Routing Protocol [7] is based on a sending node forwarding a data packet after being sure that the packet reaches the forwarding node, and acknowledgment is returned to the sending node after receiving the data packet. The forwarding decision of node is taken by applying Bayes' uncertainty theorem. Optimizing Energy through Parabola Based Routing (PBR) [8] uses parabola as forwarding region and it emphasizes on noise. This protocol cannot predict the lifetime of a node, link capacity of a path and assumption on node mobility.

The Depth Based Routing Protocol named as DBR [9] only uses the depth information of the nodes in the routing process. Depending on the depth, the data forwarding decisions are taken in this protocol. This method not only improves the overall energy consumption of the network, but also reduces the end-to-end delay. The New Energy Efficient and Depth based Routing Protocol (EEDBR) [10] only uses the depth information of the nodes in the routing process. This method not only improves the overall energy consumption of the network, but also reduces the end-to-end delay. A new distributed cross-layer Channel-aware Routing Protocol (CARP) [11] is proposed for multi-hop delivery of data to the sink. Depending on the history of successful transmission of data, nodes are selected to establish a robust link. Combination of link quality with simple topology information is the feature of CARP.

All of these discussed routing protocols for UWSNs works well in their own ways. This paper proposes a routing protocol to overcome the noise problem [12] and to increase the reliability of packet forwarding through creating a partial path that means selecting a few numbers of nodes based on the lifetime [13].

III. NOISE AWARE LEVEL BASED ROUTING PROTOCOL FOR UNDERWATER SENSOR NETWORK

In this section, the proposed Efficient Noise Aware Level Based Routing Protocol (LBRP) is presented in detail. Network aarchitecture, protocol overview and protocol design are discussed. The routing algorithm and evaluation of the proposed routing protocol is demonstrated at the end of this section.

A. Network Architecture

We assume that water depth is divided in different levels. An example of such networks is sketched in figure 01.

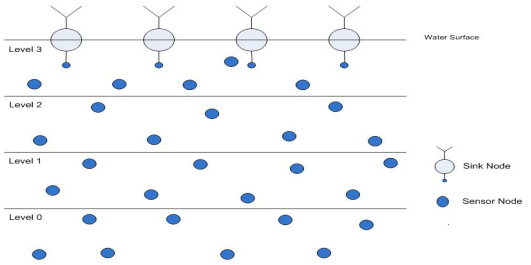


Figure 01: Noise Aware Level Based underwater sensor network architecture

Water depth is divided in different levels which are calculated by dividing the water depth by the level distance. The multiple sink nodes available in the proposed network architecture are fixed; whereas forwarding nodes are always mobile. The water surface nodes, called sink nodes are equipped with the modem that is capable of capturing both radio-frequency and acoustic signal. The nodes deployed in the underwater environment can send and receive only acoustic signal. Underwater sensor nodes with acoustic modems are placed in the interested 3 - area and each such node is assumed likely to be a data source. Underwater acoustic nodes can accumulate data and also assist to convey data to the sinks. When a sink node receives a packet from an underwater acoustic node, the sink nodes can converse with each other efficiently via radio channels.

B. Overview of Efficient Noise Aware Level Based Routing Protocol for UWSNs

The proposed protocol consists of two phases, named as- Candidate Node Selection for creating partial path and Forwarding data packets in Partial Path.

i. Candidate Node Selection for Creating Partial Path

The node selection phase follows the subsequent steps to create a partial path.

- An underwater sensor node forwards packets to the higher level sensor nodes. The forwarding node first calculates four levels in a rectangular shape having fixed height and width as shown in figure 2. The rectangle is calculated in the upward direction.
- Each level contains a number of sensor nodes. When the forwarding node broadcasts a control packet to the nodes in adjacent level, only the best fitted nodes responds to the forwarding node. This is represented in figure 3.
- The fitness probability of each node is calculated based on the lifetime of the node and noise. A threshold value is considered for each node; the fitness being less than the threshold, the node does not respond to the forwarding node.
- The best fitted node sends data about the fitness probability of selected nodes to the forwarding node to be maintained in the forwarding data table, illustrated in figure 3. If the average fitness of the nodes of a level is better, then only a few nodes are selected; if not, a large number of nodes are selected.
- Node with the highest fitness probability again broadcasts the control packet to the adjacent higher level nodes, shown in figure 4. Again the fitness probability is calculated and the data are sent to the previous level best fitted node to be maintained in the forwarding data table, represented in figure 5.
- A partial path is created from the forwarding node to the highest level bet fitted node. The partial path is shown in figure 6.

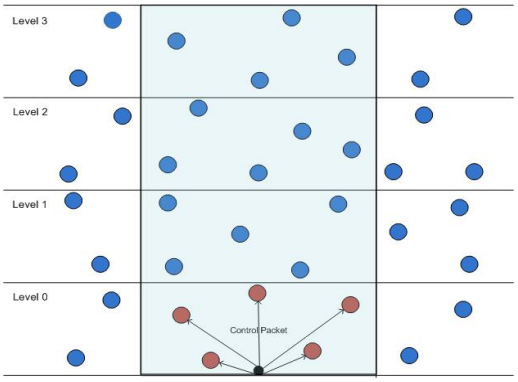


Figure 02: Broadcasting control packet

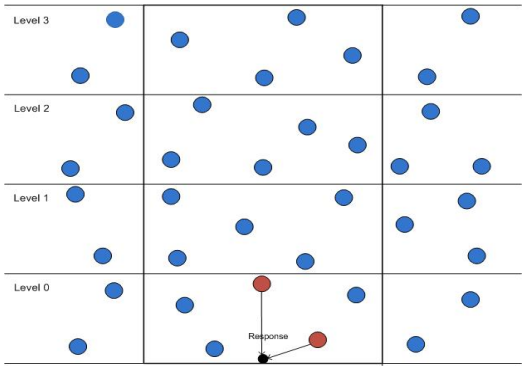


Figure 03: Response message to forwarding node

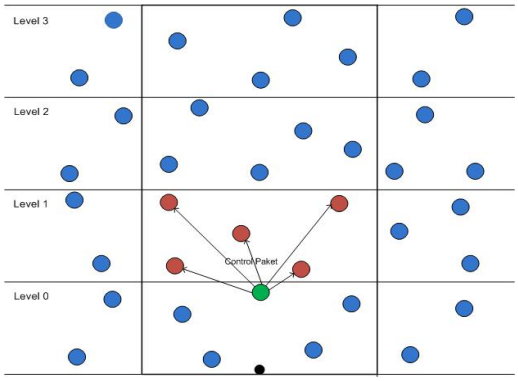


Figure 04: Broadcasting control packet to next level

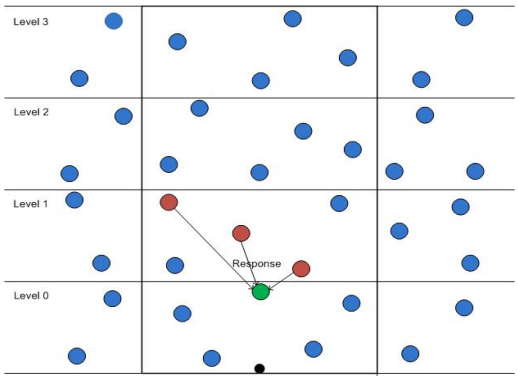


Figure 05: Response to previous level

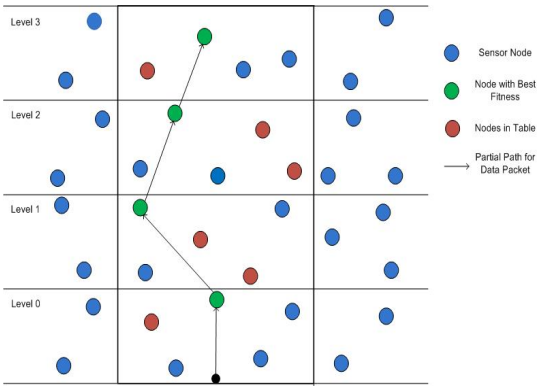


Figure 06: Partial path

ii. Forwarding Data Packets in Partial Path

As soon as this path is created, the forwarding node broadcasts the data packet to the best fitted node of the adjacent level; this node then forwards the data packet to the higher level best fitted node and so on.

Each node broadcasts an acknowledgement to the sender node after receiving the data packet. If any sender node does not get the acknowledgement after a certain time, it again broadcasts the data packet to the second best fitted node. Each level calculates TTL (Time to Life). In last level, TTL goes 4 and the best fitted node in level 4 calculates new levels in rectangular shape and the two phases are repeated.

C. Figures and Tables Fuzzy Interface Based Fitness Probability Calculation

i. Membership Function

Two input fuzzy variables named as- Noise and Lifetime is taken. Here Noise is the underwater acoustic noise and Lifetime is the assumed lifetime of sender nodes. There is only one output for this membership function- Probability. It is the probability of responding to a forwarding node's control packet. The range of each variable is taken from 0 to 100. Each variable has three membership functions- Low (0-40), MEDIUM (30-70), HIGH (60-100). The membership function plots for input and output variables are shown in figure 07, 08 and 09.

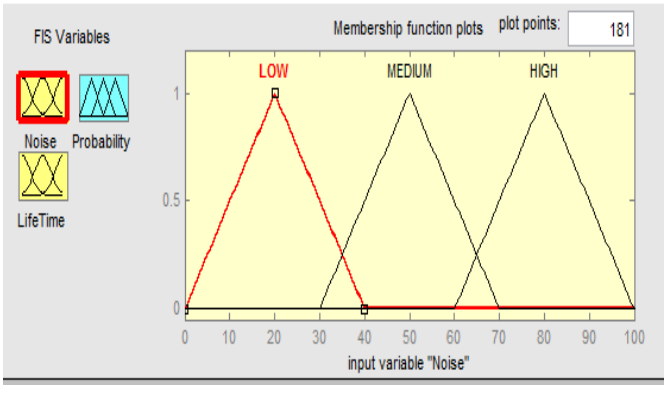


Figure 07: Fuzzy logic input using three membership function labels for Noise

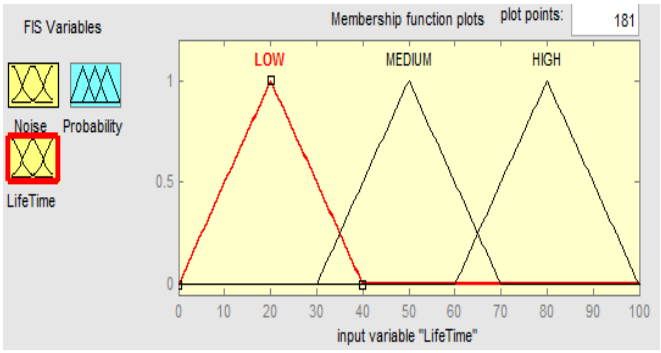


Figure 08: Fuzzy logic input using three membership function labels for Lifetime

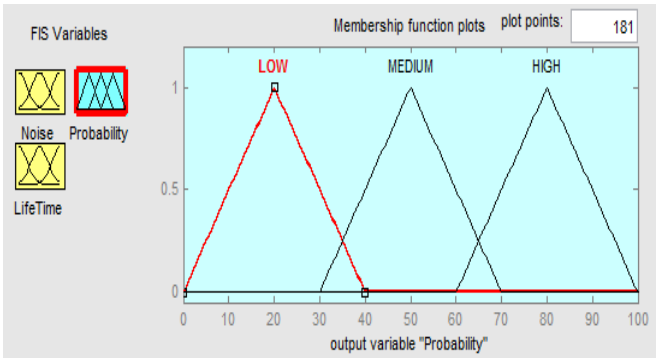


Figure 09: Fuzzy logic output using three membership function labels for Probability

ii. Rule Evaluation

In this section, the probability to response to forwarding node's control packet is calculated using Fuzzy based rules. The following rules are used to calculate the probability.

Table 01: Fuzzy Rules for Calculating Fitness Probability

No.	Rule
1	If Noise= LOW and Life Time=HIGH, then P=HIGH
2	If Noise= LOW and Life Time=MEDIUM, then P= MEDIUM
3	If Noise= LOW and Life Time=HIGH, then P= LOW
4	If Noise= MEDIUM and Life Time=HIGH, then P= MEDIUM
5	If Noise= MEDIUM and Life Time= MEDIUM, then P= MEDIUM
6	If Noise= MEDIUM and Life Time= LOW, then P= LOW
7	If Noise= HIGH and Life Time=HIGH, then P= LOW
8	If Noise= HIGH and Life Time= MEDIUM, then P= LOW
9	If Noise= HIGH and Life Time= LOW, then P= LOW

iii. Proposed Fitness Probability Calculation using MATLAB

We Consider a sensor node with estimated Lifetime = 67.7 unit that is defined as membership function- HIGH and acoustic Noise = 33.5 unit that is defined as membership function- LOW. Putting these values in MATLAB Fuzzy Logic Toolbox, we get the probability 69.2 that is defined as membership function- High. Figure 3.10 represents this scenario and it illustrates proposed Fuzzy rule number 1 in table 01.

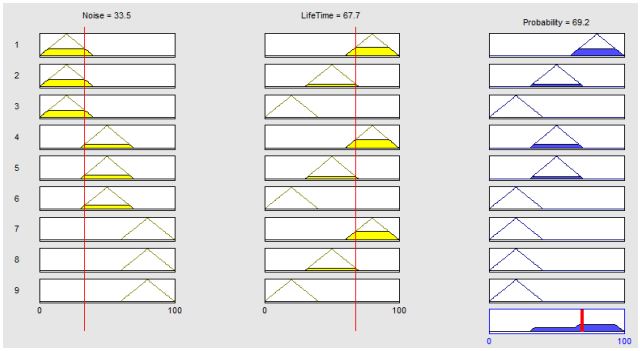


Figure 10: Output for Probability for Noise and Lifetime input variable

Figure 11 shows the overall expected outputs fro probability for all possible combinations of input Noise and Lifetime.

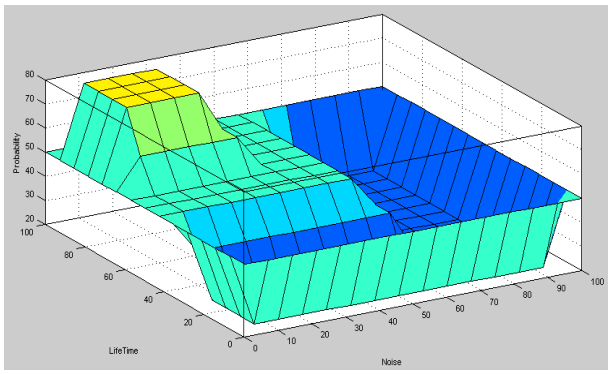


Figure 11: Total expected output results for all possible inputs

IV. PERFORMANCE EVALUATION

A. Simulation Setting

Network Simulator (ns2) [14] with an underwater sensor network simulation package (ns2-MIRACLE) extension is used to perform all the simulations. Sensor nodes are deployed in a 4kmX4kmX100m area in the simulations. Position generation of each node is done using Poisson's process. The simulation is concerned about 99 sensor nodes and a single sink node. The sensor nodes are considered to be mobile at the speed of water current while sink node is considered as stationary. Different speed of water current is considered in order to measure the performance of the proposed routing protocol. The minimum and the maximum speed of water

current are taken as 1 m/s and 10 m/s respectively. The sensor nodes move in random direction in underwater environment. The direction of each sensor nodes in 3D space is defined randomly for the ease of simulation. The control packets used in the protocol are assumed to be much shorter compared to data packets. The energy consumed for each data packet is assumed to be 1 energy unit and for each hello packet to be 0.02 unit. 300 m is considered as the transmission range of the simulation in all directions. The threshold energy of the sensor nodes is taken to be 1000 energy units. The source node is chosen from the bottom of the taken 3D space. The media access control (MAC) protocol is used for broadcasting. In this protocol, a node first senses the channel before sending a packet. It continues to broadcast the packet if the channel is free. Otherwise, it backs of. If the maximal number of back offs has been reached, the packet will be dropped.

B. Simulation Parameters and Performance Metric

Table 02 briefly shows the parameters needed to be precise for the ideal simulation of the proposed LBRP.

Table 02: Simulation Parameters for Proposed Protocol

Simulation Software	NS 2 (MIRACLE)
Deployment Area	4kmX4km
Deployment Depth	100m
Traffic Generation	Poisson's Process
Size Of Control Packet	≤6B
Data Packet Payload	3000bytes
Transmission frequency	25kHz
Maximum Range	300m (in all directions)
Bandwidth	2000Hz
Simulation Time	3600 s
Initial Energy	1000
Number of Nodes	100

The following metrics are pointers used to appraise the performance of the proposed routing protocol.

- *Throughput Efficiency*: Throughput is the measurement of the average bit rate successfully delivered to the sink node.
- *End to End Latency*: This metric calculates the average time between the packet generation and the time to deliver that packet to the sink divided by the distance between source and destination.
- *Energy per Bit*: To deliver a bit from source to destination, the energy consumed by the network is measured to evaluate performance.

C. Result and Analysis

i. Throughput Efficiency

The comparison of the performances of LBRP, CARP, FBR and DBR in terms of throughput efficiency is illustrated in figure 12. The figure clearly describes that the proposed LBRP performs a bit better than the best performing among the three protocols considered.

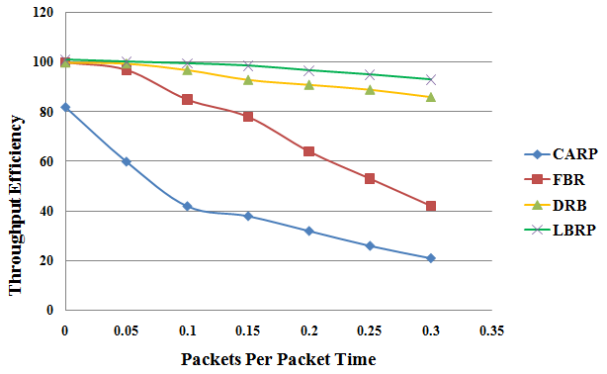


Figure 12: Comparison of Throughput Efficiency

ii End to End Latency

Figure 13 shows the average end to end latency of the considered protocols- proposed LBRP, CARP, DBR and FBR. As the time increases, end to end latency also increases. Simulation also shows that performance degrades with increment of traffic.

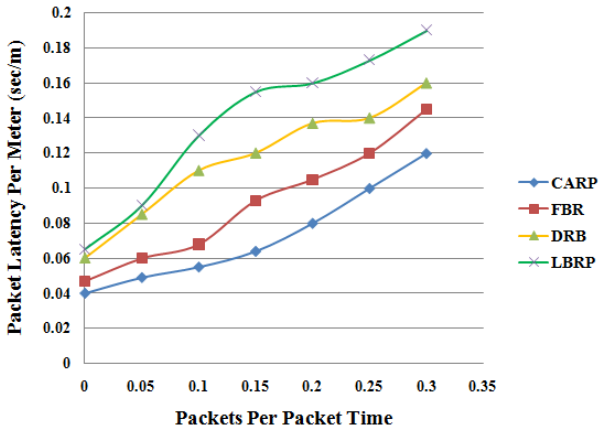


Figure 13: Comparison of End to End Latency

iii Energy per Bit

Energy consumed for each bit to deliver a single bit from source to destination for the proposed LBRP is calculated and shown in figure 14 along with CARP, FBR and DBR.

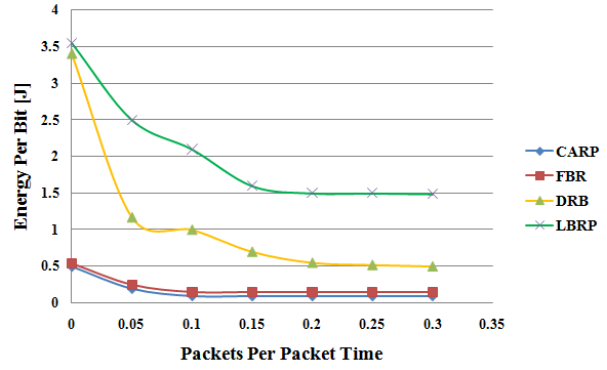


Figure 14: Comparison of Energy Consumption

V. CONCLUSION

The proposed protocol outperforms in terms of energy consumption and fitness of sensor nodes. Level wise divisions of monitoring environment reduce the energy consumption because no extra circuitry is needed in the sensor node to measure the depth under the water. Evaluation of this protocol shows better performance than other existing protocols. The proposed LBRP is expected to be used in UWSNs for energy efficiency and noise concern routing.

REFERENCES

[1] I. F. Akyildiz, D. Pompili, and T. Melodia, "Underwater acoustic sensor networks: research challenges," Elsevier: Adhoc Networks, no. 3,pp. 257–279, 2005

[2] L. Guo, P. H. Santschi, M. Baskaran, and A. Zindler, "Distribution of dissolved and particulate in seawater from the gulf of mexico and of cape hatteras as measured by sims," Earth and Planetary Science Letters ,vol.133,no.1-2,pp.117–128,1995.

[3] P. Xie, J. H. Cui, and L. Lao, "VBF: vector-based forwarding protocol for underwater sensor networks," in Networking 2006.Networking Technologies, Services, and Protocols; Performance of Computer and Communication Networks; Mobile and Wireless Communications Systems ,vol.3976,pp.1216–1221,2006

[4] N. Nicolaou, A. See, P. Xie, J. H. Cui, and D. Maggiorini, " Improving the robustness of location-based routing for underwater sensor networks," in Proceedings of the OCEANS Europe ,pp.1–6, June 2007.

[5] Mari Carmen Domingoa, Rui Prior , "Energy analysis of routing protocols for underwater wireless sensor networks " , Computer Communications 31 (2008) 1227–1238,

[6] N. Nicolaou, A. See, P. Xie, J. Cui, and D. Maggiorini, \Improving the robustness of location-based routing for underwater sensor networks," in OCEANS 2007-Europe. IEEE, 2007, pp. 1-6.

[7] Md.Asraf Uddin and Mamun-or-Rashid, "Link Expiration Time- Aware Routing Protocol for UWSNs", Hindawi Publishing Corporation volume: 2013, Article ID 625274.

[8] Sanjay K. Dhurandher, Mohammad S. Obaidat, SiddharthGoel and Abhishek Gupta "Optimizing Energy through Parabola Based Routing in Underwater Sensor Networks" In IEEE Globecom proceedings, 2011.

[9] H. Yan, Z. J. Shi and J. –H. Cui, "DBR: Depth-Based Routing for Underwater Sensor Networks", in Proceedings of the 7th International IFIP-TC6 NETWORKING 2008, pp. 72-86.

[10] Reza Javidan (Corresponding Author), HamidehRafiee, "A New Energy Efficient and Depth based Routing Protocol for Underwater Sensor Networks", British Journal of Science,January 2013, Vol. 8 (1)

[11] Stephano Basagni, Chiara Petriolo, Roberto Petroccia and Deniele Spaccini, "Channel Aware Routing for Underwater Wireless Networks", in proceedings of the OCEANS, 2012-Yeosu, IEEE, 2012.

[12] P. Raja Jurdak, Cristina Videira Lopes, "Sensor Network Operations, Battery Lifetime Estimation and Optimization for Underwater Sensor Networks", Wiley-IEEE Press, 2006, pp. 397–416.

[13] Raja Jurdak, Cristina Videira Lopes, and Pierre Baldi,"Battery lifetime estimation and optimization for underwater sensor networks",California Institute for Telecommunications and Information Technology Cal-(IT) University of California, Irvine CA 92697.

[14] The Network Simulator NS-2, http://www.isi.edu/nsnam/ns/index.html, 2002.

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

Sentiment Analysis of Students Feedback: A Study towards Optimal Tools

Mohammad Aman Ullah
Department of Computer Science and Engineering,
International Islamic University Chittagong, Chittagong-4203, Bangladesh,
ullah047@yahoo.com

***Abstract*- Educational Institutions attempts to gather feedback from students' to study their sentiments towards courses and instructors and to enhance the performance of the instructors. Basically, such feedbacks are gathered at the end of the semester with the use of survey forms. However, this technique is very tedious, slow and time consuming. With the advent of social media, especially Facebook, the collection of feedback become easier through Facebook pages and groups. But, analyzing those feedbacks is equally challenging. This paper addresses those problems and uncovers the best model for analyzing those feedbacks with the use of machine learning techniques such as Support Vector Machines (SVM), Maximum Entropy (ME), Naive Bayes (NB), and Complement Naive Bayes (CNB) and applying neutral class. And, found SVM as the highest performer with an accuracy of 97% by applying different preprocessing and feature extraction techniques and avoiding neutral class, which outperform state-of-art work by 2%.**

Keywords- Student; Feedback; Analysis; Machine; Learning

I. INTRODUCTION

With an advent of the social network, feedback system became very common and useful in every domain of interest. One of such domain is educational Institution, where, instructors used to give the lecture in the class and students give the opinions in the social media such as Facebook or twitter etc., which provides the chances and equal methodological difficulties in understanding students' behavior, but need to address those difficulties and process them carefully to derive useful information to support instructors in improvement for further lectures. One of the study supports this is known as Sentiment Analysis (SA). SA is a field of Computational Study of people's like/dislike about an entity or objects [1]. Also, it reveals the polarity of natural language texts, such as positive, negative and in some cases neutral is considered [2], [3].

As research shows, sentiment analysis could be easily done and highly useful, if it could be applied in a particular domain, such as in this case Education domain [2]. So, the purpose of this research is to find best suited tools for each step of sentiment analysis, such as preprocessing, feature extraction and Machine learning techniques in educational settings in order to improve teaching methods, faculty members, student's related facility and infrastructure. Also, this research will find the best combinations of sentiment analysis tools as a whole.

The remaining work is organized as follows: Section II includes Literature Review, Section III represents details methodology. Whereas, Section IV describes the results and Discussions and section V includes Conclusions.

II. LITERATURE REVIEW

An optimal combination of pre-processing tools was searched for performance enhancement of the sentiment analysis from student feedback by Ortigosa et al. [4], Martin et al.[6]. They found that, if the following pre-processing tools such as stop words removal, case normalization, repeated letter removal, Removing Exclamation, Punctuation and question marks, negation, and spelling check being used in combination than the performance improve to a large extent. Troussas et al. [5] used emoticons along with above tools and also found the improved performance. With the combination of pre-processing tools such as stop words removal, case normalization, repeated letter removal, Removing Exclamation, Punctuation and question mark Altrabsheh et al. [2] got improved accuracy of 95%, which is highest among all others in the same sort of analysis.

An optimal feature extractor is also necessary for improving performing of the analysis of student feedback. Among all of the available features, Term co-occurrence is widely used. It may be unigram, bigram, N-gram etc. Agarwal et al. [7] and Wang et al.[8] found that, unigrams is better in performance than bigrams and trigrams in most of the cases. For instance, Go et al. [10], found decreased performance in case of bigrams. On the other hand, a twitter analysis done by Pak and Paroubek [9] shows that bigrams gives a higher performance than unigrams. But among all, Altrabsheh et al. [2] got the highest performance with accuracy 95%, with the use of the combinations of Term co-occurrences. Also, among different machine learning algorithms Support Vector Machines (SVM) were found to be the best by Song et al. [11] in the case of student feedback analysis. Altrabsheh et al. [2] also found SVM as the best classifier tools in analyzing student feedback. Altrabsheh et al. [13] in one of their research on real time student feedback found

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

Complement Naive Bayes (CNB) as the highest performer among all the techniques with 84% accuracy. Fallakhair et al. [14] conduct an in-depth analysis of real time student feedback in class room and found SVM RB as the best performer among all the algorithms with accuracy 94%.

III. METHODOLOGY

The methodology of this research is shown in Fig. 1, and the details descriptions of each step of the methodology such as corpus collection, Data preprocessing, Feature Extraction, sentiment classification, polarity detection and performance analysis are given below:

A. Corpus Collection

One of the important tasks of any research is the collection of data. Similarly important is to identify sources of the data. This research considered Facebook as a source of data, as Facebook messages are not limited in size like tweets and become more eligible for research in our said domain. We have used the same data set of the University of Portsmouth as used in state-of-art work by Nabeela et al. [2], where they have collected data from different universities and the students are allowed to express their opinions in free of context about the lectures. The data set consists of 1036 data, among them 641 are positive, 292 are negative, and remaining are neutrally distributed. The amount of data for analysis was fixed up using trial and error method. The data are labeled into different classes as positive, negative, according to the intensity of expression, if the intensity is hard to determine, the neutral class is assigned. We have measured percent agreement (80.7%), the Fleiss kappa (0.665) and Krippendor's alpha (0.666) in order to verify reliability. The first measure is optimistic and the later measure is conservative.

B. Preprocessing

As the research aim is to find the optimal tools that help in improving the accuracy of the study. So, firstly, we have tried to find the optimal combination of pre-processing tools, as we know, the more the data set is error free, the more the accuracy is. For this, we have firstly conducted a test without pre-processing, as the result was not promising, we have done few low level pre-processing, as the result was improved, so, we have conducted different levels of pre-processing till the result exceeds the state-of-art research. The total pre-processing steps are as described below and shown in Table II, III, and IV.

1. Preprocessing W/O: At this level, the data set is tested without any preprocessing except case normalization. And the result of the test is used as the baseline for the different level of preprocessing being applied.
2. Preprocessing P1: At this level, unnecessary characters, numbers, exclamations, punctuations, question marks, and stop words that do not contain any sentiments are removed to reduce noise from the data set.
3. Preprocessing P2: At this level, the data set is the previously preprocessed one in P1. The pre-processing included at this level are, removing repeated words and replacing "n't" with negation "not" in order to increase the possibility of matching sentiment words, thus, improve the performance of analysis.
4. Preprocessing P3: This level includes the data set that being preprocessed in P2. The pre-processing done at this level is stemming, where, the short words are returned to its original words, thus improve accuracy analysis.
5. Preprocessing P4: This level includes the data set that being preprocessed in P3. The pre-processing done at this level is the removal of emoticons in order to ease and improve the results of the analysis.
6. Preprocessing P5: This level includes the data set that being preprocessed in P4. Additional preprocessing done at this level is the removal of the URL in order to reduce the complexity of the analysis.

It is to be noted that, the combination we have used in our study was little different than that was used by state-of-art research. As our study outperform the previous research with this combination of pre-processing, so, this combination could be considered as optimal pre-processing tools.

C. Feature Extraction

As the research shows, proper extraction and selection of features help improve the accuracy of sentiment analysis to a great extent. So, secondly, we have tried to find optimal feature extraction tools from among Term Co-occurrence and Opinion words or phrase. From among different Term Co-occurrence, we have used Uni-gram, Bigrams and Trigrams. But the results of precision and recall shows that, trigrams outperform the results of Uni-gram, Bigrams, and Opinion words as in Table I, so, trigram could be considered as optimal feature extraction tool.

Table I

PRECISION AND RECALL VALUES FOR FEATURE EXTRACTION

Features	**Features extraction**	
	Precision	***Recall***
Uni-gram	60	65
Bi-gram	55	70
Tri-gram	80	75
Opinion words	65	70

D. Sentiment classification Techniques

Thirdly, From among different sentiment classification techniques, we have tested our experiments by only Support Vector Machines (SVM), Naive Bayes (NB), Complement Naive Bayes (CNB), and Maximum Entropy (ME) due to their wide acceptance and better performance in many similar and

state-of-art researches. SVM also outperform other models in this case.

E. Polarity Detection

Polarities were detected after the sentiment being classified using different classifier.

F. Performance Analysis

This subsection analyzes the performance on the different combination of preprocessing, features, and machine learning algorithms and their relative performance comparison, which are discussed in details in Result and Discussion section.

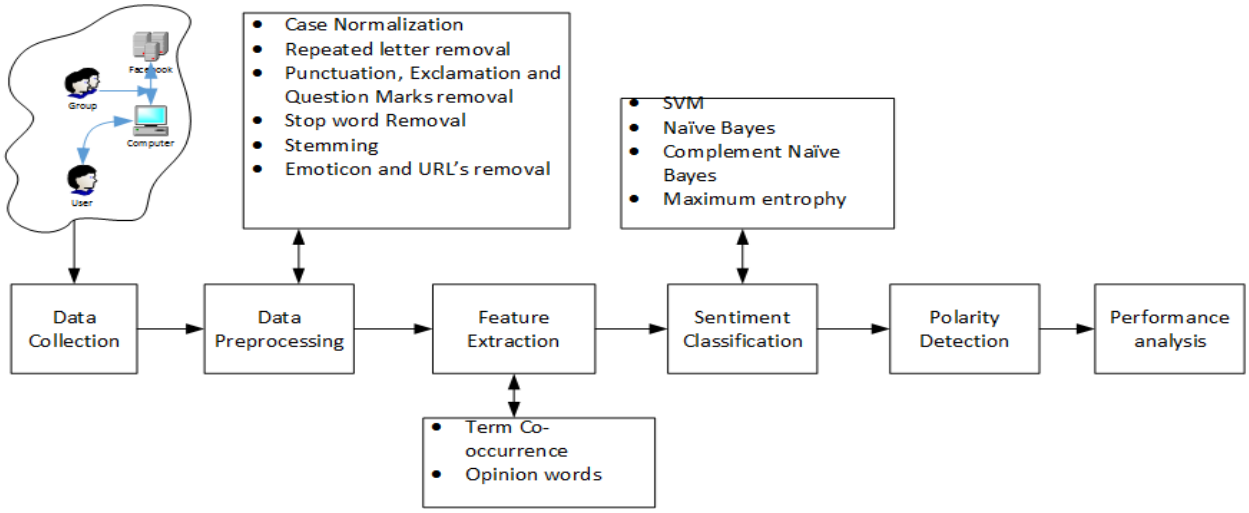


Fig. 1. Methodology of this Research

IV. RESULTS AND DISCUSSIONS

In this research, models were validated by measuring different matrix of the confusion matrix such as accuracy, recall, precision, and F-score and tested using 10-fold cross-validation. Table II represent the comparison between highest outcome, shown by the different models such as SVM, NB, CNB, and ME in the two state of art research and our study after pre-processing. From the Table II, Fig. 2 and Fig. 3, it is clear that, as per expectation, almost all the models shows better performance after preprocessing in both the study, but our study outperforms the stat-of-art researches by 2%, where our optimal model was SVM and their Optimal model was SVM with 95% accuracy [2] and CNB with 84% accuracy [13], this was due to the proper combination of pre-processing tools in our study. It was also observed that, CNB also shows good accuracy in all the research studied in this work. But, NB and ME suddenly shows improved accuracy in our work, than the previous research, which is 89% and 87% respectively.

Table II

COMPARISON IN HIGHEST RESULTS FOR EACH MODEL IN OUR STUDY AND STATE-OF-ART RESEARCH

Verification Parameters	**Result of State-of-art Research1**				**Result of State-of-art Research2**				**Result of Our Study**			
	Machine Learning Techniques				***Machine Learning Techniques***				***Machine Learning Techniques***			
	SVM	NB	CNB	ME	SVM	NB	CNB	ME	SVM	NB	CNB	ME
Accuracy	0.690	0.500	0.840	0.570	0.945	0.517	0.842	0.717	0.966	0.891	0.931	0.871
Recall	0.690	0.490	0.840	0.300	0.947	0.526	0.878	0.342	0.883	0.701	0.867	0.821
Precision	0.740	0.490	0.870	0.330	0.945	0.521	0.842	0.407	0.910	0.695	0.780	0.920
F-Score	0.570	0.480	0.840	0.310	0.944	0.499	0.848	0.372	0.910	0.705	0.885	0.920

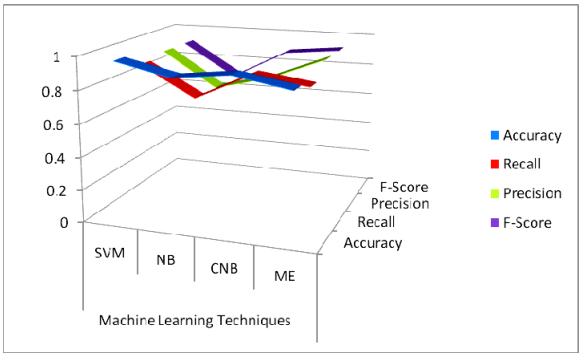


Fig. 2. Performance analysis of different Machine Learning Techniques in our Study

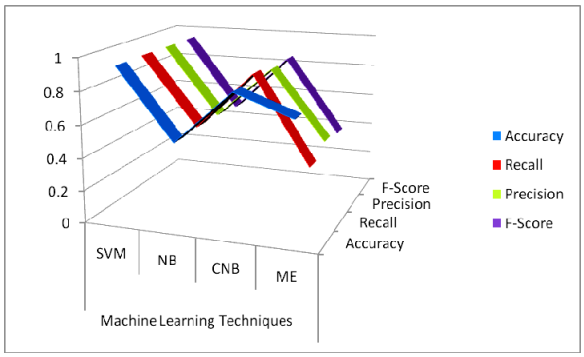


Fig. 3. Performance analysis of different Machine Learning Techniques in State-of-art Research2

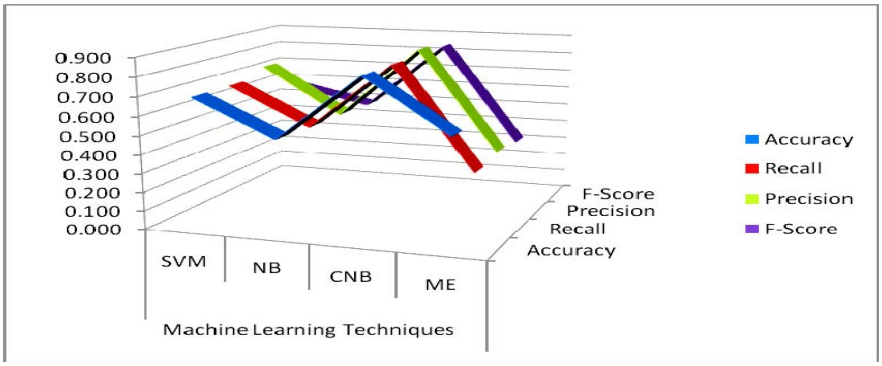


Fig. 4. Performance analysis of different Machine Learning Techniques in State-of-art Research

From the Table II, it is also seen that, SVM has a relatively high performance not only as because of the highest level of accuracy i.e.; 97% but also highest recall of SVM, which is 88%. But, in case the of precision and F-score, ME shows the highest value, which was 92% for both the cases respectively. From the analysis in our case, it was also seen that, CNB present nearby accurate result as SVM i.e.; 93%, this is because it dealt with the uneven class problem of this research. From Fig. 2, Fig. 3, and Fig. 4. Shows the performance analysis of different machine learning techniques in our study, state-of-art reserach1, and research2. In Fig 2. the peak of the curve shows nearby and very high for all the techniques, but in Fig. 3 and Fig. 4 the techniques are shown to fluctuates and differ in a great extent in their all of performance parameters.

Table III

RESULT OF DIFFERENT MODELS AFTER DIFFERENT LEVEL OF PREPROCESSING

Pre-processing	**Machine learning Techniques**							
	SVM				***ME***			
	Accuracy	***Recall***	***Precision***	***F-score***	***Accuracy***	***Recall***	***Precision***	***F-score***
P1	0.754	0.721	0.693	0.701	0.756	0.652	0.673	0.621
P2	0.823	0.772	0.792	0.810	0.773	0.755	0.736	0.711
P3	0.854	0.801	0.821	0.822	0.824	0.712	0.736	0.810
P4	0.893	0.859	0.831	0.840	0.843	0.782	0.801	0.815
P5	0.962	0.883	0.910	0.910	0.871	0.821	0.811	0.844

Table IV

RESULT OF DIFFERENT MODELS AFTER DIFFERENT LEVEL OF PREPROCESSING (CONTN.)

Pre-processing	Machine learning Techniques							
	NB				*CNB*			
	Accuracy	*Recall*	*Precision*	*F-score*	*Accuracy*	*Recall*	*Precision*	*F-score*
P1	0.698	0.444	0.432	0.435	0.701	0.650	0.555	0.624
P2	0.753	0.454	0.443	0.541	0.763	0.654	0.643	0.741
P3	0.784	0.652	0.564	0.555	0.840	0.662	0.764	0.755
P4	0.853	0.687	0.665	0.645	0.893	0.758	0.777	0.864
P5	0.891	0.701	0.695	0.705	0.931	0.867	0.780	0.885

Table III and Table IV shows the changes of performance, happens due to the different levels of preprocessing and model used. Here, we have included only the results that are remarkable. Almost every model show improved performance as the level of preprocessing proceeds. The change is slight in some cases but in most cases the preprocessing and model used, bring a great improvements of the analysis, which may be 20 to 30 percent and thus reflect that, the pre-processing are immense necessary in order to carry on feedback analysis of the students as the students in most of the cases doesn't express their opinion in plain or good English.

V. CONCLUSIONS AND FUTURE WORK

In this research, we have investigated different levels of preprocessing in combination with different feature extraction methods and machine learning algorithms on student feedback data to find out optimal tools among them for this sort of analysis. We found that, proper use of preprocessing tools, feature extraction tools and machine learning algorithms increased accuracy, which in turn shows improved performance of our analysis up to 20 to 30 percent, which we desired before analysis. It is also observed that, some level of preprocessing such as emoticons removal, in some cases loses valuable information, thus decrease the performance of the models. From the study we suggest that, the combination of preprocessing tools and feature extraction tools we have used in this study may be considered as optimal tools for this sort of analysis.

We also observed that, NB and CNB show highest performance in the case of uneven class problems, but in most of the cases it doesn't show the considerable result. But SVM and ME shows the acceptable results in most of the combinations of preprocessing, features and machine learning techniques. So, from the analysis we could say SVM and ME could be considered as best model and we suggest this could be used further for these sorts of feedback analysis.

In the future, there would be a possibility of including more preprocessing such as removing substring, feature extraction, such as syntactic dependency as well as sentiment word list extraction, and finally, more machine learning techniques such as an Artificial Neural Network (ANN) in this domain.

REFERENCES

[1] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," Ain Shams Engineering Journal, vol. 5, pp. 1093-1113, 2014.

[2] N. Altrabsheh, M. Cocea, and S. Fallahkhair, "Sentiment analysis: towards a tool for analysing real-time students feedback," in 2014 IEEE 26th international conference on tools with artificial intelligence, 2014, pp. 419-423.

[3] A. Hogenboom, D. Bal, F. Frasincar, M. Bal, F. de Jong, and U. Kaymak, "Exploiting emoticons in sentiment analysis," in Proceedings of the 28th Annual ACM Symposium on Applied Computing, 2013, pp. 703-710.

[4] A. Ortigosa, J. M. Martín, and R. M. Carro, "Sentiment analysis in Facebook and its application to e-learning," Computers in Human Behavior, vol. 31, pp. 527-541, 2014.

[5] C. Troussas, M. Virvou, K. J. Espinosa, K. Llaguno, and J. Caro, "Sentiment analysis of Facebook statuses using Naive Bayes classifier for language learning," in Information, Intelligence, Systems and Applications (IISA), 2013 Fourth International Conference on, 2013, pp. 1-6.

[6] J. M. Martín, A. Ortigosa, and R. M. Carro, "SentBuk: Sentiment analysis for e-learning environments," in Computers in Education (SIIE), 2012 International Symposium on, 2012, pp. 1-6.

[7] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau, "Sentiment analysis of twitter data," in Proceedings of the workshop on languages in social media, 2011, pp. 30-38.

[8] W. Wang and J. Wu, "Notice of Retraction Emotion recognition based on CSO&SVM in e-learning," in Natural Computation (ICNC), 2011 Seventh International Conference on, 2011, pp. 566-570.

[9] A. Pak and P. Paroubek, "Twitter based system: Using Twitter for disambiguating sentiment ambiguous adjectives," in Proceedings of the 5th International Workshop on Semantic Evaluation, 2010, pp. 436-439.

[10] A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision," CS224N Project Report, Stanford, vol. 1, p. 12, 2009.

[11] D. Song, H. Lin, and Z. Yang, "Opinion mining in e-learning system," in Network and Parallel Computing Workshops, 2007. NPC Workshops. IFIP International Conference on, 2007, pp. 788-792.

[12] B. Pang and L. Lee, "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts," in Proceedings of the 42nd annual meeting on Association for Computational Linguistics, 2004, p. 271.

[13] N. Altrabsheh, M. Cocea, and S. Fallahkhair, "Learning sentiment from students' feedback for real-time interventions in classrooms," in Adaptive and Intelligent Systems, ed: Springer, 2014, pp. 40-49.

[14] S. Fallahkhair, M. Cocea, and N. Altrabsheh, "Learning sentiment from students' feedback," in research gate, 2014, pp. 44-48.

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

Fitness Based Biogeography Based Optimization Algorithm

Harish Sharma[a], Priya Kumari Sharma[b] and Jagdish Chand Bansal[c]
[a,b]Rajasthan Technical University, Kota, India
[c] South Asian University, New Delhi, India
Email: [a]hsharma@rtu.ac.in ,[b]priyasharma02413@gmail.com, [c]jcbansal@gmail.com

***Abstract*—Biogeography Based Optimization (BBO) Algorithm is a stochastic and population-based evolutionary search method influenced by the theory of biogeography. Like other population-based evolutionary algorithms, BBO also prefers diversification at the cost of intensification. Therefore, in this article, a new variant of BBO algorithm introduced namely, Fitness Based Biogeography Based Optimization (FBBBO) Algorithm. In the proposed algorithm, a fitness learning factor based position update process is introduced. In this fitness based process, the step size of the solutions is controlled through a probability which is a function of fitness. If a solution is a superior fit, then it gives additional priority to finest solution and less to itself to prevent from local convergence. The developed algorithm compared with BBO and two other algorithms, namely Differential Evolution (DE) Algorithm and Particle Swarm Optimization (PSO) Algorithm with the experiments over 10 test problems. Obtained results confirm the competitive performance of the proposed algorithm.**

***Index Terms*—Meta-heuristic Optimization Techniques; Fitness Based Process; Evolutionary Algorithm;**

I. Introduction

Nature Inspired Algorithms (NIA) [1] are a significant source for motivations to both grow intelligent systems and provide solutions to challenging engineering optimization problems. Various evolutionary methods have been evolved to deal with difficult engineering optimization problems. Evolutionary processes influenced by the natural selection and model of the biological evolution process. Evolutionary [11] algorithms optimize a function by repetitively developing candidate solution in terms of given measure of goodness. BBO [8] is an evolutionary search method that is influenced by modeled science of habitat biogeography that contains sharing of biological species through space and time.

Research globally in the field of BBO has produced new optimization methods that have proven greater to primary approaches. Approximately, entire metaheuristic methods [4, 5, 10, 2, 9] share the identical properties such as the natural phenomenon inspires they all, use random variables and they have several attributes. Every metaheuristic method has different assets on reliability, efficiency, robustness and accuracy in a noisy atmosphere and unique problem domain. According to "no-free-lunch" theorem [12], it is impossible for any metaheuristic method to handle entire optimization problems optimally. Thus, new metaheuristic methods are continuously produced to solve distinct optimizing problems.

In this paper, to maintains a proper balance between the intensification and diversification capability of BBO algorithm, a modified variant of primary BBO [8] is studies, named as Fitness Based Biogeography Based Optimization (FBBBO) Algorithm. In FBBBO, a Fitness based position update strategy is adopted to improve a solution. In This, to explore the solution search field, less fit individuals take large step size whereas the more fit individual's step size is small for avoiding the situation of skipping the optima. This strategy maintains the proper balance between the intensification and diversification properties of the algorithm and enhances the efficiency of BBO algorithm also.

Section II of this paper gives the brief overview of BBO Algorithm and the modified algorithm portrayed in section III. Performance of FBBBO is tested with several numerical benchmark functions and presented under section IV. Finally, section V gives a summary and conclude the work.

II. Overview of BBO algorithm

BBO [8] is an evolutionary search technique that describes the dispersion of bio-diversity of biological species through space and time. The Scientific model of biogeography describes Speciation (evolution of new species), Migration of species (animal, birds and plants) among island (habitat) and Extinction of species. In BBO [8], each individual is considered to be analogs to an island and is recognized by habitat (island) suitability index (HSI). Habitat (solution) that are friendly to life are known with high HSI. HSI corresponds to the goodness of a BBO solution. Facets that combined with HSI include land area, temperature, rainfall etc. Facets that identify habitability are known as suitability index variable (SIVs). Figure 1 demonstrates a representation of species profuse in a particular island.

In terms of habitability, SIVs can be taken as independent variables and HSI as dependent variables of the island respectively. Habitat that is supported by a large number of species is called High-HSI habitat and habitat with only a few species is called Low-HSI habitat. High-HSI solutions give their facets to low-HSI solutions and new facets are accepted by low-HSI solutions from high-HSI solutions. High-HSI solutions emigrate their facets to low-HSI solutions.

The BBO algorithm starts by creating a randomly dispersed population of habitats. Then evaluate the emigration rate (μ)

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

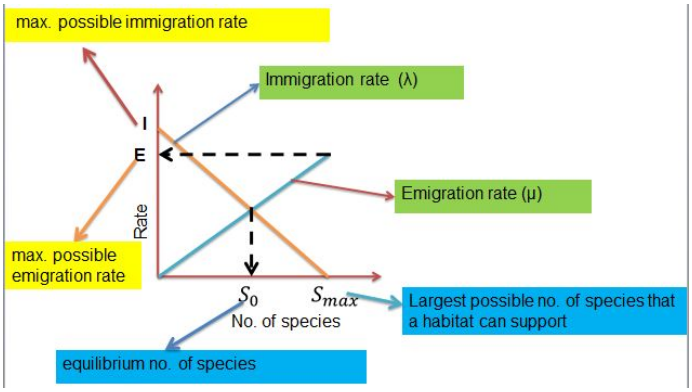


Fig. 1: Species Model [8]

and immigration rate (λ) for every individual.

$$\lambda_i = I(1 - \frac{k_i}{n}) \quad (1)$$

$$\mu_i = E(\frac{k_i}{n}) \quad (2)$$

Here, E is highest possible μ, I is highest possible λ and k_i is goodness rank of i^{th} individual after sorting goodness of i^{th} individual.

After initialisation of the BBO population, it needs the repeated iterations of the two procedures, named as migration procedures and another one is mutation procedures. Each procedure is portrayed as follows.

A. Migration

Migration procedure is probabilistically sharing of facets between habitats by taking advantage of the immigration rate (λ_i) and emigration rate (μ_j). The migration operator (α) is accountable for sharing the facets between individuals to modify goodness. The immigrating solution is selected according to the probability of (λ_i) and emigrating solution is selected according to the probability of (μ_j) of solutions. Based on these probabilities algorithm decide which of the solution feature (SIV) of immigrating solution (H_i) is needed to be modified. Then immigrating solution (H_i) SIV is replaced by emigrating solution's (H_j) SIV.

$$new(H_{ik}) = H_{ik} + \alpha(H_{jk} - H_{ik}) \quad (3)$$

here α is the user defined migration operator and k is the dimension of the solution.

B. Mutation

Mutation is accountable for keeping the variety of solutions in BBO algorithm. Mutation provides a chance to improve the goodness of solutions for low-HSI candidate solutions, and for high-HSI solutions, it can enhance the quality of solutions even more than solutions already have.

After, each generation we apply elitism in BBO for preserving the best solutions from present generation to future generation. Elitism prevents the best solution from being polluted by migration or mutation of the future generation.

From the above discussion in section II, procedure of the BBO is represented in algorithm 1.

Algorithm 1 BBO Algorithm

creating a randomly dispersed set of solutions $H1$, $H2$, ...,Hn;
Calculate goodness (HSI) of every solutions;
while termination criteria met **do**
  Sort the solutions from best to worst;
  Calculate λ and μ for each solutions based on HSI ;
  Apply migration procedure;
  Renew the probability for each solutions;
  Apply mutation procedure;
  Calculate goodness of every solutions;
  Apply elitism for maintaining best solutions;
end while

III. Fitness Based Biogeography Based Optimization Algorithm

In the evolutionary optimization algorithms, exploration and exploitation are the two vital processes. The exploration consists of probing a much larger portion of the search region with the desire of finding other promising solutions that are yet to be refined. While the exploitation consists of probing a limited region of the search space with the desire of inspiring a promising solution that we already have in hand. For achieving superior performance in optimization algorithms, both exploration and exploitation properties should be well coordinated with each other. Like other evolutionary algorithms, BBO also suffers from the problem of stagnation and prefer diversification at the cost of intensification. To avoid these issues, it is necessary that initially algorithm explore the search space while exploits in later iterations.

From the equation 3 of basic BBO, it is clear that immigrating solution accepts facets only from emigrating solution. Therefore, in this paper we modify the migration equation of the individuals as follows:

$$new(H_{ik}) = H_{ik} + (\beta)(H_{ik} - H_{jk}) + \overbrace{(2.0 - prob_i)}^{LearningFactor} \times (H_{BestSolk} - H_{ik}) \quad (4)$$

In the above equation, β is a uniformly dispersed random number in the range -1 to 1. Here for evaluating step size, ongoing individual and a randomly chosen individual are used. However, in this random component is used for updating step size. Therefore, to balance the intensification and diversification capability of an algorithm, we further modified the strategy. In this, the step size of the solutions is controlled through a probability which is a function of fitness. The $prob_i$ is a probability which is a function of individual's fitness and calculated by equation 5.

$$prob_i = \frac{0.9 \times fit(i)}{maxfit(i)} + 0.1 \quad (5)$$

Algorithm 2 FBBBO Algorithm

Generate a random set of solutions $H1$, $H2$, ...,Hn;
Calculate HSI values;
while the termination criteria is not met **do**
  MIGRATION
  Sort the solutions from best to worst;
  Compute λ and μ for each solutions based on HSI ;
  For any solution;
  For any SIV (k) (solution dimension);
  Elect solution H_i with probability $\propto \lambda_i$;
  if H_i is selected **then**
    Elect habitat H_j with probability $\propto \mu_j$;
    if H_j is elected **then**
      $new(H_{ik}) = H_{ik} + (\beta)(H_{ik} - H_{jk}) + \overbrace{(2.0 - prob_i)}^{LearningFactor} \times (H_{BestSolk} - H_{ik})$;
    end if
  end if
  MUTATION
  Elect $H_i(SIV)$ based on m_i;
  if $H_i(SIV)$ is elected **then**
    Replace $H_i(SIV)$ with a randomly created SIV;
  end if
  Recompute HSI values;
end while

The proposed strategy follows the approach that the probability will be high for superior fit individuals while the probability is low for inferior fit individuals.

Here, $fit(i)$ is the fitness of i^{th} individual. Fitness is calculated using the objective value (f_i) as equation 6.

$$\text{individual fitness } (fit(i)) = \begin{cases} \frac{1}{1+f_i}, & \text{if } f_i \geq 0 \\ 1 + abs(f_i), & \text{if } f_i < 0 \end{cases} \quad (6)$$

It expected that the true solution should be nearby to the superior fit individuals and if step sizes of superior individuals will be large then may be a chance of skipping global optima due to greater step size.
Hence, the step size is lower for superior fit individuals which are accountable for intensification and step size are higher for inferior fit individuals which are accountable for diversification.
In equation 4, the term $(2.0 - prob_i)$ is a learning factor (LF). The value of LF will be low for high fit solutions (due to a high value of $prob_i$). Hence the step size will be low.

So, the individuals having a higher probability (higher fitness) will move slowly and exploit the search space in its neighborhood. While for the lower probability individuals, the value of LF will be high (due to low value of $prob_i$) so as the step size. Therefore in the modified work, the superior fit individuals exploit the solution search space, and inferior fit individuals explore the search field. The proposed migration process described in algorithm 2.

IV. Results and Discussions

In this section, a fair experimental study will be presented to establish the competitiveness of the proposed FBBBO.

A. Test problems

To analyze the validity of the FBBBO algorithm, 10 distinct global numerical optimisation functions (f_1 to f_{10}) are used here, displayed in table I.

B. Parameter setting

To test the efficiency of FBBBO, a comparative experiment is carried out among FBBBO, BBO, DE [6] and PSO [3] over considered optimization test functions. Following experimental parameters are adopted:

- pMutation=0.01
- Number of population=50
- Total number of iteration=5000
- Total number of runs=30
- Parameter selection settings for the algorithms BBO [8], DE [6] and PSO [3] are imitated from their elementary research papers.

C. Results comparison

Table II shows the experimental results of the FBBBO, BBO, DE and PSO algorithms and also gives a description of the standard deviation (SD), mean error (ME), average number of function evaluations (AFE) and success rate (SR). Results in table II represent that maximum time FBBBO better performs in terms of reliability, efficiency as well as accuracy in comparison to the BBO, DE and PSO. Based on boxplots and acceleration rate (AR) some more analysis has been performed for results of FBBBO, BBO, DE and PSO algorithms. Besides, boxplots [1] analysis is carried out for comparing the examined algorithms in the form of combined performance though it can efficiently depict the empirical dispersion of data graphically. The boxplots for FBBBO, BBO, DE and PSO are exhibited in figure 2. The results acknowledge that interquartile range and medians of FBBBO are relatively less. Further, all examined algorithms are correspondingly

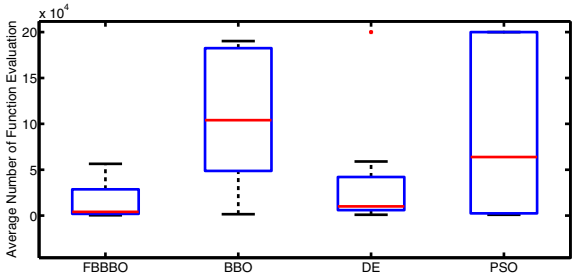


Fig. 2: Boxplots graphs (Average Number of Function Evaluation)

compared by giving weighted importance to the SR, AFE and ME. The performance indices which is portrayed in [1] is measurement for the comparison.

TABLE I: TP: Test problems, OF: Objective Function, SR: Search Range, OV: Optimum Value, D: Dimension, AE: Acceptable Error for FBBBO

TP	**OF**	**SR**	**OV**	**D**	**AE**
Ackley	$f_1(x) = -20 + e + exp(-\frac{0.2}{D}\sqrt{\sum_{i=1}^{D} {x_i}^3})$	$[-1, 1]$	$f(0) = 0$	30	$1.0E-05$
Schewel	$f_2(x) = \sum_{i=1}^{D} \lvert x_i \rvert + \prod_{i=1}^{D} \lvert x_i \rvert$	$[-10, 10]$	$f(0) = 0$	30	$1.0E-05$
Sum of different powers	$f_3(x) = \sum_{i=1}^{D} \lvert x_i \rvert^{i+1}$	$[-1, 1]$	$f(0) = 0$	30	$1.0E-05$
Brannin's Function	$f_4(x) = a(x_2 - bx_1^2 + cx_1 - d)^2 + e(1-f)\cos x_1 + e$	$[-5, 0], [10, 15]$	$f(0) = 0.3979$	2	$1.0E-04$
2D Tripod	$f_5(x) = p(x_2)(1+p(x_1)) + \lvert (x_1 + 50p(x_2)(1-2p(x_1))) \rvert + \lvert (x_2 + 50(1-2p(x_2))) \rvert$	[-100,100]	$f(0, -50) = 0$	2	$1.0E-04$
Shifted Sphere	$f_6(x) = \sum_{i=1}^{D} z_i^2 + f_{bias},\ z = x - o\ , x = [x_1, x_2, ...x_D],\ o = [o_1, o_2, ...o_D]$	[-100,100]	$f(0) = f_{bias} = -450$	10	$1.0E-05$
Shifted Schwefel	$f_7(x) = \sum_{i=1}^{D} (\sum_{j=1}^{i} z_j)^2 + f_{bias},\ z = x - o\ , x = [x_1, x_2, ...x_D],\ o = [o_1, o_2, ...o_D]$	[-100,100]	$f(0) = f_{bias} = -450$	10	$1.0E-05$
Easom's function	$f_8(x) = -\cos x_1 \cos x_2 e^{((-(x_1-\Pi)^2-(x_2-\Pi)^2))}$	$[-10, 10]$	$f(\pi, \pi) = -1$	2	$1.0E-13$
Dekkers and Aarts	$f_9(x) = 10^5 x_1^2 + x_2^2 - (x_1^2 + x_2^2)^2 + 10^{-5}(x_1^2 + x_2^2)^4$	$[-20, 20]$	$f(0, 15) = f(0, -15) = -24777$	2	$5.0E-01$
Moved axis parallel hyper-ellipsoid	$f_{10}(x) = \sum_{i=1}^{D} 5i \times x_i^2$	$[-5.12, 5.12]$	$f(x) = 0; x(i) = 5 \times i, i = 1 : D$	30	$1.0E-15$

The resultant values of PI [1] for the FBBBO, BBO, DE and PSO are computed and equivalent performance indices graphs are shown in figure 3. The weighted importance of SR, AFE and ME are characterized in figures 3(a), 3(b) and 3(c) subsequently.

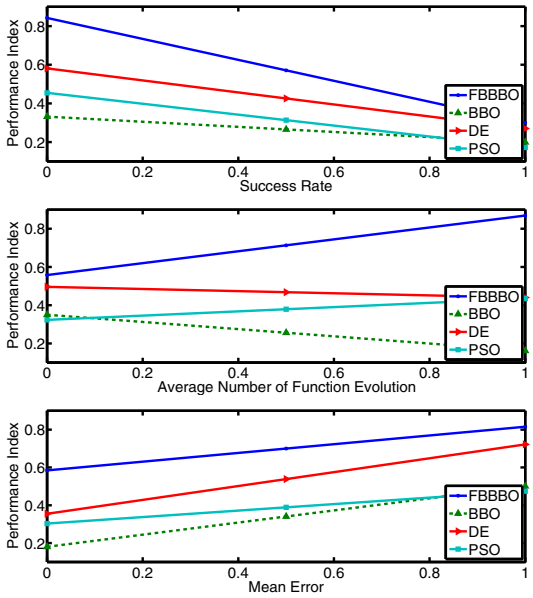


Fig. 3: PIs for test functions; (a) for weighted importance to SR, (b) for weighted importance to AFE and (c) for weighted importance to ME.

It is clear from figure 3 that PI of FBBBO are superior to the other examined algorithms in each case i.e. FBBBO outperforms on the examined test problems as compare to BBO, DE and PSO.

AFEs are tested to analyzed the examined algorithms for the convergence speed (CS) and lesser AFEs denotes the faster convergence. Acceleration Rate (AR) [7] is used to analyze CS and is calculated as in equation 7 for all examined algorithms.

$$AR = AFE_{ALG}/AFE_{FBBBO} \tag{7}$$

where $ALG \in BBO, DE, PSO$ and $AR > 1$ means that FBBBO is speedy than the other considered algorithms. To examine the AR [1] of the developed algorithm, as compared to the BBO, DE and PSO, results of table II are analyzed and the value of AR is evaluated using equation 7. Table III shows a clear comparison among FBBBO-BBO, FBBBO-DE, and FBBBO-PSO in terms of AR. It is clear from table III that FBBBO converges faster than the other examined algorithms.

V. Conclusion

In this paper, fitness based individual's position update process is developed. In the proposed position update process, the step sizes of the individuals are decided on the basis of fitness. The developed process is named as Fitness Based Biogeography Based Optimization (FBBBO) Algorithm. The competitiveness of the developed strategy is acknowledged through various statistical analyses over 10 well known complex optimization problems while comparing with DE and PSO. In future, the developed strategy may be applied to solve some complex continuous real world optimization problem.

TABLE II: Comparative result of Test Problems for FBBBO

TP	Algorithm	SD	ME	AFE	SR
f_1	FBBBO	7.98E-07	9.50E-06	28965.00	30
	BBO	1.28E-06	8.61E-06	60383.33	30
	DE	4.28E-07	9.49E-06	42043.33	30
	PSO	8.26E-01	7.79E-01	107183.33	15
f_2	FBBBO	1.69E-06	8.84E-06	28606.67	30
	BBO	9.51E-07	9.09E-06	48673.33	30
	DE	4.87E-07	9.58E-06	37901.67	30
	PSO	4.47E-02	1.32E-02	158420.00	7
f_3	FBBBO	8.40E-07	9.01E-06	3918.33	30
	BBO	5.66E-06	8.56E-06	42401.67	27
	DE	2.28E-06	7.26E-06	7665.00	30
	PSO	2.04E-06	7.49E-06	5243.33	30
f_4	FBBBO	3.20E-05	4.38E-05	3510.00	30
	BBO	1.87E-05	7.40E-05	54626.67	30
	DE	4.10E-05	6.55E+04	200000.00	0
	PSO	3.18E-05	3.43E-05	1146.67	30
f_5	FBBBO	2.46E-07	5.38E-07	1605.00	30
	BBO	1.56E-03	8.95E-04	153673.33	7
	DE	2.24E-07	6.78E-07	2938.33	30
	PSO	0.00E+00	4.00E+01	200000.00	0
f_6	FBBBO	1.17E-06	8.31E-06	4238.33	30
	BBO	5.16E-05	4.76E-05	190210.00	4
	DE	1.23E-06	8.72E-06	9996.67	30
	PSO	1.09E-11	1.75E+04	200000.00	0
f_7	FBBBO	1.48E-06	8.42E-06	4276.67	30
	BBO	2.87E-05	3.14E-05	182506.67	9
	DE	1.81E-06	8.15E-06	10013.33	30
	PSO	1.09E-11	1.96E+04	200000.00	0
f_8	FBBBO	2.49E-14	5.07E-14	1855.00	30
	BBO	4.12E-05	1.92E-05	186843.33	2
	DE	2.70E-14	4.98E-14	5938.33	30
	PSO	2.79E-14	3.61E-14	2423.33	30
f_9	FBBBO	6.46E+03	1.01E+04	348.33	30
	BBO	6.51E+03	4.52E+03	1493.33	30
	DE	5.94E+03	5.41E+03	895.00	30
	PSO	1.17E+03	1.32E+03	910.00	30
f_{10}	FBBBO	2.85E-16	7.24E-16	56391.67	30
	BBO	2.81E-16	6.13E-16	147703.33	30
	DE	6.21E-17	9.08E-16	58980.00	30
	PSO	1.69E-16	7.88E-16	20440.00	30

TABLE III: Acceleration Rate (AR) of $FBBBO$ compare to the conventional BBO, DE and PSO

TP	BBO	DE	PSO
f_1	2.0847	1.4515	3.7004
f_2	1.7015	1.3249	5.5379
f_3	10.8214	1.9562	1.3382
f_4	15.5632	56.9801	0.3267
f_5	95.7466	1.8307	124.6106
f_6	44.8785	2.3586	47.1884
f_7	42.6750	2.3414	46.7654
f_8	100.7242	3.2013	1.3064
f_9	4.2871	2.5694	2.6124
f_{10}	2.6192	1.0459	0.3625

References

[1] Jagdish Chand Bansal, Harish Sharma, KV Arya, and Atulya Nagar. Memetic search in artificial bee colony algorithm. *Soft Computing*, 17(10):1911–1928, 2013.

[2] Wenyin Gong, Zhihua Cai, Charles X Ling, and Hui Li. A real-coded biogeography-based optimization

with mutation. *Applied Mathematics and Computation*, 216(9):2749–2758, 2010.

[3] James Kennedy. Particle swarm optimization. In *Encyclopedia of machine learning*, pages 760–766. Springer, 2011.

[4] Xiangtao Li, Jinyan Wang, Junping Zhou, and Minghao Yin. A perturb biogeography based optimization with mutation for global numerical optimization. *Applied Mathematics and Computation*, 218(2):598–609, 2011.

[5] Haiping Ma and Dan Simon. Blended biogeography-based optimization for constrained optimization. *Engineering Applications of Artificial Intelligence*, 24(3):517–525, 2011.

[6] Kenneth Price, Rainer M Storn, and Jouni A Lampinen. *Differential evolution: a practical approach to global optimization*. Springer Science & Business Media, 2006.

[7] Harish Sharma, Jagdish Chand Bansal, KV Arya, and Xin-She Yang. Lévy flight artificial bee colony algorithm. *International Journal of Systems Science*, pages 1–19, 2015.

[8] Dan Simon. Biogeography-based optimization. *Evolutionary Computation, IEEE Transactions on*, 12(6):702–713, 2008.

[9] Dan Simon, Mehmet Ergezer, and Dawei Du. Population distributions in biogeography-based optimization algorithms with elitism. In *Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on*, pages 991–996. IEEE, 2009.

[10] Dan Simon, Mehmet Ergezer, Dawei Du, and Rick Rarick. Markov models for biogeography-based optimization. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 41(1):299–306, 2011.

[11] Kit-Sang Tang, Kim-Fung Man, Sam Kwong, and Qun He. Genetic algorithms and their applications. *Signal Processing Magazine, IEEE*, 13(6):22–37, 1996.

[12] David H Wolpert and William G Macready. No free lunch theorems for optimization. *IEEE transactions on evolutionary computation*, 1(1):67–82, 1997.

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

Underwater Network Cardinality Estimation Using Cross-correlation: Effect of Unequal Sensor Spacing

* B. K. Dash[1], S. A. H. Chowdhury[1], A. H. M. M. Kamal[1], M. S. Anower[2], and A. Halder[2]
[1]Department of Electronics & Telecommunication Engineering
[2]Department of Electrical & Electronic Engineering
Rajshahi University of Engineering & Technology, Rajshahi-6204, Bangladesh
* biswajit.rueten@gmail.com, arif.1968.ruet@gmail.com, prince.ete10@yahoo.com,
md.shamimanower@yahoo.com, amit.rueten@gmail.com

***Abstract*—Network cardinality is very crucial factor to ensure proper functionality of a network. Cardinality estimation of underwater network using the existing protocol based techniques is a severe challenge due to inimitable properties of underwater environment (such as strong background noise, non-negligible capture effect, limited bandwidth, long propagation delay, high path loss, node mobility etc.). For this reason, a statistical signal processing approach using cross-correlation of Gaussian signals received at multiple sensors has been investigated. Equidistant sensors have been used for estimation in three-sensor case of this technique. But, due to the hazardous condition of underwater environment, equal sensor spacing may not be possible in practical cases. To address this issue, effect of unequal sensor spacing on estimation results is analyzed in this paper.**

Keywords—Cardinality estimation; cross-correlation function; underwater acoustic channel; underwater acoustic sensor network; unequal sensor separation.

I. INTRODUCTION

More than seventy percent (70%) of earth's surface is enveloped with water. So, we have no choice but to utilize this vast resources for the welfare of mankind as a medium of reliable communication. Underwater acoustic sensor network (UASN) is one of the most desirable means of communication in aqueous environment as it has many valuable applications, such as oceanographic data collection, pollution monitoring, tactical surveillance, assisted navigation, tsunami warnings [1], oil/gas spills monitoring, and so forth. To make these applications viable, network cardinality estimation plays a vital role which assures proper network operations. So, it is very important to estimate the number of active nodes accurately within the region of interest of UASN.

A number of cardinality estimation techniques have been investigated so far for various types of network. A cardinality estimation technique for mobile peer-to-peer (P2P) network has been proposed by Chen et al. [2] involving only a subset of nodes. Another new efficient method to establish the size of an unstructured P2P network has been introduced by Evans et al. [3]. In radio frequency identification (RFID) system, number of tag ID which refers to the network cardinality has been estimated using different protocols [4], [5]. Consensus based size estimation methods have been discussed in [6], [7]. Estreme, a neighborhood cardinality estimator has been proposed by Cattani et al. [8] to address the problem of neighborhood cardinality estimation in dynamic wireless networks. To estimate the size of wireless sensor networks, novel distributed algorithms based on distributed forms of Gram-Schmidt orthogonalization algorithms have been presented by Sluciak et al. [9]. These estimation techniques show poor performances in UASN due to the unique characteristics of underwater acoustic channel (UAC) [10]. To solve this issue, an underwater network cardinality estimation approach based on intermission between the nodes has been proposed by Blouin [11]. Considering non-negligible capture effect and large propagation latency of UAC, ALOHA based delay insensitive network size estimation technique using empty slot estimator has been investigated by Howlader et al. [12], [13]. But, all of these estimation techniques ([11]–[13]) are based on protocol(s) and suffer from protocol complexity in underwater environment.

To reduce the protocol complexity, a novel estimation process based on cross-correlation of Gaussian signals received at two sensors has been introduced by Anower et al. [14] which is effective for any type of environment networks. Three-sensor version of this estimation process has been proposed by Chowdhury et al. [15], [16] to improve the estimation performance [17]. For cardinality estimation, three sensors can be arranged in two different ways: sensors in line (SL) arrangement and triangular sensors (TS) arrangement. In both arrangements, equal separation between the sensors has been assumed for estimation purpose. This might be difficult to achieve in deep water due to the waves and sea creatures. It is also troublesome to locate and retain the sensors at desired locations (to maintain equal sensor spacing) among the randomly distributed nodes.

Therefore, the assumption of equidistant sensors creates serious limitations and requires further investigation. In this paper, to eliminate these limitations unequally spaced sensors are used for cardinality estimation using SL case of three-sensor scheme and analyzed the effect of unequal sensor spacing on estimation results.

II. METHODOLOGY

A brief description of the cardinality estimation process using SL scheme [15], [16] is presented in this section. In this estimation process, a 3D spherical region has been considered as an UASN that contains N evenly distributed nodes and three

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

equally separated sensors (H_1, H_2 and H_3) as shown in Fig. 1. These probing nodes (sensors) are placed in a line inside the UASN for estimation purpose, where the middle sensor (H_2) lies at the centre of the sphere and the other two probing nodes (H_1 and H_3) are positioned such that, $d_{\mathrm{DBS}_{12}}$ (distance between H_1 and H_2) = $d_{\mathrm{DBS}_{23}}$ (distance between H_2 and H_3) = d_{DBS} (distance between the equidistant pair of sensors).

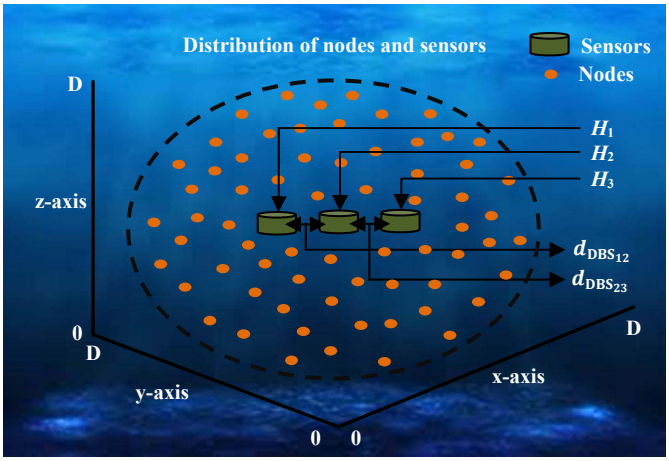


Fig. 1. Distribution of underwater network nodes with N transmitting nodes and three equally spaced sensors for SL scheme.

These three sensors are considered as receiving nodes that receive Gaussian signals from the surrounding N nodes which are considered as the transmitting nodes. These sensors initiate the estimation process by sending probe request to N number of transmitting nodes which in response transmit Gaussian signals to the sensors. These Gaussian signals are summed at each sensor location to form composite Gaussian signals. By cross-correlating these Gaussian signals two cross-correlation functions (CCFs), $C_{12}(\tau)$ (formulated by cross-correlating the composite Gaussian signals $S_{\mathrm{rc}_1}(t)$ and $S_{\mathrm{rc}_2}(t)$ received at H_1 and H_2, respectively) and $C_{23}(\tau)$ (formulated by cross-correlating the composite Gaussian signals $S_{\mathrm{rc}_2}(t)$ and $S_{\mathrm{rc}_3}(t)$ received at H_2 and H_3, respectively) is formulated in SL scheme. Due to simultaneous transmission from all nodes, the total estimation time is less for this scheme even at the presence of large propagation latency of UASN [17].

Gaussian signal has a certain characteristics that, cross-correlation of two Gaussian signals results a delta function, which is the basic idea of this estimation approach and also the reason of using Gaussian signals as transmitted signals. Thus, CCF(s) due to the composite Gaussian signals takes the form of a series of delta functions [14] as shown in Fig. 2. Bins, b in the CCF (as shown in Fig. 2) is defined as a place occupied by a delta inside a space of a width twice the distance between sensors and that place is determined by the delay difference of the signals coming to the sensors [18]. The deltas of equal delay differences are placed in that particular bin. Number of bins, b is defined as twice the number of samples between the sensors, m, minus one and can be written as [18]:

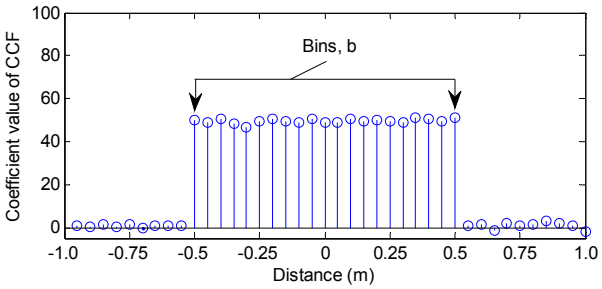


Fig. 2. Bins, b of CCF obtained with N (=1000) nodes.

$$b = \frac{2 \times d_{\mathrm{DBS}} \times S_{\mathrm{R}}}{S_{\mathrm{P}}} - 1 \tag{1}$$

Here, S_{R} is the sampling rate and S_{P} means speed of acoustic wave propagation.

Ratio of standard deviation (σ) to the mean (μ), R of CCF is chosen as the cardinality estimation parameter of this process, as it is the most preferable estimation parameter of cross-correlation based technique [19], [20]. Two estimation parameters, R_{12} and R_{23} are derived from two CCFs, $C_{12}(\tau)$ and $C_{23}(\tau)$, respectively, to calculate the final estimation parameter of SL scheme. Then, the final estimation parameter, $R^{\mathrm{2CCF}}_{\mathrm{average}}$ of SL scheme is obtained by taking the average of R_{12} and R_{23}, and can be expressed as [15]:

$$R^{\mathrm{2CCF}}_{\mathrm{average}} = \frac{R_{12} + R_{23}}{2} = \frac{\frac{\sigma_{12}}{\mu_{12}} + \frac{\sigma_{23}}{\mu_{23}}}{2} \tag{2}$$

where, σ_{12} and σ_{23} are the standard devation and μ_{12} and μ_{23} represent the mean of the CCFs, $C_{12}(\tau)$ and $C_{23}(\tau)$, respectively.

Direct calculation of standard deviation and mean of the CCFs using mathematical expressions is very complex. So, the cross-correlation problem is reframed into a probability problem using the well-known occupancy problem [21], [22]. After reframing, R of CCF can be written as [14]:

$$R = \frac{\sigma}{\mu} = \frac{\sqrt{N \times \frac{1}{b} \times \left(1 - \frac{1}{b}\right)}}{N \times \frac{1}{b}} = \sqrt{\frac{(b-1)}{N}} \tag{3}$$

Using this expression, (2) can be expressed as:

$$R^{\mathrm{2CCF}}_{\mathrm{average}} = \frac{R_{12} + R_{23}}{2} = \frac{\sqrt{\frac{b_{12}-1}{N}} + \sqrt{\frac{b_{23}-1}{N}}}{2} \tag{4}$$

Here, b_{12} and b_{23} denote the number of bins of the CCFs, $C_{12}(\tau)$ and $C_{23}(\tau)$, respectively. In this case, $b_{12} = b_{23} = b$ according to (1); as the values of S_{R} and S_{P} are fixed during the estimation process and $d_{\mathrm{DBS}_{12}} = d_{\mathrm{DBS}_{23}} = d_{\mathrm{DBS}}$. Hence, (4) becomes [15]:

$$R^{\mathrm{2CCF}}_{\mathrm{average}} = \frac{R_{12} + R_{23}}{2} = \sqrt{\frac{(b-1)}{N}} \tag{5}$$

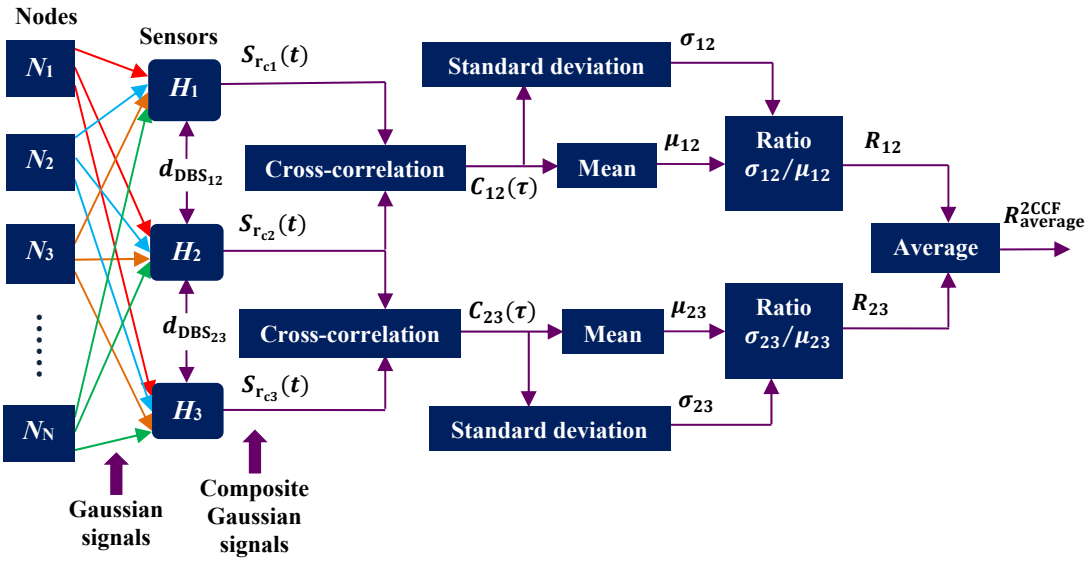


Fig. 3. Block diagram representation of the estimation process for SL scheme.

Using (5), N can be estimated, as we know b from (1) and can calculate $R^{2CCF}_{average}$ from the CCFs. The total estimation process of SL scheme is represented through a block diagram in Fig. 3.

Since all the transmitted Gaussian signals are combined at the sensors and node specific signals are irrelevant for cardinality estimation, the concept of capture effect does not apply to this method. Moreover, this technique requires a simple protocol for probing to initiate the simultaneous transmission, which eliminates the problem of using medium access control protocol in underwater network. But, limited underwater bandwidth (BW) restricts the use of infinite band Gaussian signals. For this reason, band limited Gaussian signals have been used to effectively estimate the cardinality of UASN [23], [24].

III. Effect of Unequal Sensor Spacing

In previous section, estimation is performed using SL scheme with equal distances between the sensors. This section investigates the effect of unequal sensor separation on the estimation process. In this investigation, three unequally spaced sensors denoted by H_1, H_2 and H_3, are considered along a line to perform estimation, where the middle sensor (H_2) is placed at the centre of the spherical shaped network as shown in Fig. 4. Unequal spacings between the sensors of Fig. 4 suggest that, $d_{DBS_{12}} \neq d_{DBS_{23}} \neq d_{DBS_{31}}$ (distance between H_3 and H_1).

Network cardinality estimation process with unequally separated sensors using SL scheme is similar to that of equally spaced sensors as described in Section II. Therefore, final estimation parameter of SL scheme with unequal sensor separation, $R^{2CCF}_{unequal:SL}$ can be calculated similarly as discussed in the previous section by taking the average of the estimation parameters, R_{12} and R_{23}, obtained from the CCFs, $C_{12}(\tau)$ and $C_{23}(\tau)$, recpectively. After reframing the cross-correlation problem into a probability problem using the well-known occupancy problem, $R^{2CCF}_{unequal:SL}$ of CCFs can be written using (4) as:

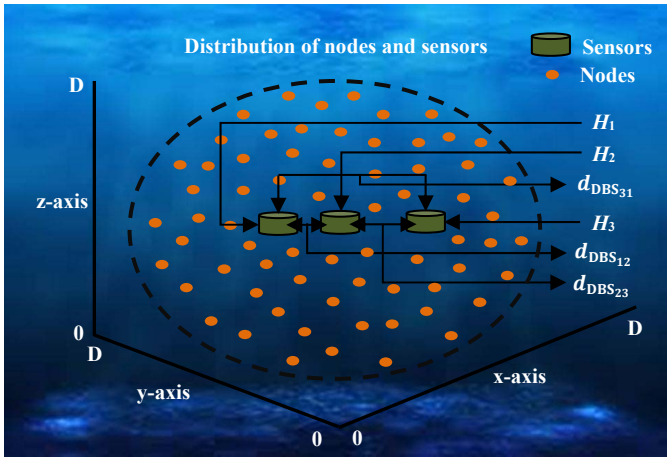


Fig. 4. Distribution of underwater network nodes with N transmitting nodes and three unequally separated sensors for SL scheme.

$$R^{2CCF}_{unequal:SL} = \frac{R_{12} + R_{23}}{2} = \frac{\sqrt{\frac{b_{12}-1}{N}} + \sqrt{\frac{b_{23}-1}{N}}}{2} \quad (6)$$

Now, b_{12} and b_{23} can be expressed using (1) as:

$$b_{12} = \frac{2 \times d_{DBS_{12}} \times S_R}{S_P} - 1 \quad (7)$$

and

$$b_{23} = \frac{2 \times d_{DBS_{23}} \times S_R}{S_P} - 1 \quad (8)$$

respectively. In this case, $b_{12} \neq b_{23}$ according to (7) and (8); as the values of S_R and S_P are fixed during the estimation process and $d_{DBS_{12}} \neq d_{DBS_{23}}$. Consequently, it is not possible to simplify (6) any further. Network cardinality, N can be estimated after rearranging (6) as follows:

$$N = \left(\frac{\sqrt{b_{12}-1} + \sqrt{b_{23}-1}}{2 \times R^{2CCF}_{unequal:SL}} \right)^2 \quad (9)$$

To obtain a relationship between equal and unequal sensor separation cases of SL scheme, ratio of estimation parameters of these cases, $R^{2CCF}_{average}$ and $R^{2CCF}_{unequal:SL}$, respectively, is taken as:

$$r = \frac{R^{2CCF}_{average}}{R^{2CCF}_{unequal:SL}} = \frac{\sqrt{\frac{(b-1)}{N}}}{\frac{\sqrt{\frac{b_{12}-1}{N}} + \sqrt{\frac{b_{23}-1}{N}}}{2}}$$

i.e., $$r = \frac{R^{2CCF}_{average}}{R^{2CCF}_{unequal:SL}} = \frac{2\sqrt{b-1}}{\sqrt{b_{12}-1} + \sqrt{b_{23}-1}} \quad (10)$$

From (10), we can write:

$$R^{2CCF}_{average} = r \times R^{2CCF}_{unequal:SL} \quad (11)$$

This is the relationship between the estimation parameters of SL scheme with equal and unequal spacings between the sensors. It is obvious from (11) that, similar $R^{2CCF}_{average}$ and $R^{2CCF}_{unequal:SL}$ is possible, if $r = 1$. Hence, we can formulate a condition to obtain similar $R^{2CCF}_{average}$ and $R^{2CCF}_{unequal:SL}$ by putting $r = 1$ in (10) as follows:

$$\frac{2\sqrt{b-1}}{\sqrt{b_{12}-1} + \sqrt{b_{23}-1}} = 1$$

i.e., $$b = \frac{\left(\sqrt{b_{12}-1} + \sqrt{b_{23}-1}\right)^2 + 4}{4} \quad (12)$$

Another CCF, $C_{31}(\tau)$ (formulated by cross-correlating the composite Gaussian signals received at H_3 and H_1) can be used for estimation in unequal sensor spacing case of SL scheme. This CCF is not considered in equal sensor separation case, as only the CCFs due to the equidistant pair of sensors are used for that case. Considering the additional CCF, $C_{31}(\tau)$, the final estimation parameter of SL scheme with unequally separated sensors can be obtained by averaging the estimation parameters, R_{12}, R_{23} and R_{31}, derived from the CCFs, $C_{12}(\tau)$, $C_{23}(\tau)$ and $C_{31}(\tau)$, respectively, and can be expressed as:

$$R^{3CCF}_{unequal:SL} = \frac{\sqrt{\frac{b_{12}-1}{N}} + \sqrt{\frac{b_{23}-1}{N}} + \sqrt{\frac{b_{31}-1}{N}}}{3} \quad (13)$$

Here, b_{31} represents the number of bins of $C_{31}(\tau)$ and can be written using (1) as:

$$b_{31} = \frac{2 \times d_{DBS_{31}} \times S_R}{S_P} - 1 \quad (14)$$

It is obvious from (7), (8) and (14) that, $b_{12} \neq b_{23} \neq b_{31}$. In SL scheme, the sensor arrangements are such that, $d_{DBS_{31}} = d_{DBS_{12}} + d_{DBS_{23}}$ as shown in Fig. 4. Now, by putting this relationship into (14), we can write:

$$b_{31} = b_{12} + b_{23} + 1 \quad (15)$$

Accordingly, (13) becomes,

$$R^{3CCF}_{unequal:SL} = \frac{\sqrt{\frac{b_{12}-1}{N}} + \sqrt{\frac{b_{23}-1}{N}} + \sqrt{\frac{b_{12}+b_{23}}{N}}}{3} \quad (16)$$

The condition of identical results for $R^{2CCF}_{average}$ and $R^{3CCF}_{unequal:SL}$ is derived in similar way as shown earlier and can be expressed as:

$$b = \frac{\left(\sqrt{b_{12}-1} + \sqrt{b_{23}-1} + \sqrt{b_{12}+b_{23}}\right)^2 + 9}{9} \quad (17)$$

It is noted that, b_{12} and b_{23} are the two different number of bins of the two CCFs for unequal sensor separation case and b is the number of bins of these two CCFs for equal sensor separation case.

IV. Results and Discussion

Matlab is used for simulation purposes. Some useful simulation results of SL scheme with unequal sensor spacing are presented in this section to verify the estimation process described in Section III. The parameters used in the simulations are shown in the Table I.

TABLE I. Simulation Parameters

Parameters	Values
Dimension of the sphere, D	2000 m
Signal length, N_s	10^6 samples
Sampling rate, S_R	30 kSa/s
Speed of propagation, S_P	1500 m/s
Absorption coefficient, a	1
Dispersion factor, k	0
SNR	20 db
Distance between the sensors, d_{DBS}, $d_{DBS_{12}}$, and $d_{DBS_{23}}$	0.8m, 0.25m, and 1.25m (can be varied)

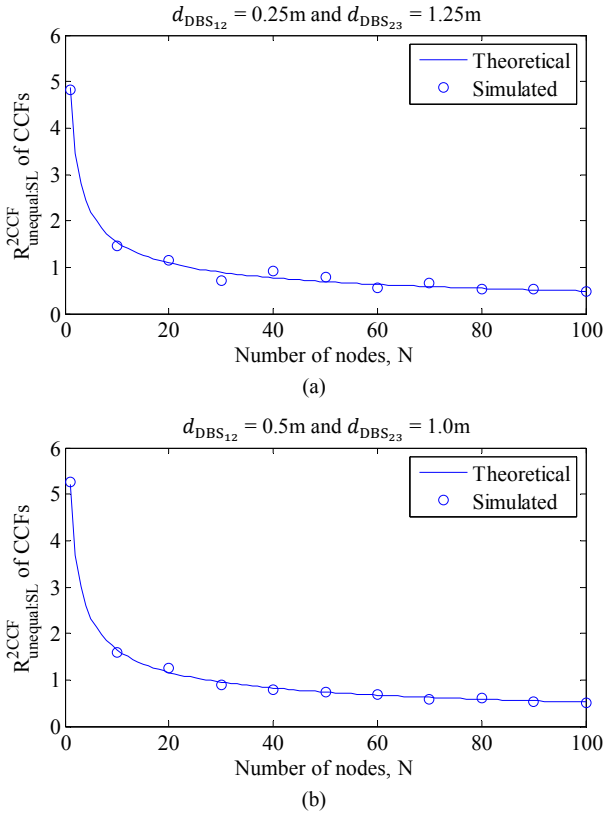


Fig. 5. $R^{2CCF}_{unequal:SL}$ versus N plots for SL scheme with unequal sensor separation using different $d_{DBS_{12}}$ and $d_{DBS_{23}}$.

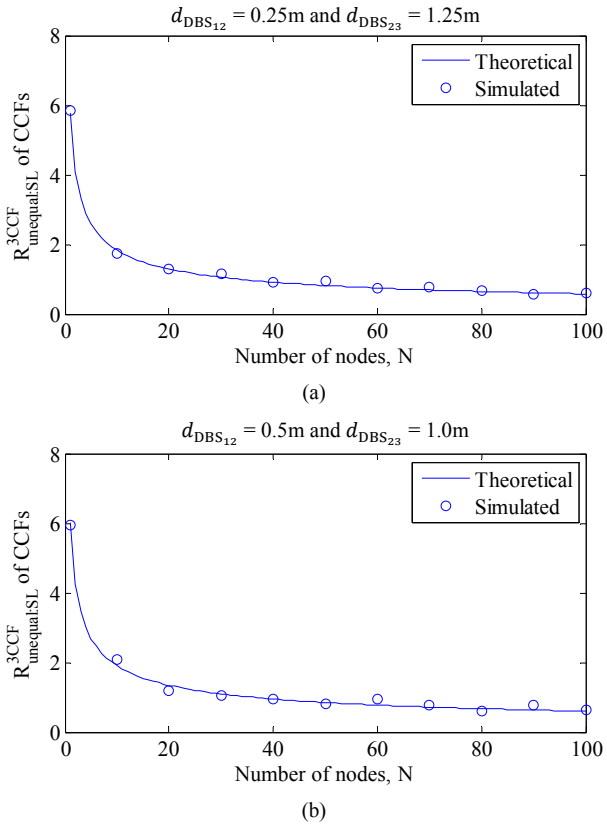


Fig. 6. $R^{3CCF}_{unequal:SL}$ versus N plots for SL scheme with unequal sensor separation using different $d_{DBS_{12}}$ and $d_{DBS_{23}}$.

Simulated results of $R^{2CCF}_{unequal:SL}$ and $R^{3CCF}_{unequal:SL}$ are plotted against N in Fig. 5 and 6, respectively, with the corresponding theoretical results (obtained using (6) and (16), respectively) for different $d_{DBS_{12}}$ and $d_{DBS_{23}}$. The values of $d_{DBS_{12}}$ and $d_{DBS_{23}}$ are: 0.25m and 1.25m in Fig. 5(a) and 6(a); and 0.5m and 1.0m in Fig. 5(b) and 6(b), respectively. In these figures, the solid lines indicate the theoretical results and the circles the corresponding simulated results. Matching theoretical and simulated results of $R^{2CCF}_{unequal:SL}$ and $R^{3CCF}_{unequal:SL}$ in Fig. 5 and 6, respectively, indicate the effectiveness of the cross-correlation based cardinality estimation process using unequally separated sensors.

Estimation is performed using simulations to validate the conditions ((12) and (17)) formulated in the previous section which result similar estimation parameters for equal and unequal sensor spacing cases. Fig. 7 and 8 show the identical results of: $R^{2CCF}_{average}$ and $R^{2CCF}_{unequal:SL}$; and $R^{2CCF}_{average}$ and $R^{3CCF}_{unequal:SL}$, respectively. In Fig. 7 and 8, the solid lines represent the theoretical results and the circles and stars represent the simulated results with equal and unequal distances between the sensors, respectively. The values used in Fig. 7 and 8 are: b = 31 (d_{DBS} = 0.8m), b_{12} = 19 ($d_{DBS_{12}}$ = 0.5m) and b_{23} = 46 ($d_{DBS_{23}}$ = 1.175m) which satisfies (12); and b = 40 (d_{DBS} = 1.025m), b_{12} = 19 ($d_{DBS_{12}}$ = 0.5m) and b_{23} = 44 ($d_{DBS_{23}}$ = 1.125m) which satisfies (17), respectively.

It is obvious from Fig. 7 and 8 that, the two simulated results agree with each other and both follow the theoretical

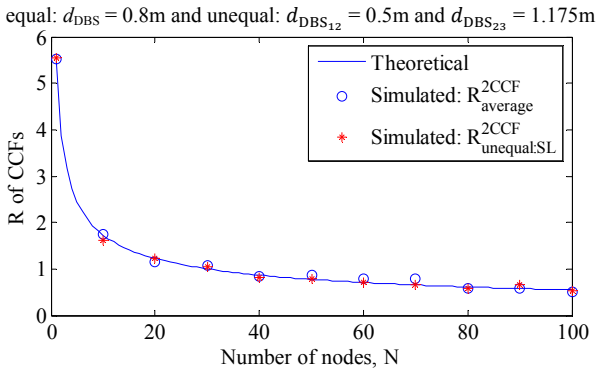


Fig. 7. R versus N plot comparisons of results for theoretical and simulated with equal ($R^{2CCF}_{average}$: b = 31) and unequal ($R^{2CCF}_{unequal:SL}$: b_{12} = 19 and b_{23} =46) sensor separation using SL scheme.

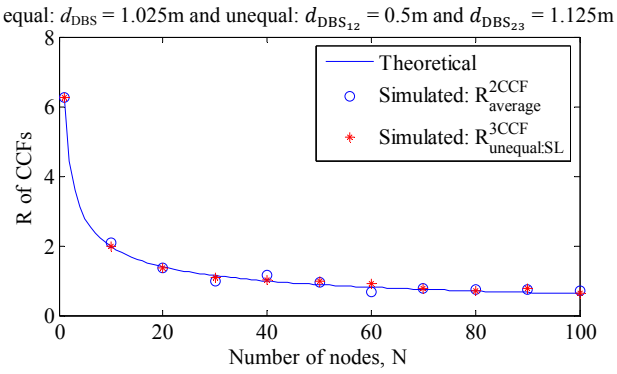


Fig. 8. R versus N plot comparisons of results for theoretical and simulated with equal ($R^{2CCF}_{average}$: b = 40) and unequal ($R^{3CCF}_{unequal:SL}$: b_{12} = 19 and b_{23} =44) sensor separation using SL scheme.

results. These results demonstrate the usefulness of the conditions derived in Section III. The results obtained in this paper are based on an assumption that, the sensors receive equal power from the nodes which can be achieved by applying proper probing technique. However, the assumption of equal received power has been further investigated with equal transmitted power of the nodes to analyze the effect of dispersion coefficient on the estimation process [26].

V. CONCLUSION

The distances between the sensors are an important factor for cross-correlation based cardinality estimation technique as number of bins of the CCFs depend on them. But, the constraint of equal sensor spacing in three-sensor schemes makes practical implementation of these schemes difficult in harsh underwater environment. So, it is preferable to have a estimation process that is effective in both equal and unequal sensor separation cases. This paper focuses on the use of unequally separated sensors using SL scheme to obtain a flexible estimation method.

The effectiveness of this method is verified by simulations in the previous section. Unequal distances between the sensors allow SL scheme to use an additional CCF which might improve the estimation performance in terms of accuracy and required estimation time [16]. Current research is going on to compare the estimation performance of equal and unequal

sensor separation cases in terms of accuracy, required estimation time and energy. In future, we intend to provide a more flexible estimation method by removing the assumption of spherical shaped network, central placement of sensors, and uniform distribution of the transmitting nodes. Moreover, we intend to conduct experimental estimation using this technique.

REFERENCES

[1] A. Zielinski, "Communications Underwater," *Journal of Hydroacoustics* (Invited Paper), vol. 7, pp. 235–252, 2004.

[2] Shipping Chen, Y. Qiao, S. Chen, and J. Li, "Estimating the cardinlaity of a mobile peer-to-peer network," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 9, pp. 359–368, September 2013. doi: 10.1109/JSAC.2013.SUP.0513032.

[3] N. Evans, B. Polot, and C. Grothoff, "Efficient and Secure Decentralized Network Size Estimation," In *Proceedings of NETWORKING 2012*, vol. 7289, pp. 304–317. doi: 10.1007/978-3-642-30045-5_23.

[4] P. Solic, J. Radic, and N. Rozic, "Energy efficient tag estimation method for ALOHA based RFID system," *IEEE Journal on Sensors*, vol. 14, no. 10, pp. 3637–3647, Oct. 2014. doi: 10.1109/JSEN.2014.2330418.

[5] Y. Zheng, and M. Li, "Towards More Efficient Cardinality Estimation for Large-Scale RFID Systems," *IEEE/ACM Transactions on Networking*, vol. 22, no. 6, pp. 1886–1896, Dec. 2014. doi: 10.1109/TNET.2013.2288352.

[6] H. Terelius, D. Varagnolo, and K. H. Johansson, "Distributed size estimation of dynamic anonymous networks," In *Proceedings of IEEE 51st Annual Conference on Decision and Control (CDC)*, Maui, Hawaii, USA, 10–13 Dec. 2012, pp. 5221–5227.
doi: 10.1109/CDC.2012.6425912.

[7] D. Varagnolo, G. Pillonetto, and L. Schenato, "Distributed Cardinality Estimation in Anonymous Networks," *IEEE Transactions on Automatic Control*, vol. 59, no. 3, pp. 645–659, March 2014. doi: 10.1109/TAC.2013.2287113.

[8] M. Cattani, M. Zuniga, A. Loukas, and K. Langendoen, "Lightweight neighborhood cardinality estimation in dynamic wireless networks," In *Proceedings of the 13th International Symposium on Information Processing in Sensor Networks (IPSN-14)*, 15–17 April 2014, Berlin, Germany, pp.179–189. doi: 10.1109/IPSN.2014.6846751.

[9] O. Sluciak, and M. Rupp, "Network size estimation using distributed orthogonalization," *IEEE Signal Processing Letters*, vol. 20, no. 4, pp. 347–350, April 2013. doi: 10.1109/LSP.2013.2247756.

[10] S. Climent, A. Sanchez, J. V. Capella, N. Meratnia, and J. J. Serrano, "Underwater Acoustic Wireless Sensor Networks: Advances and Future Trends in Physical, MAC and Routing Layers," *Sensors (Basel, Switzerland)*, vol. 14, no. 1, pp. 795–833, Jan. 2014. doi: 10.3390/s140100795.

[11] S. Blouin, "Intermission-based adaptive structure estimation of wireless underwater networks," In *Proceedings of 10th IEEE International Conference on Networking, Sensing and Control (ICNSC)*, 10–12 April, 2013, pp. 146–151. doi: 10.1109/ICNSC.2013.6548727.

[12] M. S. A. Howlader, M. R. Frater, and M. J. Ryan, "Estimation in underwater sensor networks taking into account capture," In *Proceedings of IEEE Oceans'07*, Aberdeen, Scotland, 18–21 June, 2007, pp. 1–6.

[13] M. S. A. Howlader, M. R. Frater, and M. J. Ryan, "Delay-insensitive identification of neighbors using unslotted and slotted protocols," *Wirel. Commun. Mob. Comput.*, vol. 14, no. 8, pp. 831–848, June 2014. doi: 10.1002/wcm.2242.

[14] S. Anower, M. R. Frater, and M. J. Ryan, "Estimation by cross-correlation of the number of nodes in underwater networks," In *Proc. Australasian Telecommunication Networks and Applications Conf. (ATNAC)*, 10–12 November, 2009, pp. 1–6.
doi: 10.1109/ATNAC.2009.5464716.

[15] S. A. H. Chowdhury, M. S. Anower, and J. E. Giti, "A signal processing approach of underwater network node estimation with three sensors," In *Proc. 1st Int. Conf. Electrical Engineering and Information & Communication Technology (ICEEICT 2014)*, 10–12 April, 2014, Dhaka, Bangladesh, pp. 1–6. doi: 10.1109/ICEEICT.2014.6919116.

[16] S. A. H. Chowdhury, M. S. Anower, and J. E. Giti, "Effect of sensor number and location in cross-correlation based node estimation technique for underwater communications network," In *Proceedings of 3rd International Conference on Informatics, Electronics & Vision (ICIEV 2014)*, 23–24 May, 2014, Dhaka, Bangladesh, pp. 1–6. doi: 10.1109/ICIEV.2014.6850728.

[17] S. A. H. Chowdhury, M. S. Anower, and J. E. Giti, "Performance comparison of underwater network size estimation techniques," *Int. J. Systems, Control and Communications* (in press).

[18] M. S. Anower, "Estimation using cross-correlation in a communications network," Ph.D. dissertation, SEIT, University of New South Wales at Australian Defense Force Academy, Canberra, 2011.

[19] S. A. H. Chowdhury, M. S. Anower, J. E. Giti, and M. I. Haque, "Effect of signal strength on different parameters of cross-correlation function in underwater network cardinality estimation," In *Proceedings of 17th International Conference on Computer and Information Technology (ICCIT 2014)*, 22–23 Dec. 2014, Dhaka, Bangladesh, pp. 430–434. doi: 10.1109/ICCITechn.2014.7073159.

[20] M. S. Anower, S. A. H. Chowdhury, J. E. Giti, A. S. M. Sayem, and M. I. Haque, "Underwater network size estimation estimation using cross-correlation: selection of estimation parameter," In *Proc. 9th Int. Forum on Strategic Technology (IFOST 2014)*, 21–23 Oct. 2014, Cox's Bazar, Bangladesh, pp. 170–173. doi: 10.1109/IFOST.2014.6991097.

[21] W. Feller, *An Introduction to Probability Theory and its Applications*, John Wiley, 1968.

[22] M. S. Anower, M. A. Motin, A. S. M. Sayem, and S. A. H. Chowdhury, "A node estimation technique in underwater wireless sensor network," In *Proceedings of* 2n*d International Conference on Informatics, Electronics & Vision (ICIEV 2013)*, 17–18 May, 2013, pp. 1–6.
doi: 10.1109/ICIEV.2013.6572582.

[23] S. K. Bain, S. A. H. Chowdhury, A. H. M. Asif, M. S. Anower, M. F. Pervej, and S. S. Haque, "Impact of underwater bandwidth on cross-correlation based node estimation technique," In *Proceedings of 17th International Conference on Computer and Information Technology (ICCIT 2014)*, 22–23 Dec. 2014, Dhaka, Bangladesh, pp. 494–497.
doi: 10.1109/ICCITechn.2014.7073159.

[24] M. S. Anower, S. A. H. Chowdhury, J. E. Giti, and M. I. Haque, "Effect of bandwidth in cross-correlation based underwater network size estimation," In *Proc. 8th Int. Conf. Electrical and Computer Engineering (ICECE 2014)*, 20–22 Dec. 2014, Dhaka, Bangladesh, pp. 413–416. doi: 10.1109/ICECE.2014.7026897.

[25] M. A. Hossen, S. A. H. Chowdhury, M. S. Anower, S. Hossen, M. F. Pervej, and M. M. Hasan, "Effect of Signal Length in Cross-correlation based Underwater Network Size Estimation," In *Proc. 2nd Int. Conf. Electrical Engineering and Information & Communication Technology (ICEEICT)*, 21–23 May, 2015, Dhaka, Bangladesh.

[26] M. K. Hossain, M. S. Anower, M. M. Rahman, and S. M. N. Siraj, "Effect of dispersion coefficient on underwater network size estimation," In *Proc. 2nd Int. Conf. Electrical Engineering and Information & Communication Technology (ICEEICT)*, 21–23 May, 2015, Dhaka, Bangladesh.

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

Cross-Gender Acoustic Differences in Hypernasal Speech and Detection of Hypernasality

Shahina Haque* and Md. Hanif Ali
Department of CSE
Jahangirnagar University
Savar, Dhaka-1342, Bangladesh
*Corresponding author: alphashahina@yahoo.com

A.K.M. Fazlul Haque
Department of ETE
Daffodil International University
Dhaka-1207, Bangladesh

***Abstract*—Spectrum of hypernasal speech produced by cleft palate (CP) speakers carries acoustic information regarding hypernasality (HP). This study compares the variation of acoustic features of HP in continuous read speech with gradually increasing severity of HP within and across gender and detect HP. Three point vowels /i/, /a/ and /u/ from continuous read speech produced by seven male and seven female speakers are used. In first part of this study, comparative study of variation in acoustic parameters within and across vowel category of males and females are made. Experimental results show considerable variability in spectral features among CP subjects. High back vowel /u/ shows greatest variability for both males and females. Mid low vowel /a/ shows least variation among the vowels for females and high front vowel /i/ shows least variation for males. The inter-speaker variability measurement suggests that high back vowel /u/ is mostly affected and has the highest variability. In the second part, ratio of vowel space area (VSA) of hypernasal and normal speech is used as a threshold for HP detection. Results show that VSA spanned by CP males is shrunken by 3.7 times than CP females. This indicates that VSA spanned by CP females is larger compared to CP males.**

Keywords—acoustic feature; cleft palate; continuous read speech; cross gender; hypernasality; speech analysis; velopharyngeal opening; vowel space area

I. INTRODUCTION

Cleft Palate speakers' with defective velopharyngeal (VP) mechanism produces hypernasal speech with reduced intelligibility [1]. HP and is the second most frequent congenital malformation worldwide [2]. Detection of HP using speech processing techniques is useful for it's simplicity for the diagnosis of CP speakers. It aids to decide the severity of HP and what support (surgery or speech therapy) is to be provided to the CP speakers by the physicians.

In 1971, a detailed analysis of the variation in the voice tone measurement was done [3]. Fant found that for a nasalized vowel, oral-nasal coupling introduces an additional pole-zero pair into the oral vowel [4]. HP detection is performed by many researchers by analyzing the disordered speech, synthesized hypernasal speech, and nasalized vowels of normal speech. Main features of nasalization are changes in the low-frequency regions of the speech spectrum, where there is a very low-frequency peak with wide bandwidth along with the presence of a pole-zero pair due to the acoustic coupling was shown by

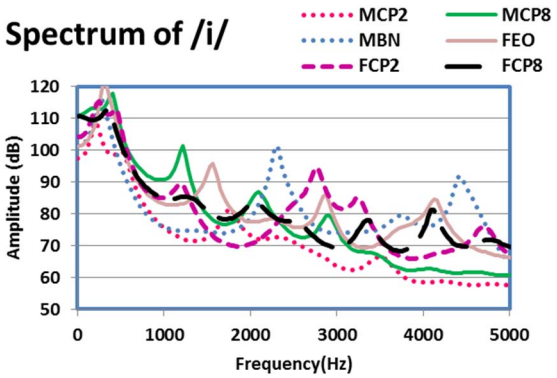


Fig. 1. Vocal tract transfer function for /i–/ and /i/ for normal and CP male (M) and female (F) speakers.

Hawkins and Stevens [5] and Glass and Zue [6]. HP detection was done by using Teager energy operator for multicomponent signals by Cairns et al. [7]. Presence of zeros in spectrum is used as a cue for the detection of HP by Rah et al [8]. From the literature, it is observed that the acoustic cues of hypernasal speech are additional formants, antiformants, formant bandwidth broadening. Study on perceptual analysis was carried out on the vowel sounds by adding nasal formants in spectrum [9-10], showed that formant at 250 Hz plays important role in the nasalization of vowels. Group delay function was used successfully used for hypernasality detection using acoustic features of speech in CP speakers [10-11].

Vowel Space Area (VSA) is the two-dimensional area bounded by lines connecting first and second formant frequency coordinates (F2/F1) of vowels [12]. VSA has been used for various purpose such as studying vowel identity, speaker characteristics, speech development, speech disorder, vowel distinctiveness and assess intelligibility that influences vowel production [13-15]. As frequencies of the first and second formants is related to the size and shape of the cavities created by mouth opening (F1) and tongue position (F2), the VSA reflects the movement of the speech articulators. In general, studies have shown that VSA is larger in speech that is clearer and more intelligible than speech associated with smaller VSAs. Abnormal vowel formant change (centralization) is a common feature of speech production deficiency, VSA estimation is used for characterizing speech motor control, including speech development, speech disorders. In a study with large database, automatic assessment of VSA was done and is reported to have good result than the traditional

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

method of VSA measurement [16]. In another study it was found that psychological distress and depression reduces the VSA [17]. Hypernasal vowel speech near a plosive of CP children were analyzed and proposed an objective measure. It was found that mean falling and rising slopes of the amplitude in the nasalized vowel are smaller than those of the oral vowel [18]. Recently, perceptual speech intelligibility and speech production variability in disordered speech was studied using various measures and found that among the various acoustic measures VSA was drastically reduced for children with cerebral palsy [19]. Spectral enhancement is done for cleft lip and palate speech and is reported to have better intelligibility than the speech after repaired surgery [20].

Most of the study on acoustic analysis of HP is based on isolated sustained vowel for HP. Variation of spectral features and HP was carried out using simple method and VSA in a study for CP males using continuous speech [21]. There is less work reported on cross gender comparative study for variation in speech parameters with defective VP opening causing hypernasal speech in continuous read speech sentences which exhibit greater complexities with respect to speech intelligibility, which formed the motivation for this study. This study aims to investigate how useful the extracted speech parameter information from continuous read speech rather than isolated sustained vowel to capture the defect in speech production articulators in terms of VSA for the detection of HP. This report presents a comparative cross gender study on the VP opening variability on speech features and the relationship between HP detection using VSA in read speech.

Nasality has similarity with HP in production which is reflected in acoustic features. Most of the study is concerned with nasalization of vowels near a nasal consonant to make a comparison with HP. In this study Bangla nasal vowels are used to make a comparison of HP with nasality. In Bangla, all the seven vowels have their nasal counterpart. Bangla is a language in which nasality is phonemic. Thus to make a comparison between nasality and HP, Bangla oral and nasal vowels are taken into account and used as reference. This study is performed to answer the following research questions:

1) To what extent vocal tract features are affected by variation of HP within and across vowels? How these effects differ across gender?

2) To what extent does the VSA of CP speakers' is affected by HP as compared to the normal speaker? How these effects differ from male to female?

The rest of the paper is organized as follows:

Section II describes about speech materials, Section III discusses about the acoustic analysis procedure, Section IV discusses about results and discussions. Section V concludes the paper.

II. Speech Materials

The aim of this section is to describe how the speech samples are acquired. Front, mid and back vowels /i/, /a/ and /u/ of English and Bangla language are selected as the speech samples of the study. The experimental part consists of recording each of the isolated Bangla oral-nasal vowels at a normal speaking rate three times in a quiet room in a DAT tape at a sampling rate of 48 kHz and 16 bit value. The best one of these three speakers sample data is used for the study. Speech data for three English vowels /i/, /a/ and /u/ of normal and CP speakers with gradually increasing severity of HP are obtained from read speech data of American Cleft Palate Craniofacial Association. A stable portion is cut from each of the selected vowel for the purpose of our work. These digitized speech sound are then downsampled to 22050Hz and normalized for the purpose of analysis. Vowels uttered by non-CP speakers are used as reference.

Three types of 60 data used are:

- 12 Isolated uttered Bangla data = 3 Bangla oral (BO) and Isolated uttered 3 Bangla Nasal (BN) vowel data obtained from 1 male and 1 female Non-CP Bangla speakers.
- 6 English data segmented from continuous English read speech = 3 English oral (EO) vowel data of 1 male and 1 female Non-CP English speakers.
- 42 EO data segmented from continuous English read speech = 3 data of 7 males and 7 females CP English speakers.

III. Acoustic Analysis of Normal and Hypernasal Speech

Speech signal is non-stationary in nature, but it can be assumed to be stationary over short duration called frames by windowing for the purpose of analysis. Speech signal is analyzed frame-wise, with a frame-rate of 50-100 frames/sec, and for each frame the duration of speech segment is taken to be 20-30 msec. A new frame is obtained by shifting the Hamming windowing function by 10msec to a subsequent time. After normalization and windowing, the speech samples are ready to be used for analysis.

Linear Predictive Coding (LPC) analysis decomposes digitized speech signal into it's fundamental frequency (F0 and it's amplitude i.e. loudness of the source) and the vocal tract is represented by all pole filters, which can be modeled by a number of coefficients known as LPC order. The vocal tract system is excited by an impulse train for voiced speech or a random noise sequence for unvoiced speech. Thus, the parameters of this model are: voiced/unvoiced classification, pitch period for voiced speech, gain parameter G, and the coefficients $\{a_k\}$ of the digital filter. Eq. 1 expresses the transfer function of the filter model in z-domain, where V(z) is the vocal tract transfer function. G is the gain of the filter and $\{a_k\}$ is a set of autoregression coefficients called Linear Prediction Coefficients. The upper limit of summation, p, is the order of the all-pole filter.

$$V(z) = \frac{G}{1-\sum_{k=1}^{p} a_k z^{-k}} \quad (1)$$

Speech samples are analyzed by LPC method using LPC order 28. The selected speech samples are windowed using hamming window of 20ms at 10ms interval. Acoustic

parameters (vocal tract parameters) are extracted from a stable portion of segmented vowels and are documented.

IV. Results and Discussion

To study the cross gender variation of acoustic features with HP and assess HP within and across CP speakers, scatter plots and VSA are utilized. Fig. 2 shows the block diagram of the working procedure. Female spectrum is not as clear and concise for formant extraction for some CP speakers as compared to male spectrum. Therefore, it was comparatively easier to extract formants from male spectrum than female spectrum. Vowels are selected for this study as these are reported to form simplest relationship between vocal tract configuration and formant frequencies during it's production.

To answer the research questions stated in section I:

First, the values of the vocal tract parameters (first and second formants, F1 and F2) of the vowels are documented.

Second, a comparison on the variation of the measured vocal tract parameters of males and females is conducted.

Finally, VSA of CP speakers is compared to Non-CP normal VSA to assess HP within gender. HP detection variation among males and females are investigated to make a comparison.

A. Comparative Study of Acoustic Feature Variation

Vocal tract transfer function for nasal vowel and various degree of HP is plotted in Fig. 3. As the severity of hypernasality increases, changes are visible as compared to the nasal vowel spectrum. New spectral peaks are visible at near about F1 and F2 noticeably around 200Hz, 500Hz, 1 kHz, 1.5 kHz depending on the vowel. Each vowel is represented by their F1 and F2 values displayed on scatter plots as shown in Fig. 4. The scatter plot of F1 and F2 for vowels reflects the inter-speaker variation within vowel and across vowels making them useful for differentiation and identification of vowels. Fig. 4 plots the individual and mean formant values (F1 and F2, in Bark) in vowel space for the vowels /i a u/. Each point represents the mean of three formant measurements per speaker.

In the considered utterance context, various degrees of inter-speaker variability is measured in terms of standard deviation about the mean. Variability of acoustic features among CP speakers are calculated and observed to be different depending on the vowel within and across gender. For pronouncing normal /i/ articulator is characterized by semi-openness, and has the highest front position among the vowels. /i/ and /u/ has the lowest F1 among the vowels. Among vowels, /u/ has the highest back position. During the production of /u/, articulators are characterized by lip-rounding, closeness, backness.

The inter-speaker variability among CP subjects for /i/ is 0.96 with mean (4.9, 13.82) for male and 1.23 with mean (3.7, 14.2) for female. For /a/ the variability is 1.13 with mean (5.65, 10.46) for male and 0.71 with mean (7.2, 10.9) for female. For /u/ variability is 2.05 with mean (4.58, 9.64) for male and 3.06 with mean (3.6, 8.6) for female. Vowel /u/ shows the highest

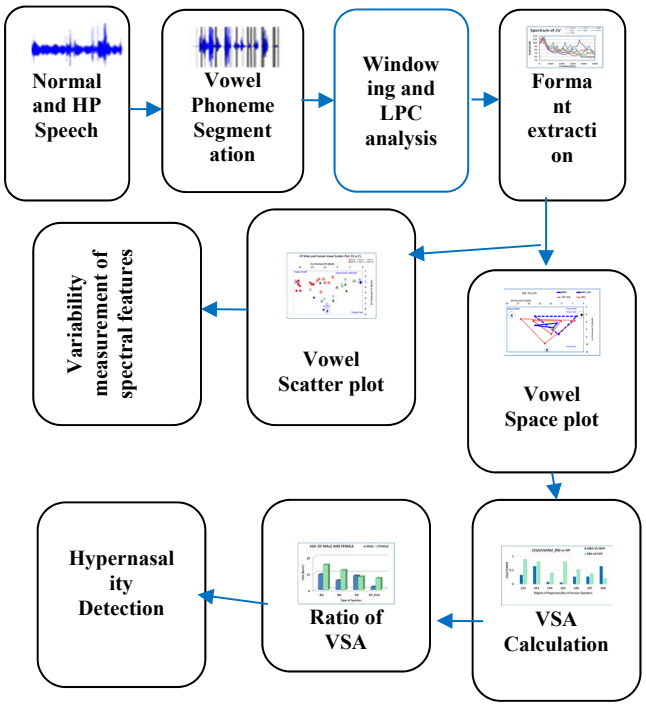


Fig. 2. Block diagram of working procedure for spectral variability and HP assessment.

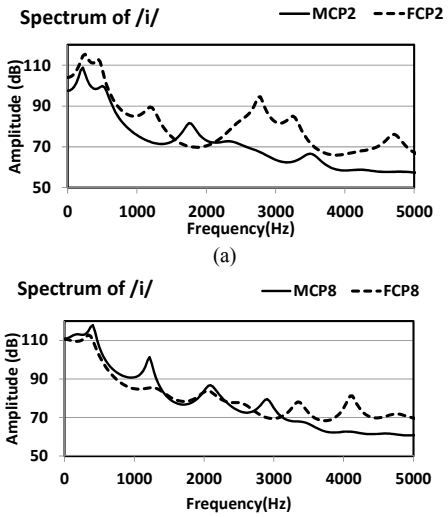


Fig. 3. Spectrum of /i/ (a) of English hypernasal CP2 of Male (MCP2) and Female (FCP2) (b) English hypernasal CP8 of Male (MCP8) and Female (FCP8).

variability within and across speakers in the concerned speech data reflecting the differences in articulatory openness for some speakers. From the inter-speaker variability measurement it is observed that high back vowel /u/ is mostly affected and has the highest variability. In males, vowel /i/ is least affected and has the lowest variability with HP. In females /a/ is least affected. The amount of inter-speaker variability of CP speakers in the high front vowels /i/ is less than open vowel /a/ for males.

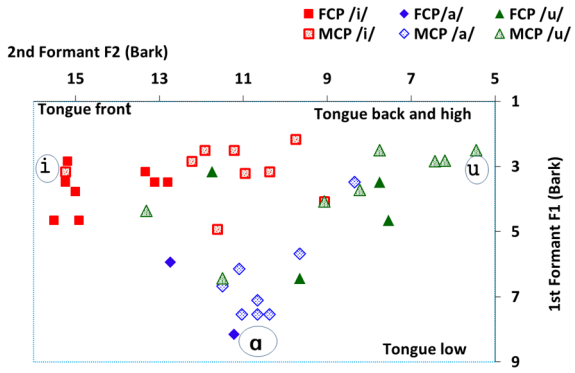


Fig. 4. Scatter plot of /i a u/ for CP Male (M) and Female (F) speakers.

B. *Variability of VSA and HP Detection*

In the second part, VSA is calculated for all 18 speakers. VSA obtained from Non-CP speakers is taken as reference. Ratio of VSA of hypernasal speech and normal speech is calculated and used as a measure for HP detection. Oral and CP hypernasal VSA and their graphical plot is shown in Fig. 5. Average value of VSA of CP male (MCP_AVG) and CP female (FCP_AVG) speaker is indicated by thick solid triangle and thin solid triangle. Fig. 5 shows average VSA of all speakers spanned by mean values for the three repetitions for each of the three vowels. The differences between isolated oral VSA, isolated nasal VSA, read speech average VSA of CP speakers is investigated.

Four types of VSA are obtained. VSA of isolated oral vowel has the highest area. Isolated nasal vowel VSA has the second highest area. Read speech English vowel of non-CP speakers has the third highest area. The lowest VSA is obtained for average of CP speakers read speech. The results show that for males

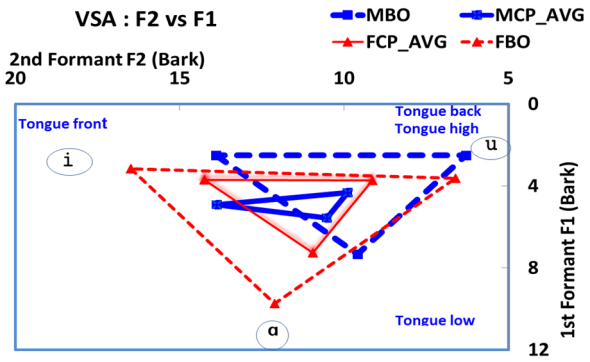
$$VSA_{isolated.oral} > VSA_{isolated.nasal} > VSA_{read.oral} > VSA_{HP}$$

For females,

$$VSA_{isolated.oral} > VSA_{isolated.nasal} > VSA_{read.oral} > VSA_{HP}$$

For both males and females, the acoustic vowel spaces show how the isolated VSA distanced from the nasal VSA, read speech VSA and the hypernasal speech VSA. As observed from Fig. 5, as the degree of HP increases, VSA changes. As the VP opening of CP speakers increases gradually VSA changes, but no gradual change is observed. This may be related to the individual vocal tract characteristics of each speaker. Ratio of vowel space (VSA_{ind}/VSA_{ref}) of individual's VSA (VSA_{ind}) and the reference VSA (VSA_{ref}) is calculated to characterize how large the individual's vowel space of CP speakers is to the reference (EO and BN) VSA.

Results plotted in Fig. 6 shows the vowel space ratio obtained across various VP opening conditions by taking EO and BN as the reference for both males and females. While taking EO as reference the maximum value of ratio came out to be 0.43 for male and 0.88 for females. While taking BN as calculated and used as a measure for HP detection. Oral and CP reference the maximum value of ratio came out to be 0.65 for



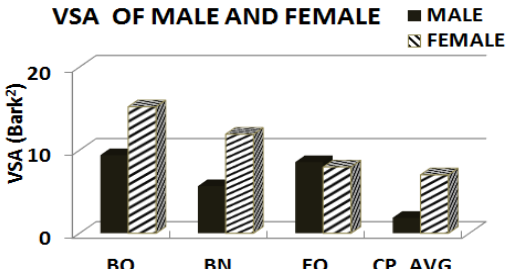


Fig. 5. Various VSA in F1xF2 plane for males and females (Top) and some graphical plot of VSA within and across gender (Bottom).

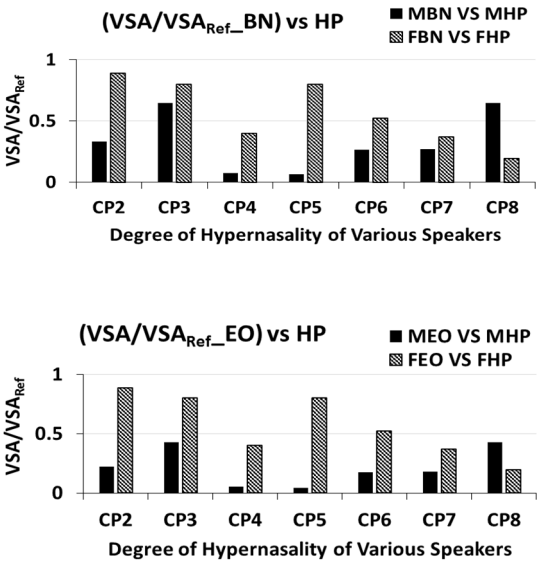


Fig. 6. Vowel space ratio across various VP opening conditions taking BN (Top) and EO (Bottom) as reference.

male and 0.74 for females. These measures may be used as threshold for determining HP. Female CP average VSA is observed to be 3.7 times more than average VSA obtained for male. Euclidean distance measure between male and female VSA in /i/ is 1.29, /a/ is 1.75 and /u/ is 0.75. This reduction in

VSA appropriately reflects the effect of HP of CP female speakers across various conditions of severity.

It is understood that due to the effect of HP speech clarity is reduced in both males and females. Therefore as observed, VSA can differentiate hypernasal speech from normal vowel articulation indicating it's relation to the sensitivity to interspeaker variability. VSA of connected read speech of CP speakers may be used for detecting HP of CP speakers across various conditions of severity.

V. Conclusion

This study reports the outcomes of the experimental observations obtained by comparing acoustic characteristics of males and females with increasing severity of HP and brings together the isolated Bangla oral and nasal vowels and continuous read speech English vowels of normal and CP speakers. The main objective of this work is to make a comparative study for the variation of acoustic features for CP male and female speakers due to various VP opening and detection of HP using VSA. For this purpose, VSA is estimated for each of the selected speakers utilizing extracted acoustic features. The evolution of VSA with the 7 degrees of hypernasal articulation is analyzed. This triangle consists of the first two formants of three vowels /i/, /a/ and /u/ represented in the vowel space. Interspeaker variability of HP among CP male and female speakers is measured by calculating mean and standard deviation of the selected vowels. /u/ shows the most variability in both males and females. Second conclusion is the significant reduction of the VSA in both CP male and female speakers as speech becomes disordered. VSA of females shows greater excursion than males.

Continuous read speech is articulated differently from isolated vowel. This study can be further be extended to make a comparison with sustained isolated hypernasal vowel data of CP speakers. The results of this study depended on a particular set of target vowel data of some continuous read speech. In future research, the data collection can be expanded to include different contexts, speech styles, include all vowels of a language.

References

[1] R. Rullo, D. Di Maggi o, V.M. Festa, and N. Mazzarella, "Speech assessment in cleft palate patients: a descriptive study," Int. J. of Pediatric otorhinolaryngology, 73(5):641-644, 2009.

[2] "Congenital malformations worldwide," International Clearing house for Birth Defects Monitoring Systems, Amsterdam, Holland. Tech. Rep., 1991.

[3] O. Fujimura and J. Lindqvist, "Sweep-tone measurements of the vocal tract characteristics," J Acoustical Soc. Am., vol. 49(2), pp. 541-58, 1971.

[4] G. Fant, Speech Sounds and Features (MIT Press, Cambridge, 1973).

[5] S. Hawkins, and K. N. Stevens, "Acoustic and perceptual correlates of the non-nasal-nasal distinction for vowels," J. Acoust. Soc. Amer., vol. 77, no. 4, pp. 1560–1574, Apr. 1985.

[6] J. R. Glass and V. W. Zue, "Detection of nasalized vowels in American English," in Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing, 1985, pp. 1569–1572.

[7] D.A. Cairns, J.H.L. Hansen and J.E. Riski, "Detection of hypernasal speech using a nonlinear operator," in Proc. of IEEE Conf. on Engineering in Medicine and Biology Society, pp. 253-4, 1994.

[8] D. K. Rah, Y. I. ko, C. Lee, and D. W. Kim, "A noninvasive estimation of hypernasality using a linear predictive model," Ann. Biomed. Eng., vol. 29, pp. 587–594, 2001.

[9] P. Vijayalakshmi and M. R. Reddy, "Analysis of hypernasality by synthesis," in Proc. of Int. Conf. Spoken Language Processing, Jeju island, South Korea, Oct. 2004, pp. 525–528.

[10] P. Vijayalakshmi, M. R. Reddy and Douglas O'Shaughnessy, "Acoustic analysis and detection of hypernasality using group delay function," IEEE Trans. Biomedical Engineering, vol. 54, no. 4, pp. 621 – 629, Apr. 2007.

[11] P. Vijayalakshmi and M. R. Reddy, "The analysis of band-limited hypernasal speech using group delay based formant extraction technique," in INTERSPEECH, Eurospeech, Lisbon, Portugal, Sep.2005, pp. 665 - 668.

[12] A. Bladon, "Two-formant models of vowel perception: Shortcomings and enhancement," Speech Commun. 2(4), 305–313 (1983).

[13] A. T. Neel, "Vowel space characteristics and vowel identification accuracy," J. Speech Lang. Hear. Res. 51(3), 574–585 (2008).

[14] S. Skodda, W. Gronheit, and U. Schlegel, "Impairment of vowel articulation as a possible marker of disease progression in Parkinson's disease," PloS ONE 7(2), e32132 (2012).

[15] L. B. Leonard, S. E. Weismer, C. A. Miller, D. J. Francis, J. B. Tomblin, and R. V. Kail, "Speed of processing, working memory, and language impairment in children," J. Speech Lang. Hear. Res. 50(2), 408 (2007).

[16] S. Sandoval, V. Berisha, R. L. Utianski, and J. M. Liss, "Automatic assessment of vowel space area," J. Acoust. Soc. Am. **134**, EL477 (2013).

[17] S. Scherer, L. Morency, J. Gratch, and J.P. Pestian, "Reduced vowel space is a robust indicator of psychological distress: a cross-corpus analysis,**"** in Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), pages 4789-4793, 2015.

[18] M. Eshghi, M. M. Alemi and M. Eshghi, "Vowel nasalization might affect the envelop of the vowel signal by reducing the magnitude of the rising and falling slope amplitude," J. Acoust. Soc. Am. **137**, 2304 (2015).

[19] L. Chen, Y. C. Lin, C.H. Katherine and D. K. Raymond,"Perceptual speech intelligibility and speech production variability in Mandarin-speaking children with cerebral palsy," J. Acoust. Soc. Am. **139**, 2045 (2016); http://dx.doi.org/10.1121/1.4950051).

[20] C.M. Vikram, A. Nagaraj, and S. R. M. Prasanna, "Spectral Enhancement of Cleft Lip and Palate Speech," in INTERSPEECH 2016, San Francisco, USA, September, 2016. DOI:10.21437.

[21] S. Haque, M. H. Ali and A.K.M. Fazlul Haque, "Variability of acoustic features of hypernasality and it's assessment," International Journal of Advanced Computer Science and Applications(IJACSA), 7(9), 2016. http://dx.doi.org/10.14569/IJACSA.2016.070928.

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

Image Based Drinks Identification for Dietary Assessment

Rubaiya Hafiz, Saiful Islam, Roksana Khanom and Mohammad Shorif Uddin

Department of Computer Science and Engineering
Jahangirnagar University, Savar, Bangladesh
Email: rubaiya1991@gmail.com, saiful.raju@gmail.com, roksana.popi@gmail.com, shorifuddin@gmail.com

Abstract— **In this paper, we propose a method to detect and recognize different types of drinks with a view to estimate its nutrient values using vision based algorithms. We use visual saliency and thresholding techniques to segment out the drink region from the image. After detection, both speed up robust features and color based visual features are used to create a Bag of Features (BoF). This BoF model is employed for recognition of drinks. Based on experimentation, our proposed method confirms that it is capable of detecting and recognizing different drinks categories with an accuracy greater than 89%.**

Keywords—Drinks classification; Color based BoF; Speed up robust features (SURF); Mean-shift segmentation.

I. INTRODUCTION

According to a survey report published in June 2016, 38% of U.S. grownups and 17% of youngsters are enduring from obesity [1]. Although the ratio is quite tremendous, the great news is that people all over the world are becoming more sensible at taking healthy foods and drinks because the quantity of calories they consume has an immediate influence on their health. Researchers found that proper food and diet patterns could guard upon diabetes, heart disease, stroke and other chronic complications. Therefore, a self-regulating system can show the nutritive quantity of different drinks that can help an individual to decide whether a particular drink is good for his/her health or not.

There are some food recognition methodologies proposed in these papers [2], [3], [4], and [5]. All these approaches focused on solid foods. Still manual assessment is required in the procedure for food recognition and finding out its nutrient value [2]. Therefore, it is a monotonous course. We use Google frequently to ascertain a particular drink composition, but it takes a lot of time.

Our methodology will provide an approach that would recognize drinks from a set of cluttered images and be able to detect and display their nutrition fact.

In contrast to existing methods as mentioned in paper [4] and [5] where only colour and shape attributes are focused, we introduced more diverse attributes that extents to colour, texture and saliency. Previously, Parisa Pouladzadeh et.al [3] proposed a method that worked with solid foods and used SVM technique on the classification steps. In our proposed system, we used Bag of Features (BoF) technique rather than SVM, since it is easy to execute, takes minor processing time, and gives approximately the identical accurateness.

The organization of this paper is as follows: section II gives the overall procedure of our proposed system that consists of two parts. The first step is pre-processing of segment drink region and then we move on to our learning and recognition step. In section III, we delineate about the data sets that we have employed and in section IV, the experimental results are presented. Eventually, our paper concludes in section V.

II. METHODOLOGY

It is very important for our method to detect the object region as accurately as possible because it classifies the segmented drinks based on its color and shape. That is why we need lots of pre-processing to determine the region that contains our object of interest. The scheme of our proposed method shown in Fig. 1.

All the steps to extract the object of interest from original image are given below.

A. Detection of Object of Interest

i. Salient area detection

Visual saliency enables us to focus on only the desired portion on an image with complex scene [6]. We employed graph based visual saliency detection method in our system. It is a bottom up technique consist of two stages. At first, it forms an activation map using features extracted over the image plane and then normalizes the activation map to generate a saliency map [7]. Then we create a binary mask and project it on the earliest image to discover the region that contains 70% saliency. Most of the background scenes are excluded from the given image using this map.

ii. Segmentation

Our saliency detection step can identify only the most salient part of the input image rather than the salient object itself. That is why we need a segmentation technique to partition the image into multiple regions. There are various image segmentation techniques available. K-means clustering

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

and mean-shift are two most used methods because of their simplicity and less computation time. In K-means technique, the number of clusters that will be used needs to be known. But as mean shift is a non-parametric algorithm, it does not assume anything about number of clusters. That is why we select mean shift segmentation for our method [8]. Mean Shift

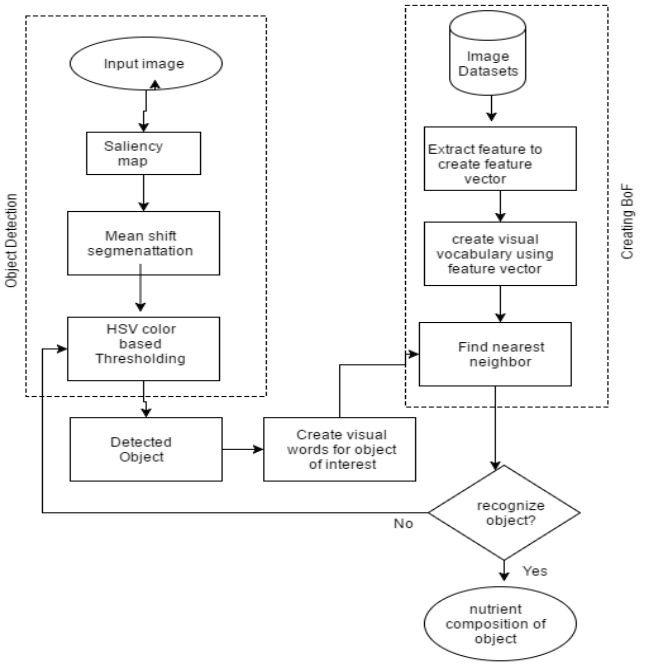


Fig. 1: Overall block diagram of our proposed method.

algorithm uses the Gaussian Kernel function presented in Eq. (1) to calculate the distance between the points.

$$K(x_i\text{-}x) = e^{(-c||x_i\text{-}x||^2)} \quad (1)$$

The algorithm will select each of the points, bound an area around them and calculate the position of the centroid in this density area using the function presented in Eq. (2).

$$m(x) = \frac{\sum_{x_i \in N(x)} K(x_i - x) x_i}{\sum_{x_i \in N(x)} K(x_i - x)} \quad (2)$$

This segmentation method partitions our image into several coherent regions based on RGB color in such a way that each region has a particular property.

iii. Color based thresholding

Color based thresholding method can be used to track an object of any color from an input image [9], [10]. Since our original image may contain a cluttered scene and have many different objects, after segmentation we may get an image that contains our object of interest as well as some other objects too. So, we use HSV color based thresholding to detect the region of drink object. HSV color model is a cylindrical geometric, with hue (H) as angular dimension, starting at the red primary at 0°, passing through the green primary at 120° and the blue primary at 240° and then wrapping back to red at 360° [11], [12].

HSV tries to capture the color components in the same way, as humans perceive color. As we have to work with lots of images, we have found that an identical object provides different RGB color information in several images due to differences in brightness, shadow etc. H is the true color and insensitive to noise and brightness change. For this reason, we chose HSV color based thresholding in our method.

The output image that we have found from our segmentation step is used as an input for the thresholding step. Once we completed thresholding, we have an image that contained only drinks of a particular color. From this image, we create a binary mask and project it on the original image to segment out the partial image containing the drinks. This is the overall process of object detection and extraction from a raw input image.

The combination of visual saliency, mean shift segmentation and thresholding process enables our system to segment out the drink from images.

B. Learning and Recognition

All the steps we have delineate above together help us to extract a drink region from a cluttered image. This section demonstrates how we use computer vision system to learn this. Bag of Features (BoF) methods is a well-established computer vision approach, which has been applied to image classification, object detection, image retrieval, and even visual localization [13]. We have selected both color and SURF based BoF-based systems because it gives better results than other approaches for image classification and image retrieval, while being computationally cheaper and conceptually simpler.

We use SURF (Speed up robust features) descriptor [13] in our method because it consists of both feature detection and representation aspect and it works faster than other methods like SIFT (Scale Invariant Feature Transform) descriptor [14]. A random subset of images from the dataset is chosen for training our classifier.

In SURF technique, at first we perform the detection of interest point (points that are likely to be found in different images of same object again) for an image. The output of this process is a set of key points that specify locations in the image with corresponding scales and orientation [13]. We employ Hessian matrix based blob detector, which is a very popular point detection approach. Then the right orientation of that point is found. A square region is extracted to describe the region around that point. This is known as feature descriptor and is centred on the point of interest.

SURF descriptor itself cannot be able to classify accurately between similar shaped objects (for example, gray-value versions of Pepsi and Coca cola cans). For this reason, we use color layout feature descriptor to select local RGB color features from each images.

We extract SURF descriptors and color features from each of the training images to train our classifier. We keep 80 percent of the strongest features through quantization from this image set. All image features are then plotted on a high dimensional space and clustered using k-means algorithm. The set of centroid coordinates of all clusters are used to create our vocabulary.

By means of Euclidean distance metric, we then compute the nearest neighbour of each of the feature vector and build a histogram that counts how many times each centroid occurs in each image. Finally, we train our classifier using the error-correcting output codes (ECOC) framework with SVM.

The output image, we have found on our object detection step, contains only the detected drink and it is used as a test image for this step. Again both SURF and color features are extracted from this image, build a histogram of visual words, normalize the histogram and predict the output using the trained model.

Once we recognize the drinks family, we can effortlessly extract its composition from our nutrient fact table.

III. Experimental Data

Our produced data set (prepared by us and it will be available on request) consists of six distant categories of drinks and beverages. We used total 389 images to create our experimental data. It consists of 80 Green coconut images, 60 Pepsi can/bottle images, 68 Coca cola can/bottle images, 55 Fanta can/bottle images, 74 Coffee cup images and 52 Mineral water bottle images. We break down all images into two sets: training set and validation set. The training set consists of 195 images and validation set consists of 194 images. We have trained our classifier engine using both color and SURF features. Some sample images from our training dataset and validation dataset are shown in Fig. 2 and Fig. 3, respectively.

IV. Experimental Result

This section describes the experimental result of our proposed method. We have segment out different types of drinks using experimental images. After segmentation, the system recognizes the drinks through matching with the training images. Finally, the nutrient information is shown from the nutrition table.

Fig. 2. Some sample images from training dataset

Fig. 3. Some sample images from validation dataset

Fig. 4 shows the confusion matrix that evaluates the recognition validation. The confusion matrix shows the experimental result containing true negative (TN) rate, false positive (FP) rate, false negative (FN) rate, and true positive (TP) rate for each type of drink.

Eq. (3) shows the formula to calculate the accuracy from this matrix.

$$Accuracy = \frac{(TN + TP)}{TP + TN + FP + FN} \qquad (3)$$

Some true detection sample outputs are shown in Fig. 5. Fig. 6 shows an example of false negative (FN) and false positive (FP) prediction of our system. This image contains a Mineral water bottle. But our system incorrectly classifies this as Pepsi. This error may not be happened if the full bottle is visualized.

(a) Confusion Matrix for Coca cola images

		Estimated class		Accuracy
Actual class		***Other drinks***	***Coca cola***	93.29%
	Other drinks	151 (TN)	9 (FP)	
	Coca cola	4 (FN)	30 (TP)	

(b) Confusion Matrix for Green coconut images

		Estimated class		Accuracy
Actual class		***Other drinks***	***Green coconut***	97.9%
	Other drinks	150 (TN)	4 (FP)	
	Green coconut	0 (FN)	40 (TP)	

(c) Confusion Matrix for Coffee images

		Estimated class		Accuracy
Actual class		***Other Drinks***	***Coffee***	95.3%
	Other Drinks	151 (TN)	6 (FP)	
	Coffee	3 (FN)	34 (TP)	

(d) Confusion Matrix for Pepsi images

		Estimated class		Accuracy
Actual class		***Other drinks***	***Pepsi***	91.2%
	Other drinks	157 (TN)	7 (FP)	
	Pepsi	10 (FN)	20 (TP)	

(e) Confusion Matrix for Mineral water images

		Estimated class		Accuracy
Actual class		***Other drinks***	***Mineral water***	89.6%
	Other drinks	150 (TN)	17 (FP)	
	Mineral water	2 (FN)	24 (TP)	

(f) Confusion Matrix for Fanta images

		Estimated class		Accuracy
Actual class		***Other drinks***	***Fanta***	93.8%
	Other drinks	159 (TN)	8 (FP)	
	Fanta	4 (FN)	23 (TP)	

Fig. 4: Confusion matrix for our dataset.

V. CONCLUSION

In this paper, we have proposed a method to segment out a drink from its image with an objective of estimating its amount of nutrition such as calorie, energy, fat, sugar etc. The combination of visual saliency, mean shift segmentation and thresholding process helps us to segment out our object of interest. From a set of training images, we extract both RGB positive rate is not high. This method is very useful for patients, especially for those, who have a major effect of food calories on their health.

In future, we will stress on robustness of the system in order to improve its recognition efficiency. In addition, this work will be an obvious opportunity for future work to cover more drinks and foods.

color features and SURF features to create a BoF. Finally, this BoF model is used to classify the drink category. From our experimentation, we have found that our system has an accuracy above 89%. The accuracy is good and the false

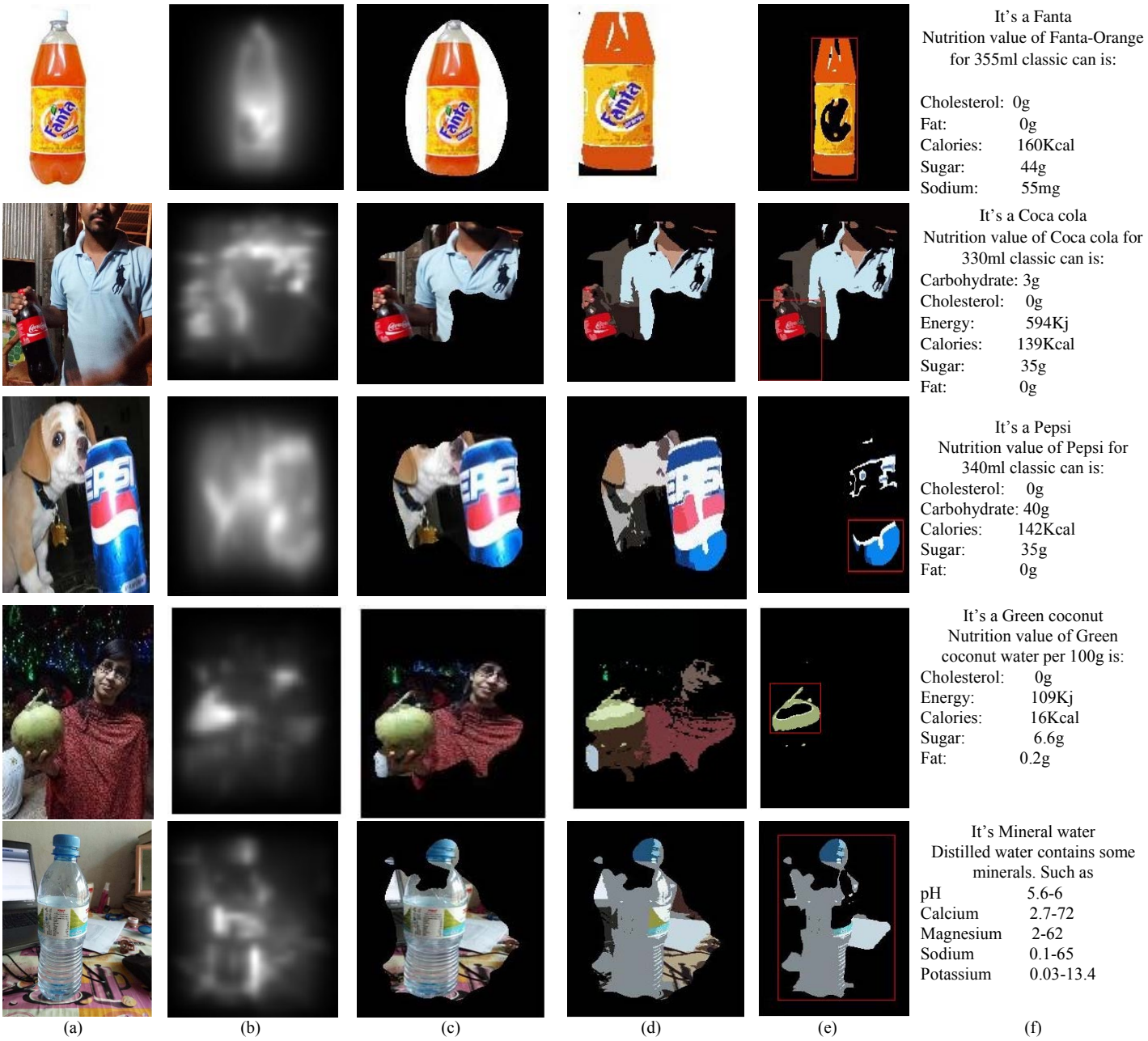


Fig. 5: (a) Input image, (b) computed saliency map, (c) partial image with 70% saliency, (d) image after segmentation, (e) detected object (f) final nutrition value after recognition.

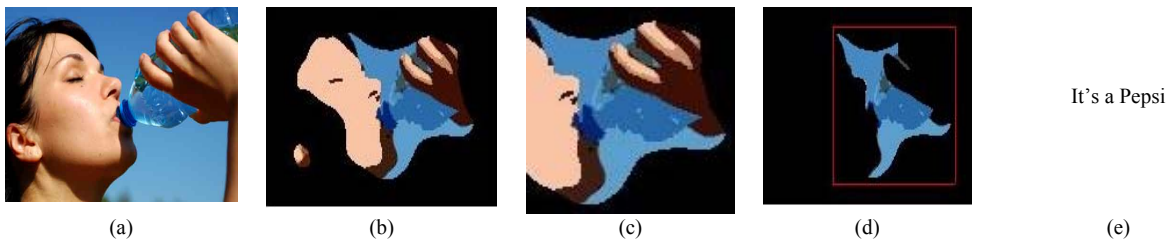


Fig. 6: False negative (FN) prediction for the Mineral water (Fig. 4 (e)) and false positive (FP) prediction for Pepsi (Fig. 4(d)) confusion matrix. (a) Input image, (b) partial image with 70% saliency, (c) image after segmentation, (d) detected object (e) output after classification.

REFERENCES

[1] Maggie Fox. "American obesity epedemic hits a new high," Available online, http://www.cnbc.com/2016/06/07/americas-obesity-epidemic-hits-a-new-high.html, last visited on Jun 7, 2016.

[2] Parisa Pouladzadeh , Shervin Shirmohammadi, and Rana Almaghrabi. "Measuring calorie and nutrition from food image," IEEE Transactions on Instrumenntation and Measurement, 63(8):1847-1956, August 2014.

[3] Parisa Pouladzadeh, Gregorio Villalobos, Rana Almaghrabi, Shervin Shirmohammadi, "A novel SVM based food recognition method for alorie measurement applications," Proc. IEEE International Conference on Multimedia and Expo Workshops (ICMEW), pp. 495-498, 2012.

[4] D.-W. Sun, T. Brosnan, "Pizza quality evolution using computer vision-Part 1 Pizza base and sause spread," *Journal of Food Engineering*, Vol. 57, No. 3, pp. 81- 89, 2003.

[5] Ferat Sahin, "A Radial Basis Function Approach to a Color Image Classification Problem in a Real Time Industrial Application," Master's thesis, Virginia polytechnic institute, Blacksburg, 1997.

[6] Xin Li, Fan Yang, Leiting Chen, and Hongbin Cai, "Saliency Transfer: An Example-Based Method for Salient Object Detection," Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16).

[7] Harel, Christof Koch , Pietro Perona, "Graph-Based Visual Saliency Jonathan," Proceedings of the Twentieth Annual Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, December 4-7, 2006.

[8] Paulo Muniz de Avila, Roan Simoes da Silva, Luiz Angelo Valota Francisco, Rodrigo Palucci Pantoni, David Buzatto and Sergio Donizetti Zorzo, "Comparing K-Means and Mean Shift Algorithms Performance Using Mahout in a Private Cloud Environment," *Journal of Communication and Computer 11* (2014) 45-51.

[9] Jayati Singh , Shubham. "Object Detection by Color Threshold Method," International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 10, October 2015.

[10] Hani Hunud A. Kadouf and Yasir Mohd Mustafah. "Colour-based Object Detection and Tracking for Autonomous Quadrotor UAV," IOP Conf. Series: Materials Science and Engineering **53** (2013) 012086 doi:10.1088/1757-899X/53/1/012086

[11] Shamik Sural, Gang Qian and Sakti Pramanik, "Segmentation and Histogram Generation Using the HSV Color Space for Image Retrieval", *IEEE ICIP 2002*, pp. II-589 to II-592.

[12] Javier Poveda Figueroa1 , Vladímir Robles Bykbaev, "Image retrieval based on the combination of RGB and HSV's histograms and Colour Layout Descriptor,*" Fourth call for Research Founds of the Universidad Politecnica,* Salesiana, Ecquador pp. 3-10.

[13] Stephen O'Hara And Bruce A. Draper, "Introduction to the Bag of Features paradism for image classification and retrieval", Computing Research Repository, vol. arXiv:1101.3354v1, 2011.

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

Feature-Based Image Stitching Algorithms

Moushumi Zaman Bonny and Mohammad Shorif Uddin
Dept. of Computer Science and Engineering, Jahangirnagar University, Dhaka, Bangladesh
Email: zaman.moushumi@gmail.com

***Abstract*—The method of joining images to make a panorama is known as image stitching. It is an enthusiastic research area in image processing and computer vision but still a challenging problem for panoramic images. A good number of researches had been carried out to develop different algorithms for image stitching in the last few years. Image stitching approaches is classified mainly in two groups: direct and feature based. Direct techniques evaluate pixel intensities of the input images and feature- based methods resolve an association among the images based on the extracted features of inputted images. A detail study on the state-of-the-art of feature-based image stitching approaches is presented in this paper. We have shown the performance of some of the feature-based image stitching approaches using images from Yale Database. In addition, we briefly discussed the challenges and possible future work of image stitching.**

Keywords—*Panorama, Image stitching, Image features, Corner and edge detection.*

I. Introduction

The procedure of unification of several images to generate a single panorama image is called image stitching. It finds general applications in representing the wide objects either in natural scene or in microscopy. In panoramic stitching, an image set will have a logical amount of overlap to prevail over the deformation of lens and must contain sufficient measurable features. This set will also have reliable disclosure between two frames to curtail the chance of seam occurrence. In addition, the images of same scene can have different intensities, scale and orientation.

If the images are taken from the different places, then it may associate with parallax errors. Blind stitching can be done using feature-based alignment methods. But it can produce erroneous result. To overcome the parallax errors, we can use large camera system in a fixed disclosure.

Image stitching is a broadly used technique in various applications such as, image stabilization features, panoramas in maps and images of satellite with high resolution, images for medical solutions, high resolution multiple images, video stitching and object insertion.

This stitching method is mainly separated into three core sequences: image calibration, image registration and alignment, and image blending [1]. The aim of calibration is to reduce the differences between camera-lens combination and an ideal lens model. These differences are calculated from optical defects [2]. The image could be reconstructed to a 3D scene using the coordinates of the pixels and the intrinsic and extrinsic parameters of camera.

The extrinsic camera parameter and the linkage among the pixel coordinates of an image point with the corresponding coordinates, is made by intrinsic camera parameters. These topics are used to describe the position and direction of the camera reference frame. This frame must be compared with a world reference frame.

The calibration step is followed by image registration and alignment [3], [4]. Alignment is useful in matching an image when it transforms to another coordinate system. Here, the transformations of an image normally go through translation, rotation and scaling. Finally image blending technique is applied to ensure seamless transition from one image to another image by removing artifacts [4].

The most likely used alpha blending receives weighted average of two images and works really well during the time of alignment of image pixels through intensity shift. After that the images are united by the Gaussian pyramid blending at different frequency bands and filtered consequently.

This research highlights the state-of-the-art of feature-based image stitching approaches and its additional objective is to discuss the challenges of image stitching.

We have organized the rest of the paper as follows: different image stitching techniques are described in Section II. The performance evaluation of feature-based techniques is analyzed in section III. Section IV demonstrates the discussion of different feature-based techniques. Then in section V, the future works and challenges are presented. Last of all, Section VI draws the conclusion of this paper.

II. Image Stitching Techniques

Image stitching technique basically divided into two basic types: direct and feature-based.

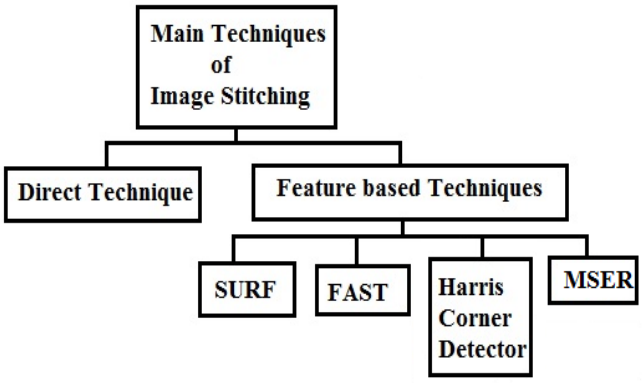


Fig. 1. Different types of image stitching techniques

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

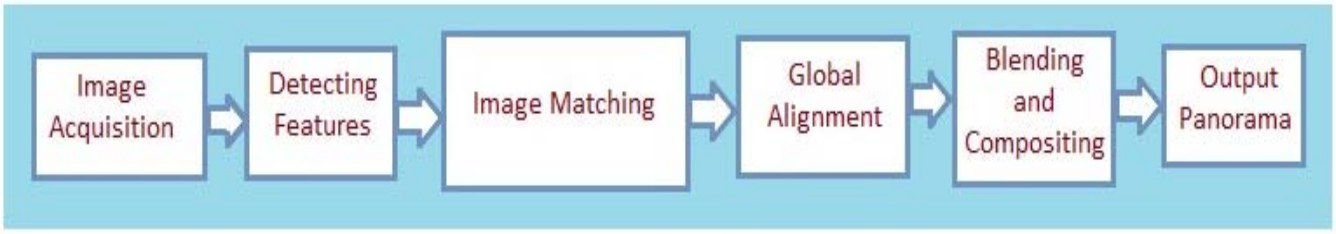


Fig. 2. Basic algorithm of feature based image stitching

A. Direct technique

The direct technique is dependent on the comparison of pixel intensities of the images and it decreases total differences among overlapping pixels [1]. The information from all the pixels is applied for the estimation of homography which is useful for the minimization of a particular cost function. Occasionally, Phase-Correlation is exploited to calculate the some homography parameters. Comparing each pixel with one another made the method as a complex one. It often uses any other available cost functions. But, in image scaling and rotation, these methods seem complex and invariant [5], [6].

There are some advantages of using direct method including optimal use of the information present in the image. This technique can evaluate the input of each pixel in the image but having an inadequate range of convergence is the biggest drawback of direct techniques [1].

B. Feature-based technique

Feature-based methods are more robust and quicker. It can recognize panoramas by automatic discovery of associations among the unsorted images [1]. To locate corresponding feature points between two images, it is necessary to compare every feature of different images of the local descriptors.

The feature-based algorithm consists of the following steps:

i. Image Acquisition

The first step of image stitching algorithm is the image acquisition which means image capturing by camera devices or acquiring from secondary sources.

ii. Feature detection and matching

For this purpose the following information are important: registration of image, stabilization of videos and reconstruction of 3D. These are necessary for matching the corresponding features between several views. Corners are matched to give quantitative measurement. It provides a rational matching for the image pairs based on rotation, translation and scaling [10], [11].

iii. Alignment

The most suitable method is photogrammetric method to merge many images of the identical scene in a perfect reconstruction of 3D scene and it is known as bundle adjustment [12], [13]. The purpose of alignment is to locate a reliable alignment parameter set which can decrease the miss-registration between every pair of images. It is very useful to widen the pair wise matching criterion to a global energy function [14].

iv. Blending and composition

Generally, when a few images are stitched together, one image is selected as the reference and then twisted all the images in the system (Coordinate system) of that location. The result is often known as a flat panorama [15], [16].

Image stitching can be applied on several projective layouts, for example, rectilinear projection, where the panorama is observed lying on a plane (two-dimensional) crossing a panosphere at a point [17], [18], [19]. Rectilinear projection uses cube faces with cubic mapping for viewing panorama and it demonstrates the cylindrical projection.

After plotting the source pixels against the ultimate composite surface, it is necessary to blend them for creating a panorama. Different blending methods are applied in image stitching. Feathering, image pyramid and gradient domains are some familiar blending procedures [19], [20].

Feathering is used for blurring the edges of features. From the two overlapping images, the average pixel values are evaluated for the blended regions [21]. Multi-band image blending is another well-known method that executes in the gradient domain.

Image pyramid blending is another important approach which represents the image using different frequency-band image set. Image pyramid gives numerous functional possessions for numerous purposes [22], [23], [24].

Below, we have represented some of the noble feature based techniques of image stitching.

i. SURF based approach

SURF (Speeded Up Robust Features) is known as local feature detector and descriptor [26], [27]. SURF descriptors are useful to find out and identify objects, faces, to restructure 3D scenes, to trail objects and to extort points of interest [28].

Juan et al [29] proposed an image stitching technique that combines the following algorithms: matching algorithm of image, modified SURF (Speeded Up Robust Features) technique, blending algorithm and multi-band blending. Using modified SURF, this technique gets feature descriptor of the image, then, finds matching pairs, verifies the neighbors using K-NN (K-nearest neighbor) and confiscates the mismatch couples using RANSAC and at last, regulates images using bundle adjustment and calculates approximately the accurate homography matrix and then blend images by multi-band blending.

Besides the above approach, Karthik et al [30] introduced a panoramic view using Invariant Moments and SURF Features. Wang et al [31] presented a unique method of image stitching using adaptive uniform distribution SURF.

ii. FAST feature-based approach

FAST (Features from Accelerated Segment Test) is a method of corner detection which extracts feature points and follows and plots objects in images [32].

Rosten Drummond introduced the FAST (Features from Accelerated Segment Test) approach [32], [33]. It identifies the interest points in images. The main reason was to work with FAST algorithm is to extend the significance point detector for the utilization of it in the real time applications of frame rate. The FAST detector evaluates the pixels around a point which is categorized as a corner depending on brightness or darkness. The FAST detector identifies too many features of neighboring which is counted as the major drawback of it. This FAST feature-based image stitching technique works with the FAST features. The system consists of following stages: detection, description and matching of features using FAST, image matching and then, application of Alpha blending method.

iii. Harris corner detector based approach

Harris corner detector used for this work because it is faster to detect the corners of the image and gives exacted corners of the images with different intensity or orientation. Extracting corner feature and using normalized cross-correlation of the local intensity values in particular points of matching them are the two main purpose of it [34].

Abdelfatah et al [35] presented an automatic image stitching process which is implemented using image sequences. It can determine a model that can convert points of one image to the other by a homography matrix because it contains correlated feature points between the two images. It gives the chances to overlap two images depending on the position of correlated feature points.

A window is positioned on an image. Then, the window is located onto a corner and it is stirred in different way which produces a huge alteration into the intensity. This mechanism is used by Harris corner detector. It is used for a probabilistic algorithm RANSAC that returns satisfactory results.

The RANSAC method is robust to estimate the homography and it is very effective even in the presence of lots of false matches. A linear gradient alpha blending can be used for blending for faster result. It removes the seams and discontinuities of the composite image and gives accurate outcomes.

iv. MSER feature based approach

MSER (Maximally Stable Extremal Regions) is a method for detecting blobs in images which was presented by Matas et al [36].

The intensity function of the region and the outer border defines the regions exclusively which directs to several key features of the regions which compile them functional. The local binarization is constant over a big variety of thresholds and the properties of it are as follows: invariance, covariance, stability and multi-scale detection.

This detector can find correspondences between image elements depending on different viewpoints. That's why, this method shows better performance in extracting a broad number of corresponding among image elements which is very useful for image matching. We have simulated this method as the feature detector for the feature based image stitching.

III. Performance Evaluation Of Feature Based Approaches

For performance evaluation of different feature-based image stitching techniques, we have done a simulation. For this purpose, we have collected a panorama image of size 1500×397 pixels from Yale Database [38].

We have made five input images (640×480 pixels each) from this original image using Microsoft Office Picture Manager. These images have some overlapping regions as well as repetitions. Then, we have inputted these images to generate the stitched image using four feature-based image stitching techniques [39]. After that we have compared theses stitched image with the original image.

In the table 1, we have demonstrated the accuracy rate of different feature-based image stitching techniques.

TABLE I. Performance Of Different Image Stitching Techniques Based On Accuracy Rate

Image stitching techniques	Extraction Detector	Accuracy Rate (%)
SURF based Approach	SURF	95.99%
FAST based approach	FAST	50.74%
Harris corner detector based approach	Harris corner detector	94.68%
MSER based Approach	MSER	90.53%

The accuracy of the SURF based approach is 95.99%, FAST based approach is 50.74%, Harris corner detector based approach is 94.68% and MSER based approach is 90.53%.

So, the results confirmed the superiority of SURF based method. The Harris corner detector and MSER have given comparatively better results after SURF. But, FAST has shown the poorest output among all these approaches. Besides this, in Figure 4, we have presented all the output stitched images of these feature-based approaches.

IV. Discussions

A. Critical comments on SURF based Approach

The SURF based approach is not good enough at the illumination changes and viewpoint changes in images. So, development is needed for SURF during this kind of changes.

B. Critical comments on FAST based Approach

The main drawback of FAST detector is, it detects too many features. But, it is very important to concentrate about the main features of the images.

C. Critical comments on Harris Corner Detector based Approach

Harris corner detector can detect the features from noisy images. But, noise can corrupt a natural image during the time of capturing and transmitting. As a result, it becomes difficult to get the desired output. So, before detecting corners, de-noising should be done to get better results.

D. Critical comments on MSER based Approach

MSER has shown the highest sensitiveness to the changes due to blurring. It is needed to overcome this sensitiveness of MSER.

So, in the future researcher should concentrate on illumination and viewpoint changes, noise removal and changes due to blurring. We have applied the feature-based techniques for 5 panorama images but we think this number of images is not good enough. In the future, we will apply these methods on more images.

Fig. 3. Original image from the Yale Database

(a)

(b)

(c)

(d)

(e)

Fig. 4. Five input images which are generated from the original image of Fig. 3.

(a)

(b)

(c)

(d)

Fig. 5. Output stitched images determined by: (a) SURF based approach (b)FAST based approach (c) Harris corner detector based approach (d) MSER based approach

V. Future Works And Challenges Of Image Stitching

After analyzing several approaches, we have come to know that, in the future, the image stitching can be improved and developed by taking hybrid technique which will use different approaches for better result [35]. Besides this, we can focus on the improvement of accuracy. On the other hand, there are many challenges in image stitching. Images are frequently changed by noise during the transmission and acquisition. So, filtering can be used for noise removal. Another problem is the elimination of the seams. For the removal of seam different methods have been proposed in the last two decades and we have a vision to work with seam removal.

VI. Conclusion

In this paper, we have briefly discussed about the image stitching technique and its three key steps calibration, registration and blending. Then, two main image stitching techniques namely, direct techniques and feature-based techniques are analyzed. We have also discussed the basic image stitching algorithm which consists of acquisition of images, detection of features, image matching, and global alignment and blending and composition. After that, we have discussed about different feature-based image stitching techniques: SURF based approach, FAST based approach, Harris corner detector based approach and MSER based approach. Then, we have done simulation and compared those techniques based on the performance and accuracy rate. The simulation results confirmed that SURF based method is better than any other methods. We have also conversed about the drawbacks of these techniques and discussed about the future works and challenges of our research.

References

[1] Ebtsam Adel, Mohammed Elmogy, HazemElbakry, "Image Stitching based on Feature Extraction Techniques: A Survey," *International Journal of Computer Applications*, vol.99, no.6, pp.1-8, August 2014.

[2] Zhengyou Zhang, "A Flexible New Technique for Camera Calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1330-1334, Nov 2000.

[3] Szeliski, "Image Alignment and Stitching," Tech. rep., December 10, 2006.

[4] S. Pravenaa, Dr. R. Menaka, "A Methodical Review on Image Stitching and Video Stitching Techniques," *International Journal of Applied Engineering Research*, vol.11, pp. 3442-3448, 2016.

[5] Preeti Mandle, Bharat Pahadiya, "An Advanced Technique of Image Matching Using SIFT and SURF," *International Journal of Advanced Research in Computer and Communication Engineering*, Vol. 5, Issue 5, May 2016.

[6] Pranoti Kale, K.R.Singh, "A Technical Analysis of Image Stitching Algorithm," International Journal of Computer Science and Information Technologies, vol. 6, pp. 284-288, 2015.

[7] Deng.Y. andZhang.T., "Generating Panorama Photos", *International Society for Optics and Photonics*, pp. 270-279, 2003.

[8] Brown M. and Lowe, "Recognising Panoramas," Ninth IEEE International Conference on Computer Vision, vol. 2, pp. 1218-1225,

2003.

[9] Hyung Il Koo and Nam Ik Cho, "Feature-based Image Registration Algorithm for Image Stitching Applications on Mobile Devices," *IEEE Transactions on Consumer Electronics*, vol. 57, no. 3, pp. 1303-1310, Aug 2011.

[10] Xianyong Fang, "Feature Based Stitching of a Clear/Blurred Image Pair," *IEEE International Conference on Multimedia and Signal Processing*, vol. 1, pp. 146-150, 2011.

[11] Mrs.Hetal M. Patel, "Comprehensive Study and Review of Image Mosaicing Methods," International Journal of Engineering Research & Technology (IJERT*)*, vol. 1, pp.1-9, nov 2013.

[12] K.Shashank, N.SivaChaitanya, G.Manikanta,Ch.N.V.Balaji, V. V. S.Murthy, "A Survey and Review Over Image Alignment and Stitching Methods," International Journal of Electronics & Communication technology, vol.5, pp.50-52, march 2014.

[13] Oh-Seol Kwon and Yeong-Ho Ha, "Panoramic Video using Scale-Invariant Feature Transform with Embedded Color-Invariant Values," IEEE Transactions on Consumer Electronics, vol. 56, No. 2, May 2010.

[14] A. Levin, A. Zomet, S. Peleg, Y. Weiss, "Seamless image stitching in the gradient domain," In Eighth European Conference on Computer Vision (ECCV 2004), (Prague), vol. IV, pp. 377-389, Springer-Verlag, May 2004.

[15] J. Y. Jia, C. K. Tang, "Image Stitching Using Structure Deformation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, No. 4, pp. 617-631, 2008.

[16] A.Zomet, A.Levin, S.Peleg, Y.Weiss, "Seamless Image Stitching by minimizing false edges, " IEEE Transactions on Image Processing, vol. 15, No. 4, pp. 969-977, April 2006.

[17] Feature Based Panoramic Image Stitching, Retrieved from http://www.mathworks.com/help/vision/examples/feature-based-panoramic-image-stitching.html, last visited: 22.10.2016

[18] Mclauchlan, P. F., Jaenicke, A., Xh, G. G., "Image mosaicing using sequential bundle adjustment," Proc. BMVC, pp. 751-759, 2000.

[19] Chao Tao, Hanqiu Sun, Changcai Yang, Jinwen Tian, "Efficient Image Stitching in the Presence of Dynamic Objects and Structure Misalignment," *Journal of Signal and Information Processing*, vol. 2, pp. 205-210 , 2011.

[20] Rosten. E. and Drummond.T., "Machine Learning for High-speed Corner Detection," *European Conference on Computer Vision*, vol.1, pp. 430-443, 2006.

[21] Ashley Eden, Matthew Uyttendaele, Richard Szeliski, "Seamless Image Stitching of Scenes with Large Motions and Exposure Differences," Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 2498-2505, June 17-22, 2006.

[22] Deepak Kumar Jain, Gaurav Sexena, Vineet Kumar Singh, "Image mosaicing using corner technique," *IEEE International Conference on Communication Systems and Network Technologies*, pp. 79-84, may 2012.

[23] Tao-Cheng Chang, Cheng-An Chien, Jia-Hou Chang, and Jiun-In Guo, "A Low-Complexity Image Stitching Algorithm Suitable for Embedded Systems," *IEEE International Conference on Consumer Electronics (ICCE),* pp. 197-198, jan 2011.

[24] Patrik Nyman,"Image Stitching using Watersheds and Graph Cuts," Research Article, Lund University, Sweden.

[25] Panoramic Mosaic Stitching, http://pages.cs.wisc.edu/~lizhang/courses/cs766-2008f/projects/mosaic/students/bfield/report.html, Last visited Date: 22.10.2016.

[26] Mathew Brown, David G. Lowe, "Automatic Panoramic Image Stitching using Invariant Features," *International Journal of Computer Vision*, vol. 74, pp.59 – 73, Aug 2007.

[27] Ji-Hun Mun, G. Oubong, "Panorama Image Stitching based on 'SIFT' Feature Points," *International Conference on Embedded Systems and Intelligent Technology*, pp. 55-56.

[28] Zhen Hua, Yewei Li, Jinjiang Li, "Image Stitch Algorithm Based on SIFT and MVSC," *IEEE 7th International Conference on Fuzzy Systems and Knowledge Discovery,* vol. 6, pp. 2628-2632, Aug 2010.

[29] L. Juan, G. Oubang, "SURF applied in panorama image stitching," *Image Processing Theory Tools and Applications (IPTA)*, 2010 2nd International Conference, pp. 495 – 499, Jul 2010.

[30] R.Karthik, A.AnnisFathima,V.Vaidehi, "Panoramic View Creation using Invariant Moments and SURF Features," *IEEE International Conference on Recent Trends in Information Technology (ICRTIT)*, pp. 376-382, July 2013.

[31] Ze-lang Wang, Fang-hua Yan, Ya-yuZheng, "An Adaptive Uniform Distribution SURF for Image Stitching," *IEEE 6th International conference on Image and Signal Processing (CISP)*, vol. 2, pp. 888-892, 2013.

[32] Rosten, E., & Drummond, T., "Machine Learning for High-speed Corner Detection," 9th European Conference on Computer Vision, Vol. Part I, pp. 430-443, Berlin, Heidelberg: Springer-Verlag, 2006.

[33] Ebtsam Adel, Mohammed Elmogy, Hazem Elbakry, "Image Stitching System Based on ORB Feature- Based Technique and Compensation Blending," *International Journal of Advanced Computer Science and Applications*, vol. 6, No. 9, 2015.

[34] A.V.Kulkarni, J.S.Jagtag, V.K.Harpale, "Object recognition with ORB and its Implementation on FPGA," *International Journal of Advanced Computer Research,* pp. 164-169, 2013.

[35] Russol Abdel fatah, Dr. Haitham Omer, "Automatic Seamless of Image Stitching," *Journal of Computer Engineering and Intelligent systems*, vol. 4, No. 11, 2013.

[36] J. Matas, O. Chum, M. Urban, T. Pajdla,"Robust wide baseline stereo from maximally stable extremal regions," *Proc. of British Machine Vision Conference*, pp. 384-396, 2002.

[37] W. Zhao, S. Gong, C. Liu, and X. Shen, "A self-adaptive Harris corner detection algorithm," *Computer Engineering*, vol. 34, no. 10, pp. 212–214, 2008.

[38] Panorama of University from Water Tower Looking North, http://www.findit.library.yale.edu/?page=3&q=Panorama+image&search_field=all_fields&utf8=?, Last visited: 31.10.2016.

[39] Feature Based Panoramic Image Stitching, https://www.mathworks.com/examples/matlab-computer vision/mw/visionproductFeatureBasedPanoramicImageStitchingExample-feature-basedpanoramic-image-stitching, last visited: 22.10.2016.

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

A Framework For Dynamic Vehicle Pooling and Ride-Sharing System

Nusrat Jahan Farin*†, Md. Nur Ahsan Ali Rimon†, Sifat Momen†, Mohammad Shorif Uddin* and Nafees Mansoor†
*Computer Science and Engineering Department
Jahangirnagar University, Savar, Dhaka
*†Computer Science and Engineering Department
University of Liberal Arts Bangladesh (ULAB), Dhanmondi, Dhaka
Email: nusratfarin89@gmail.com, rimon2009@outlook.com, sifat.momen@ulab.edu.bd,
shorifuddin@gmail.com and nafees.mansoor@ulab.edu.bd

***Abstract*—Due to the ever-advancing technology, usage of information technology in modern life is increasing at a fast pace. Hence, ICT based systems have enormous potential providing accessibility and affordability to the urban inhabitants in developing countries. In this paper, a framework for a dynamic vehicle pooling system for Dhaka city is proposed. Comparing to existing systems, some new concepts have been incorporated in the proposed framework. One of the unique features in the proposed framework is that the system is not limited to any particular type of vehicle. Thus, any type of vehicle such as car, bus or even lorry can be pooled using the proposed system. Moreover to ensure payment security, the system is designed as a prepaid system. Hence, rider has to ensure payment before getting into the ride, where the drive's account is credited once the rider reaches his/her destination.**

Keywords - dynamic; ride-sharing; vehicle pooling.

I. Introduction

Ride sharing is a service, where people can share their riding sources. Now-a-days it has become very popular in developed country, where this concept is not much popular in developing country. However, it is very beneficial for the developing country. The intention of the ride-sharing system is to minimize the abuse of the transport's fuel as well as relieving from traffic jams.

App-based pooling vehicle system has increased exponentially over the past few years. Online application for ride-sharing is a tool that allows riders and drivers to be informed about the present status and exact location of rider and driver. It also helps to know about the current situation of the traffic. Mainly, app-based system of ride sourcing services is generally utilized as a part of finding the nearest vehicle in a very short time. Hence, app-based Ride sourcing services are introduced by many researchers. In some cases, they have different name as Uber [1], Lyft, and their competitors also known as Transportation Network Companies (TNCs) or ride sharing promise to increase reliability and reduce waiting time for moving transport.

Ride-sharing service is more efficient to take the service of vehicles without much hassle. There are many factors in ride sourcing services such as vehicle information, information of drivers and riders, services and so on. Ride-sharing services gather an extraordinary arrangement of log information about closest vehicles and drivers. People may choose this app-based ride sourcing system for many reasons, such as proving the information of nearest vehicle, driver information such as name, contact number etc, riders comfort to find the vehicle easily, reducing waiting time and so forth.

Considering the benefits of ride-sharing services, an app-based pooling service of vehicles is developed and proposed in this paper. Though there exists some application of pooling service of vehicles, however all of these systems are failed to add the maximum benefit of ride sourcing services in their model. For example, in many systems, it is not possible to find all types of vehicle such as bike, private car, taxi cab, bus etc. Therefore, this paper proposes a novel application to find the ride-sharing services. Moreover, this system provides the security, accessibility, identification of the users. The design aspects and features of the proposed system are described thoroughly in the paper.

This paper is organized as follows. The background study on the existing system is presented in section II. In section II, we also discussed about different ride-sharing services application with their pros and cons. Proposed model is presented in section III. The implementation of the proposed system is described in IV. Paper is concluding in section V. Section V also contains the future scopes.

II. Background Study

In recent years, advantages of information and communication engineering have qualified new services that give a wide variety of real time and fabulous tours. Different companies such as Lyft [2], UberX [3], Sidecar [4], carpool [5] have flourished offering smart-phone apps to link up the riders with the community of vehicles driver. Passengers can request for a ride-sharing to a licensed driver through the application. Passenger can communicate with drivers via the application to be informed the current location and status.

Some debates have gone into defining these types of services. There exist some terminologies named as Transportation Network Companies (TNCs), real-time ride sharing, parataxis, ride matching, on-demand rides, app-based rides etc. Similar

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

types of car pooling systems are GoLoco [6], Ridegrid, Hitchsters [7] and Rideamigos [8]. The web based application like WoTCoMS [9] are also very popular in recent years. There are also some other applications which are developed considering positioning system based like as [10–15]. Real time system has become very popular and many systems are developed for this.

A web-based real-time carpooling system [7] is introduced by Dejan. This is a web based system. There also exists many mobile application related with car pooling. A distributed optimized approach based on the multi agent concept for the implementation of a real time carpooling service with an optimization aspect on siblings [7], and Orchestrating Yahoo! Fire Eagle location based service for carpooling [8] are related with the vehicle pooling system. All of these services are online based application.

In recent year, dynamic car and taxi pooling systems are increased because of the availability of IoT (Internet of things), WoT (Web of Things) and cellular mobile phone or mini electrical devices (e.g., notebook, tab etc.). By utilizing the digitized and easy communication system people use the existing car pooling system.

Most of the existing cars pooling systems have to implement automatic payment system. Moreover, those systems are not design for all kind of vehicles. Hence, the proposed system is designed for all kinds of vehicles (e.g., bus, track, motorcycle, taxi, car etc.), with a unique payment system. Users of the system are able to pay the bill through this system. The proposed system has combined with the ICT-based solution both in the secured payments and the dynamic resources allocation.

III. Proposed Model

The vehicle pooling system is designed with some key features while identifying any key features of any established system. The key features of the proposed system are quite unique than any other existing systems. The flowchart of the proposed system is described in Figure 1.

A. Dynamic

In this paper, a dynamic system is designed and implemented. This system works for different types of vehicles. This system is designed for the most available vehicles such as car, taxi, track, bus, motor bike, CNG etc. User can access and can send request to the driver of the vehicles. The system works dynamically for all the vehicles. There is no carpooling system available for dynamic vehicles. This proposed system is designed for dynamic vehicles.

B. Pre-paid System

In this paper we propose and implement a top-up or pre-paid method to make payment through the system. The user or passenger can make a request to the driver. The request when accepted and passenger ride on their vehicle, passenger can make a payment to the driver via bKash [16] account or any other bank account. Online transaction system gives the security to transfer money, since online transaction is safe and a secured process. So, the user feels free to use the system. The system has also a payment assurance option and searching technique that gives security, accessibility to the users.

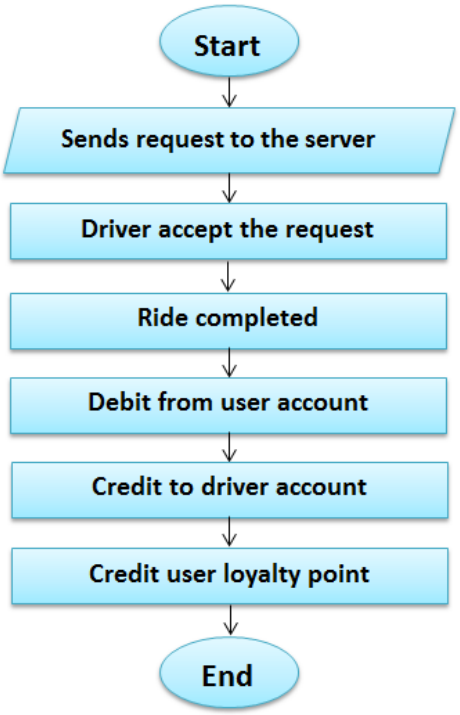


Fig. 1: Flowchart of the proposed system

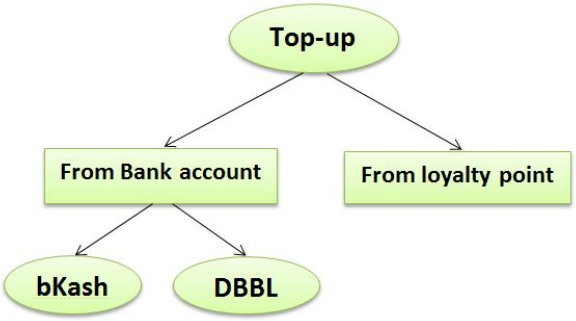


Fig. 2: Top-up system architecture

C. Searching Technique

In this system the user can look for different types of vehicle. By the GPS the user location is tracked by the system and then nearest vehicle is suggested to the user. The nearest vehicle is identified by measuring the euclidean distance [17]. Equation 1 & 2 shows mechanism of the distance measurement of the passenger and the driver or vehicle. Then the system suggests the passenger the nearest vehicle. To implement the searching technique an algorithm is used here. Algorithm 1

described the detail of the used algorithm.

$$Distance(X,Y) = \frac{\sum_{i=1}^{n} X_i.Y_i}{\sqrt{\sum_{i=1}^{n}(X_i)^2}\sqrt{\sum_{i=1}^{n}(Y_i)^2}} \quad (1)$$

$$Distance_{cosine} = 1 - Distance(X,Y). \quad (2)$$

Algorithm 1 Algorithm of the system

```
Collect the position from GPS of the user and the vehicle
Set P_user = value from the tracking location of the user
Set distance [ ] = 0
Set minimum [ ] = 0
Set V_n = [x_1,....., x_n ]
loop
    distance [ ] = distance between P_user and vehicle
end loop
loop
    if distance [i] < distance [i + 1] then
        minimum [i] = distance [i]
    else
        go to step 6
    end if
end loop
```

D. User Royalty

One of the key features of the proposed system is the royalty scheme, where user earns royalty points for using the system. Once a user receives the royalty points, he/she can use these points for payment purpose. As highlighted from the exiting literatures, none of the existing systems for vehicle pooling provides such reward system. It is expected that the number of the current users and number of potential users will be increased once the royalty scheme is applied properly and configured successfully. Hence, this feature is anticipated to attarct more users. Furthermore, once applied, this may become a very helpful solution for the urban inhabitants.

E. Authentication System

In order to use this system, a user has to register an account. Next, the user can activate this registered account by providing personal bank account information to the system. The system authenticates user through National Identification Number (NID) with a two factors authentications policy. Thus, during registration, the user must provide his/her cellular phone number. Since all the mobile phone numbers have been already authenticated using NID, SMS based authentication is used to verify the credibility of the users. Moreover, providing the banking information, the user assures prepaid and automatic payment.

IV. Framework Implementation

A. Architectural Design

The integration process of the proposed system depends on the three tier architecture. Three tiered architecture is used to enable the proposed system to be more robust and to provide flexibility to its applications. Thus, the proposed system blends different protocols and softwares effectively.

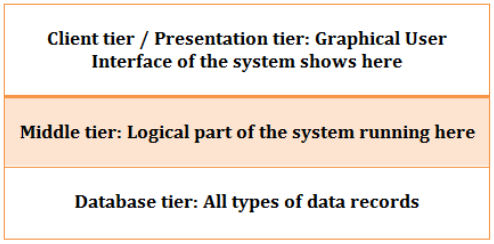


Fig. 3: 3-tier Architecture of Application

Figure 3 shows the overview of the 3-tier architectural view of the proposed system.

1) The Client Tier: The interface acts as the client tier of the three tier architecture model. The web browser and smart phone's touch screen process and display interface of the system. The graphical user interfaces of the proposed system operates in this tier. These interfaces of the system are designed for the user. The system provides the user a very simple and friendly GUI. Thus, user from any age group can easily access the system smoothly. The interfaces of the system are presented in Figure 4 and Figure 5. After the client tier, middle tier of the proposed system starts, which is also an immediate tier.

2) The Middle Tier: In three tier Web Application systems, the majority of the applicational logics are at the middle tier. The client tier presents data and collects data from the user as the database tier stores and receives the data. The middle tier serves most of the remaining roles and merges the other tiers. This tier determines the structure and the contents to be displayed to a user and also processes user input. The user inputs are formed into queries on the database to read or to write data. This middle tier application logic integrates the users with the database management system. The components of the middle tier are interacting with the database management system. For the proposed vehicle pooling system, the middle tier plays an important role of the system. The middle tier executes and supervises all the necessary things for the system. Figure 3 shows all using tier of this system. The database tier is the immediate tier of the system.

3) The Database Tier: The database tier is the base tier of the proposed system. Designing databases and building tier is the first step. Based on the entity relationship diagram which is done in the system design phase, database of the proposed system is created in MySQL.

B. Roles of User

The proposed system can be considered to be a simple process. Initially, the user has to create an account in the vehicle pooling website. First level of user validation is done when the user confirms his/her account via email verification. Next, user can login and can access the dashboard panel. The dashboard panel shows the options to the user. Afterwards,

user chooses to send a request for a ride. Once the request is accepted, user goes to the payment option. Upon completion of the payment successfully, user sees the driver details. In the proposed system, driver is another type of user. Therefore, similar to user registration, a driver also requires opening a new account and requires completing email verification. Upon logging into the system, driver can access the dashboard panel and can see the options. On a drivers dashboard, all the currently posted requests for riding are displayed.

1) Passenger: The first role of a passenger in Vehicle pooling is to open an account and to confirm the account via email verification. Next, passenger sends request for a new ride and also tracks history. Considering a passengers roles and activities, there are several required features in the proposed vehicle pooling system. However, we assigned to meet few of these requirements, which are, (1) Create new account to access the system (2) Can log in the system (3) Can send Request (4) Payment Option (5) Help

2) Driver: The first role of Driver in Vehicle pooling is also to open an account and confirms the account via email verification. Afterwards, a driver can view the current requests for new rides and can also track history. Considering a drivers roles and activities, there are several required features in the proposed vehicle pooling project. Few included requirements are as follows, (1) Can open new account to access the system (2) Can log in the system (3) Can view current requests (4) Can view the received payment details of an accepted request (5) Help

3) Admin: For the proposed system, an admin manages the database. Admin keeps the tracking number and also maintains the records. Considering an admins roles and activities, there are several required features in the proposed vehicle pooling project. However, few of these requirements included in this system, which are as follows, (1) Log in as admin (2) View all current and previous requests (3) Monitor and track the ongoing request

4) Physical Design: After finishing the system analysis, the designing phase of SDLC is considered. Here the system is designed both logically and physically. As mentioned earlier, for the logical design the process is modeling by DFD (Data Flow Diagram) and data modeling by ER Diagram (Entity-Relationship Diagram). All the necessary materials like DFD, Logical ERD, Physical ERD and Data Dictionary of system design or developments are already discussed. In this section the findings from the system analysis phase while designing the entire system. How the system would work and how the output will be given were the main concerns at this level.

C. Application

The home page of vehicle pooling system is depicted in Figure 4. When a user goes to the vehicle pooling website and this page is displayed.

Sequence diagram is a powerful tool that makes understand the dynamics of a use case. Sequence diagrams represent possible interaction scenarios.

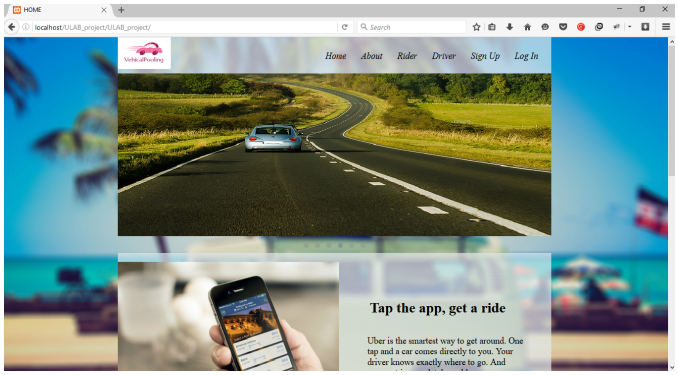


Fig. 4: Vehicle pooling system home page

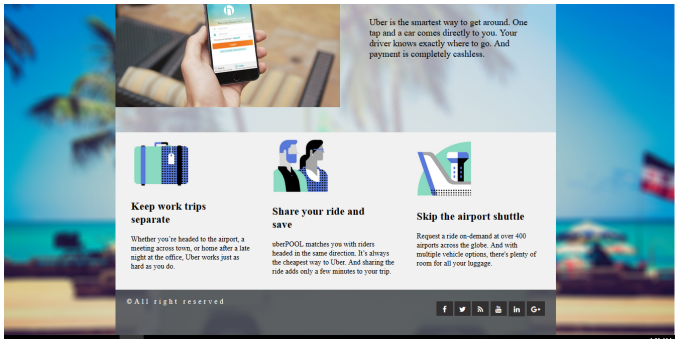


Fig. 5: Proposed System homepage

1) Sequence Diagram for Sign up: In the sign up sequence diagram we can assure the procedure of the sign up process. The user will request for the sign up page and subsequently the sign up page will open. Then user will provide information and upon clicking the submit button, the information will be saved in the database [11].

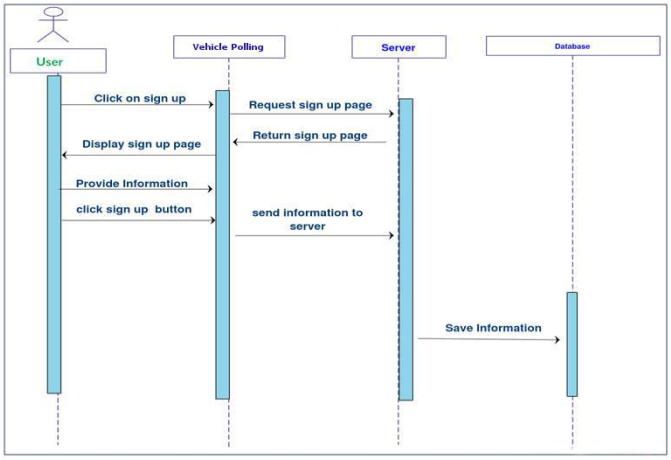


Fig. 6: Sequence diagram for sign up

2) Sequence Diagram for Sign in: For sign in, user will request for the sign in page. In the sign in page, user will provide user name and password and submit to the system. After clicking submit button, users existence in the database

will be checked. A message will be shown to user regarding the validity of the user. By signing to the system user can use

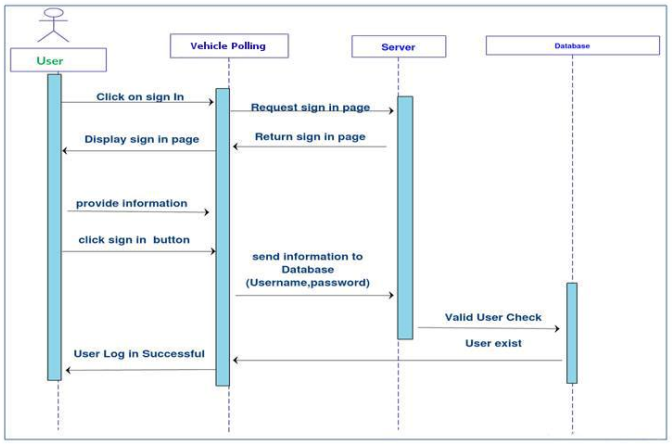


Fig. 7: Sequence diagram for sign in

the facilities and find benefits from the unique key features of the system.

Due to the scope of the testing results, user feedback and details of evolution are not included in the paper. Authors' next research step is to incorporate few other features to the proposed vehicle pooling system and implement the system.

V. Conclusion and Future works

In this paper, a system for dynamic vehicle pooling for urban area is proposed. The main intention is to build the framework to use/utilize the empty seats in any type of vehicles. Moreover, the proposed system also aims to eliminate the space misuseages in vehicles. Thus, the proposed system is considered to be an augmented system that matches the transportation demands of the user based on the location and destination. The system also offers security to the user. Hence, the NID number is used for registration and dual-level verification is used for authentication. Thus, the proposed system assures more security to the user than any other existing systems. The framework has social as well as economical impacts to the society. It may also help to reduce the traffic jams by the effective usage of vehicles. The proposed scheme is in the initial stage. In future, complete acceptance tests are planned to be performed. Features for all types of user (e.g., admin, driver, user) are not fully addressed in the current system.

References

[1] Jochen Wirtz and Christopher Tang. Uber: Competing as market leader in the us versus being a distant second in china. In *SERVICES MARKETING: People Technology Strategy*, pages 626–632. 2016.

[2] Brian W Powers, Scott Rinefort, and Sachin H Jain. Nonemergency medical transportation: Delivering care in the era of lyft and uber. *JAMA*, 316(9):921–922, 2016.

[3] Judd Cramer and Alan B Krueger. Disruptive change in the taxi business: The case of uber. *The American Economic Review*, 106(5):177–182, 2016.

[4] Joseph Donald Manor. Reverse mechanism for a motorcycle with a sidecar, April 28 2016. US Patent 20,160,114,856.

[5] Sheng-Kai Chou, Ming-Kai Jiau, and Shih-Chia Huang. Stochastic set-based particle swarm optimization based on local exploration for solving the carpool service problem. 2016.

[6] Jae-Pyo Jeon, Dhananjay P Thakur, Jin-bin Tian, Insuk So, and Michael X Zhu. Regulator of g-protein signalling and goloco proteins suppress trpc4 channel function via acting at gαi/o. *Biochemical Journal*, 473(10):1379–1390, 2016.

[7] Ken Taylor, Herman Jaya So, and Jim Lilley. A system for sharing taxis with multiple interfaces.

[8] Evan Meyer, Jeffrey Chernick, and Ben Dalton. Matching system for ride reservation platforms, October 9 2012. US Patent 8,285,570.

[9] Nusrat Jahan Farin, Atiqur Rahman, Nafees Manoor, and Sazzad Hossain. Wotcoms: A novel cross-layered web-of-things based framework for course management system, January 9 2016. ICAICT, 2016.

[10] Douglas Oliveira Santos and Eduardo Candido Xavier. Dynamic taxi and ridesharing: A framework and heuristics for the optimization problem. In *IJCAI*, volume 13, pages 2885–2891, 2013.

[11] Sergio Di Martino, Clemente Giorio, and Raffaele Galiero. A rich cloud application to improve sustainable mobility. In *International Symposium on Web and Wireless Geographical Information Systems*, pages 109–123. Springer, 2011.

[12] Jianhua Shao and Chris Greenhalgh. Dc2s: a dynamic car sharing system. In *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks*, pages 51–59. ACM, 2010.

[13] Nianbo Liu, Yong Feng, Feng Wang, Bang Liu, and Jinchuan Tang. Mobility crowdsourcing: toward zero-effort carpooling on individual smartphone. *International Journal of Distributed Sensor Networks*, 2013, 2013.

[14] Eugénie Lioris, Guy Cohen, and Arnaud de La Fortelle. Overview of a dynamic evaluation of collective taxi systems providing an optimal performance. In *Intelligent Vehicles Symposium (IV), 2010 IEEE*, pages 1110–1115. IEEE, 2010.

[15] Douglas O Santos and Eduardo C Xavier. Taxi and ride sharing: a dynamic dial-a-ride problem with money as an incentive. *Expert Systems with Applications*, 42(19):6728–6737, 2015.

[16] Gregory Chen and Stephen Rasmussen. bkash bangladesh: A fast start for mobile financial services. Technical report, The World Bank, 2014.

[17] Jan Draisma, Emil Horobeţ, Giorgio Ottaviani, Bernd Sturmfels, and Rekha R Thomas. The euclidean distance degree of an algebraic variety. *Foundations of Computational Mathematics*, 16(1):99–149, 2016.

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

Modified Spider Monkey Optimization

Garima Hazrati[a], Harish Sharma[b], Nirmala Sharma[c] and Jagdish Chand Bansal[d]
[abc]Department of Computer Science and Engineering
Rajasthan Technical University, Kota
[d]Department of Mathematics
South Asian University, Delhi
Email: [a]ghazrati9@gmail.com, [b]hsharma@rtu.ac.in, [c]nsharma@rtu.ac.in, [d]jcbansal@gmail.com

***Abstract*—Spider Monkey Optimization is a well known meta-heuristic in the arena of nature inspired algorithms. It is basically known for its stagnation removal power in its original design. To balance the meta-heuristics mechanisms while preserving premature convergence, a new variant is developed which is named as Modified spider monkey optimization. In this paper, metropolis principle is used from simulated annealing which improves the global search capability of algorithm. In addition to this strength of spider monkey is used for maintaining the step-size to enhance the convergence speed. The intended algorithm is tested over 10 benchmarks functions and compared with Spider monkey optimization, particle swarm optimization and one of its recent variant Self-adaptive spider monkey optimization.**

***Index Terms*—Swarm Intelligence, Nature Inspired Algorithms, Particle Swarm Optimization, Spider Monkey Optimization**

I. Introduction

Nature inspired algorithms(NIA) are those probabilistic based approaches that are inspired by natural phenomenons. NIA are classified into several categories like evolutionary algorithms [1], swarm intelligence (SI) based algorithms [10], physics based algorithms [5], bio-inspired algorithms [11]. Evolutionary algorithms basically depends on evolution of species based on Darwinian principle like differential evolution [14], cuckoo search [18] etc. Particle swarm optimization (PSO) [9], Artificial bee colony [3], Ant colony optimization [6] are those which fall in the category of SI based algorithms as in these particles use their persistence and collective intelligence to update their positions. Gravitational search algorithm [12], Magnetic optimization algorithm [15] are inspired by laws of physics. Bio-inspired algorithms are those which are inspired by biological processes like immune algorithm [16], krill herd [7] etc. Spider Monkey Optimization (SMO) is a latest entry to the arena of SI based algorithms developed in 2014 by J.C. Bansal et. al. [4] which imitates the food foraging behavior of spider monkeys. SMO is a well balanced algorithm among other rooted algorithms by removing their various flaws. Besides, this SMO needs more balancing of meta-heuristic mechanisms.

In this paper, a modified variant of SMO is designed to overcome these flaws and is named as Modified Spider Monkey Optimization (MSMO). In this variant, exploration and exploitation is properly balanced in all phases and it also removes premature convergence by updating location of global leader.

Further, the paper is structured as follows: Section II gives an overview about spider monkey optimization and Section III presents the modifications in various phases of SMO. Section IV presents the experimental outcomes followed by conclusion in Section V.

II. Spider Monkey Optimization

SMO is a latest advancement in the arena of nature inspired algorithms generated by researchers by imitating the foraging process of spider monkeys. This algorithm is a mixture of six phases that make this algorithm a balanced one and are described below:

A. Local Leader Phase

After being initialized in initialization phase, SMs update their position by taking knowledge from their local leader and neighbour while using its own persistence. This is presented by equation 1:

$$Mnew_{ij} = M_{ij} + a \times (LL_{kj} - M_{ij}) + b \times (M_{rj} - M_{ij}) \quad (1)$$

In equation 1, $Mnew_{ij}$ and M_{ij} is the updated and old position of i^{th} SM. LL_{kj} presents itself as local leader of k^{th} group in j^{th} dimension. M_{rj} represents the randomly taken neighbour. a and b are randomly distributed random numbers in range of [0,1] and [-1,1] respectively.

B. Global Leader Phase

Based on the fitness SMs again get a chance to update themselves and reach global optima in this phase. Here, SMs are inspired from their own persistence, random neighbour and global leader of the bevy. Equation 2 show location update strategy at this stage:

$$Mnew_{ij} = M_{ij} + a \times (GL_j - M_{ij}) + b \times (M_{rj} - M_{ij}) \quad (2)$$

GL_j is the global optima of bevy.

C. Global Leader Learning Phase

This phase learns about the existence of global leader in the whole bevy and it is checked that leader is updating its position or not to a certain threshold for further action.

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

D. Local Leader Learning Phase

After knowing about global optima, algorithm needs to find out the local leader of small clusters to find local optima as well. This phase also checks that wether local leaders are updating themselves or not by checking the counter for threshold.

E. Local Leader Decision Phase

Now, if local leaders are not updating themselves to a certain threshold then all SMs of that bevy update their location inspiring either from global leader or by random initialization depending on perturbation rate. Equation 3 presents inspiration from global leader.

$$Mnew_{ij} = M_{ij} + a \times (GL_j - M_{ij}) + a \times (M_{ij} - LL_{kj}) \quad (3)$$

F. Global Leader Decision Phase

If global leader doesn't get updated to a particular threshold i.e. global leader limit then algorithm enters in this phase where fission and fusion of whole bevy takes place.

III. Modified Spider Monkey Optimization

A. Modified Local Leader Phase

In local leader phase, new location of SM is selected on the basis of greedy approach i.e. is the strength of SM is high at new location then it is elected else rejected. Due to this approach, sometimes when a low strength SM is near global optima doesn't get a chance to update itself and algorithm moves in other direction. For avoiding this problem a position update chance is also provided to SM having low strength with some probability factor so that, it can also reach global optima. This probability factor is taken from simulated annealing [8] and applied with greedy approach for selection of new location of SM. By this factor solutions with high objective value and low fitness also get chance to update itself and is defined in equation

$$\Delta Obj = Obj_N - Obj_O \quad (4)$$

$$P(\Delta Obj) = \exp(-\Delta Obj/T) \quad (5)$$

In equation 5, ΔObj is the difference between the objective value at new position (Obj_N) and old (Obj_O) position which is calculated in equation 4. This ΔObj is then passed to the exponential factor for resulting in probability factor i.e. shown in equation 5 where T is a temperature constant that is used to maintain the equilibrium of the phase. This probability factor is an exponential factor of difference objective values by high temperature value which helps in decreasing step size at later iterations while giving chance to low fit solutions to reach global optima. This probability factor becomes the decision maker of the life of SM and is defined as shown in algorithm III-A. This probability factor step is known as metropolis step in simulated annealing forming it's backbone.

Now, greedy selection of modified local leader phase is shown as:

By the above selection approach, low strength spider monkeys also get a chance to update themselves and help in reaching global optima. When a SM is having high strength then it is elected by greedy election as it takes the algorithm towards global optima but if a SM has low strength then metropolis step come into existence and probability factor show its magic. This exponential probability factor helps the SM to explore the search space well so that it helps the algorithm in reaching new optimal values. This modified version helps the phase for bettering the explorative capability of algorithm due to which algorithm can't get stagnated and new opportunities are formed to reach optimal value.

if (fit_N (SM)> fit_O)(SM) $\|$ r > $\exp(-(f_N - f_O)/T)$
select *updated solution*
else
select *previous solution*

B. Modified Global Leader Phase

Every position update equation consists of individual′s old position and step-size which is shown in equation 2. Step-size of any solution is used to calculate its new position to reach the optima. The length of step-size decides wether SM explore the search area or exploit it and results a SM to prominent and non-prominent solution. The random numbers ϕ and ψ are sometimes the decision maker of step-size of SM as they lead SM in diverse directions. To maintain the balance and diversity of algorithm a probability factor is introduced in random numbers. With the help of probability factor in the step-size SM's can't skip the true solution lying near to it. By introducing the probability factor, the newly formed position update equation is as equation 6:

$$Mnew_{ij} = M_{ij} + U((0, 2.2) - (prob_i) \times (GL_j - M_{ij}) + U(-(2.2 - prob_i), (2.2 - prob_i)) \times (M_{rj} - M_{ij}) \quad (6)$$

By introducing the probability factor in equation 6, SM's having high probability lie close to optimal solution and low probability SM's lies far from it. Since, probability is a function of strength i.e directly proportional to strength SM's having good strength get more chance to update themselves and comes near to global optima. From this factor, it is clear that high strength SM's have small step-size and resulting in exploitation of the area whereas low strength SM's have larger step-size and results in exploration of the area. In addition to this greedy approach is changed with the help of metropolis step as shown in above modified local leader phase. The algorithm for modified global leader phase is shown in algorithm 1.

It is apparent from algorithm 1 that a SM which is having better probability factor results in less step-size and reach global optima by exploiting the search area. When a SM is having lesser probability factor then it results in larger step-size due to which it explores the search area and helps in finding new optima. Besides, this controlled movement of step-size if any SM is not having good strength then also it is

Algorithm 1 Position update process in Global Leader Phase (GLP):

$count = 0$;
while $count < group\ size$ **do**
  for each member $SM_i \in$ group **do**
    if $U(0,1) < prob_i$ **then**
      $count = count + 1$.
      Randomly select $j \in \{1...D\}$.
      Randomly select $M_r \in$ group s.t. $r \neq i$.
      $Mnew_{ij} = M_{ij} + U(0, 2.2 - prob_i) \times (GL_j - M_{ij}) +$
      $U((2.2 - prob_i), -(2.2 - prob_i)) \times (M_{rj} - M_{ij})$
    end if
    Metropolis Step
  end for
end while

accepted with the exponential probability factor of metropolis step i.e. described in modified local leader phase.

C. Modified Global Leader Decision Phase

This phase presents the perfect behavior of fission-fusion social structure as there is a need to split the group or combine it. This structure is well understood by its algorithm as it depends on location of global leader and global leader limit. Sometimes, this presence of fission-fusion social structure leads to premature convergence of algorithm. So, to maintain the convergence speed and avoiding stagnation global leader is initialized randomly if global limit count reaches global leader limit.

Algorithm 2 Global Leader Decision Phase:

if $GlobalLimitCount > GlobalLeaderLimit$ **then**
  $GlobalLimitCount = 0$
  random initialization of global leader
  if Number of groups $< MG$ **then**
    Divide the population into groups.
  else
    Combine all the groups to make a single group.
  end if
  Update Local Leaders position.
end if

IV. Experimental Outcomes

This section deals with the benchmark functions used for analysis of developed variant and its outcomes.

A. Test problems considered

The modified variant is tested over 10 benchmarks to prove its existence among other established meta-heuristics. These 10 benchmarks are shown in Table I.

TABLE I: Test problems

Test Problem	Objective function	Search Range	Optimum Value	D	Acceptable Error
Alpine	$f_1(x) = \sum_{i=1}^{D} \lvert x_i \sin x_i + 0.1x_i \rvert$	[-10, 10]	$f(0) = 0$	30	$1.0E-05$
Michalewicz	$f_2(x) = -\sum_{i=1}^{D} sinx_i(sin(\frac{i.{x_i}^2}{\pi})^{20})$	$[0, \pi]$	$f_{min} = -9.66015$	10	$1.0E-05$
Cosine Mixture	$f_3(x) = \sum_{i=1}^{D} {x_i}^2 - 0.1(\sum_{i=1}^{D} cos5\pi x_i) + 0.1D$	[-1, 1]	$f(0) = -D \times 0.1$	30	$1.0E-05$
Schewel	$f_4(x) = \sum_{i=1}^{D} \lvert x_i \rvert + \prod_{i=1}^{D} \lvert x_i \rvert$	[-10 10]	$f(0) = 0$	30	$1.0E-05$
Sum of different powers	$f_5(x) = \sum_{i=1}^{D} \lvert x_i \rvert^{i+1}$	[-1 1]	$f(0) = 0$	30	$1.0E-05$
Quartic	$f_6(x) = \sum_{i=1}^{n} i.x_i^4 +$ random$[0, 1)$	[-1.28 1.28]	$f(0) = 0$	30	$1.0E-05$
Rotated hyper-ellipsoid	$f_7(x) = \sum_{i=1}^{D} \sum_{j=1}^{i} x_j^2$	[-65.536 65.536]	$f(0) = 0$	30	$1.0E-05$
Levy montalvo 1	$f_8(x) = \frac{\Pi}{D}(10sin^2(\Pi y_1) + \sum_{i=1}^{D-1}(y_i - 1)^2 \times (1 + 10sin^2(\Pi y_{i+1})) + (y_D - 1)^2)$, where $y_i = 1 + \frac{1}{4}(x_i + 1)$	[-10, 10]	$f(-1) = 0$	30	$1.0E-05$
Beale	$f_9(x) = [1.5 - x_1(1 - x_2)]^2 + [2.25 - x_1(1 - x_2^2)]^2 + [2.625 - x_1(1 - x_2^3)]^2$	[-4.5,4.5]	$f(3, 0.5) = 0$	2	$1.0E-05$
Shifted Schwefel	$f_{10}(x) = \sum_{i=1}^{D} (\sum_{j=1}^{i} z_j)^2 + f_{bias},\ z = x - o\ ,x = [x_1, x_2, ...x_D],\ o = [o_1, o_2, ...o_D]$	[-100,100]	$f(o) = f_{bias} = -450$	10	$1.0E-05$

B. Experimental settings

To verify that MSMO is a contended member in field of swarm intelligence algorithms, comparative analysis is done among MSMO, SMO [4], PSO [9] and a recent variant SaSMO [13]. Following experimental setup is done:

- Total number of Spider Monkeys (N) =50;
- Maximum no. of groups=5;
- pr ϵ [0.1, 0.8]
- LLL =1500
- GLL =50
- Rest settings of SMO, PSO and SaSMO are taken from their elementary papers [4], [9], and [13].

C. Outcomes

Table II unfolded the attained outcomes of all the taken algorithms SMO, PSO, SaSMO and MSMO based on above parameter settings. Results are shown in the form of standard deviation (SD), mean error (ME), average function evaluation (AFE) and success rate (SR).

Results in above table II show that MSMO is a better variant than SMO, PSO and SaSMO regarding accuracy, reliability and efficiency. In addition to above results box-plot analysis of compared algorithms in terms of average function evaluations (AFE) is presented. Box-plot analysis [17] of SMO, PSO, and SaSMO is shown in figure 1 representing the empirical distribution of data graphically. Figure 1 shows that variation,

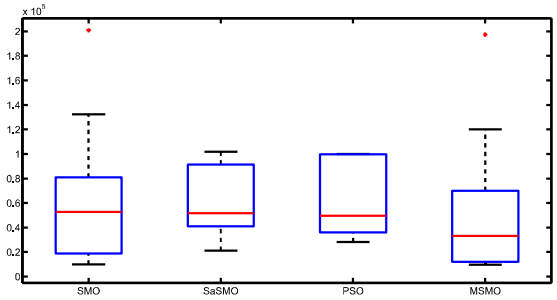


Fig. 1: Boxplots graph for average number of function evaluation

interquartile range and medians of developed MSMO is less than other two. After this, a comparison is made by using the performance indices (PI) [2] based on ME, SR, and AFE. The computed values of PI for SMO, PSO, MSMO and SaSMO are portrayed in figure 2.

Figure 2(a), 2(b), and 2(c) show the performance index of success rate, average function evaluations and mean error respectively. Figure 2 shows that PI of MSMO is notable as compared to other variants.

V. Conclusion

Modified SMO is newly formed variant of SMO elucidated in this paper. It enhances the meta-heuristic components that are exploration and exploitation more balanced by using metropolis principle and probability in local leader phase and global leader phase respectively. Further, to reduce the chances of premature convergence global leader is again initialized in global leader decision phase. From this modification it is unblemished that metropolis principle results in enhancement of global search capability while use of probability make step-size adaptable so that global optima can't be skipped and if global leader get stuck in local optima then it is re-initialized. This intended variant is tested over 10 benchmark functions and its analysis show that it is a spell variant. In future it can be applied to real world optimization problems for further use.

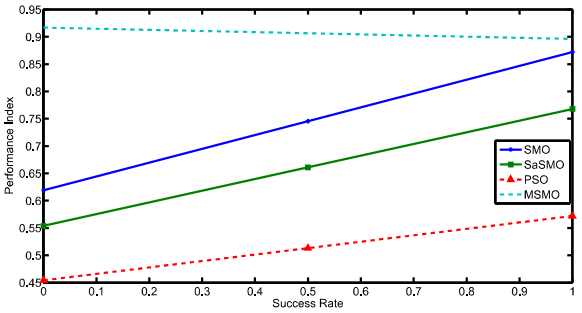
TABLE II: Comparison of the outcomes of test problems

Test Function	Algorithm	SD	ME	AFE	SR
f_1	SMO	4.21E-03	4.95E-04	56524.47	98
	SaSMO	4.88E-04	5.45E-05	54914.81	98
	PSO	4.20E-01	4.20E-01	99402.5	2
	MSMO	2.34E-04	2.85E-05	59763.4	99
f_2	SMO	4.43E-02	1.48E-02	52779.84	90
	SaSMO	1.76E-06	8.44E-06	48378.63	100
	PSO	6.29E-02	2.51E-02	49744	85
	MSMO	3.41E-06	6.07E-06	24131.95	100
f_3	SMO	9.16E-07	8.72E-06	9876.24	100
	SaSMO	1.51E-06	8.34E-06	41028.84	100
	PSO	6.08E-07	9.32E-06	28182.5	100
	MSMO	8.68E-07	8.84E-06	9714.87	100
f_4	SMO	2.55E-02	1.93E-01	200862.84	7
	SaSMO	1.35E-01	1.56E+00	101746.98	0
	PSO	8.01E-02	3.98E-01	100003	1
	MSMO	3.13E-02	1.89E-01	197366.57	10
f_5	SMO	0.00E+00	0.00E+00	16239.24	100
	SaSMO	0.00E+00	0.00E+00	21085.75	100
	PSO	0.00E+00	0.00E+00	36092.5	100
	MSMO	0.00E+00	0.00E+00	10408.94	100
f_6	SMO	5.21E-02	5.25E-03	80817.68	99
	SaSMO	1.56E-01	5.05E-02	91340.72	45
	PSO	6.05E-01	1.58E+00	99659.5	2
	MSMO	7.82E-02	7.87E-03	69868.25	97
f_7	SMO	1.03E-02	1.05E-03	18723.73	99
	SaSMO	1.74E-06	8.12E-06	39102.07	100
	PSO	6.00E-07	9.30E-06	33252.5	100
	MSMO	9.14E-07	8.90E-06	12102.75	100
f_8	SMO	1.53E-03	2.28E-04	19319.56	98
	SaSMO	1.74E-06	7.99E-06	43708.22	100
	PSO	4.03E-03	1.77E-03	46168	84
	MSMO	1.63E-06	8.58E-06	14553.34	100
f_9	SMO	1.98E-04	8.10E-04	52573.96	100
	SaSMO	1.34E-02	8.16E-03	92630.76	39
	PSO	4.68E-04	8.79E-04	49273.5	98
	MSMO	1.84E-04	8.42E-04	42396.4	100
f_{10}	SMO	4.79E-03	1.79E-03	132298.41	81
	SaSMO	2.83E-03	9.19E-04	59303.68	86
	PSO	5.61E-02	6.59E-02	100050	0
	MSMO	1.77E-03	4.94E-04	119920.16	90

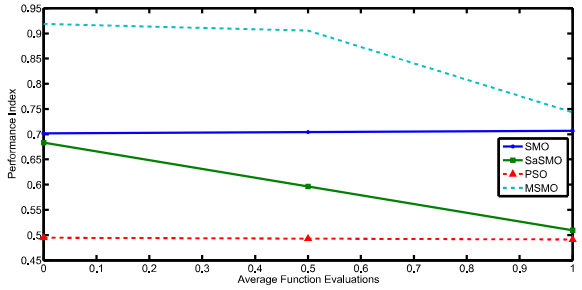
References

[1] Thomas Back, David B Fogel, and Zbigniew Michalewicz. *Handbook of evolutionary computation*. IOP Publishing Ltd., 1997.

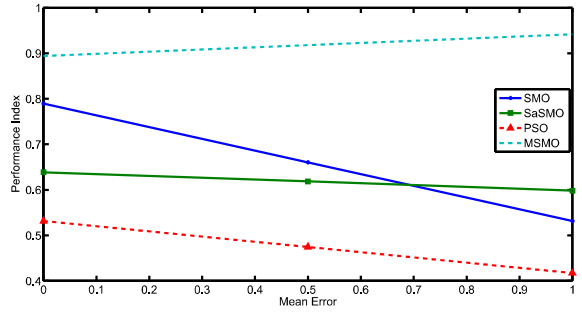
[2] Jagdish Chand Bansal and Harish Sharma. Cognitive learning in differential evolution and its application to



(a)



(b)



(c)

Fig. 2: Performance index for test problems; (a) for case (1), (b) for case (2) and (c) for case (3).

model order reduction problem for single-input single-output systems. *Memetic Computing*, 4(3):209–229, 2012.

[3] Jagdish Chand Bansal, Harish Sharma, and Shimpi Singh Jadon. Artificial bee colony algorithm: a survey. *International Journal of Advanced Intelligence Paradigms*, 5(1):123–159, 2013.

[4] Jagdish Chand Bansal, Harish Sharma, Shimpi Singh Jadon, and Maurice Clerc. Spider monkey optimization algorithm for numerical optimization. *Memetic computing*, 6(1):31–47, 2014.

[5] Anupam Biswas, KK Mishra, Shailesh Tiwari, and AK Misra. Physics-inspired optimization algorithms: a survey. *Journal of Optimization*, 2013, 2013.

[6] Christian Blum. Ant colony optimization: Introduction and recent trends. *Physics of Life reviews*, 2(4):353–373, 2005.

[7] Amir Hossein Gandomi and Amir Hossein Alavi. Krill herd: a new bio-inspired optimization algorithm. *Communications in Nonlinear Science and Numerical Simulation*, 17(12):4831–4845, 2012.

[8] Chii-Ruey Hwang. Simulated annealing: theory and applications. *Acta Applicandae Mathematicae*, 12(1):108–111, 1988.

[9] James Kennedy. Particle swarm optimization. In *Encyclopedia of Machine Learning*, pages 760–766. Springer, 2010.

[10] James Kennedy, James F Kennedy, Russell C Eberhart, and Yuhui Shi. *Swarm intelligence*. Morgan Kaufmann, 2001.

[11] Efrén Mezura-Montes and Blanca Cecilia Lopez-Ramirez. Comparing bio-inspired algorithms in constrained optimization problems. In *2007 IEEE Congress on Evolutionary Computation*, pages 662–669. IEEE, 2007.

[12] Esmat Rashedi, Hossein Nezamabadi-Pour, and Saeid Saryazdi. Gsa: a gravitational search algorithm. *Information sciences*, 179(13):2232–2248, 2009.

[13] Rajani Kumari Sandeep Kumar, Vivek Kumar Sharma. Self-adaptive spider monkey optimization algorithm for engineering optimization problems. *International Journal of Information, Communication and Computing Technology*, Vol II, Issue II (July-Dec2014), 2014.

[14] Harish Sharma, Jagdish Chand Bansal, and KV Arya. Fitness based differential evolution. *Memetic Computing*, 4(4):303–316, 2012.

[15] Mohammad-H Tayarani-N and MR Akbarzadeh-T. Magnetic optimization algorithms a new synthesis. In *2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence)*, pages 2659–2664. IEEE, 2008.

[16] Lei Wang, Jin Pan, and Licheng Jiao. The immune algorithm. *Acta Electronica Sinica*, 28(7):74–78, 2000.

[17] David F Williamson, Robert A Parker, and Juliette S Kendrick. The box plot: a simple visual method to interpret data. *Annals of internal medicine*, 110(11):916–921, 1989.

[18] Xin-She Yang and Suash Deb. Engineering optimisation by cuckoo search. *International Journal of Mathematical Modelling and Numerical Optimisation*, 1(4):330–343, 2010.

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

Embedding of (i,j)-Regular Signed Graphs in $(i+k,j+l)$-Regular Signed Graphs

Deepa Sinha
Department of Mathematics
South Asian University,
Akbar Bhawan New Delhi
Email:deepa_sinha2001@yahoo.com

Anita Kumari Rao
Department of Mathematics
South Asian University,
Akbar Bhawan New Delhi
Email: anita.rao31@gmail.com

Pravin Garg
Department of Mathematics
University of Rajasthan,
Rajasthan, India
Email: garg.pravin@gmail.com

***Abstract*—A *signed graph* is an ordered pair $S=(S^u,\sigma)$, where $S^u=(V,E)$ is a graph and $\sigma: E\to\{+,-\}$ is a function from the edge set E of S^u into the set $\{+,-\}$. A signed graph S is called *regular* if the number of positive (negative) edges, $d^+(v)$ ($d^-(v)$) incident at a vertex v in S, is independent of the choice of v in S. In this paper, we settle the problem of embedding (i,j)-regular signed graphs in $(i+k,j+l)$-regular signed graphs with the minimum number of vertices.**

2010 Mathematics Subject Classification: $5C22, 05C60$

Key words:Signed graph, regular signed graph, embedding.

I. Introduction

General areas that employ embedded systems covers every branch of day to day science and technology, namely communications, automotive, military, medical, consumer, machine control etc...for example cell phone, digital camera, microwave oven, MP3 player, portable digital assistant and automobile antilock brake system etc. Seeing its importance, it has been worked out for theoretically for embedding signed-regular graphs in higher order signed-regular graphs.

For the preliminary notation and terminology in graph theory we refer Harary [9] and West [20]; and for signed graphs we refer Zaslavsky [21], [22], [23]. A *signed graph* is an ordered pair $S=(S^u,\sigma)$, where $S^u=(V,E)$ is a graph, called the *underlying graph* of S and $\sigma: E\to\{+,-\}$ is a function from the edge set E of S^u into the set $\{+,-\}$, called the *signature* of S. Let $E^+(S)=\{e\in E(S^u):\sigma(e)=+\}$ and $E^-(S)=\{e\in E(S^u):\sigma(e)=-\}$. The elements of $E^+(S)$ and $E^-(S)$ are called *positive* and *negative* edges of S, respectively. A signed graph is said to be *homogeneous* if all its edges have the same sign and *heterogeneous* otherwise.

Two edges of a graph are said to be *adjacent* if they are incident with a common vertex. A set of edges in a graph is *independent* if no two of them are adjacent. The *complement* $\overline{S^u}$ of a graph S^u is a graph with the same vertex set as S^u and two vertices are adjacent in $\overline{S^u}$ if and only if they are not adjacent in S^u. A graph is called r$-$*regular* if all its vertices are of degree r. A signed graph S is called *regular* if the number of positive (negative) edges, $d^+(v)$ ($d^-(v)$) incident at a vertex v in S, is independent of the choice of v in S. *i.e.* S is (i,j)-regular, where $i=d^+(v)$ is the positive degree of v in S and $j=d^-(v)$ is the negative degree of v in S (*see* [25]).

A *u-v path* in a signed graph S is an alternating sequence of vertices and edges $u,e_1,v_1,e_2,...,e_n,v$, beginning and ending with vertices, in which each edge is incident with the two vertices immediately preceding and following it and no vertex is repeated. A path is called *homogeneous* if all its edges have the same sign and *heterogeneous* otherwise. The number of edges in a path is called the *length* of the path. P_n denotes a path of length n.

The *cartesian product* ${S^u}_1\Box{S^u}_2$ of two graphs ${S^u}_1$ and ${S^u}_2$ is a graph with vertex set $V({S^u}_1)\times V({S^u}_2)$ and two vertices (u_1,u_2) and (v_1,v_2) are adjacent in ${S^u}_1\Box{S^u}_2$ if and only if u_1 is adjacent to v_1 in ${S^u}_1$ and $u_2=v_2$ or u_2 is adjacent to v_2 in ${S^u}_2$ and $u_1=v_1$. Since in tensor product of graphs ${S^u}_1$ and ${S^u}_2$, the degree of a vertex (u,v) is defined as $deg(u,v)=deg(u)deg(v)$. Thus if ${S^u}_1$ and ${S^u}_2$ are regular, so is ${S^u}_1\Box{S^u}_2$.

Let $S=(S^u,\ \sigma)$ be a signed graph. S is called *cartesian product* of two signed graphs $S_1=(S_1^u,\ \sigma_1)$ and $S_2=(S_2^u,\ \sigma_2)$ if $S^u\cong S_1^u\Box S_2^u$ and for any edge $(u_1,u_2)(v_1,v_2)$ of S^u,

$$\sigma((u_1,u_2)(v_1,v_2))=\begin{cases}\sigma_1(u_1v_1) & \text{if } u_2=v_2\ ,\\ \sigma_2(u_2v_2) & \text{if } u_1=v_1.\end{cases}$$

II. Embedding (i,j)-regular signed graphs in $(i+k,j+l)$-regular signed graphs

A graph G is said to be *embedded* in a graph G' written as $G\subseteq G'$, if there exists a subgraph of G' which is isomorphic to G. If $G\subseteq G'$ it is some times convenient to regard G itself as a subgraph of G' (*see* [1]). Further, if $G\subseteq G'$ and $G\ncong G'$ then the embedding is said to be *proper or strict*; in this case we agree to write $G\subset G'$. In this section, we present the results related to embedding of (i,j)-regular signed graphs into $(i+k,j+l)$-regular signed graphs, $k,l\geq 0$

Gardiner (*see* [8]) has proved that every r-regular graph G can be embedded in the $(r+1)$-regular graph $G\Box P_2$. This result can be lifted to signed graphs as follows:

Theorem 1. *Every (i,j)-regular signed graph $S=(S^u,\sigma)$ can be embedded in the $(i,j+1)$-regular signed graph $S\Box P_2^-$, where P_2^- denotes the path of length 2 with a negative edge.*

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

The above theorem can be written alternatively as follows:

Theorem 2. *Every (i,j)-regular signed graph $S=(S^u,\sigma)$ can be embedded as an induced sub-signed graph in the $(i,j+1)$-regular signed graph $S\Box P_2^-$.*

Corollary 3. *Every (i,j)-regular signed graph $S=(S^u,\sigma)$ can be embedded as an induced subsigned graph in the $(i+1,j)$-regular signed graph $S\Box P_2^+$.*

Due to Theorem 1 and Corollary 3, we see that an (i,j)-regular signed graph S always embeds in an $(i+k,j+l)$-regular signed graph. Now, we wish to minimize number of extra vertices needed to achieve such an embedding in the remaining part of this paper. Throughout the paper, let $\ell(S)$ be the minimum number of extra vertices required to substantiate such an embedding.

III. Embedding of an (i,j)-regular signed graph in $(i,j+1)$-regular signed graph

Theorem 4. *For an (i,j)-regular signed graph $S=(S^u,\sigma)$, if $\overline{S^u}$ has a 1-factor, then $\ell(S)=0$.*

Corollary 5. *For an (i,j)-regular signed graph $S=(C_{2n},\sigma)$, where C_n denotes cycle with n vertices, $\ell(S)=0$*

Corollary 6. *For an (i,j)-regular signed graph $S=(K_{n,n},\sigma)$, if n is even, then $\ell(S)=0$*

Theorem 7. *For an (i,j)-regular signed graph $S=(S^u,\sigma)$ of order n, if $\overline{S^u}$ has no 1-factor, then*

$$i+1\le\ell(S)\le n.$$

We denote by $\delta(G)$ the minimum degree and by $\Delta(G)$ the maximum degree in a graph G. We use $\lceil k\rceil$ to denote the least integer not less than k in the following theorem.

Theorem 8. [6] *For a graph G of order n, if $\delta=\Delta$ is even, then G has at least $\lceil\frac{n\Delta}{2(\Delta+1)}\rceil$ independent edges.*

Now, we improve the upper bound of the inequality of Theorem 7 in the following theorems. $K_{i+j+1}^{(i,j)}$ denotes an (i,j)-regular complete signed graph of order $i+j+1$ in the subsequent theorems.

Theorem 9. *For an (i,j)-regular signed graph $S=(S^u,\sigma)$ of order n, if $\overline{S^u}$ has no 1-factor and n and $i+j$ are of opposite parity, then*

$$\ell(S)\le i+j+1.$$

Proof. Suppose we take $i+j+1$ new vertices and form an (i,j)-regular complete signed graph $K_{i+j+1}^{(i,j)}$. If n and $(i+j)$ are of opposite parity, then $\overline{S^u}$ is $(n-i-j-1)$-regular graph and $(n-1-i-j)$ is even. Now using result in Theorem 8, the number of independent edges in $\overline{S^u}\ge\frac{n-i-j-1}{2}$.

Thus, select $\frac{n-(i+j+1)}{2}$ independent edges in $\overline{S^u}$ and make them adjacent with a negative edge in S. Now, join remaining $i+j+1$ vertices of S to the vertices of $K_{i+j+1}^{(i,j)}$ with a negative edge in one to one manner, i.e. a vertex of S is adjacent to a vertex of $K_{i+j+1}^{(i,j)}$ and the another vertex of S is adjacent to the another vertex of $K_{i+j+1}^{(i,j)}$. Thus, the resulting signed graph is $(i,j+1)$-regular. Hence,

$$\ell(S)\le i+j+1.$$

□

Theorem 10. *For an (i,j)-regular signed graph $S=(S^u,\sigma)$ of order n, where n and $(i+j)$ are of opposite parity, if $\overline{S^u}$ has no 1-factor and $n\ge(i+1)(j+1)$, then $\ell(S)=i+1$.*

Theorem 11. *For an (i,j)-regular signed graph $S=(S^u,\sigma)$ of order n, where n is even and $i+j$ is odd, if $\overline{S^u}$ has no 1-factor and $i\ge1$, then*

$$\ell(S)=\begin{cases} i+j+1 & \text{if } n=i+j+1,i+j+3,\ldots,2i+2j-2\ ,\\ i+j & \text{if } n=2i+2j,\ i=2,4,\ldots,\frac{n}{2}-j\ ,\\ i+j+1 & \text{if } n=2i+2j,\ i=3,5,\ldots,\frac{n}{2}-j.\end{cases}$$

Proposition 12. *For an (i,j)-regular signed graph $S=(K_{n,n},\sigma)$, if n is odd, then*

$$\ell(S)=\begin{cases}2 & \text{if } i=1\ ,\\ i+j & \text{if } i=2,4,\ldots,i+j-1,\\ i+j+1 & \text{if } i=3,5,\ldots,i+j-2.\end{cases}$$

Theorem 13. [20] *If G is a simple connected graph of order n with $n\ge3$ and $d(v)\ge\frac{n}{2}\ \forall\ v\in V(G)$, then G is Hamiltonian.*

Theorem 14. *For an (i,j)-regular signed graph $S=(S^u,\sigma)$ of order n, if $n\ge2i+2j+2$ is even, then $\ell(S)=0$.*

Theorem 15. [8] *Let G be a r-regular graph on n vertices. If $\overline{G}$ has no 1-factor and n, r have the same parity, then $n<2r$ and $\ell(S)=r+2$.*

Theorem 16. *For an (i,j)-regular signed graph $S=(S^u,\sigma)$ of order n, where n is odd and $i+j$ is even, if $\overline{S^u}$ has no 1-factor and $i\ge1$, then*

$$\ell(S)=\begin{cases}i+j+1 & \text{if } n<3(i+j-1)\ ,\\ i+j-(r-2) & \text{if } n=r(i+j-k),\ k \text{ and } r \text{ are odd}\end{cases}$$

IV. Embedding of an (i,j)-regular signed graphs in $(i+1,j+1)$-regular signed graphs

Lemma 17. *A graph G is 2-factorable if and only if G is r-regular for some positive even integer r.*

Theorem 18. *For an (i,j)-regular signed graph $S=(S^u,\sigma)$ of order n, where both n and $i+j$ are of opposite parity, when n is even and $i+j$ is odd, then*

$$\ell(S)=\begin{cases}i+j+1 & \text{if } n=i+j+1,\\ 0 & \text{otherwise}\end{cases}$$

when n is odd and $i+j$ is even, then

$$\ell(S)=\begin{cases}i+j+1 & \text{if } n=i+j+1\ ,\\ 1 & \text{otherwise}\end{cases}$$

Theorem 19. *For an (i,j)-regular signed graph $S=(S^u,\sigma)$ of order n, where both n and $i+j$ are of same parity, then*

$$\ell(S)=\begin{cases} i+j+2 & \text{if } i+j+2\le n\le 2(i+j-2)\ , \\ 0 & \text{if } n\ge 2(i+j)\end{cases}$$

V. Embedding of an (i,j)-regular signed graphs in $(i+1,j+1)$-regular signed graphs

Lemma 20. *A graph G is 2 -factorable if and only if G is r-regular for some positive even integer r.*

Theorem 21. *For an (i,j)-regular signed graph $S=(S^u,\sigma)$ of order n, where both n and $i+j$ are of opposite parity, when n is even and $i+j$ is odd, then*

$$\ell(S)=\begin{cases} i+j+1 & \text{if } n=i+j+1\ , \\ 0 & \text{otherwise}\end{cases}$$

when n is odd and $i+j$ is even, then

$$\ell(S)=\begin{cases} i+j+1 & \text{if } n=i+j+1\ , \\ 1 & \text{otherwise}\end{cases}$$

Proof. Case 1: When n is even and $i+j$ is odd.

If $n=i+j+1$, then $\overline{S^u}$ is a null graph and has no 1-factor. Now we construct an $(i,j-1)$-regular complete signed graph $K_{i+j}^{(i,j-1)}$. To obtain an $(i+1,j+1)$-regular signed graph n should be at least $2(i+j)$, a contradiction to assumption, therefore

$$\ell(S)\ge i+j+1. \tag{1}$$

Further, we construct an (i,j)-regular complete signed graph $K_{i+j+1}^{(i,j)}$ and join the vertex v of S with two vertices of $K_{i+j+1}^{(i,j)}$ in such a way that one is joined by a positive edge to a vertex v of S and the other one is joined by a negative edge to a vertex v of S. Thus, the resulting signed graph is $(i+1,j+1)$-regular. Therefore

$$\ell(S)\le i+j+1. \tag{2}$$

By equations (1) and (2), we get

$$\ell(S)=i+j+1.$$

If $n\neq i+j+1$, then by lemma 20 $\overline{S^u}$ has 2-factor, since n is even and $i+j$ is odd, therefore $n-(i+j+1)$ is even and hence $\overline{S^u}$ is $n-(i+j+1)$-regular.

Next, we select edges of 2-factor in $\overline{S^u}$ and make them adjacent with alternative positive and negative edge in S. Then the resulting graph is $(i+1,j+1)$-regular. Hence

$$\ell(S)=0.$$

Case 2: When n is odd and $i+j$ is even. If $n=i+j+1$, then $\overline{S^u}$ is a null graph and has no 1-factor. Now we construct an $(i,j-1)$-regular complete signed graph $K_{i+j}^{(i,j-1)}$. To obtain an $(i+1,j+1)$-regular signed graph n should be at least $2(i+j)$, a contradiction to assumption, therefore

$$\ell(S)\ge i+j+1. \tag{3}$$

Next, we construct an (i,j)-regular complete signed graph $K_{i+j+1}^{(i,j)}$ and join the vertex v of S with two vertices of $K_{i+j+1}^{(i,j)}$ in such a way that one is joined by a positive edge to a vertex v of S and the other one is joined by a negative edge to a vertex v of S, so that the degree of every vertex in S and $K_{i+j+1}^{(i,j)}$ is increased by one positive degree and one negative degree. Hence the resulting graph obtained in such a manner is $(i+1,j+1)$-regular. Therefore

$$\ell(S)\le i+j+1. \tag{4}$$

By equations (3) and (4), we get

$$\ell(S)=i+j+1.$$

If $n\neq i+j+1$, then $\overline{S^u}$ has 2-factor. Now select $\frac{n-i-1}{2}$ alternate edges of 2-factor in $\overline{S^u}$ and join them in S with positive edge and then select another $\frac{n-j-1}{2}$ alternate edges of 2-factor with no common vertices to the above selected vertices except one and join them in S with negative edge. Now introduce a vertex w and join $i+1$ vertices of positive degree i in S to w with positive edge and $j+1$-vertices of negative degree j in S to w with negative edge. The resulting graph is $(i+1,j+1)$-regular. Hence

$$\ell(S)=1.$$

□

Theorem 22. *For an (i,j)-regular signed graph $S=(S^u,\sigma)$ of order n, where both n and $i+j$ are of same parity, then*

$$\ell(S)=\begin{cases} i+j+2 & \text{if } i+j+2\le n\le 2(i+j-2)\ , \\ 0 & \text{if } n\ge 2(i+j)\end{cases}$$

VI. Conclusion

In graph theory, it is well known that every graph can be embedded as an induced subgraph of a regular graph (*see* [5]; Theorem 1.4, p.16). Acharya [2] raised a question that can this result be lifted to regular signed graphs by solving the following conjecture? In this regard we have tried to exhaust all the cases of embedding signed regular graph into a signed regular graph.

There is a much older notion of regularity in signed graphs proposed by Chartrand, *et al*. (*see* [7]), viz., a signed graph S is *regular* (we call it $net-regular$) if the $net-degree$ $d_S(u)=d_S^+(u)-d_S^-(u)$ of u is independent of the choice of $u\in V(S)$. Acharya has conjectured that every signed graph can be embedded as an induced sub-signed graph of a regular signed graph. However, this problem seems to be related to the analogous problem for regular signed graphs, for if S is an (i,j)-regular signed graph then it is also $i-j$-net-regular; note that the converse of this statement is not true in general. That is, net-regularity of a signed graph according to Chartrand *et al*. is a generalization of regularity of a signed graph.

ACKNOWLEDGEMENT

The authors are very much thankful to late Dr. B.D. Acharya for encouraging them to obtain more results on this idea of regularity of signed graphs.

REFERENCES

[1] B.D. Acharya, *Construction of certain infinite families of graceful graphs from a given graceful graph*, Def. Sci. J., 32(3)(1982), 231-236.
[2] B.D. Acharya, *Personal communication*, 2012.
[3] B. Mohar, *Embedding graphs in an arbitrary surface in linear time*, *In Proc.* 28th Ann. ACM STOC, Philadelphia, ACM Press, 1996, 392397.
[4] D. Archdeacon, *The complexity of the graph embedding problem, in Topics in Combinatorics and Graph Theory*, R. Bodendiek and R. Henn (Eds.), Physica-Verlag, Heidelberg, 1990, 5964.
[5] M. Behzad and G. Chartrand, *Introduction to the Theory of Graphs*, Allyn and Bacon, Inc., Boston, 1971.
[6] B. Bollobas and S.E. Eldridge, *Maximal matchings in graphs with given minimal and maximal degrees*, Math. Proc. Cambridge Philos. Soc., 79(1976), 221-234.
[7] G. Chartrand, H. Gavlas, F. Harary and M. Schultz, *On signed degrees in signed graphs*, Czechoslovak Math. J., 44(1994), 677-690.
[8] A. Gardiner, *Embedding* k−*regular graphs in* k+1−*regular graphs*, J. London Math. Soc., (2)28(1983), 393-400
[9] F. Harary, *Graph Theory*, Addison-Wesley Publ. Comp., Reading, Massachusetts 1969.
[10] M. Juvan, B. Mohar, *A linear time algorithm for the* 2 − *M obius band embedding extension problem*, SIAM J. Discrete Math., (1997).
[11] Ken-ichi Kawarabayashi, Bojan Mohar, Bruce Reed, *A simpler linear time algorithm for embedding graphs into an arbitrary surface and the genus of graphs of bounded tree− width*, 0272-5428/08 2008 IEEE DOI 10.1109/FOCS.2008.53.
[12] A. Kotzig, 1−*factorization of cartesian product of regular graphs*, J. Graph Theory 43(1979), 23-24.
[13] V. Mishra, Graphs associated with $(0,1)$ and $(0,1,-1)$ matrices, Ph.D. Thesis, IIT Bomby, India, 1974.
[14] Bojan Mohar, *A linear time algorithm for embedding of graphs in an arbitrary surface*, SIAM J. Discrete Math.,12 (1999) 626.
[15] Bojan Mohar, *Obstructions for the disk and the cylinder embedding extension problems*, Comb.Probab. Comput., 3 (1994) 375-406.
[16] Jessica Pereira, T. Singh and S. Arumugam, *Graceful embedding of graphs with property P*, submitted 2015.
[17] Jessica Periera, T. Singh and S. Arumugam, *Graceful embedding of sigraphs with graph property P*, submitted 2015.
[18] T. Singh and Mukti Acharya, *Construction of graceful signed graphs*, Defense Science Journal, 56 (5) (2006), 801-808.
[19] T. Singh and Mukti Acharya, *Embedding of Sigraphs in Graceful Sigraphs*, ARS Combinatoria, 111(2013), 421-426.
[20] D.B. West, *Introduction to Graph Theory*, Prentice-Hall of India Pvt. Ltd. 1996.
[21] T. Zaslavsky, *Signed graphs*, Discrete Math., 4(1982), 47-74.
[22] T. Zaslavsky, *Glossary of signed and gain graphs and alliedareas*, II Edition, Electron. J. Combin., #DS9(1998).
[23] T. Zaslavsky, *A mathematical bibliography of signed and gain graphs and allied areas*, VIII Edition, Electron. J. Combin., #DS8(2009).
[24] T. Zaslavsky, *Orientation embedding of signed graphs*, Journal of Graph Theory, 16(5)(1992), 399-422.
[25] T. Zaslavsky, *Matrices in the theory of signed simple graphs*, Advances in Discrete Mathematics and Applications (Proceedings of the International Conference on Discrete Mathematics-2008, Mysore, India), eds.: B.D. Acharya, G.O.H. Katona and J. Nesetril, Ramanujan Mathematical Society Lecture Notes Series, 13(2010), 207-229.

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

On 2-path signed graphs

Deepa Sinha
Department of Mathematics
South Asian University,
Akbar Bhawan New Delhi
Email:deepa_sinha2001@yahoo.com

Deepakshi Sharma
Department of Mathematics
South Asian University,
Akbar Bhawan New Delhi
Email: deepakshi.sharma1990@gmail.com

Abstract—A signed graph (sigraph) is a graph where each edge receives positive or negative sign. In this paper, we report our investigation on a new signed graph $(S)_2$ called 2-path sigraphs of a given sigraph S defined as that sigraph whose vertex set is the vertex set of S and two vertices in $(S)_2$ are adjacent if there exist a path of length two between them in S and negatively signed if all the paths of length two between them are all-negative, otherwise positively signed.

Keywords— Sigraph; marked sigraph; 2-path sigraph; open neighborhood; balance; sign-compatibility; signed-regularity; clusterability.

I. Introduction

For standard terminology and notation in graph theory one can refer Harary [4] and West [6] and for signed graphs literature one can refer (Zaslavsky [7]). Throughout the text, we consider finite, undirected graph with no loops or multiple edges. A *signed graph* (*sigraph*) is an ordered pair $S = (S^u, \sigma)$, where S^u is a graph $G = (V,E)$, called the *underlying graph* of S and $\sigma : E \to \{+,-\}$ is a function from the edge set E of S^u into the set $\{+,-\}$, called the signature (or *sign* in short) of S. A sigraph is $all-positive$ (respectively, $all-negative$) if all its edges are positive (negative). Further, it is said to be *homogeneous* if it is either all-positive or all-negative and *heterogeneous* otherwise.

A sigraph S is said to be $signed-regular$ if the number of positive edges (negative edges), $d^+(v)(d^-(v))$ incident at a vertex v in S, is independent of the choice of v in S, that is S is (i,j)-signed-regular, where $i = d^+(v)$ is the positive degree of every vertex v in S and $j = d^-(v)$ is the negative degree of every vertex v in S.

By a *negative section* of a cycle or a path Z we mean a maximal set D of vertices of Z such that the subsigraph consisting of the edges of Z joining vertices in D is all-negative and connected. A *marked sigraph* is an ordered pair $S_\mu = (S, \mu)$ where $S = (S^u, \sigma)$ is a sigraph and $\mu : V(S^u) \to \{+,-\}$ is a function from the vertex set $V(S^u)$ of S^u into the set $\{+,-\}$, called a *marking* of S. We define, $(V(S))_\mu = \{v_i^+, v_i^- : \forall\ v_i \in V(S)\}$. Let v be an arbitrary vertex of a graph S. We denote the set consisting of all the vertices of S adjacent with v by $N(v)$. This set is called the *neighborhood set* of v and sometimes we call it as *neighborhood* of v. Next we define two marked neighborhoods as: $N_*(t) = \{v^\mu \in (V(S))_\mu : tv$ is an edge with sign $\mu\}$, $N_*{}^-(t) = \{v^- \in (V(S))_\mu : tv$ is an edge$\}$. For each $N(t)$ of S^u, there exist $N_*(t)$ in S and vice versa.

The *clique* of a graph S^u is a subset of vertices such that every two vertices in the subset are connected by an edge. $\delta(N(t))$ is a clique generated by vertices in $N(t)$. A *cycle* in a sigraph S is said to be *positive* if the product of the signs of its edges is positive or, equivalently, if the number of negative edges in it is even. A cycle which is not positive is said to be *negative*. A sigraph is *balanced* if all its cycles are positive. A sigraph is said to be *clusterable* if its vertex set can be partitioned into pairwise disjoint subsets, called *clusters*, such that every negative edge joins vertices in different clusters and every positive edge joins vertices in the same clusters.

Property 1. *A 2-subset $\{v_i, v_j\}$ in a neighborhood of a vertex v_k has property P if $\{v_i{}^-, v_j{}^-\} \subset N_*(v_k)$ for some i, j, k then for each $N(t)$ containing v_i, v_j, $\{v_i{}^-, v_j{}^-\} \subset N_*(t)$.*

The $2-path\ sigraph$[3] $(S)_2 = (V, E', \sigma')$ of a sigraph $S = (V,E,\sigma)$ is defined as follows: The vertex set is same as the original sigraph S and two vertices $u, v \in V((S)_2)$, are adjacent if and only if there exist a path of length two in S. The edge $uv \in V((S)_2)$ is negative if and only if all the edges in all the two paths in S are negative otherwise the edge is positive. The 2-subset $\{v_i, v_j\}$ having property P are named as P *pairs* and the set of all P pairs is denoted by P_*. A negative section $v_1, v_2, \ldots, v_k$ in cycle or path is said to be P *section* if every alternate vertices forms P pair. If $v_k = v_1$ in the P section then such a P section is called P *cycle*.

II. 2-path sigraphs

In this section we give a characterization of 2-path sigraph to detect whether a given sigraph is a 2-path sigraph of some sigraph and find its underlying sigraph. Following characterization of 2-path graphs was given by Acharya and Vartak.

Lemma 2. *[1] A connected graph G with vertices v_i, $i = 1,\ldots,n$ is a 2-path graph of some graph H if, and only if, G contains a collection of complete subgraphs $G_1, G_2, \ldots, G_n$ such that for each $i, j = 1,\ldots,n$, the following hold: (i)$v_i \notin G_i$, (ii) $v_i \in G_j \Leftrightarrow\ v_j \in G_i$, (iii) $v_i v_j \in G$ there exists G_k containing $v_i v_j$.*

Theorem 3. *A connected sigraph S with vertices v_i, $i = 1,\ldots,n$ is a 2-path sigraph of some sigraph S' if, and only if, S contains a collection of complete subsigraphs $S_1, S_2, \ldots, S_n$ with marked vertices $v_i{}^\mu$, $\mu \in \{+,-\}$ such that for each $i, j = 1,\ldots,n$, the following hold:*

(i) $v_i{}^\mu \notin S_i$,

(ii) $v_i{}^{\mu_1} \in S_j \Leftrightarrow v_j{}^{\mu_2} \in S_i, i \neq j, \mu_1 = \mu_2$,

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

(*iii*) *$v_iv_j \in E(S)$ with sign σ then there exists S_k containing $v_i^{\mu_i}, v_j^{\mu_j}$ where $\mu_i, \mu_j \in \{+,-\}$ and if $\sigma(v_i,v_j) = -$ then $\{v_i, v_j\}$ is a P pair in S_k.*

Proof: Necessity: Suppose S is a 2-path sigraph of sigraph S' with vertices $v_1, v_2, ..., v_n$. To show that there exist a collection of n complete subsigraphs such that (i), (ii) and (iii) hold. Consider the marked neighborhood $N_*(v_i)$ of each vertex v_i, $i = 1$ to n in S'. By definition, each $N_*(v_i)$, $i = 1$ to n, generates a complete subsigraph in 2-path sigraph S of S'. Next, $v_i^{\mu} \notin N_*(v_i)$ since sigraph S' is simple. Let $v_i^{\mu} \in N_*(v_j)$ for some i, j $i \neq j, 1 \leq i, j \leq n$. Then v_iv_j is an edge in S' with sign μ. Clearly, $v_j^{\mu} \in N_*(v_i)$. If v_iv_j is an edge in 2-path sigraph S and thus in S^u then $v_i, v_j \in N(v_k)$ for some k. Let $v_i, v_j \in S$ with a sign '$-$', then v_iv_k and v_kv_j are negative edges in S'. Thus $\{v_i^-, v_j^-\} \in N_*(v_k)$ and hence $\{v_i, v_j\}$ is P pair.

Sufficiency: Let S contains a collection of complete subsigraphs $S_1, S_2, ..., S_n$ satisfying the three properties $(i), (ii)$ and (iii). Associate a vertex v_i with each set S_i generating complete subsigraph such that $v_i \notin S_i$ and join each v_i to the vertices of S_i, the obtained sigraph S' (say) will be such that $(S')_2 \cong S$.

■

Observation 4. *Neighborhood of each vertex of S generates a clique in $(S)_2$, such that the union of such cliques of S gives $(S)_2$.*

Observation 5. *The cycle of order n, $n \neq 3$ in $(S)_2$ for a given sigraph S is due to a cycle in S.*

Observation 6. *The cycle of order three in $(S)_2$ of a sigraph S is either due to*

- *a cycle of length three or*
- *$K_{1,3}$ in S or*
- *a cycle of order six or*
- *due to clique generated by neighborhood of a vertex in S.*

III. Balanced 2-path sigraph

In this section we provide with a characterization of balanced 2-path sigraphs. We refer a following lemma given by Zaslvasky.

Lemma 7. *[7] A sigraph in which every chordless cycle is positive, is balanced.*

Theorem 8. *For a sigraph S of order n, the following statements are equivalent:*

(i) *$(S)_2$ is balanced.*

(ii) *For all sequences of vertices $v_1, v_2, ..., v_r$; $1 \leq r \leq n$ in S such that $v_1^{\mu_1}, v_2^{\mu_2} \in N_*(t_1)$; $v_2^{\mu_2}, v_3^{\mu_3} \in N_*(t_2)$; .... ; $v_1^{\mu_1}, v_r^{\mu_r} \in N_*(t_r)$ for some $t_1, t_2, ..., t_r \in V(S)$ and $\mu_1, \mu_2, ..., \mu_r \in \{+,-\}$, $v_i^{\mu_i}, v_{i+1}^{\mu_{i+1}} \in N_*(t_i)$; $1 \leq i \leq r$ having property **P** are even in each sequence.*

(iii) *(a) Each homogeneous cycle in S is either positive or is a **P** cycle of length 4k, for some positive integer k. (b) Each heterogeneous odd cycle in S does not contain **P** section of even length and even cycle do not contain **P** section.*

*(c) For each vertex u in S with $d(u) \geq 3$, $N_*(u)$ contains even number of **P** pairs.*

IV. Clusterability of 2-path sigraphs

In this section, we discuss the clusterablity of a 2-path sigraph.

Lemma 9. *[2] A sigraph S is clusterable if and only if S contains no cycle with exactly one negative edge.*

Theorem 10. *For a given sigraph S of order n, following conditions are equivalent:*

(i) *$(S)_2$ is clusterable.*

(ii) *if for all sequence of vertices $x_1, x_2, ..., x_r$; $1 \leq r \leq n$ in S such that $x_1, x_2 \in N(t_1)$; $x_2, x_3 \in N(t_2)$; ... ; $x_1, x_r \in N(t_r)$ for some $t_1, t_2, ..., t_r \in V(S)$ if there exist a pair of vertices in sequence $x_i, x_{i+1} \in N(t_i)$ having property P then the sequence has atleast one pair of vertices $x_l, x_{l+1} \in N(t_l)$ satisfying property P, for some l, $1 \leq l \leq r$.*

(iii) (a) *If S contains a heterogeneous cycle then no even cycle in S contains a P section of length < 5 and no odd cycle contains a P section of length 3.*
(b) *No neighborhood of vertex v in S with $d(v) \geq 3$ contains a single P pair.*

Proof: $(i) \Rightarrow (ii)$. Let for a given sigraph S, $(S)_2$ be clusterable. Then no cycle in $(S)_2$ has exactly one negative edge. Let $x_1, x_2, ..., x_r$; $1 \leq r \leq n$ be a sequence of vertices in S such that $x_1, x_2 \in N(t_1)$; $x_2, x_3 \in N(t_2)$; ... ; $x_1, x_r \in N(t_r)$ for some $t_1, t_2, ..., t_r \in V(S)$. Let $x_i, x_{i+1} \in N(t_i)$ be a pair in sequence $x_1, x_2, ..., x_r$; $1 \leq r \leq n$ having property P. Clearly x_i, x_{i+1} will form an negative edge in $(S)_2$. Thus there is atleast one pair of vertices $x_l, x_{l+1} \in N(t_l)$ satisfying property P.

$(ii) \Rightarrow (i)$. Let for all sequences of vertices $x_1, x_2, ..., x_r$; $1 \leq r \leq n$ in S such that $x_1, x_2 \in N(t_1)$; $x_2, x_3 \in N(t_2)$; ... ; $x_1, x_r \in N(t_r)$ for some $t_1, t_2, ..., t_r \in V(S)$ if there exist a pair of vertices in sequence $x_i, x_{i+1} \in N(t_i)$ having property P then the sequence has atleast one pair of vertices $x_l, x_{l+1} \in N(t_l)$ satisfying property P. This sequence of vertices generates cycles in $(S)_2$ such that no cycle has exactly one negative edge. Thus by Lemma 9, $(S)_2$ is clusterable.

$(i) \Rightarrow (iii)$. Let us assume that $(S)_2$ is clusterable. To prove that (a) and (b) hold. Let if possible (a) does not hold. That is there exist a even heterogeneous cycle in S with P section of length < 5. Then if cycle is of length 2k, then it will be a cycle of length k in $(S)_2$, and clearly one of the cycle will contain exactly one negative edge, which is a contradiction to hypothesis by Lemma 9. Next if there exist an odd heterogeneous cycle with P section of length 3, then it will correspond to a single negative edge in $(S)_2$, which is again not possible.

Next let a neighbourhood of vertex v in S with $d(v) \geq 3$, contains a single P pair. Since each neighborhood $N(v)$ gives rise to $\delta(N(v))$ clique generated by $N(v)$, and the cycle so formed will contain exactly one negative edge in $(S)_2$, which is a contradiction to the hypothesis.

$(iii) \Rightarrow (i)$. By (a) and (b) it is clear that no cycle in $(S)_2$ contains a single negative edge, hence by Lemma 9 $(S)_2$ is clusterable. ■

V. SIGNED-REGULARITY

In this section, we establish a characterization of a signed-regular 2-path sigraph.

Theorem 11. *For a sigraph S, $(S)_2$ is signed-regular if and only if*

(i) *For all set if vertices $v_i, v_j \in V(S)$ such that $v_i \in N(v_j)$, $|\cup N(v_j) - \{v_i\}|$ is identical $\forall v_i \in V(S)$.*

(ii) *If P_* is collection of all pairs of vertices satisfying property P then $\sum |v_i| \in P_*$ is identical.*

Proof: Necessity: Let for a given sigraph S, $(S)_2$ is signed-regular. Then number of edges incident to each vertex in $(S)_2$ is identical. The cycle of order $n, n \neq 3$ in $(S)_2$ for a given sigraph S is due to a cycle in S. Also neighborhood of each vertex of S generates a clique in $(S)_2$, such that the union of such cliques of S gives $(S)_2$, thus (i) holds. Since P_* consists of all pair of vertices satisfying property P then as $(S)_2$ is signed-regular, number of negative edges incident to each vertex in $(S)_2$ must be equal for all vertices. Thus each vertex v_i appears in P same number of times. Therefore, $\sum |v_i|$ belonging to P_* is identical.

Sufficiency: Let (i) and (ii) hold for the given sigraph S. Then by definition of 2-path sigraphs and neighborhood of a given sigraph S, $(S)_2$ is signed regular. ■

VI. SIGN-COMPATIBILITY

This section deals with sign-compatibility of a 2-path sigraph of a given sigraph.

Theorem 12. *[5] A sigraph S is sign-compatible if and only if S does not contain a subsigraph isomorphic to either of the two sigraphs, S_1 formed by taking the path $P_4 = (x,u,v,y)$ with both the edges xu and vy negative and the edge uv positive and S_2 formed by taking S_1 and identifying the vertices x and y .*

Theorem 13. *For a given sigraph S, $(S)_2$ is sign-compatible if and only if for no sequence of vertices x_1, x_2, x_3, x_4 in S such that $x_1, x_2 \in N(t_1)$; $x_2, x_3 \in N(t_2)$; $x_3, x_4 \in N(t_3)$ for some $t_1, t_2, t_3 \in V(S)$ has x_1, x_2 and x_3, x_4 as P pairs and $x_2, x_3 \notin P$.*

Proof: Necessity: Let for a given sigraph S, $(S)_2$ be sign-compatible. Then by Theorem 12 $(S)_2$ will not be a subsigraph which is a path, since such a path in $(S)_2$ is given by a sequence x_1, x_2, x_3, x_4 in S so that $x_1, x_2 \in N(t_1)$; $x_2, x_3 \in N(t_2)$; $x_3, x_4 \in N(t_3)$ for some $t_1, t_2, t_3 \in V(S)$. Hence for no sequence of vertices x_1, x_2, x_3, x_4 in S such that $x_1, x_2 \in N(t_1)$; $x_2, x_3 \in N(t_2)$; $x_3, x_4 \in N(t_3)$ for some $t_1, t_2, t_3 \in V(S)$ has x_1, x_2 and x_3, x_4 as P pairs.

Sufficiency: Let for no sequence of vertices x_1, x_2, x_3, x_4 in S such that $x_1, x_2 \in N(t_1)$; $x_2, x_3 \in N(t_2)$; $x_3, x_4 \in N(t_3)$ for some $t_1, t_2, t_3 \in V(S)$ has x_1, x_2 and x_3, x_4 as P pairs. Thus $(S)_2$ does not contain forbidden subsigraphs mention in Theorem 12. Hence by Theorem 12 $(S)_2$ is sign-compatible. ■

REFERENCES

[1] B.D. Acharya and M.N. Vartak, Open Neighbourhood Graphs, Research report of the Indian Institute of Technology, Bombay, India, 1973.

[2] J.A. Davis, Clustering and structural balance in graphs, Human Relations, *20*(1967), 181-187.

[3] M.K. Gill, Contribution to some topics in graph theory and its applications, Ph.D. Thesis, Indian Institute of Technology, Bombay, 1983.

[4] F. Harary, Graph Theory, Addison-Wesley Publ. Comp., Reading, Massachusetts, 1969.

[5] D. Sinha and A. Dhama, Sign-Compatibility of some derived signed graphs, Indian Journal of Mathematics, *55*(2013), 23-40.

[6] D.B. West, Introduction to Graph Theory, Prentice-Hall of India Pvt. Ltd., 1996.

[7] T. Zaslavsky, A mathematical bibliography of signed and gain graphs and allied areas, Electron. J. Combin., VIII Edition,1998, 289.

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

Stability analysis of Differential Evolution

Anshul Gopal
Department of Mathematics
South Asian University
New Delhi, India
Email: anshul.gopal@gmail.com

Jagdish Chand Bansal
Department of Mathematics
South Asian University
New Delhi, India
Email: jcbansal@gmail.com

***Abstract*—The stability analysis of evolutionary algorithms is considered to be an important research topic. DE is one of the most popular and efficient, continuous optimization algorithm. In this paper, the stability analysis for DE has been carried out. As outcome of this stability analysis, DE parameters' optimal choice has been suggested.**

Keywords: Differential Evolution (DE); Stability; Finite Difference Scheme.

I. Introduction

Differential Evolution (DE) algorithm is presently among the highly rated population based stochastic algorithm. Its simplified and efficient approach enables it in solving problems pertaining to multi objective, multi-modal, dynamics and constrained optimization problem. Since the inception of first variant of DE algorithm by R.Storm and K.Price in 1995 [5], quite many variants have been proposed so far. These variants are developed by hybridizing the original DE with other optimization algorithms, by using self adaptive schemes, by manipulating the equations at Mutation phase of the algorithm and many more. Surveys containing current status of research in the field of DE were recently published [1], [3]. In recent research each parameter vector of DE algorithm is modelled as a search-agent moving under the directives of the DE algorithm, over the fitness landscape in continuous time searching for the optima. Based on the proposed model, the stability and convergence of the DE-vectors in a small neighbourhood centered on an isolated equilibrium point, has been investigated with Lyapunov's stability theorem [2].
This paper presents the stability analysis of two strategies of DE, namely DE/rand/1 and DE/target-to-best/1. The optimal choice of scaling factor F has also been recommended based on the von Neumann stability analysis of finite difference scheme [6].

II. DE Algorithm

Differential Evolution (DE) algorithm is a branch of evolutionary programming for optimizing problems over continuous domains. Differential Evolution is a design tool of great utility that is immediately accessible for practical applications. DE has various strategies depending upon the selection of target vector, count of difference vectors used and the variety of crossover. Similar to other population based search algorithms, in DE a population of potential individuals searches the solution. In a N-dimensional search space, an individual is represented by a N-dimensional vector $(x_{i_1}, x_{i_2}, ..., x_{i_N})$, i = 1, 2, . . ., NP where, NP is the population size.
The DE algorithm have three operators: Mutation, Crossover and Selection operator. Initially using uniform distribution a population is generated randomly. Trial vector generation plays a pivotal role in the algorithm. To generate the trial vectors we use mutation and crossover operators whereas selection operator is used to select the best trial vector for the later generation. DE operators are explained briefly in the following subsections.

A. Mutation Operator

For each individual of the current population a trial vector is generated using the mutation operator. A target vector is mutated with a weighted differential for the generation of the trial vector. Using the newly produced trial vector an offspring is produced in the crossover operations. If G is the index for generation counter, the mutation operator for generating a trial vector $v_i(G)$ from the parent vector $x_i(G)$ is defined as follows:

- A target vector $x_{i_1}(G)$ is selected from the population such that $i \neq i_1$.
- From the population, two individuals x_{i_2} and x_{i_3} are selected randomly such that i, i_1, i_2, i_3 are distinct from each other.
- For calculating the trial vector the target vector is mutated as follows :

$$\vec{v_i}(G) = \vec{x_{i_1}}(G) + F(\vec{x_{i_2}}(G) - \vec{x_{i_3}}(G))$$

where, $F \in [0, 1]$ is the mutation scale factor which is used in controlling the amplification of the differential variation.

B. Crossover Operator

Using the crossover of parent vector $x_i(G)$ and the trial vector $v_i(G)$ the offspring is generated as follows :

$$x'_{i,j}(G) = \left\{ \begin{array}{ll} v_{i,j}(G), & \text{if } j \in J \\ x_{i,j}(G), & \text{otherwise} \end{array} \right\}$$

where, J is the set of cross over points, $x_{i,j}(G)$ is the j^{th} element of the vector $x_i(G)$.

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

Binomial crossover and exponential crossover are used most frequently to determine the set J. In this algorithm, CR is the probability that the considered crossover point will be included. Larger the value of CR, more the crossover points will be selected. Here, J is a set of crossover points, CR is crossover probability, U(1, N) is a uniformly distributed random integer between 1 and N.

C. Selection Operator

There are two functions of the selection operator: First it selects the individual for the mutation operation to generate the trial vector and second, it selects the best, between the parent and the offspring based on their fitness value for the next generation. If fitness of parent is greater than the offspring then parent is selected otherwise offspring is selected:

$$x_i(G+1) = \left\{ \begin{array}{ll} x_i^{'}(G), & \text{if } f(x_i^{'}(G)) > f(x_i(G)). \\ x_i(G), & \text{otherwise} \end{array} \right\}$$

This helps in controlling the deterioration of average fitness of the population [4].

III. Stability Analysis of DE Algorithm

Consider the two mutation strategies namely DE/rand/1 and DE/target-to-best/1 :

DE/rand/1 (Strategy 1) : $\vec{V_i} = \vec{X_{r_1}} + F(\vec{X_{r_2}} - \vec{X_{r_3}})$
where, $\vec{V_i}$ is the donor vector, $\vec{X_{r_j}}$; j = 1, 2, 3 are randomly selected target vectors from the population and F is the scaling factor.

DE/target-to-best/1 (Strategy 2) : $\vec{V_i} = \vec{X_i} + F_1(\vec{X_{best}} - \vec{X_i}) + F_2(\vec{X_{r_1}} - \vec{X_{r_2}})$
where, $\vec{V_i}$ is the donor vector, $\vec{X_i}$ is the target vector, $\vec{X_{best}}$ is the best target vector from the population, $\vec{X_{r_1}}$ and $\vec{X_{r_2}}$ are the randomly selected target vectors from the population, F_1 and F_2 are the scaling factors.

Since, right hand side of **Strategy 1** is a vector addition of a randomly selected solution vector $\vec{X_{r_1}}$ and a vector whose initial and terminal points are $\vec{X_{r_3}}$ and $\vec{X_{r_2}}$ respectively. Also, $\vec{X_{r_3}}$ and $\vec{X_{r_2}}$ are randomly selected individuals from the current population, without loss of generality we can assume F to be always non-negative.
Similarly, in case of **Strategy 2** we can consider both F_1 and F_2 to be non-negative always.

First, consider the **Strategy 1** of DE algorithm in the following form:

$$x_k(t+1) = wx_{r_1}(t) + F(x_{r_2}(t) - x_{r_3}(t)) \quad (1)$$

where, w is the introduced inertial weight (originally w = 1), k and t represents population index and time respectively. r_1, r_2 and r_3 are randomly generated numbers.
Without loss of generality, we can take $r_1 = k \pm a$, $r_2 = k \pm b$, $r_3 = k \pm c$ where a, b and c are randomly generated integers. Also, $1 \leq r_1$, r_2 , $r_3 \leq$ NP

Hence,

$$x_k(t+1) = wx_{k\pm a}(t) + F(x_{k\pm b}(t) - x_{k\pm c}(t)) \quad (2)$$

If the actual solution to a problem in an k-t computational domain is given by $x = x(k,t)$, the approximation in the nodes of a computational grid will be given by $x_{i,j} = x(k_i, t_j)$. Let the m^{th} component of the complex Fourier series solution to the given equation be given by :

$$x_m(k,t) = A_m e^{\iota(\sigma_m k - \beta_m t)} \quad (3)$$

where, $\iota = \sqrt{-1}$, A_m is the amplitude of the m^{th} component, β_m is the angular frequency of the m^{th} component and σ_m represents the wave number of the m^{th} component.

In terms of grid points (i, j), equation (2) can be written as

$$x_{i,j+1} = wx_{i\pm a,j} + F(x_{i\pm b,j} - x_{i\pm c,j}) \quad (4)$$

Again, the m^{th} component of the solution of the equation (4) at point (i, j) is given by

$$x_{m_{i,j}} = A_m e^{-\iota\beta_m j\Delta t} e^{\iota\sigma_m i\Delta k} \quad (5)$$

where, $k_i = i\Delta k$ and $t_j = j\Delta t$
by putting the value of $x_{m_{i,j}}$ in equation (4), we get the Amplification factor as :

$$e^{-\iota\beta_m\Delta t} = we^{\pm\iota\sigma_m a\Delta k} + F(e^{\pm\iota\sigma_m b\Delta k} - e^{\pm\iota\sigma_m c\Delta k}) \quad (6)$$

For marginal stability, the magnitude of Amplification factor must be equal to unity [6].

$$\text{i.e. } |e^{-\iota\beta_m\Delta t}| = 1$$

Now,
$|e^{-\iota\beta_m\Delta t}| = |we^{\pm\iota\sigma_m a\Delta k} + F(e^{\pm\iota\sigma_m b\Delta k} - e^{\pm\iota\sigma_m c\Delta k})|$
or $|e^{-\iota\beta_m\Delta t}| \leq |we^{\pm\iota\sigma_m a\Delta k}| + |F(e^{\pm\iota\sigma_m b\Delta k} - e^{\pm\iota\sigma_m c\Delta k})|$
or $|e^{-\iota\beta_m\Delta t}| \leq |w| + 2|F|$

The algorithm will be marginally stable when $|e^{-\iota\beta_m\Delta t}| = |w| + 2|F| = 1$
i.e., $|w| = 1 - 2|F|$
Now, $|w| \geq 0$, i.e. $1 - 2|F| \geq 0$
or $|F| \leq 1/2$
or $F \in [-1/2, 1/2]$
In other words, algorithm will be marginally stable when $F \in [-1/2, 1/2]$. This recommendation matches with the

general setting of F also, i.e. F = 1/2.

Secondly, consider the **Strategy 2** of DE algorithm in the following form:

$$x_k(t+1) = wx_k(t)+F_1(x_{best}(t)-x_k(t))+F_2(x_{r1}(t)-x_{r2}(t)) \quad (7)$$

where, w is the introduced inertial weight (originally w = 1), k and t represents population index and time respectively. r_1 and r_2 are randomly generated numbers. F_1 and F_2 are scaling factors.

Without loss of generality we can take $best = k \pm a$, $r_1 = k \pm b$, $r_2 = k \pm c$ where a is fixed integer, b and c are randomly generated integers. Also, $1 \leq best$, r_1 , $r_2 \leq$ NP

Hence,

$$x_k(t+1) = wx_k(t)+F_1(x_{k\pm a}(t)-x_k(t))+F_2(x_{k\pm b}(t)-x_{k\pm c}(t)) \quad (8)$$

If the actual solution to a problem in an k-t computational domain is given by $x = x(k,t)$, the approximation in the nodes of a computational grid will be given by $x_{i,j} = x(k_i, t_j)$. Let the m^{th} component of the complex Fourier series solution to the given equation be given by

$$x_m(k,t) = A_m e^{\iota(\sigma_m k-\beta_m t)} \quad (9)$$

where, $\iota = \sqrt{-1}$, A_m is the amplitude of the m^{th} component, β_m is the angular frequency of the m^{th} component and σ_m represents the wave number of the m^{th} component.

In terms of grid points (i,j), equation (8) can be written as:

$$x_{i,j+1} = wx_{i,j}+F_1(x_{i\pm a,j}-x_{i,j})+F_2(x_{i\pm b,j}-x_{i\pm c,j}) \quad (10)$$

Again, the m^{th} component of the solution of the equation (10) at point (i,j) is given by

$$x_{m_{i,j}} = A_m e^{-\iota\beta_m j\Delta t} e^{\iota\sigma_m i\Delta k} \quad (11)$$

where, $k_i = i\Delta k$ and $t_j = j\Delta t$
by putting the value of $x_{m_{i,j}}$ in equation (10) we get the Amplification factor as

$$e^{-\iota\beta_m\Delta t} = w+F_1(e^{\pm\iota\sigma_m a\Delta k}-1)+F_2(e^{\pm\iota\sigma_m b\Delta k}-e^{\pm\iota\sigma_m c\Delta k}) \quad (12)$$

For marginal stability, the magnitude of Amplification factor must be equal to unity [6].

$$\text{i.e. } |e^{-\iota\beta_m\Delta t}| = 1$$

Now,
$|e^{-\iota\beta_m\Delta t}| = |w + F_1(e^{\pm\iota\sigma_m a\Delta k} - 1) + F_2(e^{\pm\iota\sigma_m b\Delta k} - e^{\pm\iota\sigma_m c\Delta k})|$
or $|e^{-\iota\beta_m\Delta t}| \leq |w| + |F_1(e^{\pm\iota\sigma_m a\Delta k} - 1)| + |F_2(e^{\pm\iota\sigma_m b\Delta k} - e^{\pm\iota\sigma_m c\Delta k})|$
or $|e^{-\iota\beta_m\Delta t}| \leq |w| + 2|F_1| + 2|F_2|$

The algorithm will be marginally stable when, $|e^{-\iota\beta_m\Delta t}| = |w| + 2|F_1| + 2|F_2| = 1$
i.e., $|w| = 1 - 2(|F_1| + |F_2|)$. Now, $|w| \geq 0$
i.e. $1 - 2(|F_1| + |F_2|) \geq 0$
or $|F_1| + |F_2| \leq 1/2$
or $(F_1 + F_2) \leq 1/2$ (as previously discussed)
or $(F_1 + F_2) \in [-1/2, 1/2]$
In other words, algorithm will be marginally stable when $(F_1 + F_2) \in [-1/2, 1/2]$.

Finally, we see that for **Strategy 1** the algorithm is marginally stable when the scaling factor F lies in the interval [0,1/2], considering that $0 \leq |w| \leq 1$.
In case of **Strategy 2**, if we consider $0 \leq |w| \leq 1$, the sum of scaling factor i.e. $(F_1 + F_2)$ should be in the range [0,1/2].

IV. Conclusion

In order to find the suitable value of scaling factor F in Differential Evolution (DE) algorithm, stability analysis of DE has been carried out using von Neumann stability analysis of finite difference scheme. Two strategies DE/rand/1 and DE/target-to-best/1 were considered. It is found that the algorithm is stable when the sum of scaling parameters is within the range [0,1/2].

References

[1] Swagatam Das and Ponnuthurai Nagaratnam Suganthan. Differential evolution: a survey of the state-of-the-art. *IEEE transactions on evolutionary computation*, 15(1):4–31, 2011.

[2] Sambarta Dasgupta, Swagatam Das, Arijit Biswas, and Ajith Abraham. On stability and convergence of the population-dynamics in differential evolution. *Ai Communications*, 22(1):1–20, 2009.

[3] Ferrante Neri and Ville Tirronen. Recent advances in differential evolution: a survey and experimental analysis. *Artificial Intelligence Review*, 33(1-2):61–106, 2010.

[4] Harish Sharma, Jagdish Chand Bansal, and KV Arya. Fitness based differential evolution. *Memetic Computing*, 4(4):303–316, 2012.

[5] Rainer Storn and Kenneth Price. *Differential evolution-a simple and efficient adaptive scheme for global optimization over continuous spaces*, volume 3. ICSI Berkeley, 1995.

[6] RF Warming and BJ Hyett. The modified equation approach to the stability and accuracy analysis of finite-difference methods. *Journal of computational physics*, 14(2):159–179, 1974.

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

Grey Wolf Gravitational Search Algorithm

Susheel Joshi
Department of Mathematics
South Asian University
New Delhi, India
Email: sushil4843@gmail.com

Jagdish Chand Bansal
Department of Mathematics
South Asian University
New Delhi, India
Email: jcbansal@gmail.com

***Abstract*—Gravitational search algorithm is a nature inspired optimization algorithm, inspired by newton's law of gravity and law of motion. In this paper, a new variant of Gravitational search algorithm is presented. The exploration and exploitation capability of GSA is balanced by splitting the whole swarm into two groups. The search process is modified so that one group better exploits and one group becomes responsible for better exploration. This proposed algorithm is tested over some benchmark functions. The results show that our approach gives a better balance between exploration and exploitation to get the optimal solution. A comparative study of this algorithm with GSA and some well-known swarm based meta-heuristic search methods like Bio-geography based optimization (BBO), Differential evolution (DE) and Artificial bee colony (ABC) confirm its efficiency and robustness.**

Keywords: Gravitational Search Algorithm (GSA), Gravitational Constant, Grey Wolf Optimizer, Exploration, Exploitation.

I. Introduction

Solving rigid and highly non linear optimization problems has become an emerging area in the field of science and engineering. Due to non-linear, non-differentiable, multimodal and dynamic in nature, these problems are usually difficult for traditional or exhaustive search methods.

Therefore swarm based nature inspired algorithms are one of the better methods to deal these problems due to its flexibility, simplicity, derivation-free mechanism, and local optima avoidance.

Gravitational search algorithm (GSA) [7] is a relatively new population based stochastic search technique in the field of nature inspired algorithms. GSA is inspired by the movement of agents under the influence of the gravitational forces. Due to these forces, a global movement generates which drives all agents towards the agents having heavies masses [2].

The effectiveness and robustness of these meta heuristic algorithms depend upon two fundamental process which navigate the swarm in the search space: the exploration process which explores the large search space and ensures that solution does not converge in local optima while exploitation process concentrates on best solution for convergence to optimality [12]. A proper balance between exploration and exploitation makes a algorithm perfect.

So far, a lot of work have been done by researcher to achieve a proper balance between exploration and exploitation. Seyedali Mirjalili et al. [5] introduced the memory feature in GSA, namely PSOGSA. Sarafrazi et al. [8] proposed disruption operator to improve the exploration and exploitation ability of GSA. Doraghinejad et al. [1] suggested a new operator inspired by some of the characteristics of the black hole to prevent from premature convergence and to improve the exploration and exploitation abilities of GSA.

To improve the exploitation and exploration properties of basic Gravitational search algorithm, a new variant called Grey Wolf Gravitational search algorithm (GWGSA) is proposed in this paper.

The remaining paper is organized as follows: Section II provides a overview of standard GSA. In Section III, we introduce a brief description of Grey Wolf GSA. In Section IV, experiment results and comparative study is presented and finally the paper is concluded in section V.

II. STANDARD GSA

Gravitational Search Algorithm (GSA) is a new population-based intelligence technique for optimization developed by Rashedi et al [7]. This algorithm has been motivated by the law of gravity and the law of motion.

The GSA algorithm can be described as follows:

Consider the swarm of N agents, in which each agent X_i in search space is defined as:

$$X_i = (x_i^1, ..., x_i^d, ..., x_i^n), \quad \forall\ i = 1, 2, ..., N \tag{1}$$

Here, x_i^d shows the position of i^{th} agent in d dimensional space, n is the dimension of the search space and N is swarm size. Since mass of each agent is dependent upon its fitness value therefore, computation of masses are done after computing the current population fitness as follows:

$$q_i(t) = \frac{fit_i(t) - worst(t)}{best(t) - worst(t)} \tag{2}$$

$$M_i(t) = \frac{q_i(t)}{\sum_{j=1}^{n} q_j(t)} \tag{3}$$

Here,

$fit_i(t)$ is the fitness value of agent i at iteration t,

$M_i(t)$ is the mass of agent i at iteration t.

Worst(t) and best(t) are worst and best fitness of the current population.

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

The computation of acceleration of i^{th} agent is denoted by $a_i^d(t)$ and defined as:

$$a_i^d(t) = \frac{F_i^d(t)}{M_i(t)} \tag{4}$$

where, $M_i(t)$ is the mass of i^{th} agent and $F_i^d(t)$ is the total force acting on i^{th} agent by a set of Kbest heavier masses in d^{th} dimension at iteration t. $F_i^d(t)$ is calculated as:

$$F_i^d(t) = \sum_{j \in Kbest, j \neq i} rand_j F_{ij}^d(t) \tag{5}$$

Here, Kbest is first K agents with the best fitness values and biggest masses and $rand_j$ is a uniform random number between 0 and 1. Kbest is a function of time having linearly decreasing property. The value of Kbest will reduce in each iteration and at the end only one agent will apply force to the other agents. At the t^{th} iteration, the force applied on agent i from agent j in the d^{th} dimension is defined:

$$F_{ij}^d(t) = G(t)\frac{M_i(t)M_j(t)}{R_{ij} + \epsilon}(x_i^d(t) - x_j^d(t)) \tag{6}$$

Finally, the acceleration of an agent is calculated as:

$$a_i^d(t) = \sum_{j \in Kbest, j \neq i} rand_j G(t)\frac{M_j(t)}{R_{ij} + \epsilon}(x_i^d(t) - x_j^d(t)), \tag{7}$$

$d = 1, 2, ..., n$ and $i = 1, 2, ..., N$.
Here, $R_{ij}(t)$ is the euclidean distance between two agents, i and j. ϵ is a small value.
G(t) is Gravitational constant and is a decreasing function of time :

$$G(t) = G_0 exp^{-\alpha \frac{t}{T}} \tag{8}$$

G_0 and α are constants which are initialized at the beginning. the value of $G(t)$ will be reduced with time to control the search accuracy. In GSA the value of G_0 and α are set to 100 and 20 respectively. t is the current iteration and T is the total number of iterations.

The velocity and the position update equations of an agent is calculated as :

$$v_i^d(t+1) = rand_i \times v_i^d(t) + a_i^d(t) \tag{9}$$

$$x_i^d(t+1) = x_i^d(t) + v_i^d(t+1) \tag{10}$$

where $v_i^d(t)$ and $x_i^d(t)$ present the velocity and position of agent i in dimension d respectively. $rand_i$ is uniform random number in the interval [0,1].

III. Grey Wolf Gravitational Search Algorithm

The balanced trade off between exploration and exploitation is the most important aspect for the robustness and effectiveness of a swarm based meta-heuristic algorithm.
Exploration of the whole search space and exploitation of the near optimal solution region may be balanced by maintaining the diversity in early and later iterations of any random number based search algorithm [9].

In GSA, it is clear from the position update equations (9) and (10) that the next move of an agent depends upon its velocity which further depends upon its acceleration. Acceleration of an agent is decided by the gravitational constant. Therefore the step size of an agent is directly proportional to gravitational constant. Hence gravitational constant is the key entity which makes the balance between exploration and exploitation ability of search space. Equation (8) shows that gravitational constant is a decreasing function of time, however the value of G(t) may cause either pre mature convergence or skipping the local optima.
To overcome this drawback and improving the exploitation and exploration ability, in this paper two position update strategies are proposed in which first position update strategy enhances the exploitation while second position update strategy improves exploration property of the search space.
To improve the exploitation ability of the search space, half agents of the population having less fitness probabilities encircle the best agent of the population (agent of maximum fitness / prey) using encircling behavior of grey wolf [6]. In grey wolf optimizer, grey wolves (agents) encircle the prey (best agent) during the hunt. The encircling behavior can be written as:

$$H = abs(C \cdot X_{max} - x_i(t)) \tag{11}$$

$$x_i(t+1) = X_{max} - A \cdot H \tag{12}$$

where t indicates the iteration counter, A and C are coefficient vectors, x_i is the position vector of the i^{th} grey wolf, and X_{max} indicates the position vector of a prey. The vectors A and C are calculated as follows:
$A = 2 \cdot G_{new}(t) \cdot r_1 - G_{new}(t)$,
$C = 2 \cdot r_2$.
Here r_1 and r_2 are two different uniformly random numbers between 0 and 1. $G_{new}(t)$ is the new gravitational constant defined as:

$$G_{new}(t) = G_0 exp^{-\alpha \frac{t}{T}} \tag{13}$$

Here the value of G_0 and α are set to 0.5 and 35 respectively. The difference between $G(t)$ and $G_{new}(t)$ has been demonstrated in Figure 1.

Equations (11) and (12) define the next move of a grey wolf (agent) according to the position of the prey (best agent). Different places around the best agent can be reached with respect to the current position by adjusting the value of A and C vectors.

The fitness probability of each agent x_i is calculated as:

$$prob_i = \frac{0.9 \times fit(i)}{max fit(i)} + 0.1 \tag{14}$$

$fit(i)$ is the fitness value of i^{th} agent. Fitness is calculated with the help of objective value and using the following formula:

$$fit(i) = \begin{cases} \frac{1}{1+f_i}, & \text{if } (f_i \geq 0) \\ 1 + abs(f_i), & \text{if } f_i < 0. \end{cases} \tag{15}$$

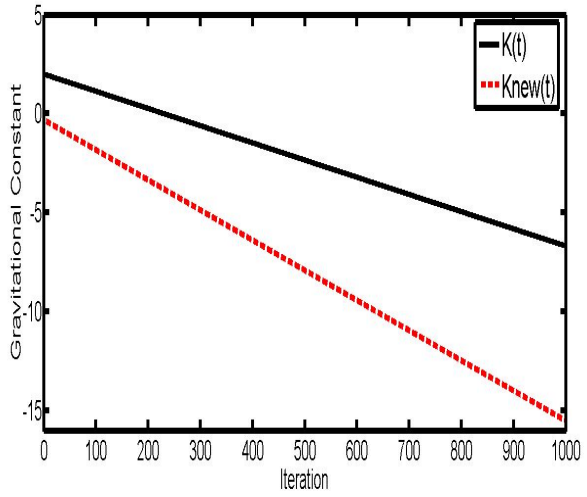


Fig. 1: Difference between $G(t)$ and $G_{new}(t)$ with respect to iterations, where $K(t) = log_{10}G(t)$ and $K_{new}(t) = log_{10}G_{new}(t)$

To improve the exploration ability of search space, rest half of agents having higher fitness probabilities are diversified in the search space with their respective randomized velocities in according to the following equation:

$$x_i^d(t+1) = x_i^d(t) + rand \times v_i^d(t+1) \tag{16}$$

The position update strategies of GWGSA is shown in Algorithm 1.

Algorithm 1 Position update strategies in $GWGSA$:

Evaluate the fitness probability of each individual of swarm;
Sort agents with respect to the order of sorted probability;
Find X_{max} (agent having maximum fitness / prey);
Divide swarm into two groups each having (swarm size/2) agents;
Gp_1 is a first group of agents having less fitness;
Gp_2 is a second group of agents having higher fitness;
for each agent $x_i \in Gp_1$, encircles X_{max} **do**
  $r_1 = rand_1$;
  $r_2 = rand_2$;
  $A = 2 \cdot G_{new}(t) \cdot r_1 - G_{new}(t)$;
  $C = 2 \cdot r_2$;
  $H = abs(C \cdot X_{max} - x_i(t))$;
  $x_i(t+1) = X_{max} - A \cdot H$;
end for
for each agent $x_i \in Gp_2$ **do**
  $x_i^d(t+1) = x_i^d(t) + rand \times v_i^d(t+1)$
end for

IV. EXPERIMENTAL RESULTS AND DISCUSSION

In the Table II, 19 standard benchmark functions [7] with their respective range, optimum value and dimension are used to evaluate the performance of proposed algorithm.

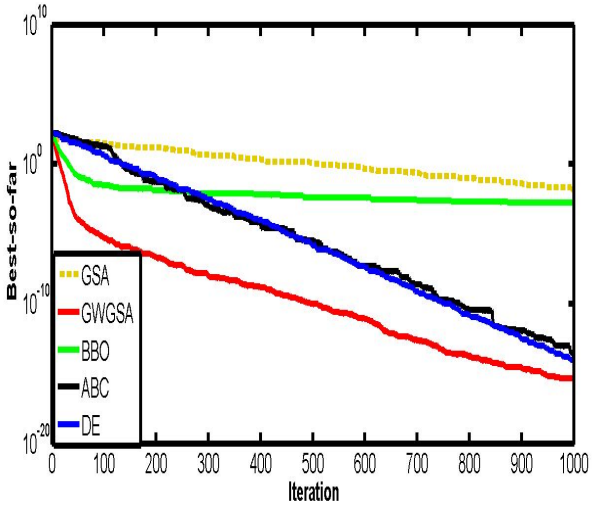
An empirical comparative analysis is performed among GWGSA, standard GSA, BBO, DE and ABC for justification of the efficiency and robustness of the proposed algorithm. Following GWGSA parameters have been considered for numerical experiments:

Test Problem	U rank sum test with GWGSA			
	GSA	BBO	DE	ABC
f_1	+	+	+	+
f_2	+	+	+	+
f_3	+	+	+	+
f_4	+	+	+	+
f_5	+	+	-	-
f_6	+	+	+	+
f_7	+	+	+	+
f_8	+	+	+	+
f_9	+	+	+	+
f_{10}	+	+	+	+
f_{11}	+	+	+	+
f_{12}	+	+	+	+
f_{13}	+	+	+	+
f_{14}	+	+	+	+
f_{15}	+	=	+	+
f_{16}	+	+	+	+
f_{17}	+	+	+	-
f_{18}	+	+	+	+
f_{19}	+	+	+	+

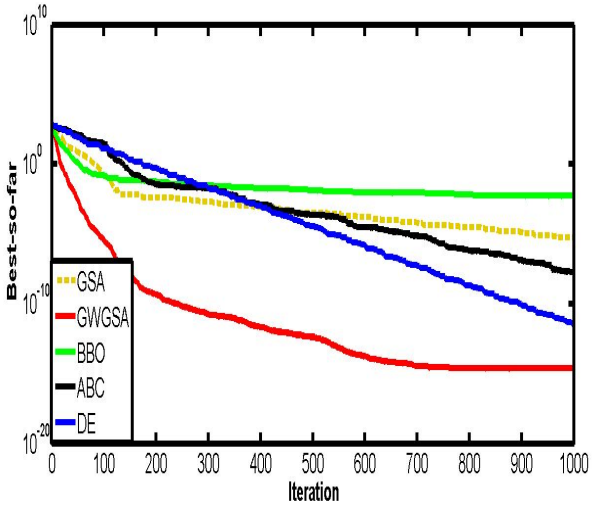
TABLE I: Comparison based on mean function evaluations and the Mann-Whitney U rank sum test at a $\alpha = 0.05$ significance level ('+' indicates GWGSA is significantly better, '-' indicates GWGSA is worse and '=' indicates that there is no significant difference), TP: Test Problem.

- The number of simulations/run =30,
- Individuals swarm size=50,
- For GSA, $G_0 = 100$ and $\alpha = 20$ [7],
- The stopping criteria is either acceptable error (referred in Table II) has been achieved or maximum number of function evaluations (which is set to be 200000) is reached,
- Parameters for the algorithms GSA [7], BBO [10], DE [11] and ABC [3] are initialized from their primary research papers respectively.

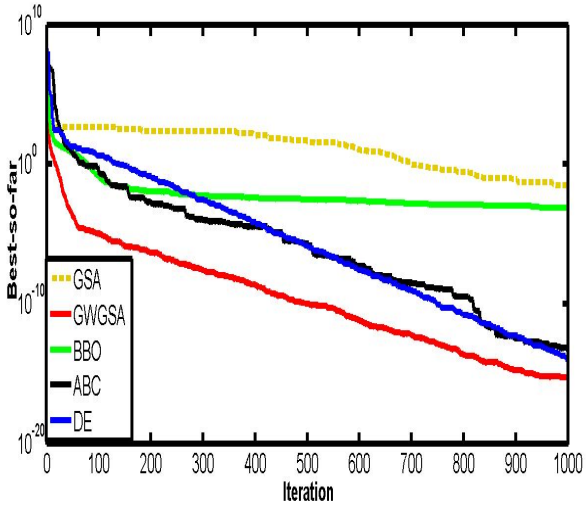
The experimental results of the mentioned algorithms are shown in Table III, which describe a comprehensive analysis about the standard deviation (SD), mean error (ME), average number of function evaluations (AFE) along with the success rate (SR). The statistical study of Table III replicates that most of the time GWGSA dominates in terms of efficiency, reliability as well as accuracy as compare to the other considered algorithms.



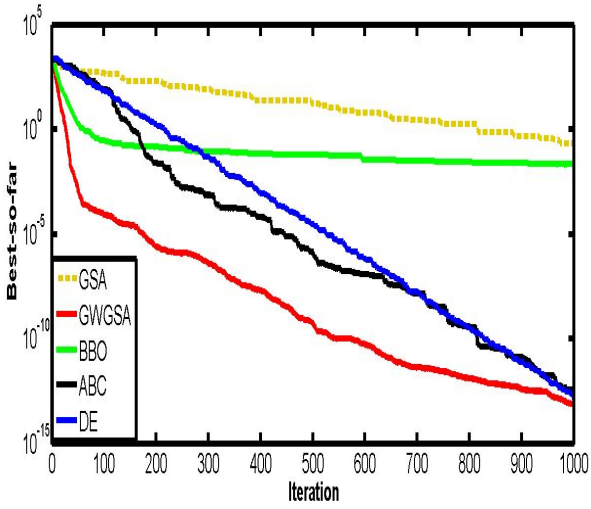
(a) Comparative behaviour for f_1



(b) Comparative behaviour for f_3



(c) Comparative behaviour for f_8



(d) Comparative behaviour for f_{10}

Fig. 2: Average best for benchmark functions f_1, f_3, f_8 and f_{10}

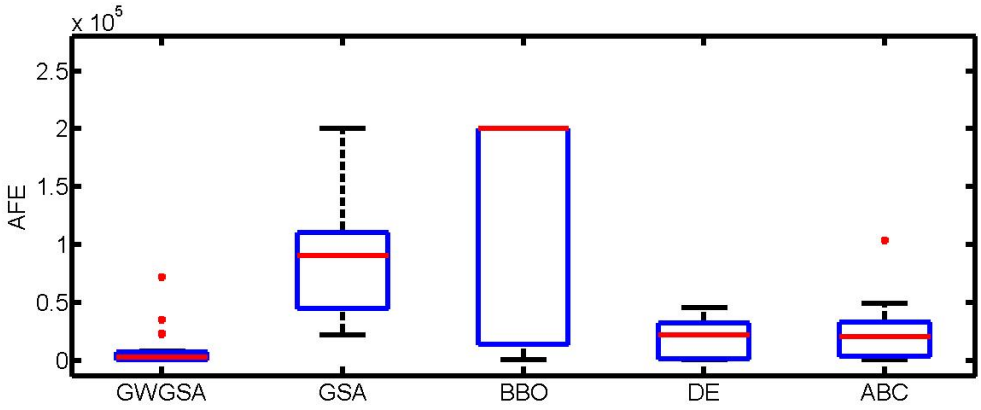


Fig. 3: Boxplots graphs (Average number of function evaluations)

Test problem	Objective function	Search Range	Optimum Value	Dim (n)	AE
Sphere	$f_1(x) = \sum_{i=1}^{n} x_i^2$	$[-5.12, 5.12]$	$f(\vec{0}) = 0$	30	$1.0E-05$
De Jong f4	$f_2(x) = \sum_{i=1}^{n} i.(x_i)^4$	$[-5.12, 5.12]$	$f(\vec{0}) = 0$	30	$1.0E-05$
Griewank	$f_3(x) = 1 + \frac{1}{4000}\sum_{i=1}^{n} x_i^2 - \prod_{i=1}^{n} \cos(\frac{x_i}{\sqrt{i}})$	$[-600, 600]$	$f(\vec{0}) = 0$	30	$1.0E-05$
Ackley	$f_4(x) = -20 + e + exp(-\frac{0.2}{n}\sqrt{\sum_{i=1}^{n} x_i^3})$ $-exp(\frac{1}{n}\sum_{i=1}^{n} \cos(2\pi.x_i)x_i)$	$[-1, 1]$	$f(\vec{0}) = 0$	30	$1.0E-05$
Cosine Mixture	$f_5(x) = \sum_{i=1}^{n} {x_i}^2 - 0.1(\sum_{i=1}^{n} \cos 5\pi x_i) + 0.1n$	$[-1, 1]$	$f(\vec{0}) = -n \times 0.1$	30	$1.0E-05$
Exponential	$f_6(x) = -(exp(-0.5\sum_{i=1}^{n} {x_i}^2)) + 1$	$[-1, 1]$	$f(\vec{0}) = -1$	30	$1.0E-05$
Cigar	$f_7(x) = {x_0}^2 + 100000\sum_{i=1}^{n} {x_i}^2$	$[-10, 10]$	$f(\vec{0}) = 0$	30	$1.0E-05$
brown3	$f_8(x) = \sum_{i=1}^{n-1}({x_i}^{2^{(x_{i+1})^2+1}} + {x_{i+1}}^{2^{x_i{}^2+1}})$	$[-1, 4]$	$f(\vec{0}) = 0$	30	$1.0E-05$
Schewel	$f_9(x) = \sum_{i=1}^{n} \lvert x_i\rvert + \prod_{i=1}^{n} \lvert x_i\rvert$	$[-10, 10]$	$f(\vec{0}) = 0$	30	$1.0E-05$
Axis parallel hyper-ellipsoid	$f_{10}(x) = \sum_{i=1}^{n} i.x_i^2$	$[-5.12, 5.12]$	$f(\vec{0}) = 0$	30	$1.0E-05$
Sum of different powers	$f_{11}(x) = \sum_{i=1}^{n} \lvert x_i\rvert^{i+1}$	$[-1, 1]$	$f(\vec{0}) = 0$	30	$1.0E-05$
Rotated hyper-ellipsoid	$f_{12}(x) = \sum_{i=1}^{n}\sum_{j=1}^{i} x_j^2$	$[-65.536, 65.536]$	$f(\vec{0}) = 0$	30	$1.0E-05$
Beale	$f_{13}(x) = [1.5 - x_1(1-x_2)]^2 + [2.25 - x_1(1-x_2^2)]^2 +$ $[2.625 - x_1(1-x_2^3)]^2$	$[-4.5, 4.5]$	$f(3, 0.5) = 0$	2	$1.0E-05$
Branins's	$f_{14}(x) = a(x_2 - bx_1^2 + cx_1 - d)^2 + e(1-f)\cos x_1 + e$	$-5 \le x_1 \le 10$, $0 \le x_2 \le 15$	$f(-\pi, 12.275)$ $= 0.3979$	2	$1.0E-05$
Gear train	$f_{15}(\vec{x}) = \left(\frac{1}{6.931} - \frac{x_1 x_2}{x_3 x_4}\right)^2$	$[12, 60]$	$f(19, 16, 43, 49)$ $= 2.7 \times 10^{-12}$	4	$1.0E-15$
Six-hump camel back	$f_{16}(x) = (4 - 2.1x_1^2 + x_1^4/3)x_1^2 + x_1x_2 + (-4 + 4x_2^2)x_2^2$	$[-5, 5]$	$f(-0.0898, 0.7126)$ $= -1.0316$	2	$1.0E-05$
Hosaki Problem	$f_{17}(x) = (1 - 8x_1 + 7x_1^2 - 7/3x_1^3 + 1/4x_1^4)x_2^2\exp(-x_2)$ subject to $0 \le x_1 \le 5, 0 \le x_2 \le 6$	$[0, 5]$, $[0, 6]$	-2.3458	2	$1.0E-05$
McCormick	$f_{18}(x) = \sin(x_1 + x_2) + (x_1 - x_2)^2 - \frac{3}{2}x_1 + \frac{5}{2}x_2 + 1$	$-1.5 \le x_1 \le 4$, $-3 \le x_2 \le 3$	f(-0.547,-1.547) $= -1.9133$	30	$1.0E-04$
Temp	$f_{19}(x) = x_1^2 + x_2^2$	$[-5, 5]$	$f(0) = 0$	2	$1.0E-05$

TABLE II: Benchmark functions

For scrutinized comparison of these algorithms, boxplot analysis has also been carried out with respect to average function evaluation (AFE). Boxplot analysis is a well known tool to describe the empirical distribution of data graphically. From Figure 3 it is clear that the boxplot of GWGSA have less interquartile range and medians as compare to other considered algorithms.

It is possible that these differences may occur due to some randomness therefore another statistical test is required to find out the real cause behind these differences.

The Mann-Whitney U rank sum test [4] is a non-parametric test for comparison among the data which are not normally distributed. In this paper, this test is performed at 5% level of significance ($\alpha = 0.05$) between GWGSA - GSA, GWGSA - BBO, GWGSA - DE, and GWGSA - ABC.

If significant difference between two data sets does not occur, it implies that the null hypothesis is accepted therefore sign '=' appears. On the contrary, the null hypothesis is rejected therefore '-' or '+' signs appears accordingly the first data set is more or less than the second.

In this paper, the data sets are the average function evaluations of a particular algorithm. '-' or '+' signs shows that a particular algorithm have more or less number of function evaluations compare than other. In the Table I, 72 '+' signs out of 76 comparisons assure that GWGSA requires less number of function evaluation implies that less cost effective compare than considered algorithms.
The convergence speed of the considered algorithms over a test problem is measured by function of evaluations, i.e. how many times a algorithm use the fitness function regarding to solve

TP	Algorithm	SD	ME	AFE	SR
	GWGSA	1.8241E-07	9.41445E-06	3866.66	30
	GSA	7.28788E-07	9.01169E-06	95195	30
f_1	BBO	3.52935E-05	9.98244E-05	200000	0
	DE	8.6284E-07	9.15079E-06	22300	30
	ABC	2.01936E-06	8.16768E-06	20508.33	30
	GWGSA	1.08307E-06	6.74016E-06	2650	30
	GSA	1.15202E-06	8.38172E-06	62898.33	30
f_2	BBO	8.00792E-07	8.9837E-06	46193.33	30
	DE	0.01904842	0.003545973	20077.58	29
	ABC	3.11143E-06	4.9033E-06	9931.66	30
	GWGSA	7.88368E-07	8.94589E-06	5233.33	30
	GSA	2.5317E-16	0.758961623	200000	0
f_3	BBO	0.001129065	0.762959217	200000	0
	DE	0.005951412	0.002633154	31741.66	24
	ABC	0.001263587	0.000228381	41050.5	30
	GWGSA	1.51572E-07	9.70134E-06	23016.66	30
	GSA	3.23539E-07	9.426E-06	160631.66	30
f_4	BBO	0.010015867	0.045344509	200000	0
	DE	5.61227E-07	9.27488E-06	42663.33	30
	ABC	1.56408E-06	8.27669E-06	49570	30
	GWGSA	1.058762768	0.74866815	71933.33	30
	GSA	1.16214E-06	8.815E-06	111108.33	30
f_5	BBO	0.186672121	0.157684431	200000	0
	DE	0.04433252	0.01478668	23050	27
	ABC	2.62508E-06	7.21939E-06	23021.66	30
	GWGSA	3.32014E-07	8.86804E-06	2983.33	30
	GSA	9.71668E-07	8.90081E-06	90871.66	30
f_6	BBO	3.61922E-07	9.53932E-06	93826.66	30
	DE	9.20808E-07	8.7858E-06	16905	30
	ABC	2.22462E-06	7.45824E-06	16988.33	30
	GWGSA	3.93638E-07	9.58879E-06	23366.66	30
	GSA	4.585612147	3.681823346	200000	0
f_7	BBO	7.825705935	22.28235959	200000	0
	DE	7.03313E-07	9.07831E-06	39768.33	30
	ABC	1.94844E-06	7.95431E-06	34818.33	30
	GWGSA	8.67848E-07	8.66318E-06	4383.33	30
	GSA	7.47565E-07	8.82557E-06	99023.33	30
f_8	BBO	1.6474E-05	4.25953E-05	200000	0
	DE	9.31329E-07	8.87821E-06	22293.33	30
	ABC	2.28625E-06	7.62743E-06	20741.66	30
	GWGSA	2.41195E-08	9.92469E-06	35200	30
	GSA	4.79823E-07	9.37587E-06	181635	30
f_9	BBO	0.008899536	0.056087836	200000	0
	DE	3.24203E-07	9.62591E-06	45363.33	30
	ABC	1.12129E-06	9.04997E-06	41708.33333	30
	GWGSA	1.08127E-07	9.88779E-06	8200	30
	GSA	9.19945E-07	8.94862E-06	110045	30
f_{10}	BBO	0.000433073	0.001282362	200000	0
	DE	6.06354E-07	8.95798E-06	25803.33	30
	ABC	1.99482E-06	7.90469E-06	22730	30
	GWGSA	2.50331E-06	7.25032E-06	1550	30
	GSA	2.40313E-06	7.11963E-06	46356.66	30
f_{11}	BBO	2.20301E-06	7.48081E-06	3241.66	30
	DE	2.10875E-06	7.04301E-06	7698.33	30
	ABC	2.83237E-06	5.23398E-06	16141.66	30

TP	Algorithm	SD	ME	AFE	SR
	GWGSA	7.11315E-07	8.90979E-06	4766.66	30
	GSA	1.19562E-06	8.75074E-06	95633.33	30
f_{12}	BBO	0.005393584	0.014373538	200000	0
	DE	8.16356E-07	9.12408E-06	32888.33	30
	ABC	2.04172E-06	7.61228E-06	27908.33	30
	GWGSA	9.59991E-07	4.45508E-06	683.33	30
	GSA	2.93149E-06	5.48556E-06	71990	30
f_{13}	BBO	0.228618561	0.07621397	36078.33	27
	DE	2.97731E-06	4.48602E-06	1416.66	30
	ABC	1.19621E-06	8.83894E-06	17239.33	30
	GWGSA	2.90363E-05	2.74038E-05	550	30
	GSA	3.65196E-05	4.21047E-05	40393.33	30
f_{14}	BBO	2.8596E-05	5.76747E-05	6991.66	30
	DE	6.28338E-06	5.81868E-06	2100	25
	ABC	6.88936E-06	6.09223E-06	1623.73	30
	GWGSA	7.07973E-14	1.88383E-12	1116.66	30
	GSA	7.76657E-13	1.88817E-12	22308.33	30
f_{15}	BBO	8.24545E-13	1.78926E-12	1090	30
	DE	7.59234E-13	1.86588E-12	39365	30
	ABC	0.01174734	0.0128677	103825.16	2
	GWGSA	1.29405E-05	5.02016E-06	700	30
	GSA	1.12012E-05	1.19688E-05	46130	30
f_{16}	BBO	0.399829919	0.326463166	83296.66	18
	DE	1.4363E-05	1.64462E-05	1546.66	15
	ABC	1.13941E-05	1.17778E-05	955	30
	GWGSA	7.87028E-06	5.08423E-06	816.66	30
	GSA	5.91694E-06	5.22298E-06	41056.66	30
f_{17}	BBO	3.34298E-16	1.218005973	200000	0
	DE	6.34868E-06	6.52551E-06	972.91	24
	ABC	6.30488E-06	6.19834E-06	663.33	30
	GWGSA	5.59563E-06	8.68309E-05	516.66	30
	GSA	7.45542E-06	8.71969E-05	44575	30
f_{18}	BBO	6.54881E-06	8.96289E-05	2600	30
	DE	6.09609E-06	8.73884E-05	990	30
	ABC	6.64539E-06	8.89613E-05	1105	30
	GWGSA	1.04981E-06	2.77983E-06	533.33	30
	GSA	3.10667E-06	5.14861E-06	44628.33	30
f_{19}	BBO	2.78005E-06	4.76596E-06	750	30
	DE	2.66975E-06	5.11721E-06	985	30
	ABC	0.07366599	0.235931	696.66	30

TABLE III: Minimization results of benchmark functions

a problem or fulfill the stopping criteria. To reduce the effect of random behavior of the algorithms, the average function of evaluations is used for each test problem over 30 runs. The convergence speed of a algorithm is inversely proportional to AFEs used by it to solve a test problem.

In order to compare the convergence speed of proposed algorithm over the considered algorithm, Acceleration rate (*AR*) is used. *AR* is based on the ratio between the AFEs of proposed (GWGSA) and considered algorithm (ALGO) defined as:

$$AR = \frac{AFE_{ALGO}}{AFE_{GWGSA}}, \tag{17}$$

where, ALGO $\in$ {GSA, BBO, DE, and ABC} and $AR > 1$ (equation (17)) means GWGSA is faster than the compared algorithm (ALGO). Table IV represents a comparison results between GWGSA and GSA, GWGSA and BBO, GWGSA and DE, and GWGSA and ABC in terms of AR. These results confirm that GWGSA have better convergence speed compare

TP	GSA	BBO	DE	ABC
f_1	24.62	51.72	5.77	5.30
f_2	23.74	17.43	7.58	3.75
f_3	38.22	38.22	6.07	7.84
f_4	6.98	8.69	1.85	2.15
f_5	1.54	2.78	0.32	0.32
f_6	30.46	31.45	5.67	5.69
f_7	8.56	8.56	1.70	1.49
f_8	22.59	45.63	5.09	4.73
f_9	5.16	5.68	1.29	1.18
f_{10}	13.42	24.39	3.15	2.77
f_{11}	29.91	2.09	4.97	10.41
f_{12}	20.06	41.96	6.90	5.85
f_{13}	105.35	52.80	2.07	25.23
f_{14}	73.44	12.71	3.82	2.95
f_{15}	19.98	0.98	35.25	92.98
f_{16}	65.90	119.00	2.21	1.36
f_{17}	50.27	244.90	1.19	0.81
f_{18}	86.27	5.03	1.92	2.14
f_{19}	83.68	1.41	1.85	1.31

TABLE IV: Acceleration Rate (AR) of GWGSA as compared to the GSA, BBO, DE and ABC, TP: Test Problems

than considered algorithms. Figure 2 also establishes the fast convergence nature of the proposed variant.

V. Conclusion

In this paper, a novel concept of Grey Wolf GSA (GWGSA) is presented. GWGSA provides a better balance between exploration and exploitation by dividing the whole swarm into two groups according to agent's fitness probabilities. First group refines exploitation whereas second group improves exploration properties of the search space. The proposed GWGSA is tested over 19 benchmark functions and proved its effectiveness.

The obtained results motivate that the proposed algorithm could be used very effectively for the real world optimization problems having high complexity.

References

[1] Mohammad Doraghinejad and Hossein Nezamabadi-pour. Black hole: A new operator for gravitational search algorithm. *International Journal of Computational Intelligence Systems*, 7(5):809–826, 2014.

[2] David Holliday, Robert Resnick, and Jearl Walker. Fundamentals of physics. 1993.

[3] Dervis Karaboga and Bahriye Basturk. A powerful and efficient algorithm for numerical function optimization: artificial bee colony (abc) algorithm. *Journal of global optimization*, 39(3):459–471, 2007.

[4] Henry B Mann and Donald R Whitney. On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, pages 50–60, 1947.

[5] Seyedali Mirjalili and Siti Zaiton Mohd Hashim. A new hybrid psogsa algorithm for function optimization. In *Computer and information application (ICCIA), 2010 international conference on*, pages 374–377. IEEE, 2010.

[6] Seyedali Mirjalili, Seyed Mohammad Mirjalili, and Andrew Lewis. Grey wolf optimizer. *Advances in Engineering Software*, 69:46–61, 2014.

[7] Esmat Rashedi, Hossein Nezamabadi-Pour, and Saeid Saryazdi. Gsa: a gravitational search algorithm. *Information sciences*, 179(13):2232–2248, 2009.

[8] S Sarafrazi, H Nezamabadi-Pour, and S Saryazdi. Disruption: a new operator in gravitational search algorithm. *Scientia Iranica*, 18(3):539–548, 2011.

[9] Harish Sharma, Jagdish Chand Bansal, and K. V. Arya. Fitness based differential evolution. *Memetic Computing*, 4(4):303–316, 2012.

[10] Dan Simon. Biogeography-based optimization. *Evolutionary Computation, IEEE Transactions on*, 12(6):702–713, 2008.

[11] Rainer Storn and Kenneth Price. Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, 11(4):341–359, 1997.

[12] Xin-She Yang. *Engineering optimization: an introduction with metaheuristic applications*. John Wiley & Sons, 2010.

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

A third (four)-order accurate nine-point compact EEM-FDM for coupled system of mildly non-linear elliptic equations

Navnit Jha
Faculty of Mathematics and Computer Science
South Asian University, Chanakyapuri, New Delhi,
India-110021 Email: navnitjha@sau.ac.in

***Abstract*—In this paper, we formulate a new algorithm for the approximate solution values of coupled mildly non-linear elliptic boundary value problems in two-space dimensions. The method is obtained on a non-uniformly spaced mesh points in such a manner that fourth-order accuracy of the scheme is a particular case of third-order method. The special nine-point discretization makes it easier to solve and applicable to the problems possessing layer behaviour. Applications to stationary Navier-Stokes equation and biharmonic equations have been given to illustrate the efficiency and robustness of the scheme.**

I. Introduction

We consider the two-space dimensional coupled mildly non-linear elliptic boundary value problems (EBVPs) with first order partial derivatives as non-linear term:

$$\begin{cases} \Delta F(x,y) \equiv F^{xx} + F^{yy} = \Phi(x,y,F,G,F^x,G^x,F^y,G^y), \\ \Delta G(x,y) \equiv G^{xx} + G^{yy} = \Psi(x,y,F,G,F^x,G^x,F^y,G^y), \end{cases} \tag{I.1}$$

where Δ is a two-space dimensional Laplacian operator in Cartesian coordinate, $(x,y) \in \Omega = (0,1) \times (0,1)$ and the associated Dirichlet's boundary values of second kind are specified for $0 \le x \le 1, 0 \le y \le 1$ as

$$\begin{cases} F(x,0) = p_1, F(x,1) = p_2, G(x,0) = p_3, G(x,1) = p_4, \\ F(0,y) = q_1, F(1,y) = q_2, G(0,y) = q_3, G(1,y) = q_4, \end{cases} \tag{I.2}$$

and $p_i = p_i(x), q_i = q_i(y), i = 1(1)4$ are known smooth functions. The notation F^x stands for $\partial F(x,y)/\partial x$ etc.

Coupled system of mildly non-linear elliptic partial differential equations (EPDEs) appears in the application areas of fluid flow, elasticity, chemical reaction, graphics design and mathematical biology [10], [19], [1], [7], [20]. Various techniques for determining numerical solution values of coupled EPDEs have been considered in the past, because of non availability of the theoretical solutions to the arbitrary choice of $\Phi(x,y,F,G,F^x,G^x,F^y,G^y)$ and/or $\Psi(x,y,F,G,F^x,G^x,F^y,G^y)$. Therefore, it is worthwhile to develop an efficient and numerically accurate algorithm for the coupled mildly non-linear elliptic equations. With the wide range of applications, we mention some of the literature in the context of approximate solution method to coupled EPDEs. Rigal [17], discussed solution of Navier-Stokes and biharmonic equations in the form of coupled elliptic equations using successive-over-ralaxation method. Watanabe [22], discussed fixed-point method for the steady-state reaction-diffusion coupled EPDEs appearing in biological sciences. The method of lower and upper solution values and monotone iterative sequence for coupled semi-linear EPDEs has been developed by Boglaev [2]. A finite difference approximations (FDA) has been formulated to solve biharmonic equation using coupled EPDEs by Mohanty [13], [14] and the convergence property of coupled elliptic equations using FDA has been elaborated by Pao [15].

In this paper, we have acquired a new nine-point compact scheme based on sequence of exponential expanding meshes and finite difference discretization. Such kind of mesh network has been considered in digital simulation of electrochemistry [3] and axis symmetric incompressible flow of fluid [6]. We have defined two compact operators in X- and Y-directions based on central and averaging operators and received a new scheme which provides in general a third-order of truncation error. The application of the proposed schemes has been described by solving stationary Navier-Stokes fluid flow problems and biharmonic equations in two-space dimensions.

II. Exponential expanding meshes (EEM) and compact operators

The two-dimensional closed domain $\Omega \cup \partial\Omega = [0,1] \times [0,1]$ is discretized as $\{0 = x_0 < x_1 < ... < x_{L+1} = 1\} \otimes \{0 = y_0 < y_1 < ... < y_{M+1} = 1\}$, with non-equidistant mesh spacing defined by $h_l = x_l - x_{l-1}, k_m = y_m - y_{m-1}, l = 1(1)L+1, m = 1(1)M+1$, the mesh-ratio parameter $\xi_{l,m} = k_m/h_l$ and the subsequent step sizes are $h_{l+1} = \alpha h_l, k_{m+1} = \beta k_m, l = 1(1)L, m = 1(1)M$, where L, M are positive integers and α, β are exponential expanding mesh parameters in X- and Y-directions respectively. Thus, the first mesh spacing in X-direction can be calculated by the formula $h_1 + ... + h_{L+1} = 1$, which gives $h_1 = (\alpha - 1)/(\alpha^{L+1} - 1)$, if $\alpha > 1$ and $h_1 = (1-\alpha)/(1-\alpha^{L+1})$, if $\alpha < 1$. When $\alpha = 1$, it is a constant mesh spacing and $h_l = 1/(L+1), l = 1(1)L+1$. In a similar manner, we can obtain $k_1 = (\beta - 1)/(\beta^{M+1} - 1)$, if $\beta > 1$, $k_1 = (1-\beta)/(1-\beta^{M+1})$, if $\beta < 1$ and $k_m = 1/(M+1), m = 1(1)M+1$ if $\beta = 1$.

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

The numerical solution of (I.1) by means of finite difference method (FDM) requires discretization of first- and second-order partial derivatives, thus we need to define the following operators:

$$\begin{aligned}\mathcal{A}_x F_{l,m} &= \alpha^{-1}(1+\alpha)^{-1}F_{l+1,m} + (1-\alpha^{-1})F_{l,m} \\ &\quad - \alpha(1+\alpha)^{-1}F_{l-1,m},\end{aligned} \tag{II.1}$$

$$\begin{aligned}\mathcal{B}_x F_{l,m} &= 2\alpha^{-1}(1+\alpha)^{-1}F_{l+1,m} - 2\alpha^{-1}F_{l,m} \\ &\quad + 2(1+\alpha)^{-1}F_{l-1,m},\end{aligned} \tag{II.2}$$

$$\begin{aligned}\mathcal{A}_y F_{l,m} &= \beta^{-1}(1+\beta)^{-1}F_{l,m+1} + (1-\beta^{-1})F_{l,m} \\ &\quad - \beta(1+\beta)^{-1}F_{l,m-1},\end{aligned} \tag{II.3}$$

$$\begin{aligned}\mathcal{B}_y F_{l,m} &= 2\beta^{-1}(1+\beta)^{-1}F_{l,m+1} - 2\beta^{-1}F_{l,m} \\ &\quad + 2(1+\beta)^{-1}F_{l,m-1},\end{aligned} \tag{II.4}$$

where $F_{l,m} = F(x_l, y_m)$, $F_{l+1,m} = F(x_l + \alpha h_l, y_m)$, $F_{l-1,m} = F(x_l - h_l, y_m)$, etc.

With the help of finite series expansion, it is easy to observed that

$$\mathcal{A}_x F_{l,m} = h_l F^x_{l,m} + O(h_l^3), \quad \mathcal{B}_x F_{l,m} = h_l^2 F^{xx}_{l,m} + O(h_l^3),$$
$$\mathcal{A}_y F_{l,m} = k_m F^y_{l,m} + O(k_m^3), \quad \mathcal{B}_y F_{l,m} = k_m^2 F^{yy}_{l,m} + O(k_m^3),$$

In a similar manner, it is possible to obtain mixed derivatives in terms of composite compact operators.

Following the technique of Mohanty [12], discussed for uniform mesh spacing, the exponential expanding mesh finite difference method (EEM-FDM) for the system of non-homogeneous Poisson's equation

$$\begin{cases}\Delta F = \Phi(x,y) \\ \Delta G = \Psi(x,y)\end{cases} \tag{II.5}$$

is given by

$$\begin{cases}\Delta_{h_l,k_m} F_{l,m} = 2h_l^2 k_m^2 \sum_{(p,q)\in H} e_{p,q}\Phi_{p,q} + T_{l,m}, \\ \Delta_{h_l,k_m} G_{l,m} = 2h_l^2 k_m^2 \sum_{(p,q)\in H} e_{p,q}\Psi_{p,q} + T_{l,m},\end{cases} \tag{II.6}$$

where

$$\begin{aligned}\Delta_{h_l,k_m} &= 3h_l^2[12 - 4(1-\alpha)A_x + (1-\alpha+\alpha^2)B_x]B_y \\ &\quad + 3k_m^2[12 - 4(1-\beta)A_y + (1-\beta+\beta^2)B_y]B_x\end{aligned} \tag{II.7}$$

is the non-uniform mesh discretization of the Laplacian operator $\Delta = \partial_{xx} + \partial_{yy}$ and

$$T_{l,m} = \begin{cases}O(h_l^2k_m^5 + h_l^3k_m^4 + h_l^4k_m^3 + h_l^5k_m^2), & \alpha \neq 1, \beta \neq 1 \\ O(h_l^2k_m^6 + h_l^4k_m^4 + h_l^6k_m^2), & \alpha = 1, \beta = 1\end{cases} \tag{II.8}$$

is the seventh(eight) order local truncation error (LTE) and

$$\begin{aligned}e_{l,m} &= [(1+\alpha^2)(2-\beta+2\beta^2) - \alpha(1-8\beta+\beta^2)]/[\alpha\beta], \\ e_{l+1,m} &= [3\alpha^2\beta - (1-\alpha)(2-\beta+2\beta^2)]/[(1+\alpha)\alpha\beta], \\ e_{l-1,m} &= [3\beta + \alpha(1-\alpha)(2-\beta+2\beta^2)]/[(1+\alpha)\beta], \\ e_{l,m+1} &= [\alpha(1-\beta+3\beta^2) - 2(1-\beta)(1+\alpha^2)]/[\alpha\beta(1+\beta)], \\ e_{l,m-1} &= [\alpha(3-\beta+\beta^2) + 2\beta(1-\beta)(1+\alpha^2)]/[\alpha(1+\beta)], \\ e_{l+1,m+1} &= [2(1-\alpha)(1-\beta)]/[\alpha\beta(1+\alpha)(1+\beta)], \\ e_{l+1,m-1} &= -\beta^2 e_{l+1,m+1}, \quad e_{l-1,m+1} = -\alpha^2 e_{l+1,m+1}, \\ e_{l-1,m-1} &= \alpha^2\beta^2 e_{l+1,m+1}.\end{aligned}$$

are coefficients on the nine stencils. The summations in equation (II.6) runs over the set $H = \{l+1, l, l-1\} \otimes \{m+1, m, m-1\}$.

III. EEM-FINITE DIFFERENCE METHOD FOR MILDLY NON-LINEAR EBVPS

In this section, we present the method for solving coupled equation (I.1) that includes first-order partial derivatives as a non-linear term. We define

$$\begin{aligned}\begin{bmatrix}\widehat{F}^x_{l,m+\eta} \\ \widehat{F}^x_{l+1,m+\eta} \\ \widehat{F}^x_{l-1,m+\eta}\end{bmatrix} &= \frac{1}{h_l}\mathcal{M}(\alpha)\begin{bmatrix}F_{l+1,m+\eta} \\ F_{l,m+n} \\ F_{l-1,m-\eta}\end{bmatrix}, \\ \begin{bmatrix}\widehat{F}^y_{l+\eta,m} \\ \widehat{F}^y_{l+\eta,m+1} \\ \widehat{F}^y_{l+\eta,m-1}\end{bmatrix} &= \frac{1}{k_m}\mathcal{M}(\beta)\begin{bmatrix}F_{l+\eta,m+1} \\ F_{l+\eta,m} \\ F_{l+\eta,m-1}\end{bmatrix},\end{aligned} \tag{III.1}$$

and

$$\begin{aligned}\begin{bmatrix}\widehat{G}^x_{l,m+\eta} \\ \widehat{G}^x_{l+1,m+\eta} \\ \widehat{G}^x_{l-1,m+\eta}\end{bmatrix} &= \frac{1}{h_l}\mathcal{M}(\alpha)\begin{bmatrix}G_{l+1,m+\eta} \\ G_{l,m+n} \\ G_{l-1,m-\eta}\end{bmatrix}, \\ \begin{bmatrix}\widehat{G}^y_{l+\eta,m} \\ \widehat{G}^y_{l+\eta,m+1} \\ \widehat{G}^y_{l+\eta,m-1}\end{bmatrix} &= \frac{1}{k_m}\mathcal{M}(\beta)\begin{bmatrix}G_{l+\eta,m+1} \\ G_{l+\eta,m} \\ G_{l+\eta,m-1}\end{bmatrix},\end{aligned} \tag{III.2}$$

where $\eta = 0, \pm 1$ and for $\theta = \alpha, \beta$,

$$\mathcal{M}(\theta) = \frac{1}{\theta(1+\theta)}\begin{bmatrix}1 & -1+\theta^2 & -\theta^2 \\ 1+2\theta & -(1+\theta)^2 & \theta^2 \\ -1 & (1+\theta)^2 & -\theta(2+\theta)\end{bmatrix}.$$

For $(p,q) \in H \backsim \{(l,m)\}$, we construct the functional approximations

$$\begin{cases}\hat{\Phi}_{p,q} = \Phi(x_p, y_q, F_{p,q}, G_{p,q}, \hat{F}^x_{p,q}, \hat{G}^x_{p,q}, \hat{F}^y_{p,q}, \hat{G}^y_{p,q}) \\ \hat{\Psi}_{p,q} = \Psi(x_p, y_q, F_{p,q}, G_{p,q}, \hat{F}^x_{p,q}, \hat{G}^x_{p,q}, \hat{F}^y_{p,q}, \hat{G}^y_{p,q}).\end{cases} \tag{III.3}$$

With the help of equations (III.1) and (III.2), the multi-dimensional Taylor's expansion to the functionals (III.3) yields

$$\begin{bmatrix}\hat{\Phi}_{l\pm1,m-1} \\ \hat{\Phi}_{l\pm1,m} \\ \hat{\Phi}_{l\pm1,m+1}\end{bmatrix} = \begin{bmatrix}\Phi_{l\pm1,m-1} \\ \Phi_{l\pm1,m} \\ \Phi_{l\pm1,m+1}\end{bmatrix} - \frac{1}{6}\boldsymbol{\mathcal{M}}^{\pm}\boldsymbol{P} + \begin{bmatrix}O(h_l^3) \\ O(h_l^3) \\ O(h_l^3)\end{bmatrix}, \tag{III.4}$$

$$\begin{bmatrix}\hat{\Phi}_{l,m-1} \\ \hat{\Phi}_{l,m+1}\end{bmatrix} = \begin{bmatrix}\Phi_{l,m-1} \\ \Phi_{l,m+1}\end{bmatrix} + \frac{1}{6}\boldsymbol{\mathcal{M}}^{\mathbf{0}}\boldsymbol{P} + \begin{bmatrix}O(h_l^3) \\ O(h_l^3)\end{bmatrix}, \tag{III.5}$$

$$\begin{bmatrix}\hat{\Psi}_{l\pm1,m-1}\\ \hat{\Psi}_{l\pm1,m}\\ \hat{\Psi}_{l\pm1,m+1}\end{bmatrix}=\begin{bmatrix}\Psi_{l\pm1,m-1}\\ \Psi_{l\pm1,m}\\ \Psi_{l\pm1,m+1}\end{bmatrix}-\frac{1}{6}\boldsymbol{\mathcal{M}}^{\pm}\boldsymbol{Q}+\begin{bmatrix}O(h_l^3)\\ O(h_l^3)\\ O(h_l^3)\end{bmatrix}, \tag{III.6}$$

$$\begin{bmatrix}\hat{\Psi}_{l,m-1}\\ \hat{\Psi}_{l,m+1}\end{bmatrix}=\begin{bmatrix}\Psi_{l,m-1}\\ \Psi_{l,m+1}\end{bmatrix}+\frac{1}{6}\boldsymbol{\mathcal{M}}^{\mathbf{0}}\boldsymbol{Q}+\begin{bmatrix}O(h_l^3)\\ O(h_l^3)\end{bmatrix}, \tag{III.7}$$

where, for $(p,q)\in H$,

$$\begin{cases}\Phi_{p,q}=\Phi(x_p,y_q,F_{p,q},G_{p,q},F^x_{p,q},G^x_{p,q},F^y_{p,q},G^y_{p,q})\\ \Psi_{p,q}=\Psi(x_p,y_q,F_{p,q},G_{p,q},F^x_{p,q},G^x_{p,q},F^y_{p,q},G^y_{p,q}),\end{cases}$$

$$\boldsymbol{\mathcal{M}}^{+}=\begin{bmatrix}\alpha(1+\alpha) & 1+\beta\\ \alpha(1+\alpha) & -\beta\\ \alpha(1+\alpha) & \beta(1+\beta)\end{bmatrix},\boldsymbol{\mathcal{M}}^{-}=\begin{bmatrix}1+\alpha & 1+\beta\\ 1+\alpha & -\beta\\ 1+\alpha & \beta(1+\beta)\end{bmatrix},$$

$$\boldsymbol{\mathcal{M}}^{0}=\begin{bmatrix}\alpha & -(1+\beta)\\ \alpha & -\beta(1+\beta)\end{bmatrix}$$

and

$$\boldsymbol{P}=\begin{bmatrix}h_l^2\left(\left(\frac{\partial\Phi}{\partial F}\right)F^{xxx}+\left(\frac{\partial\Phi}{\partial G}\right)G^{xxx}\right)\\ \\ k_m^2\left(\left(\frac{\partial\Phi}{\partial F^x}\right)F^{yyy}+\left(\frac{\partial\Phi}{\partial G^x}\right)G^{yyy}\right)\end{bmatrix},$$

$$\boldsymbol{Q}=\begin{bmatrix}h_l^2\left(\left(\frac{\partial\Psi}{\partial F}\right)F^{xxx}+\left(\frac{\partial\Psi}{\partial G}\right)G^{xxx}\right)\\ \\ k_m^2\left(\left(\frac{\partial\Psi}{\partial F^x}\right)F^{yyy}+\left(\frac{\partial\Psi}{\partial G^x}\right)G^{yyy}\right)\end{bmatrix}$$

are 2×1 matrices computed at the central mesh point (x_l,y_m).

We need to construct additional approximations of first-order partial derivatives in X- and Y-directions with the help of functional approximations (III.4)-(III.7) and approximations of second-order derivatives. Thus, we define

$$\begin{aligned}\widehat{F}^{xx}_{l,m\pm1}=2[&F_{l+1,m\pm1}-(\alpha+1)F_{l,m\pm1}\\&+\alpha F_{l-1,m\pm1}]/[\alpha(\alpha+1)h_l^2],\end{aligned} \tag{III.8}$$

$$\begin{aligned}\widehat{G}^{xx}_{l,m\pm1}=2[&G_{l+1,m\pm1}-(\alpha+1)G_{l,m\pm1}\\&+\alpha G_{l-1,m\pm1}]/[\alpha(\alpha+1)h_l^2],\end{aligned} \tag{III.9}$$

$$\begin{aligned}\widehat{F}^{yy}_{l\pm1,m}=2[&F_{l\pm1,m+1}-(\beta+1)F_{l\pm1,m}\\&+\beta F_{l\pm1,m-1}]/[\beta(\beta+1)k_m^2],\end{aligned} \tag{III.10}$$

$$\begin{aligned}\widehat{G}^{yy}_{l\pm1,m}=2[&G_{l\pm1,m+1}-(\beta+1)G_{l\pm1,m}\\&+\beta G_{l\pm1,m-1}]/[\beta(\beta+1)k_m^2],\end{aligned} \tag{III.11}$$

$$\widetilde{F}^{x}_{l,m}=\widehat{F}^{x}_{l,m}+\sigma h_l(\widehat{\Phi}_{l+1,m}-\widehat{\Phi}_{l-1,m}-\widehat{F}^{yy}_{l+1,m}+\widehat{F}^{yy}_{l-1,m}), \tag{III.12}$$

$$\widetilde{G}^{x}_{l,m}=\widehat{G}^{x}_{l,m}+\sigma h_l(\widehat{\Psi}_{l+1,m}-\widehat{\Psi}_{l-1,m}-\widehat{G}^{yy}_{l+1,m}+\widehat{G}^{yy}_{l-1,m}), \tag{III.13}$$

$$\widetilde{F}^{y}_{l,m}=\widehat{F}^{y}_{l,m}+\rho k_m(\widehat{\Phi}_{l,m+1}-\widehat{\Phi}_{l,m-1}-\widehat{F}^{xx}_{l,m+1}+\widehat{F}^{xx}_{l,m-1}), \tag{III.14}$$

$$\widetilde{G}^{y}_{l,m}=\widehat{G}^{y}_{l,m}+\rho k_m(\widehat{\Psi}_{l,m+1}-\widehat{\Psi}_{l,m-1}-\widehat{G}^{xx}_{l,m+1}+\widehat{G}^{xx}_{l,m-1}), \tag{III.15}$$

where σ and ρ are free parameters to be obtained in such a way that the LTE of the final scheme tally with the pattern (II.8) in terms of order and accuracy.

Therefore, with the help of equations (III.8)-(III.15), the new functionals defined at the central mesh-point (x_l,y_m):

$$\begin{cases}\widehat{\Phi}_{l,m}=\Phi(x_l,y_m,F_{l,m},G_{l,m},\widetilde{F}^x_{l,m},\widetilde{G}^x_{l,m},\widetilde{F}^y_{l,m},\widetilde{G}^y_{l,m}),\\ \widehat{\Psi}_{l,m}=\Psi(x_l,y_m,F_{l,m},G_{l,m},\widetilde{F}^x_{l,m},\widetilde{G}^x_{l,m},\widetilde{F}^y_{l,m},\widetilde{G}^y_{l,m}),\end{cases} \tag{III.16}$$

for the following values

$$\begin{aligned}\sigma=&\alpha\beta(1+\alpha+\alpha^2)/[2(1+\alpha)\{\alpha(1-8\beta+\beta^2)\\&-(1+\alpha^2)(2-\beta+2\beta^2)\}],\\ \rho=&\alpha\beta(1+\beta+\beta^2)/[2(1+\beta)\{\alpha(1-8\beta+\beta^2)\\&-(1+\alpha^2)(2-\beta+2\beta^2)\}],\end{aligned}$$

gives rise to the following equations

$$\begin{aligned}\begin{bmatrix}\hat{\Phi}_{l,m}\\ \hat{\Psi}_{l,m}\end{bmatrix}=&\frac{1}{6}\begin{bmatrix}h_l^2(\sigma\rho(1+\alpha)+\alpha) & k_m^2(\sigma\rho(1+\beta)+\beta)\end{bmatrix}\boldsymbol{E}\\&+\begin{bmatrix}\Phi_{l,m}\\ \Psi_{l,m}\end{bmatrix}+\begin{bmatrix}O(h_l^3)\\ O(h_l^3)\end{bmatrix},\end{aligned} \tag{III.17}$$

where

$$\boldsymbol{E}=\begin{bmatrix}\frac{\partial\Phi}{\partial F}F^{xxx}+\frac{\partial\Phi}{\partial G}G^{xxx} & \frac{\partial\Psi}{\partial F}F^{xxx}+\frac{\partial\Psi}{\partial G}G^{xxx}\\ \\ \frac{\partial\Phi}{\partial F^x}F^{yyy}+\frac{\partial\Phi}{\partial F^y}G^{yyy} & \frac{\partial\Psi}{\partial F^x}F^{yyy}+\frac{\partial\Psi}{\partial G^y}G^{yyy}\end{bmatrix}_{(x_l,y_m)}.$$

With the help of equations (II.6), (III.3) and (III.16), a new nine-point compact EEM-FDM for the numerical solution of coupled EBVPs (I.1) is given by the system of difference equations for $l=1(1)L, m=1(1)M$ as

$$\begin{cases}\Delta_{h_l,k_m}F_{l,m}=2h_l^2k_m^2\sum_{(p,q)\in H}e_{p,q}\hat{\Phi}_{p,q}+T_{l,m},\\ \Delta_{h_l,k_m}G_{l,m}=2h_l^2k_m^2\sum_{(p,q)\in H}e_{p,q}\hat{\Psi}_{p,q}+T_{l,m},\end{cases} \tag{III.18}$$

The scheme (III.18) is third-order accurate for $\alpha\neq1$, $\beta\neq1$, and turn into fourth-order of accuracy for $\alpha=\beta=1$. For the numerical implementations, the system of equation (III.18) must be combined with the boundary values defined in (I.2).

Now, we shall discuss the method for solving system of equation (III.18). Let $\boldsymbol{F}=[F_{11},..,F_{L1},..,F_{1M},..,F_{LM}]^T$, $\boldsymbol{G}=[G_{11},..,G_{L1},..,G_{1M},..,G_{LM}]^T$, then the system of equation (III.18) in matrix-vector notation can be written as

$$\begin{cases}\boldsymbol{R}(\boldsymbol{F},\boldsymbol{G})=\boldsymbol{0}\\ \boldsymbol{S}(\boldsymbol{F},\boldsymbol{G})=\boldsymbol{0}\end{cases} \tag{III.19}$$

Assume that $(\boldsymbol{F}^{(I)},\boldsymbol{G}^{(I)})$ is an initial approximation to the exact solution values $(\boldsymbol{F}^{(I)}+\boldsymbol{\delta F},\boldsymbol{G}^{(I)}+\boldsymbol{\delta G})$. Then, we have

$$\begin{cases}\boldsymbol{R}(\boldsymbol{F}^{(I)}+\boldsymbol{\delta F},\boldsymbol{G}^{(I)}+\boldsymbol{\delta G})=\boldsymbol{0}\\ \boldsymbol{S}(\boldsymbol{F}^{(I)}+\boldsymbol{\delta F},\boldsymbol{G}^{(I)}+\boldsymbol{\delta G})=\boldsymbol{0}\end{cases} \tag{III.20}$$

The multi-dimensional truncated Taylor series expansion about $(\boldsymbol{F}^{(I)}, \boldsymbol{G}^{(I)})$ gives us

$$\begin{cases} \boldsymbol{R}(\boldsymbol{F}^{(I)}, \boldsymbol{G}^{(I)}) + \boldsymbol{\delta F}\left(\dfrac{\partial \boldsymbol{R}}{\partial \boldsymbol{F}}\right)_{(\boldsymbol{F}^{(I)}, \boldsymbol{G}^{(I)})} \\ +\boldsymbol{\delta G}\left(\dfrac{\partial \boldsymbol{R}}{\partial \boldsymbol{G}}\right)_{(\boldsymbol{F}^{(I)}, \boldsymbol{G}^{(I)})} = \boldsymbol{0}, \\ \boldsymbol{S}(\boldsymbol{F}^{(I)}, \boldsymbol{G}^{(I)}) + \boldsymbol{\delta F}\left(\dfrac{\partial \boldsymbol{S}}{\partial \boldsymbol{F}}\right)_{(\boldsymbol{F}^{(I)}, \boldsymbol{G}^{(I)})} \\ +\boldsymbol{\delta G}\left(\dfrac{\partial \boldsymbol{S}}{\partial \boldsymbol{G}}\right)_{(\boldsymbol{F}^{(I)}, \boldsymbol{G}^{(I)})} = \boldsymbol{0}, \end{cases} \tag{III.21}$$

Solving equation (III.21) for $\boldsymbol{\delta F}$ and $\boldsymbol{\delta G}$, we get

$$\begin{bmatrix} \boldsymbol{\delta F} \\ \boldsymbol{\delta G} \end{bmatrix} = -\boldsymbol{D}^{-1} \cdot \begin{bmatrix} \boldsymbol{R}\dfrac{\partial \boldsymbol{S}}{\partial \boldsymbol{G}} - \boldsymbol{S}\dfrac{\partial \boldsymbol{R}}{\partial \boldsymbol{G}} \\ \boldsymbol{S}\dfrac{\partial \boldsymbol{R}}{\partial \boldsymbol{F}} - \boldsymbol{R}\dfrac{\partial \boldsymbol{S}}{\partial \boldsymbol{F}} \end{bmatrix}_{(\boldsymbol{F}^{(I)}, \boldsymbol{G}^{(I)})}$$

where $\boldsymbol{D} = \begin{bmatrix} \dfrac{\partial \boldsymbol{R}}{\partial \boldsymbol{F}} & \dfrac{\partial \boldsymbol{R}}{\partial \boldsymbol{G}} \\ \dfrac{\partial \boldsymbol{S}}{\partial \boldsymbol{F}} & \dfrac{\partial \boldsymbol{S}}{\partial \boldsymbol{G}} \end{bmatrix}_{(\boldsymbol{F}^{(I)}, \boldsymbol{G}^{(I)})}$ is the Jacobian of the function $\boldsymbol{R}(\boldsymbol{F}, \boldsymbol{G})$ and $\boldsymbol{S}(\boldsymbol{F}, \boldsymbol{G})$.

Thus, an improved approximation of the solution values $(\boldsymbol{F}^{(I+1)}, \boldsymbol{G}^{(I+1)})$ is given by

$$\begin{bmatrix} \boldsymbol{F}^{(I+1)} \\ \boldsymbol{G}^{(I+1)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{F}^{(I)} \\ \boldsymbol{G}^{(I)} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\delta F} \\ \boldsymbol{\delta G} \end{bmatrix}, I = 0, 1, 2, ... \tag{III.22}$$

This system of equation can be solved by Newton-iterative procedure and converges, if the initial vector $(\boldsymbol{F}^{(0)}, \boldsymbol{G}^{(0)})$ is sufficiently close to the solution values [8], [18].

TABLE 1A
FOURTH-ORDER ACCURACY TO THE SOLUTION OF EQUATION (IV.2)

$L+1$	α	β	I	$\mathcal{L}_2^F$	$\mathcal{L}_2^G$	Θ_2^F	Θ_2^G
4	1.00	1.00	26	3.39e-04	2.45e-04	—	—
8	1.00	1.00	83	2.24e-05	1.33e-05	3.9	4.2
16	1.00	1.00	241	1.40e-06	7.79e-07	4.0	4.1
32	1.00	1.00	601	8.81e-08	4.71e-08	4.0	4.0

TABLE 1B
THIRD-ORDER ACCURACY TO THE SOLUTION OF EQUATION (IV.2)

$L+1$	α	β	I	$\mathcal{L}_2^F$	$\mathcal{L}_2^G$	Θ_2^F	Θ_2^G
4	0.83	1.15	26	3.79e-04	1.95e-04	—	—
8	0.91	1.08	81	2.57e-05	1.20e-05	3.9	4.0
16	0.96	1.03	234	1.38e-06	6.42e-07	4.2	4.2
32	0.99	1.00	556	5.10e-08	2.78e-08	4.8	4.5

IV. APPLICATION TO BIHARMONIC EQUATION

The two-space dimensional biharmonic equations are used to discuss deflection of plates and stokes flow of fluids [5], [9], [11]. The biharmonic equation of second kind considered here is

$$\Delta^2 F(x, y) = \Psi(x, y), \quad (x, y) \in \Omega \tag{IV.1}$$

TABLE 2A
SECOND-ORDER ACCURACY TO THE SOLUTION OF EQUATION (IV.2)

$L+1$	α	β	I	$\mathcal{L}_2^F$	$\mathcal{L}_2^G$	Θ_2^F	Θ_2^G
4	1.00	1.00	33	4.95e-03	3.43e-03	—	—
8	1.00	1.00	123	1.39e-03	8.22e-04	1.8	2.1
16	1.00	1.00	426	3.67e-04	1.97e-04	1.9	2.1
32	1.00	1.00	1417	9.23e-05	4.79e-05	2.0	2.0

TABLE 2B
FIRST-ORDER ACCURACY TO THE SOLUTION OF EQUATION (IV.2)

$L+1$	α	β	I	$\mathcal{L}_2^F$	$\mathcal{L}_2^G$	Θ_2^F	Θ_2^G
4	0.94	0.93	29	4.22e-04	2.37e-04	—	—
8	0.98	0.97	116	4.16e-04	2.44e-04	0.02	0.04
16	0.99	0.98	378	6.64e-05	3.38e-05	2.6	2.9
32	1.00	0.99	1333	4.25e-05	2.19e-05	0.6	0.6

where $\Psi(x, y) = e^{xy}[(x-1)^2 + (y+1)^2][(x+1)^2 + (y-1)^2]$. In a coupled manner, we can represent equation (IV.1) as

$$\begin{cases} \Delta F(x, y) = G(x, y) \\ \Delta G(x, y) = \Psi(x, y), \quad (x, y) \in \Omega \end{cases} \tag{IV.2}$$

The associated Dirichlet's boundary values along all the four boundary lines of Ω are obtained from the analytical solution values $F(x, y) = e^{xy}$ as a test function.

Initially, we solve the system of equation (IV.2) using the difference operators defined by the equations (II.1)-(II.4). An immediate EEM-FDM for (IV.2) are given by

$$\begin{cases} k_m^2 \mathcal{B}_x F_{l,m} + h_l^2 \mathcal{B}_y F_{l,m} = h_l^2 k_m^2 G_{l,m} \\ k_m^2 \mathcal{B}_x G_{l,m} + h_l^2 \mathcal{B}_y G_{l,m} = h_l^2 k_m^2 \Psi_{l,m}. \end{cases} \tag{IV.3}$$

The truncation error of the discretization scheme (IV.3) is calculated as $O(h_l + k_m)$, if $\alpha \neq 1 \vee \beta \neq 1$ and $O(h_l^2 + h_l k_m + k_m^2)$ if $\alpha = 1 \wedge \beta = 1$ respectively. Thus, we have only first order scheme in case of exponential expanding meshes and second order scheme for constant meshes spacing. Therefore, we have a set of first, second order compact scheme (IV.3) and third, fourth-order compact schemes (III.18). To test the benefit of proposed schemes, we have computed root mean squared error $(\mathcal{L}_2)$ of the analytical solution values $(F_{l,m}, G_{l,m})$ and numerical solution values $(f_{l,m}, g_{l,m})$ using the metrics

$$\mathcal{L}_2^F = \sqrt{\frac{1}{LM}\sum_{l=1}^{L}\sum_{m=1}^{M}|\epsilon_{l,m}|^2}, \mathcal{L}_2^G = \sqrt{\frac{1}{LM}\sum_{l=1}^{L}\sum_{m=1}^{M}|\mu_{l,m}|^2},$$

$$\Theta_2^F = \log_2\left[\frac{(2L+1)(2M+1)}{LM}\frac{\sum_{l=1}^{L}\sum_{m=1}^{M}|\epsilon_{l,m}|^2}{\sum_{l=1}^{2L+1}\sum_{m=1}^{2M+1}|\epsilon_{l,m}|^2}\right]^{1/2},$$

$$\Theta_2^G = \log_2\left[\frac{(2L+1)(2M+1)}{LM}\frac{\sum_{l=1}^{L}\sum_{m=1}^{M}|\mu_{l,m}|^2}{\sum_{l=1}^{2L+1}\sum_{m=1}^{2M+1}|\mu_{l,m}|^2}\right]^{1/2}$$

where $\epsilon_{l,m} = F_{l,m} - f_{l,m}$ and $\mu_{l,m} = G_{l,m} - g_{l,m}$ are the point-wise errors.

We have used Gauss-Seidel method to solve the system of equations (IV.3) with zero initial data and iterations is

discontinued after achieving $max|F_{l,m}^{(I+1)} - F_{l,m}^{(I)}| \leq 10^{-12}$ and $max|G_{l,m}^{(I+1)} - G_{l,m}^{(I)}| \leq 10^{-12}$, where I is the iterative count. For simplicity, we have taken $L = M$ in all the computational results. Table 1A-1B and Table 2A-2B shows the $\mathcal{L}_2^F$ and $\mathcal{L}_2^G$ along with iterative count using the methods (III.18) and (IV.3) respectively.

It is natural to observe that high order (third/fourth) scheme (Table 1A-1B) is superior than (first/second) order scheme (Table 2A-2B) in terms of $\mathcal{L}_2$-norm. Comparing Table 1A-1B and Table 2A-2B, the iteration count is minimum in case of third-order EEM-FDM compared to first order exponential meshes and second, fourth-order uniform meshes FDM schemes.

TABLE 3A
ACCURACIES OF SOLUTION VALUES OF EQUATION (V.1) AT $\lambda = 0.02$

$L+1$	α	β	I	$\mathcal{L}_2^F$	$\mathcal{L}_2^G$	Θ_2^F	Θ_2^G
4	1.00	1.00	21	9.44e-04	5.08e-04	—	—
8	1.00	1.00	23	1.21e-04	4.23e-05	3.0	3.6
16	1.00	1.00	40	9.33e-06	2.60e-06	3.7	4.0
32	1.00	1.00	100	5.93e-07	1.55e-07	4.0	4.1

TABLE 3B
ACCURACIES OF SOLUTION VALUES OF EQUATION (V.1) AT $\lambda = 0.02$

$L+1$	α	β	I	$\mathcal{L}_2^F$	$\mathcal{L}_2^G$	Θ_2^F	Θ_2^G
4	0.70	1.10	15	5.11e-04	2.82e-04	—	—
8	0.83	1.02	19	6.78e-05	2.71e-05	2.9	3.4
16	0.91	1.01	39	7.65e-06	2.18e-06	3.1	3.6
32	0.95	1.00	97	8.24e-07	1.60e-07	3.2	3.8

TABLE 4A
ACCURACIES OF SOLUTION OF EQUATION (V.3), $\Phi(y) = -\cos(y)$

$L+1$	α	β	I	$\mathcal{L}_2^F$	$\mathcal{L}_2^G$	Θ_2^F	Θ_2^G
4	1.00	1.00	23	9.23e-05	6.22e-05	—	—
8	1.00	1.00	67	2.33e-05	1.20e-05	2.0	2.4
16	1.00	1.00	137	2.83e-06	1.19e-06	3.0	3.3
32	1.00	1.00	156	2.13e-07	7.93e-08	3.7	3.9

TABLE 4B
ACCURACIES OF SOLUTION OF EQUATION (V.3), $\Phi(y) = -\cos(y)$

$L+1$	α	β	I	$\mathcal{L}_2^F$	$\mathcal{L}_2^G$	Θ_2^F	Θ_2^G
4	1.59	1.06	19	4.01e-05	2.18e-05	—	—
8	1.49	1.06	40	8.96e-06	4.24e-06	2.2	2.4
16	1.48	1.03	60	6.90e-07	2.55e-07	3.7	4.1
32	1.30	1.01	65	8.21e-08	1.41e-08	3.1	4.2

V. APPLICATION TO NAVIER-STOKES EQUATIONS

In this section, the proposed EEM-FDM is applied to solve two-space dimensional stationary Navier-Stokes equation for viscous incompressible fluid [21]. Let us define the stream function $W(x,y)$ by the formula $F(x,y) = \partial W/\partial y$ and $G(x,y) = -\partial W/\partial x$, then the equation of fluid motion

$$W^y(\Delta W)^x - W^x(\Delta W)^y = \lambda\Delta^2 W, \quad (x,y) \in \Omega \qquad \text{(V.1)}$$

TABLE 5A
ACCURACIES OF SOLUTION OF EQUATION (V.3), $\Phi(y) = e^y + e^{-y}$

$L+1$	α	β	I	$\mathcal{L}_2^F$	$\mathcal{L}_2^G$	Θ_2^F	Θ_2^G
4	1.00	1.00	23	5.70e-05	4.02e-05	—	—
8	1.00	1.00	67	1.41e-05	8.46e-06	2.0	2.2
16	1.00	1.00	144	2.00e-06	1.06e-06	2.8	3.0
32	1.00	1.00	189	1.56e-07	7.81e-08	3.7	3.8

TABLE 5B
ACCURACIES OF SOLUTION OF EQUATION (V.3), $\Phi(y) = e^y + e^{-y}$

$L+1$	α	β	I	$\mathcal{L}_2^F$	$\mathcal{L}_2^G$	Θ_2^F	Θ_2^G
4	1.14	0.97	22	2.91e-05	1.96e-05	—	—
8	1.09	0.97	65	8.40e-06	3.08e-06	1.8	2.7
16	1.03	0.98	111	1.90e-06	6.78e-07	2.1	2.2
32	1.03	0.99	115	1.23e-07	4.56e-08	3.9	3.9

is equivalent to the coupled system of equation

$$\begin{cases} \lambda\Delta F = FF^x + GF^y, \\ \lambda\Delta G = FG^x + GG^y - 2e^{-2y}, \\ F^x + G^y = 0, \quad (x,y) \in \Omega \end{cases} \qquad \text{(V.2)}$$

In order to test the computational convergence order and accuracies of solution values, we have taken $F(x,y) = -[\sin(x) + \cos(x)]e^{-y}$ and $G(x,y) = [\sin(x) - \cos(x)]e^{-y}$, so that $W(x,y) = [\sin(x) + \cos(x)]e^{-y}$ satisfy the equation (5.1). The Dirichlet's boundary values on $F(x,y)$ and $G(x,y)$ along four boundary lines of Ω can be obtained from the analytical solutions as a test procedure. The accuracies of solution values at $\lambda = 0.02$ both in case of third- and fourth-order method are enumerated in Table 3A-3B.

Next, consider the Navier-Stokes equation appearing in plane flow of a viscous incompressible fluid subjected to the transverse forces [16],

$$\lambda\Delta^2 F = F^y(\Delta F)^x - F^x(\Delta F)^y - \Phi(y) \qquad \text{(V.3)}$$

where $F(x,y)$ is stream function.

If $\Delta F = G$, then

$$\lambda\Delta G = F^y G^x - F^x G^y - \Phi(y) \qquad \text{(V.4)}$$

We discuss the numerical solution for $\Phi(y) = -\cos(y)$ with the analytic solution values $F(x,y) = (1+\lambda^2)^{-1}[\sin(y) + \lambda\cos(y)] - x + y$, and $\Phi(y) = e^y + e^{-y}$ with the analytic solution values $F(x,y) = e^{-x} + (1-\lambda)^{-1}e^{-y} - (1+\lambda)^{-1}e^y + x - \lambda y$. Table 4A-4B and Table 5A-5B represents the accuracies and computational order of convergence both in case of third and fourth-order FDM for $\lambda = 0.01$.

As a last example, we consider stationary Navier-Stokes equations describing the motion of a viscous incompressible fluid induced by two parallel disks, moving towards each other [4]. The governing equation is given by

$$(F^y+\tau x)(\Delta F)^x - (F^x - \tau y)(\Delta F)^y + 2\tau\Delta F = \lambda\Delta^2 F, \qquad \text{(V.5)}$$

Introducing $\Delta F = G$, gives us

$$\lambda\Delta G = (F^y+\tau x)G^x - (F^x-\tau y)G^y + 2\tau G, (x,y) \in \Omega \qquad \text{(V.6)}$$

TABLE 6A
ACCURACIES OF SOLUTION OF EQUATION (V.5) AT $\tau = -1$

$L+1$	α	β	I	$\mathcal{L}_2^F$	$\mathcal{L}_2^G$	Θ_2^F	Θ_2^G
4	1.00	1.00	46	5.42e-02	3.16e-02	—	—
8	1.00	1.00	160	3.43e-03	1.75e-03	4.0	4.2
16	1.00	1.00	492	2.21e-04	1.02e-04	4.0	4.1
32	1.00	1.00	1328	1.38e-05	6.20e-06	4.0	4.0

TABLE 6B
ACCURACIES OF SOLUTION OF EQUATION (V.5) AT $\tau = -1$

$L+1$	α	β	I	$\mathcal{L}_2^F$	$\mathcal{L}_2^G$	Θ_2^F	Θ_2^G
4	0.68	1.00	43	1.96e-02	1.00e-02	—	—
8	0.84	1.00	141	1.26e-03	5.79e-04	4.0	4.1
16	0.92	1.00	414	8.31e-05	3.52e-05	3.9	4.0
32	0.95	1.00	863	3.89e-06	1.55e-06	4.4	4.5

The theoretical solution values are given by

$$F(x,y) = \begin{cases} x[\tau y + e^{-2y\sqrt{\tau/\lambda}} + e^{2y\sqrt{\tau/\lambda}}], \tau > 0 \\ x[\tau y + \cos(2y\sqrt{-\tau/\lambda}) + \sin(2y\sqrt{-\tau/\lambda})], \tau < 0 \end{cases}$$

Table 6A-6B and Table 7A-7B represent the accuracies of solution values for $\tau = -1$ and $\tau = 1$ respectively for $\lambda = 0.25$. In all the tabulated values, the exponential expanding mesh parameters α and β tends towards unity with diminution in mesh spacing. The root mean squared errors in case of EEM-FDM is less than that computed with constant mesh-FDM. Moreover, the iteration count and consequently the computing time is also less, if we apply third-order EEM-FDM. The experiments of Navier-Stokes equations in each cases discussed above with first(second) order compact scheme gives rise to unstable results and therefore not mentioned as a tabular values.

VI. CONCLUSION

In this article, we have developed a class of first, second, third and fourth-order accurate compact schemes for the numerical solution values of mildly non-linear EBVPs. The first and third-order methods are based on a non-uniform mesh network while second and fourth-order schemes are based on uniform mesh spacing. It is expected that higher (four) order method dipicts better results but due to layer behaviour, it has been shown that third-order EEM-FDM is more efficient both in terms of solution error and computational efficiency. The proposed method can be easily extended to the system of mildly non-linear hyperbolic and parabolic partial differential equations.

TABLE 7A
ACCURACIES OF SOLUTION OF EQUATION (V.5) AT $\tau = 1$

$L+1$	α	β	I	$\mathcal{L}_2^F$	$\mathcal{L}_2^G$	Θ_2^F	Θ_2^G
4	1.00	1.00	27	3.40e-01	1.95e-01	—	—
8	1.00	1.00	62	2.17e-02	1.15e-02	4.0	4.1
16	1.00	1.00	188	1.42e-03	6.93e-04	3.9	4.1
32	1.00	1.00	555	9.05e-05	4.23e-05	4.0	4.0

TABLE 7B
ACCURACIES OF SOLUTION OF EQUATION (V.5) AT $\tau = 1$

$L+1$	α	β	I	$\mathcal{L}_2^F$	$\mathcal{L}_2^G$	Θ_2^F	Θ_2^G
4	1.00	0.92	25	2.81e-02	1.61e-02	—	—
8	1.00	0.96	58	3.09e-03	1.25e-03	3.2	3.7
16	1.00	0.98	152	1.99e-04	7.91e-05	4.0	4.0
32	1.00	0.99	376	1.26e-05	4.90e-06	4.0	4.0

REFERENCES

[1] Aris, R., *The mathematical theory of diffusion and reaction in permeable catalysts*, Oxford University Press, London, 1975.
[2] Boglaev, I., *Numerical solution of coupled systems of non-linear elliptic equations*, Numer. Meth. Part. D. E., **28**(2012), 621-640.
[3] Britz, D., *Digital simulation in electrochemistry*, Springer Berlin, Heidelberg, 2005.
[4] Craik, A.D.D., *The stability of unbounded two- and three-dimensional flows subjected to body forces: Some exact solutions*, J. Fluid Mech., **198**(1989), 275-292.
[5] Ingham, D.B., Kelmanson, M.A., *Boundary integral equation analysis of singular, potential and biharmonic problems*, Springer-Verlag, Berlin, 1984.
[6] Jha, N., Mohanty, R.K., Chauhan, V., *Efficient algorithms for fourth and sixth-order two-point non-linear boundary value problems using non-polynomial spline approximations on a geometric mesh*, Comput. Appl. Math., (2014), 1-16.
[7] Juan, M., Ugail, H., *A general 4^{th}- Order PDE method to generate a Bezier surfaces from the boundary*, Comput. Aided Geom. D., **23**(2006), 208-225.
[8] Kelly, C.T., *Iterative methods for linear and non-linear equations*, SIAM, Philadelphia, PA, 1995.
[9] Landau, M.D., Lifshitz, E.M., *Theory of elasticity*, Pergamon Press, Oxford, 1986.
[10] Loitsyanskiy, L.G., *Mechanics of liquids and gases*, Begell House, New York, 1996.
[11] Martin, L., Lesnic, D., *The method of fundamental solutions for inverse boundary value problems associated with the two-dimensional biharmonic equation*, Math. Comput. Model., **42**(2005), 261-278.
[12] Mohanty, R.K., *Fourth order finite difference methods for the system of 2-D nonlinear elliptic equations with variable coefficients*, Int. J. Comput. Math., **46**(1992), 195-206.
[13] Mohanty, R.K., *A new high accuracy finite difference discretization for the solution of 2D non-linear biharmonic equations using coupled approach*, Numer. Meth. Part. D. E., **26**(2010), 931-944.
[14] Mohanty, R.K., Dai, W., Han, F., *A new high accuracy method for two-dimensional biharmonic equation with non-linear third derivative terms: Application to Navier-Stokes equations of motion*, Int. J. Comput. Math., **92**(2015), no. 8, 1574-1590.
[15] Pao, C.V., *Convergence of coupled system of non-linear finite difference elliptic equations*, Differential Integ Equations, **4**(1990), 783-798.
[16] Polyanin, A.D., *Handbook of linear partial differential equations for engineers and scientists*, CRC Press, 2010.
[17] Rigal, A., *SOR method for coupled elliptic partial differential equations*, J. Comput. Phys., **71**(1987), no. 1, 181-193.
[18] Saad, Y., *Iterative methods for sparse linear system*, SIAM, Philadelphia, PA, 2003.
[19] Selvadurai, A.P.S., *Partial differential equations in mechanics 2*, Springer Science & Business Media, 2000.
[20] Sweers, G., Troy, W.C., *On the bifurcation curve for an elliptic system of FitzHugh-Nagumo type*, Physica D, **177**(2003), 1-22.
[21] Tian, Z.F., Dai, S.Q., *High-order compact exponential finite difference methods for convection-diffusion type problems*, J. Comput. Phys., **220**(2007), no. 2, 952-974.
[22] Watanabe, Y., *A numerical verification method for two-coupled elliptic-partial differential equations*, Japan J. Indust. Appl. Math., **26**(2009), 233-247.

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

Analysis of a Delayed HIV Infection Model

Saroj Kumar Sahani
South Asain University
New Delhi,110021
India
Email: sarojkumar@sau.ac.in

Yashi
South Asain University
New Delhi,110021
India
Email: yashiraj19@gmail.com

***Abstract*—A multiple delayed model of HIV infection has been proposed in this paper. The first delay explains the time for an uninfected cell to become infected. This delay is termed as intracellular delay. The second delay known as the immunological delay describes the immune activation time. The basic reproduction number for the model has been given and the local stability analysis has been performed. The inclusion of the viral loss term, that is the loss of virus due to the bursting of the infected cells has also been included.**

I. Introduction

With the advances in mathematical modeling, a greater understanding of the characteristics of various disease and biological phenomena has been possible. This comprehensive understanding has enabled people to formulate various model governing the interactions between different biological components, and thus generate computer simulations to further understand the mechanism governing these interactions. Mathematical study of the epidemiology of various diseases provides an insight into its dynamics, and thus helps to propose strategies to control its spread.

Human immunodeficiency virus (HIV-1) is a retrovirus that causes acquired immunodeficiency syndrome (AIDS). HIV virus affects the CD4+ T cells of the human body. It replicates itself in the tissues and cells of the body. As soon as the body is infected by these viruses, the immune response of the body is activated. It has been observed that CTLs are the principal host immune agents which restrict viral duplication in vivo and thus, regulate the viral load [1], [2].

The HIV infected patients are treated with different kinds of drugs. Reverse Transcriptase Inhibitor(RTI) and Protease Inhibitor(PI) are the two very important and common drugs being used. The function of RTI is to block the function of the reverse trasciptase and thus prevent the conversion of viral RNA to DNA. Protease inhibitor on the other hand counters viral replication. Thus inhibiting viral reproduction. These drugs together form the therapy called Highly Active Anti Retro-Viral Therapy(HAART). In HAART both the drugs are administered together to enhance the effect on HIV virus and maintain its load at a minimum level. Mathematical modeling has played an important role in understanding the dynamics of HIV infection [2]–[8]. In recent times models with drug therapy have gained prominence and are being developed [9]–[12]. A major drawback of the previous models is, that most of the models ignored the effect of delays on the system. Herz et. al [13] proposed the first delayed model of HIV infection to account for the time lag between entry of the virus to viral production termed as intracellular delay. Various other models have been proposed since then [14]–[19]. The delay in the generation of immune response is termed as immunological delay. This delay has been discussed in many papers [19]–[22]. However only some authors have considered the effect of multiple delays on the system [20], [23]–[26]. The incorporation of multiple delays generates a better dynamics of the disease. However the analysis of such models is much complex and cumbersome. A drawback of the previous studies is the ignorance of viral loss in the course of infection. It was argued that loss of the virus was minuscule and can be included into the virus removal term [27]. However Heffernan et al [28], [29] concluded that this loss has a major role in determining the covariance between helper T cells and viral load. In the paper by Wang et al [30] they have concluded that viral waning term affects the reproduction number as well as the infected equilibrium points, thus affecting the overall viral dynamics of the system. This paper takes into account the loss of uninfected T cells, loss of viruses, CTL immune response and also includes the intracellular and immunological delays. The model has been made more practical by incorporating combined drug therapy in the system. The organization of the paper is follows. In section 2 the proposed model has been discussed. Section 3 deals with equilibrium states and threshold dynamics of the system. Section 4 deals with positivity and boundedness of the model. Local stability analysis have been discussed in section 5. The numerical simulation has been discussed in section 7 followed by conclusion and discussion in section 8.

II. Model

In this paper a four dimensional model with anti-retroviral therapy, multiple delays and viral loss term has been proposed.

$$\begin{aligned}
x_1'(t) &= s - dx_1(t) - b(1-\epsilon_r)x_1(t)x_3(t) - \frac{kx_1(t)x_3(t)}{x_1(t)+B} \\
x_2'(t) &= b(1-\epsilon_r)x_1(t-\tau_1)x_3(t-\tau_1) - \delta x_2(t) - d_1x_4(t)x_2(t) \\
x_3'(t) &= N\delta(1-\epsilon_p)x_2(t) - cx_3(t) - \eta b x_3(t)x_1(t) \\
x_4(t) &= px_2(t-\tau_2)x_4(t-\tau_2) - d_2x_4(t).
\end{aligned} \tag{1}$$

This model deals with uninfected T cells (x_1), infected T cells (x_2), viruses (x_3) and CTL immune response cells (x_4).

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

The intracellular delay (τ_1) and immunological delay (τ_2) have also been included in the model. The descriptions of various parameters has been given in Table I.

The initial conditions of (1) are as follows:

$$x_1(\theta) = \xi_1(\theta), x_2(\theta) = \xi_2(\theta), x_3(\theta) = \xi_3(\theta), x_4(\theta) = \xi_4(\theta), \tag{2}$$

$\theta \in [-\tau, 0]$, where $\phi = (\phi_1, \phi_2, \phi_3, \phi_4) \in C([-\tau, 0], R^4_+)$
with $\xi_i(0) > 0, (\theta \in [-\tau, 0], i = 1, 2, 3, 4)$

III. Equilibrium states and Threshold dynamics

The system (1) is analyzed by determining its steady states. There exists two steady states of the given system (1). The infection free steady state I_0 and the infected steady state I_1.

$$\begin{aligned} I_0 =& (x_1^0, 0, 0, 0) = \left(\frac{s}{d}, 0, 0, 0\right) \\ I_1 =& (x_1^*, x_2^*, x_3^*, x_4^*) \quad \text{where} \\ x_2^* =& \frac{d_2}{p}, x_3^* = \frac{N\epsilon_p \delta d_2}{p(c + bx_1^*\eta)} \\ x_4^* =& \frac{x_1^*(bN\delta\epsilon_r\epsilon_p - b\delta\eta) - c\delta}{(c + bx_1^*\eta)d_1} \end{aligned} \tag{3}$$

the value of x_1^* can be calculated from the following cubic equation

$$\begin{aligned} x_1^{*3} bdp\eta + x_1^{*2}(cdp - bBdp\eta - bps\eta + bN\epsilon_p\epsilon_r\delta d_2) \\ + x_1^*(Bcdp - cps - bBps\eta + kN\epsilon_p\delta d_2 - bBN\epsilon_p\epsilon_r\delta d_2) \\ - Bcps = 0 \end{aligned}$$

As the steady state values are positive, so from Eq (3) it is clear that $x_1^* > c/(bN\epsilon_r\epsilon_p - b\eta)$, also $x_1^* < s/d$. Thus, using these two relations, a required conditions for the steady states to exist is $c/(bN\epsilon_r\epsilon_p - b\eta) < s/d$. Analogously, the reproduction number $R_0 > 1$, where

$$R_0 = \frac{s(bN(1-\epsilon_r)(1-\epsilon_p) - b\eta)}{cd}$$

IV. Positivity and boundedness

Theorem 4.1: Suppose $X(t, \phi)$ is the solution of the system (1) with initial conditions given by (2). Then $X(t, \phi)$ is nonnegative and ultimately bounded.

A. Stability of I_0

This section is concerned with the local stability analysis of the equilibrium I_0. For the sake of simplicity $1 - \epsilon_r = \epsilon_r'$ and $1 - \epsilon_p = \epsilon_p'$ unless stated otherwise in the following section.

Theorem 4.2: If $R_0 < 1$ then infection free equilibrium I_0 is locally asymptotically stable and unstable if $R_0 > 1$.

Proof: The characteristic equation at I_0 is given by

$$\begin{aligned} (d + \lambda)(d_2 + \lambda)\left(\lambda^2 + \lambda\left(c + \delta + \frac{bs\eta}{d}\right)\right. \\ \left. + c\delta - \frac{bNs\epsilon_p'\epsilon_r'\delta e^{-\lambda\tau_1}}{d} + \frac{bs\delta\eta}{d}\right) = 0 \end{aligned} \tag{4}$$

Clearly (4) has two negative roots $-d$ and d_2. The remaining two roots are determined from the transcendental equation

$$\lambda^2 + \lambda\left(c + \delta + \frac{bs\eta}{d}\right) + c\delta + \frac{bs\delta\eta}{d} - \frac{bNs\epsilon_p'\epsilon_r'\delta e^{-\lambda\tau_1}}{d} = 0 \tag{5}$$

When $\tau_1 = 0$, (5) reduces to an algebraic equation

$$\lambda^2 + \lambda\left(c + \delta + \frac{bs\eta}{d}\right) + c\delta(1 - R_0) = 0 \tag{6}$$

If $R_0 < 1$, then clearly (6) has two negative real roots. And so I_0 is asymptotically stable for $\tau_1 = 0$.
Let $\lambda = i\omega$, $\omega > 0$ be root of (4), separating real and imaginary parts, it follows that

$$\begin{aligned} c\delta + \frac{bs\eta\delta}{d} - \omega^2 =& \frac{bNs\epsilon_p'\epsilon_r'\delta}{d}\cos(\omega\tau_1) \\ \left(c + \delta + \frac{bs\eta}{d}\right)\omega =& -\frac{bNs\epsilon_p'\epsilon_r'\delta}{d}\sin(\omega\tau_1) \end{aligned} \tag{7}$$

Squaring and adding the two equations of (7)

$$\begin{aligned} \omega^4 + \left(\frac{b^2s^2\eta^2}{d^2} + \frac{2cbs\eta}{d} + \delta^2 + c^2\right)\omega^2 \\ + c^2\delta^2\left(1 - \frac{b^2N^2s^2\epsilon_p'^2\epsilon_r'^2}{c^2d^2} + \frac{2bs\eta}{cd} + \frac{b^2s^2\eta^2}{c^2d^2}\right) = 0 \end{aligned} \tag{8}$$

If $R_0 < 1$, then $\left(1 - \frac{b^2N^2s^2\epsilon_p'^2\epsilon_r'^2}{c^2d^2} + \frac{2bs\eta}{cd} + \frac{b^2s^2\eta^2}{c^2d^2}\right) > 0$. Thus (8) has no positive roots. So there does not exist any root of (8) which can cross the imaginary axis as τ_1 increases. Hence it is concluded that I_0 is locally asymptotically stable for $\tau_1 \geq 0$ if $R_0 < 1$. Now consider the case for $R_0 > 1$. Then from (5), let

$$f(\lambda) = \lambda^2 + \lambda\left(c + \delta + \frac{bs\eta}{d}\right) + c\delta + \frac{bs\delta\eta}{d} - \frac{bNs\epsilon_p'\epsilon_r'\delta e^{-\lambda\tau_1}}{d} \tag{9}$$

$$f(0) = c\delta(1 - R_0) < 0, \qquad \lim_{\lambda\to\infty} f(s) = +\infty \tag{10}$$

Hence $f(\lambda) = 0$, has at least one positive real root. Thus it can be concluded that, if $R_0 > 1$, I_0 is unstable. ■

B. Stability of I_1

Characteristic equation of (1) at the infected equilibrium is

$$P(\lambda) + e^{-\lambda\tau_2}Q(\lambda) + e^{-\lambda\tau_1}R(\lambda) + e^{-\lambda(\tau_2+\tau_1)}S(\lambda) = 0 \tag{11}$$

That is,

$$\begin{aligned} \lambda^4 + p_3\lambda^3 + p_2\lambda^2 + p_1\lambda + p_0 + e^{-\lambda\tau_2}(q_3\lambda^3 + q_2\lambda^2 + q_1\lambda + q_0) \\ + e^{-\lambda\tau_1}(r_2\lambda^2 + r_1\lambda + r_0) + e^{-\lambda(\tau_2+\tau_1)}(s_1\lambda + s_0) = 0 \end{aligned} \tag{12}$$

Taking $\tau_1 = \tau_2 = 0$, (12) gives

$$\begin{aligned} \lambda^4 + (p_3 + q_3)\lambda^3 \\ + (p_2 + q_2 + r_2)\lambda^2 + (p_1 + q_1 + r_1 + s_1)\lambda \\ + p_0 + q_0 + r_0 + s_0 = 0 \end{aligned} \tag{13}$$

Using Routh-Hurwitz criteria, all the roots of (13) have negative real parts if and only if

(a) $(p_3+q_3)>0,(p_2+q_2+r_2)>0,$
$(p_1+q_1+r_1+s_1)>0,(p_0+q_0+r_0+s_0)>0$
$(H1)$

(b) $(p_3+q_3)(p_2+q_2+r_2)$
$-(p_1+q_1+r_1+s_1)>0$

(c) $(p_3+q_3)(p_2+q_2+r_2)(p_1+q_1+r_1+s_1)$
$-(p_3+q_3)^2(p_0+q_0+r_0+s_0)-(p_1+q_1+r_1+s_1)^2>0$

So the following theorem can be stated.

Theorem 4.3: Let $R_0>1$ and $\tau_1=\tau_2=0$, then I_1 is locally asymptotically stable if $(H1)$ holds.

Let $\tau_2=0$ in (12), Then the equation reduces to

$$\lambda^4+(p_3+q_2)\lambda^3+(p_2+q_2)\lambda^2+(p_1+q_1)\lambda+p_0+q_0 + e^{-\lambda\tau_1}(r_2\lambda^2+(r_1+s_1)\lambda+r_0+s_0)=0 \tag{14}$$

For $\lambda=0$, (14) reduces to $q_0+s_0+r_0+s_0\neq 0$. Hence $\lambda=0$ is not a root of (14).

Let $\lambda=i\omega$ be a root of (14), Separating the real and imaginary part gives

$$\begin{aligned} &p_0+q_0-(p_2+q_2)\omega^2+\omega^4= \\ &(r_2\omega^2-r_0-s_0)\cos\omega\tau_1-(r_1\omega+s_1\omega)\sin\omega\tau_1 \\ &(p_1+q_1)\omega-(p_3+q_3)\omega^3= \\ &-(r_2\omega^2-r_0-s_0)\sin\omega\tau_1-(r_1\omega+s_1\omega)\cos\omega\tau_1 \end{aligned} \tag{15}$$

Squaring and adding both the equations of (15) and taking $\omega^2=y$, gives

$$\begin{aligned} F(y)=&y^4+((p_3+q_3)^2-2(p_2+q_2))y^3 \\ &+((p_2+q_2+r-2)(p_2+q_2-r_2) \\ &-(2p_1+q_1)(q_3+p_3)+2(p_0+q_0))y^2 \\ &+((p_1+q_1+r_1+s_1)(p_1+q_1-r_1-s_1) \\ &-2(p_2+q_2)(p_0+q_0)+2r_2(r_0+s_0))y \\ &+(p_0+q_0+r_0+s_0)(p_0+q_0-r_0-s_0) \end{aligned} \tag{16}$$

Using Routh-Hurwitz criteria, it can be concluded that roots of (16) have negative real parts if

(a) $((p_3+q_3)^2-2(p_2+q_2))>0$ $(H2)$

(b) $((p_2+q_2+r-2)(p_2+q_2-r_2)$
$-(2p_1+q_1)(q_3+p_3)+2(p_0+q_0))>0,$
$((p_1+q_1+r_1+s_1)(p_1+q_1-r_1-s_1)$
$-2(p_2+q_2)(p_0+q_0)+2r_2(r_0+s_0))>0$

(c) $(p_0+q_0-r_0-s_0)>0$

Thus based on the above discussion, the following theorem can be stated.

Theorem 4.4: Let $R_0>1$ and $\tau_2=0,\tau_1\neq 0$, then I_1 is locally asymptotically stable if $(H1)$ and $(H2)$ holds.

It is clear that if $(p_0+q_0-r_0-s_0)<0$, holds then (16) has positive real roots. Corresponding to the positive roots of (16), the values of the delay τ_1 can be given by, for $i=1,2,3,4$

$$\tau_{1_i}=\frac{1}{\omega}\cos^{-1}\left\{\frac{\triangle_n}{\bigtriangledown_d}\right\}+\frac{2n\pi}{\omega}\quad n=0,1,2..... \tag{17}$$

where

$$\begin{aligned} \triangle_n=&(p_0+q_0)(r_0+s_0)-q_1(r_0+s_0)\omega \\ &+((p_0+q_0)(-r_2+p_1(r_1+s_1))-(p_2+q_2)(r_0+s_0))\omega^2 \\ &+(q_1r_2+(p_3+q_3)(r_0+s_0))\omega^3 \\ &+(r_0+s_0+(p_2+q_2)(r_2-p_1(r_1+s_1))\omega^4-r_2(p_3+q_3))\omega^5 \\ &+(p_1(r_1+s_1)-r_2)\omega^6 \end{aligned}$$

$$\bigtriangledown_n=-(r_0+s_0)^2+2r_2(r_0+s_0)\omega^2-p_1(r_1+s_1)^2\omega^3-r_2^2\omega^4$$

The critical value τ_1^* of τ_1 is given by taking the min of all the values of τ_1 given by (17). Thus,

$$\tau_1^*=\min\tau_{1_i}\ \text{ for all }\ i \tag{18}$$

Differentiating (14) with respect to τ_1 and then rearranging gives,

$$\begin{aligned} &\left(\frac{d\lambda}{d\tau_1}\right)^{-1}= \\ &-\frac{4\lambda^3+3(p_3+q_3)\lambda^2+2(p_2+q_2)\lambda+(p_1+q_1)}{\lambda(\lambda^4+(p_3+q_3)\lambda^3+(p_2+q_2)\lambda^2+(p_1+q_1)\lambda+(p_0+q_0))} \\ &+\frac{2r_2\lambda+(r_1+s_1)}{\lambda(r_2\lambda^2+(r_1+s_1)\lambda+r_0+s_0)}-\frac{\tau_1}{\lambda} \end{aligned} \tag{19}$$

For Hopf bifurcation, the tranversability conditions needs to be verified. Thus to analyze the crossing of the imaginary axis by the eigenvalues, substituting $\lambda(\tau_1)=i\omega^*$ in Eq (19) and taking its real part,

$$\begin{aligned} &\text{sign}\left(\frac{d\Re(\lambda)}{d\tau_1}\bigg|_{\lambda=i\omega^*}\right) \\ =&\text{sign}\left(\Re\left(\frac{d\lambda}{d\tau_1}\right)^{-1}\bigg|_{\lambda=i\omega^*}\right) \\ =&\text{sign}F'(\omega*^2) \end{aligned} \tag{20}$$

Thus based on the above discussion the following theorem can be stated

Theorem 4.5: Let $R_0>1,\tau_2=0,\ \tau_1\neq 0$, and $(p_0+q_0-r_0-s_0)<0$ holds, then there exits a critical value τ_1^* and $\omega*$ such that the infected equilibrium I_1 is asymptotically stable in the interval $[0,\tau_1^*)$. Also if $(p_0+q_0-r_0-s_0)<0$ holds and $F'(\omega*^2)\neq 0$, then Hopf bifurcation is exhibited by system (1) at the critical value τ_1^*.

Taking $\tau_1=0,\tau_2\neq 0$, (12) is reduced to

$$\lambda^4+p_3\lambda^3+(p_2+r_2)\lambda^2+(p_1+r_1)\lambda+p_0+r_0 + e^{-\lambda\tau_2}(q_3\lambda^3+q_2\lambda^2+(q_1+s_1)\lambda+q_0+s_0)=0 \tag{21}$$

For $\lambda=0$, (21) reduces to $p_0+r_0+q_0+s_0\neq 0$. Hence $\lambda=0$ is not a root of (21).

Let $\lambda=i\omega$ be a root of (21), Separating the real and imaginary part gives

$$\begin{aligned} p_0+q_0-(p_2+q_2)\omega^2+p_4\omega^4=&\cos\omega\tau_2(-s_0-r_0+r_2\omega^2) \\ &+\sin\omega\tau_2(-s_1\omega-r_1\omega+r_3\omega^3) \\ p_1\omega+q_1\omega-p_3\omega^3=&\cos\omega\tau_2(-s_1\omega-r_1\omega+r_3\omega^3) \\ &-\sin\omega\tau_2(-s_0-r_0+r_2\omega^2) \end{aligned} \tag{22}$$

Squaring and adding the above equations, and taking $\omega^2 = z$ (22) is equal to

$$\begin{aligned} G(z) &= z^4 + (p_3^2 - q_3^2 - 2(r_2 + p_2))z^3 + ((p_2 + r_2)^2 \\ &- q_2^2 - 2p_3(p_1 + r_1) + 2q_3(q_1 + s_1) + 2r_0 + 2p_0)z^2 \\ &+ ((p_1 + r_1)^2 - (q_1 + s_1)^2 - 2(p_2 + r_2)(p_0 + r_0) \\ &+ 2q_2(q_0 + s_0))z + (p_0 + r_0 + s_0 + q_0)(p_0 + r_0 - s_0 - q_0) \end{aligned} \tag{23}$$

Using Routh-Hurwitz criteria, it can be concluded that roots of (23) have negative real parts if

(a) $(p_3^2 - q_3^2 - 2(r_2 + p_2)) > 0$, $(K1)$
(b) $(p_2+r_2)^2-q_2^2-2p_3(p_1+r_1)+2q_3(q_1+s_1)+2r_0+2p_0 > 0, (p_1 + r_1)^2 - (q_1 + s_1)^2 - 2(p_2 + r_2)(p_0 + r_0) + 2q_2(q_0 + s_0) > 0,$
(c) $p_0 + r_0 - s_0 - q_0 > 0$

Thus based on the above discussion the following theorem can be stated

Theorem 4.6: Let $R_0 > 1, \tau_1 = 0$ and $\tau_2 \neq 0$, then the infected equilibrium I_1 is asymptotically stable if $H1$ and $K1$ holds

Now (23) has positive real roots if $p_0+r_0-s_0-q_0 < 0$ holds. Corresponding to the positive roots of (23), the value of the delay τ_2 can be given by, for $i = 1, 2, 3, 4$.

$$\tau_{2_i} = \frac{1}{\omega}\cos^{-1}\left\{\frac{\Delta_n}{\Delta_d}\right\} + \frac{2n\pi}{\omega} \qquad n = 0, 1, 2..... \tag{24}$$

where

$$\begin{aligned} \Delta_n =& (r_0 + s_0)(p_0 + q_0) - q_1(s_0 + r_0)\omega^2 \\ &+ ((p_0 + q_0)(p_1(s_1 + r_1) - r_2) \\ &- (p_2 + q_2)(r_0 + s_0))\omega^2 + (p_3(s_0 + r_0) + q_1r_2)\omega^3 \\ &- p_3r_2\omega^5((p_2 + q_2)(r_2 - p_1(s_1 + r_1) - p_1r_3(p_0 + q_0) \\ &+ p_4(s_0 + r_0)))\omega^4(p_1p_4(s_1 + r_1) - p_4r_2 \\ &+ p_1r_3(p_2 + q_2))\omega^6 - p_1p_4r_3\omega^8 \\ \Delta_d =& -(s_0 + r_0)^2 + 2r_2(s_0 + r_0)\omega^2 - p_1(s_1 + r_1)^2\omega^3 \\ &- r_2^2\omega^4 + 2p_1r_3(s_1 + r_1)^5 - p_1r_3^2\omega^7 \end{aligned}$$

The critical value τ_2^* of τ_2 is given by taking the min of all the values of τ_2 given by (24). Thus,

$$\tau_2^* = \min \tau_{2_i} \text{for all} \quad i \tag{25}$$

Differentiating (21) with respect to τ_2 gives

$$\begin{aligned} &\left\{\frac{d\lambda}{d\tau_2}\right\}^{-1} \\ &= \frac{4\lambda^3 + 3p_3\lambda^2 + 2p_2\lambda + p_1}{\lambda(4\lambda^4 + p_3\lambda^3 + (p_2 + r_2)\lambda^2 + (p_1 + r_1)\lambda + p_0 + r_0)} \\ &- \frac{3q_3\lambda^2 + 2q_2\lambda + q_1}{\lambda(q_3\lambda^3 + q_2\lambda^2 + (q_1 + s_1)\lambda + q_0 + s_0)} - \frac{\tau_2}{\lambda} \end{aligned} \tag{26}$$

For hopf bifurcation the tranversability conditions needs to be verified. Thus to analyze the crossing of the imaginary axis by the eigenvalues. Substituting $\lambda(\tau_2) = i\omega_*$ in Eq (26) and taking its real part.

$$\begin{aligned} &\text{sign}\left(\frac{d\Re(\lambda)}{d\tau_2}\bigg|_{\lambda=i\omega_*}\right) \\ =&\text{sign}\left(\Re\left(\frac{d\lambda}{d\tau_2}\right)^{-1}\bigg|_{\lambda=i\omega_*}\right) \\ =&\text{sign}G'(\omega_*^2) \end{aligned} \tag{27}$$

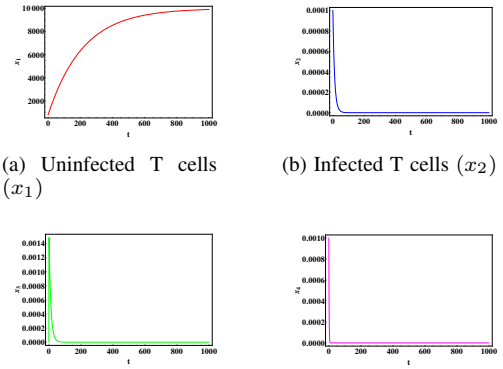
Thus based on the above discussion the following theorem can be stated.

Theorem 4.7: Let $R_0 > 1, \tau_1 = 0$ and $\tau_1 \neq 0$, and $p_0+r_0-s_0-q_0 < 0$ holds, then there exits a critical value τ_2^* and ω_* such that the infected equilibrium I_1 is asymptotically stable in the interval $[0, \tau_2^*)$. Also if $p_0+r_0-s_0-q_0 < 0$ holds and $G'(\omega_*^2) \neq 0$, then hopf bifurcation is exhibited by system (1) at the critical value τ_2^*.

V. Numerical Simulation

The preceding sections were devoted to theoretical analysis of the proposed model. In this section numerical simulation have been performed to numerically verify the results so obtained. The data sets have mostly been taken from [3], [30]–[34].

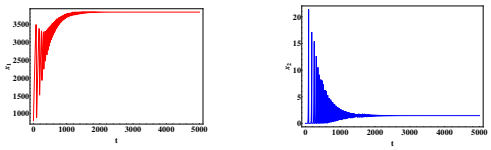
The value of the basic reproduction number R_0 was calculated to be 0.784 when ϵ_r and ϵ_p were taken as 0.9 and 0.8 respectively, the rest of the values of the parameters are given in Data 2 in Table (I). In this case the infection free steady state I_0 was stable as is clear from Fig (1). Next the values of the parameters ϵ_r and ϵ_p were changed to 0.6 and 0.6 respectively, for these values R_0 was calculated to be 10.784. In this case the infected steady state is stable as is clear from Fig (2)



(a) Uninfected T cells (x_1)
(b) Infected T cells (x_2)
(c) Viruses (x_3)
(d) Immune response cells (x_4)

Fig. 1: The plot shows stability of the uninfected equilibrium I_0 when $R_0 < 1$

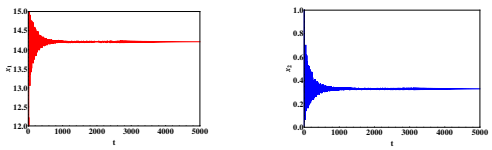
Varying the value η in Data 1 given in Table (I), it was demonstrated how viral loss term η affects the dynamics of the virus. When $\eta = 0.5$ the first viral load peak was seen overlapping with the case when $\eta = 0$. However as the value



(a) Uninfected T cells (x_1) (b) Infected T cells (x_2)

(c) Viruses (x_3) (d) Immune response cells (x_4)

Fig. 2: The plot shows stability of the infected equilibrium I_1 when $R_0 > 1$

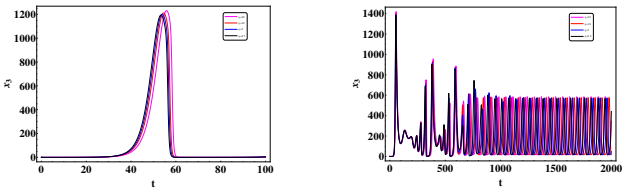


(a) Uninfected T cells (x_1) (b) Infected T cells (x_2)

(c) Viruses (x_3) (d) Immune response cells (x_4)

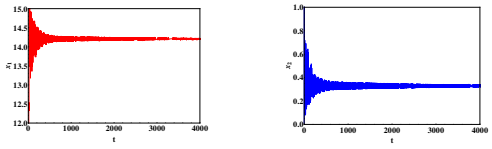
Fig. 4: The plot shows the asymptotic stability of the infected equilibrium I_1 when $\tau_1 < \tau_1^*$

of η was increased a shift in the first peak was seen. The first viral load peak at $\eta = 25$ is much higher than the peak observed at $\eta = 0.5$ Fig (3a). Initially the oscillations seemed to be overlapping, but with time there was a slight shift in the oscillations with an increase in the viral loss term. In this case also sustained oscillations could be seen in the system Fig (3b).



(a) First viral peak for different values of η (b) Virus concentration for different values of η

Fig. 3: Effect of viral loss term



(a) Uninfected T cells (x_1) (b) Infected T cells (x_2)

(c) Viruses (x_3) (d) Immune response cells (x_4)

Fig. 5: The plot shows the occurrence of periodic solutions of the infected equilibrium I_1 when $\tau_1 > \tau_1^*$

Data 1 given in Table (I) was used to calculate the critical values of the delays. Taking τ_2 as 0, calculations were performed to determine the critical value of τ_1. The critical value was calculated to be 11.5 days. Fig (4) showed that the infected equilibrium I_1 was stable when $\tau_1 < 11.5$ and Fig (5) showed the existence of Hopf bifurcation as τ_1 crosses the critical value.

VI. Conclusion

In the present article, a multiple delayed model of HIV-1 infection has been proposed. Drug therapy has also been included to make the model more practical. The delays in the model symbolize the time lapses during the course of infection namely the time lag in the production of virus by the infected cells and the activation of the CTL immune response after the detection of the infected cells in the body.

Firstly the positivity and boundedness of the solution of the model has been shown. By calculating the reproduction number R_0 of the model the local stability analysis of the various steady states has been discussed next. It has been shown that the infection free steady state is locally stable when $R_0 < 1$ and the infected steady state is locally stable when $R_0 > 1$.

Bifurcation analysis has been performed on the infected steady state and the critical value of the delays has also been calculated. Taking the immunological delay τ_2 as zero the critical value of τ_1 has been calculated as 11.5 days.

Plotting the graphs for different values of the viral loss term η, it was observed that the graphs overlap for smaller values of the virus waning term but as the value is increased the difference in the viral peak can be observed. The model presented in this paper thus generates some rich dynamics with respect to the delay and viral loss term. What remains to be seen is that how will the dynamics of the system when pharmacological delays are introduced and all the delays are taken to be nonzero. This provides a background for future work.

VII. Acknowledgment

This research is supported by UGC Grant for Junior Research Fellowship, Government of India.

Parameter	Definition	Data 1	Data 2
s cells ml-day$^-1$	production rate of uninfected	46	46
d day$^-1$	death rate of uninfected T cells	0.0046	0.0046
b viron ml-day^{-1}	rate of infection of T cells by virus	4.8×10^{-6}	4.8×10^{-6}
ϵ_r	effectiveness of reverse transcriptase inhibitor	0.5	0.9
k cell^{-1} day$^-1$	removal rate of uninfected T cells	0.36	0.36
B day^{-1}	bystander apoptosis	350	0.02
δ day^{-1}	death rate of infected T cells	0.1	0.1
d_1 viron ml -day^{-1}	rate of action of CTL cells	9.4×10^{-4}	9.4×10^{-4}
ϵ_p	effectiveness of protease inhibitor	0.6	0.8
N virion-cell^{-1}	burst rate of infected cells	4000	2500
c day$^-1$	death rate of virus	3	3
η	loss of virus	20	1
p day$^-1$	generation rate of CTL cells	0.46	0.46
d_2 day$^-1$	death rate of CTL cells	0.67	0.67
τ_1 day	intracellular delay	0.09	0
τ_2 day	immunological delay	2	0

TABLE I: The value of the various parameters

REFERENCES

[1] M. Nowak and R. M. May, *Virus Dynamics: Mathematical Principles of Immunology and Virology*. Oxford University Press, 2000.

[2] M. A. Nowak and C. R. Bangham, "Population dynamics of immune responses to persistent viruses," *Science*, vol. 272, no. 5258, pp. 74–79, 1996.

[3] A. S. Perelson, A. U. Neumann, M. Markowitz, J. M. Leonard, and D. D. Ho, "Hiv-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time," *Science*, vol. 271, no. 5255, pp. 1582–1586, 1996.

[4] A. S. Perelson, D. E. Kirschner, and R. De Boer, "Dynamics of hiv infection of cd4+ t cells," *Mathematical biosciences*, vol. 114, no. 1, pp. 81–125, 1993.

[5] J. M. Coffin, "Hiv population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy," *Science*, vol. 267, no. 5197, pp. 483–489, 1995.

[6] D. D. Ho, A. U. Neumann, A. S. Perelson, W. Chen, J. M. Leonard, M. Markowitz *et al.*, "Rapid turnover of plasma virions and cd4 lymphocytes in hiv-1 infection," *Nature*, vol. 373, no. 6510, pp. 123–126, 1995.

[7] D. Kirschner, "Using mathematics to understand hiv immune dynamics," *AMS notices*, vol. 43, no. 2, 1996.

[8] S. J. Merrill, "Modeling the interaction of hiv with cells of the immune system," in *Mathematical and statistical approaches to AIDS epidemiology*. Springer, 1989, pp. 371–385.

[9] P. Srivastava, M. Banerjee, and P. Chandra, "Modeling the drug therapy for hiv infection," *Journal of Biological Systems*, vol. 17, no. 02, pp. 213–223, 2009.

[10] J. M. Murray, S. Emery, A. D. Kelleher, M. Law, J. Chen, D. J. Hazuda, B.-Y. T. Nguyen, H. Teppler, and D. A. Cooper, "Antiretroviral therapy with the integrase inhibitor raltegravir alters decay kinetics of hiv, significantly reducing the second phase," *Aids*, vol. 21, no. 17, pp. 2315–2321, 2007.

[11] L. Rong, Z. Feng, and A. S. Perelson, "Mathematical analysis of age-structured hiv-1 dynamics with combination antiretroviral therapy," *SIAM Journal on Applied Mathematics*, vol. 67, no. 3, pp. 731–756, 2007.

[12] R. M. Granich, C. F. Gilks, C. Dye, K. M. De Cock, and B. G. Williams, "Universal voluntary hiv testing with immediate antiretroviral therapy as a strategy for elimination of hiv transmission: a mathematical model," *The Lancet*, vol. 373, no. 9657, pp. 48–57, 2009.

[13] A. Herz, S. Bonhoeffer, R. M. Anderson, R. M. May, and M. A. Nowak, "Viral dynamics in vivo: limitations on estimates of intracellular delay and virus decay," *Proceedings of the National Academy of Sciences*, vol. 93, no. 14, pp. 7247–7251, 1996.

[14] J. Tam, "Delay effect in a model for virus replication," *Mathematical Medicine and Biology*, vol. 16, no. 1, pp. 29–37, 1999.

[15] H. Zhu and X. Zou, "Dynamics of a hiv-1 infection model with cell-mediated immune response and intracellular delay," *Discrete Contin. Dyn. Syst. Ser. B*, vol. 12, no. 2, pp. 511–524, 2009.

[16] X. Jiang, X. Zhou, X. Shi, and X. Song, "Analysis of stability and hopf bifurcation for a delay-differential equation model of hiv infection of cd4+ t-cells," *Chaos, Solitons & Fractals*, vol. 38, no. 2, pp. 447–460, 2008.

[17] M. A. Safi, "Global dynamics of treatment models with time delay," *Computational and Applied Mathematics*, pp. 1–17, 2014.

[18] K. Wang, W. Wang, H. Pang, and X. Liu, "Complex dynamic behavior in a viral model with delayed immune response," *Physica D: Nonlinear Phenomena*, vol. 226, no. 2, pp. 197–208, 2007.

[19] T. Wang, Z. Hu, and F. Liao, "Stability and hopf bifurcation for a virus infection model with delayed humoral immunity response," *Journal of Mathematical Analysis and Applications*, vol. 411, no. 1, pp. 63–74, 2014.

[20] K. A. Pawelek, S. Liu, F. Pahlevani, and L. Rong, "A model of hiv-1 infection with two time delays: mathematical analysis and comparison with patient data," *Mathematical biosciences*, vol. 235, no. 1, pp. 98–109, 2012.

[21] Z. Wang and R. Xu, "Stability and hopf bifurcation in a viral infection model with nonlinear incidence rate and delayed immune response," *Communications in Nonlinear Science and Numerical Simulation*, vol. 17, no. 2, pp. 964–978, 2012.

[22] H. Zhu, Y. Luo, and M. Chen, "Stability and hopf bifurcation of a hiv infection model with ctl-response delay," *Computers & Mathematics with Applications*, vol. 62, no. 8, pp. 3091–3102, 2011.

[23] R. Ouifki and G. Witten, "Stability analysis of a model for hiv infection with rti and three intracellular delays," *BioSystems*, vol. 95, no. 1, pp. 1–6, 2009.

[24] Y. Wang, F. Brauer, J. Wu, and J. M. Heffernan, "A delay-dependent model with hiv drug resistance during therapy," *Journal of Mathematical Analysis and Applications*, vol. 414, no. 2, pp. 514–531, 2014.

[25] A. Alshorman, X. Wang, M. Joseph Meyer, and L. Rong, "Analysis of hiv models with two time delays," *Journal of biological dynamics*, pp. 1–25, 2016.

[26] H. Song, W. Jiang, and S. Liu, "Virus dynamics model with intracellular delays and immune response," *Mathematical biosciences and engineering*, vol. 12, pp. 185–208, 2015.

[27] H. L. Smith and P. De Leenheer, "Virus dynamics: a global analysis," *SIAM Journal on Applied Mathematics*, vol. 63, no. 4, pp. 1313–1327, 2003.

[28] J. M. Heffernan and L. M. Wahl, "Natural variation in hiv infection: Monte carlo estimates that include cd8 effector cells," *Journal of theoretical biology*, vol. 243, no. 2, pp. 191–204, 2006.

[29] J. Heffernan, R. Smith, and L. Wahl, "Perspectives on the basic reproductive ratio," *Journal of the Royal Society Interface*, vol. 2, no. 4, pp. 281–293, 2005.

[30] Y. Wang, Y. Zhou, F. Brauer, and J. M. Heffernan, "Viral dynamics model with ctl immune response incorporating antiretroviral therapy," *Journal of mathematical biology*, vol. 67, no. 4, pp. 901–934, 2013.

[31] D. Burg, L. Rong, A. U. Neumann, and H. Dahari, "Mathematical modeling of viral kinetics under immune control during primary hiv-1 infection," *Journal of Theoretical Biology*, vol. 259, no. 4, pp. 751–759, 2009.

[32] M. Ciupe, B. Bivort, D. Bortz, and P. Nelson, "Estimating kinetic parameters from hiv primary infection data through the eyes of three different mathematical models," *Mathematical biosciences*, vol. 200, no. 1, pp. 1–27, 2006.

[33] G. Magombedze, W. Garira, E. Mwenje, and C. P. Bhunu, "Optimal control for hiv-1 multi-drug therapy," *International Journal of Computer Mathematics*, vol. 88, no. 2, pp. 314–340, 2011.

[34] G. Magombedze, W. Garira, and E. Mwenje, "Modeling the tb/hiv-1 co-infection and the effects of its treatment," *Mathematical Population Studies*, vol. 17, no. 1, pp. 12–64, 2010.

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

Vehicle Speed Determination from Video Streams using Image Processing

Mohammad Reduanul Haque

Department of Computer Science and Engineering
Daffodil International University
Dhaka, Bangladesh

reduan.cse@diu.edu.bd

Md. Golam Moazzam, Saiful Islam, Rony Das, Mohammad Shorif Uddin

Department of Computer Science and Engineering
Jahangirnagar University
Dhaka, Bangladesh

khokanju@yahoo.com, saiful.raju@gmail.com, rony.cse07@gmail.com, shorifuddin@gmail.com

Abstract—**Recently it receives a great attention to reduce road accident and controlling traffic by limiting the speed of vehicles. Moving vehicle speed is usually determined by using Radar, Photo Radar, Drone Radar, LIDAR Speedometer Clocks, Average Speed Computers, Aircraft or Video frame methods, etc. Most of these methods are very expensive and also their accuracy is not quite satisfactory. In this paper, a video based vehicle speed calculation method is presented, which has the capability of calculating the speed with higher accuracy but at relatively low cost. The method consists of three stages: object recognition, noise removal and speed calculation. The method shows satisfactory performance on standard data sets.**

***Keywords*—Vehicle detection; speed calculation; background subtraction; vehicle tracking.**

I. INTRODUCTION

Sustainable development goal 3.6 [1] aims to halve the number of global deaths and injuries from road traffic accidents by 2020. Bangladesh is one of the top countries in the world where road accident rate is very high. Though concerned authorities have taken many initiatives to minimize the road accident rate and increasing the road safety, still every year many people get killed due to road accidents. According to statistics of road accidents and casualties from BRTA (Bangladesh Road Transport Authority), in the year of 2014, about 2027 dangerous accidents occurred and that caused 2067 deaths [2]. In fact the actual number of accident is very high as many accidents are not reported to concern authorities. The speed of vehicle is considered as one of the main factors for road accidents and also it is an important traffic parameter, so detection of speed of a vehicle [3-13] is very significant for more smooth traffic management. For this reason, there has been many efforts from the concern authorities to reduce speeding by traffic lights and vehicle speed detection control based on RADAR (Radio Detection and Ranging) or LIDAR (Laser Infrared Detection and Ranging) based methods. Although RADAR and LIDAR based methods are high in performances, these are also very high in price. Due to higher cost these methods are not scalable for large scale installment in roads around Bangladesh. Existing speed cameras are based on RADAR equipment for detecting vehicle speed. Speed cameras are usually placed behind trees, road signs and do not require human to operate unlike speed gun technology. So, at a first glance, these speed cameras seem to be the best method for speed detection but equipment for this method is still based on the RADAR approach. This method has still some disadvantages because its performance is not so high. Therefore, image processing approach was introduced instead of RADAR or LIDAR system in vehicle speed detection. Image processing only needs a single video camera and a computer that is less expensive compared to RADAR approach. For vehicle speed detection, device capable of recording video along with a computer system can serve the requirement for low cost investment. In this paper, we develop an imaged-based vehicle speed detection method where a physics-based velocity theory is used to calculate the speed of moving vehicle from the recorded video. This approach consists of five major steps such as image acquisition and enhancement, segmentation, centroid calculation, shadow removal and speed calculation.

We organized the rest of this paper as follows. Section II presents the methodology and theory of the system along with its implementation. In Section III, experimental result and discussions are described. Finally, in Section IV conclusions are drawn.

II. VEHICLE SPEED DETECTION METHOD

A generic schematic flow diagram of our vehicle speed detection algorithm is shown in Fig. 1.

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

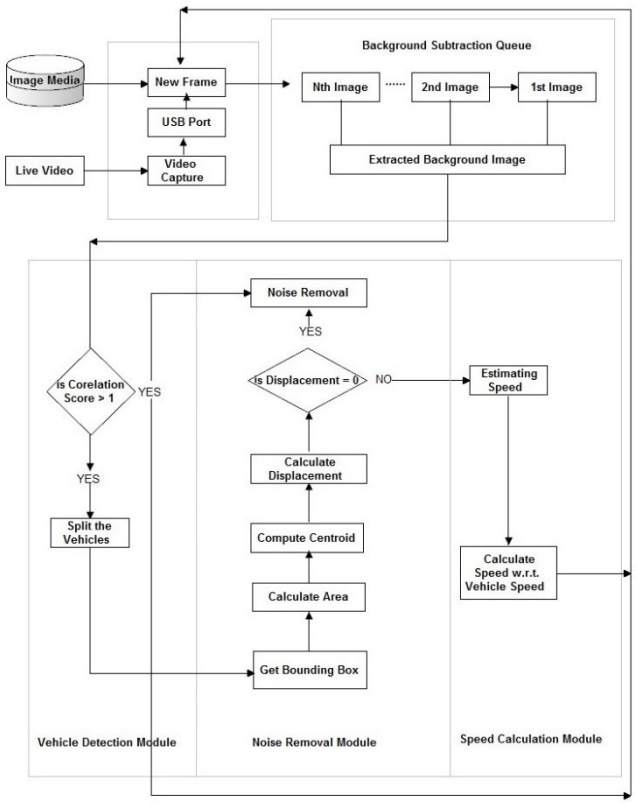


Fig. 1: Flowchart of the vehicle speed detection method.

A brief explanation of the investigated vehicle speed detection algorithm is given below:

A. Vehicle Detection Module

For calculating the moving vehicle speed, the primary task is to detect the moving vehicles from the video correctly. There are numerous approaches for detecting moving objects such as temporal differencing algorithm, optical flow algorithm, background subtraction algorithm, etc. Two adjacent frames are taken in temporal differencing algorithm. It cannot detect slow changes accurately. Camera motion is used to detect object independently in optical flow algorithm and it is not applicable in real time. To detect moving object using background subtraction algorithm, the absolute difference of the background model and the current frame is taken.

In our work, a hybrid algorithm is used for detecting moving objects [8] in which background subtraction algorithm has overcome the drawbacks' of the previous method. Here, an adaptive background subtraction technique is combined with a three-frame differencing method. The current approach consists of two successive steps, which are described below:

At first, we construct a matrix corresponds to the current frame to decide whether the pixels are in motion or stationary. A three-frame differencing technique is applied to know motion as shown in Fig. 2. After that adaptive background subtraction method is applied to extract the entire moving region. Here, the stationary pixel represents the background pixel and the moving pixel represents the foreground pixel.

Fig-2: A scene with two moving objects.

Let us take a stationary camera for capturing video. Assume for a captured pixel at position (x,y) and time t=n, its intensity value becomes IIn(x,y). By three frame differencing technique an object is moving if the current image (IIn) and its position in the last frame (IIn-1) is significantly different than its position at (IIn-2) than the pixel will be a moving pixel (x,y). Mathematically, is detected if

$$(II_n(x,y) - II_{n-1}(x,y) > Th_n(x,y) \text{ and}$$
$$(II_n(x,y) - II_{n-2}(x,y) > Th(x,y)$$

Where $Th_n(x,y)$ is the threshold value at pixel position (x, y).

The second step is to extract background from the above image. It is achieved by the subtraction of the background $Bn(x,y)$ frame from the current frame $IIn(x, y)$.

$$SI_n(x,y) = (II_n(x,y) - B_n(x,y)$$

Where, $SIn(x,y)$ be the subtracted image.

B. Noise Removal and Objects Tracking Module

To remove noise from the frame, we are using trial and error method. Using this method, we have concluded that if the pixel intensity value is less than 0.15 then we considered this pixel as shadow otherwise a foreground.

After detecting moving objects and removing noise from the frame, the next task is to track the objects accurately. It is also important for higher level analysis and extraction of temporal information. Object tracking is done by segmenting and labeling objects.

The aim of object segmentation is to discover objects and detect boundaries based on the connectivity of the objects. So we have to partition the objects from the image correctly. For this reason connected region of the objects must be assured so that every object is connected as one part. Otherwise, counting of objects may not accurate, as a result many duplicate objects will be counted. After separating the objects correctly, we calculate the area of each objects. Basically we map each of the objects as a rectangle for smooth calculation. To do this, we iterate through the frame horizontally and vertically. The frame is scanning horizontally whose starting point is the top most left pixel. Then, the whole frame is scanned. It marks the whole scan line if no foreground pixel (white pixel) is found.

Otherwise, it will skip the line and do this until reach the last line. The same process is done for vertical scanning of the frame. After two or three successive iteration, we get the accurate segmentation result and discard the regions that are not an actual object.

For keeping track of each object individually, labelling is performed. We track the frames when an object enters the scene (at frame SI_0) and when it leaves the scene (at frame *SIn*).

C. Compute Detected Centroid of the Object

After tracking the objects, the next step is to compute the centroid of each object. We can track and map easily from the center as it represents the whole object. In this section, we show how can we track the center and correct its label. The center is calculated by taking the average of the corner points of the object.

$$(x_c, y_c) = (\frac{(x_1 + x_2)}{2}, \frac{y_1 + y_2}{2})$$

where, (x_c, y_c) is the center of the vehicle.

D. Speed Calculation

Now speed calculation is done by counting the frame numbers needed to enter and leave the object to pass the scene and the distance it covers. The Euclidean distance of the centroid of nth frame and (n-1)th frame gives the distance travels by objects. The frame rate of the video is then multiplied by the total number of frames. From this total time and distance, speed is measured and mapped in real time. The distance is calibrated in km and time in hr to calculate the vehicle speed in km/hr.

$$Distance = \sqrt{(x_{n-1} - x_n)^2 + (y_{n-1} - y_n)^2}$$

Where, $((x_n, y_n), (x_{n-1}, x_{n-1}))$ is the coordinates of the centroid pixel in nth and (n-1)th frames, respectively.

Speed = α*Distance/ ((Frame-n – Frame-0)*Frame rate)

Where, α is the calibration co-efficient.

α = real height of the vehicle / image height of the vehicle.

III. Experimental Results and Discussions

For our experiment, we have used QMUL Junction Dataset [14] and a dataset (developed by us) by collecting videos from the internet.

QMUL Junction Dataset is a traffic dataset whose video length is 1 hour and it has 90000 frames. The size of each frame is 360 × 288 pixels and the frame rate is 25 Hz.

It is observed that the speed of the vehicle calculated by our system is almost same as compared to real speed of the vehicle. Table I shows the measurement result.

TABLE I

Actual Speed	Variation in measured value	Error
10 km/hr	9 – 11 km/hr	±.01%
15 km/hr	14.5 – 16 km/hr	±.005%
30 km/hr	27 – 32 km/hr	±.03%

IV. Conclusion

We described an image-based vehicle speed determination system, which is a good alternative to the traditional radar system. From the experimental results, we have found that the system works well with good accuracy as the error rate is very low. We hope that this system will find extensive application in real time traffic management field and many other purposes.

References

[1] United nation's Sustainable development goal 3.6, available online, last accessed on September 14, 2016, https://sustainabledevelopment.un.org/sdg3

[2] BRTA report, available online: http://www.brta.gov.bd/newsite/en/statistics-of-accident-casualties/
(last accessed on September 14, 2016).

[3] F. W. Cathey and D. J. Dailey, "A novel technique to dynamically measure vehicle speed using uncalibrated roadway cameras," in Proc. of the IEEE Intelligent Vehicles Symposium, Las Vegas, NV, USA, 6-8 June 2005, pp.777-782.

[4] Douxchamps, Macq and Chihara, "High accuracy traffic monitoring using road-side line scan cameras," in Proc. of the IEEE Intelligent Transportation Systems Conference (ITLS 2006), Toronto, Canada, September 2006, pp. 875-878.

[5] P. L. Rosin and Tim Ellis, "Image difference Threshold strategies and shadow detection," City University London, UK.

[6] Y. Matsushita, E. Ofek, X. Tang, and H. Shum, "Full-Frame Video Stabilization," Microsoft Research Asia Beijing Sigma Center, No.49.

[7] R. P. Avery, G. Zhang, Y. Wang, and N. L. Nihan, "An Investigation into Shadow Removal from Traffic Images," Department of Civil and Environmental Engineering University of Washington, November 15, 2006.

[8] O. Ibrahim, H. ElGendy, and A. M. ElShafee "Towards Speed Detection Camera System for a RADAR Alternative," in Proc. 11th International Conf. on ITS Communications, Sait-Petersburg, Russia, August 2011.

[9] Y. Dedeo˘glu, "Moving Object Detection, Tracking and Classification for Smart Video Surveillance," Department of Computer Engineering, Bilkent University, August, 2004.

[10] R. T. Collins, A. J. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto, O. Hasegawa, P. Burt, and L. Wixson, "A System for Video Surveillance

and Monitoring," The Robotics Institute, Carnegie Mellon University, Princeton, NJ.

[11] Muhammad Akram Adnan, Nor Izzah Zainuddin, "Vehicle Speed Measurement Technique Using Various Speed Detection," Instrumentation‖ IEEE Business Engineering and Industrial Applications Colloquium (BEIAC), 2013.

[12] Huang Chieh-Ling and Ma Heng-Ning, "A Moving Object Detection Algorithm for Vehicle Localization‖," in Proc. 6th International Conf. on Genetic and Evolutionary Computing, 2012

[13] P. Daniel Ratna Raju and G. Neelima, "Image Segmentation by using Histogram Thresholding", IJCSET Vol 2, Issue 1 pp:776-779, 2012.

[14] QMUL Junction Dataset, available online: http://personal.ie.cuhk.edu.hk/~ccloy/downloads_qmul_junction.html (last accessed on September 14, 2016).

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

A New Spline in Compression Method of Order four in space and two in time for the Solution of 1D wave equation in polar coordinates.

Gunjan Khurana
Department of Mathematics
South Asian University
New Delhi, India
Email: gunjankhurana84@gmail.com

R.K. Mohanty
Department of Mathematics
South Asian University
New Delhi, India
Email: rmohanty@sau.ac.in

***Abstract*—In this paper, we propose a new three level implicit method based on half step spline in compression approximation, of order two in time and four in space for the solution of one-space dimensional non-linear wave equation of the form $w_{tt} = w_{rr} + f(r, t, w, w_r)$. We describe spline in compression approximations and its properties. The new method for 1D wave equation is obtained directly from the consistency condition. In this method we use three grid points for the unknown function $w(r, t)$ and two half step points for the known variable 'r' in r- direction. To assess the validity and accuracy, the method is applied to wave equation in polar coordinates and numerical results are provided to demonstrate the usefulness of our method.**

Keywords: Spline in compression approximations; One-dimensional wave equation in polar coordinates; Maximum absolute errors.

I. Introduction

We consider one-space dimensional quasi-linear hyperbolic equation of the type

$$w_{tt} = w_{rr} + f(r, t, w, w_r), 0 < r < 1, t > 0 \tag{1}$$

with the following initial conditions

$$w(r, 0) = p(r), w_t(r, 0) = q(r), 0 \leq r \leq 1 \tag{2}$$

and boundary conditions

$$w(0, t) = g_0(t), w(1, t) = g_1(t), t \geq 0 \tag{3}$$

We assume that $w(r, t)$ is sufficiently smooth and required higher order partial derivatives of $w(r, t)$ exist in the solution domain $\Omega \equiv [(r, t)|0 < r < 1, t > 0]$ and the conditions (2) and (3) are given with sufficient smoothness to maintain the order of accuracy in the numerical method under consideration. Further, we assume that the initial and boundary value problem (1)-(3) has a unique smooth solution $w(r, t)$ in the solution domain Ω. The details of existence and uniqueness of the above initial boundary value problem has already been discussed in [4]

Wave equation is an important second-order linear partial differential equation for the description of waves as they occur in real life such as ripples on a lake, wind waves on water, tidal surges in estuaries, transverse waves traveling on a long string, transverse vibrations of strings and membranes, traffic density waves, seismic waves, acoustic waves, electromagnetic waves currents in coaxial cables.

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

Problems involving the propagation of non-linear waves have become of increasing interest in various branches of science and engineering. There has been a consistent effort in developing efficient and high accuracy finite difference methods to solve quasi-linear hyperbolic equations including wave equation in polar coordinates.(see [1],[8],[9],[3],[7],[5],[6],[10],[2])

Our paper is arranged as follows;
In section 2, we discuss the properties of spline in compression. In section 3, we give a new half-step three level implicit method based on spline in compression approximations. In section 4, we discuss the direct applicability of our method to wave equation in polar coordinates. In section 5, we give numerical results of proposed high accuracy method with the corresponding second order accuracy non-polynomial spline in compression method. In section 6, we give concluding remarks.

II. Spline in compression approximations

Let us discretize the solution domain $[0, 1] \times [t > 0]$ into $(N + 1) \times J$ by a set of grid points (r_l, t_j), where $0 = r_0 < r_1 < ..... < r_{N+1} = 1, 0 = t_0 < t_1 < ..... < t_J = J$, N being a positive integer with uniform mesh spacing $h = r_l - r_{l-1}$, $\tau = t_j - t_{j-1}$; $l = 1(1)N + 1, j = 1(1)J$. Let w_l^j and W_l^j be the approximate and exact solutions of $w(r, t)$ at the grid point(r_l, t_j), respectively.

Now for each sub interval $[r_{l-1}, r_l]$, $l = 1(1)N + 1$, we define the non polynomial spline in compression function $P_j(r)$ of the function $w(r, t)$ at the mesh point (r_l, t_j) as follows

$$P_j(r) = a_l^j + b_l^j(r - r_l) + c_l^j sin\omega(r - r_l) + d_l^j cos\omega(r - r_l), \quad l = 1(1)N + 1, r \in [r_{l-1}, r_l] \tag{4}$$

Here,

$$M_l^j = P_j''(r_l) = W_{rr_l}^j, M_{l\pm1/2}^j = P_j''(r_{l\pm1/2}) = W_{rr_{l\pm1/2}}^j, \tag{5}$$

and a_l^j, b_l^j, c_l^j, d_l^j and ω are given by

$$a_l^j = U_l^j + \frac{M_l^j}{\omega^2}, b_l^j = \frac{U_l^j - U_{l-1}^j}{h} + \frac{M_l^j}{\omega\theta} - \frac{M_{l-1/2}^j}{\omega\theta} cos\theta,$$
$$c_l^j = \frac{M_{l-1/2}^j - M_l^j cos\theta}{\omega^2 sin\theta}, d_l^j = -\frac{M_l^j}{\omega^2}, \theta = \frac{\omega h}{2}, \tag{6}$$

From condition of continuity

$$P_j'(r_l-) = P_j'(r_l+) \tag{7}$$

we obtain the following relation

$$\frac{W_{l+1}^j - 2W_l^j + W_{l-1}^j}{h^2} = \alpha M_{l+1/2}^j + 2\beta M_l^j + \alpha M_{l-1/2}^j, \quad l = 1(1)N \tag{8}$$

where

$$\alpha = \frac{1}{2\theta^2}(\frac{\theta}{sin\theta} - cos\theta), \beta = \frac{1}{2\theta^2}(1 - \theta cot\theta). \tag{9}$$

From (9) the consistency condition is given by

$$tan\frac{\theta}{2} = \frac{\theta}{2}, \tag{10}$$

The above equation has infinitely many roots, the smallest positive non-zero root given by

$$\theta = 8.986818916 \tag{11}$$

Now, we give two important properties of non-polynomial spline in compression

$$P_j^{'}(r_{l-1/2}) = \frac{W_l^j - W_{l-1}^j}{h} + \frac{h(2\beta M_{l-1/2}^j - \alpha M_l^j)}{4} \tag{12}$$

$$P_j^{'}(r_{l+1/2}) = \frac{W_{l+1}^j - W_l^j}{h} + \frac{h(\alpha M_l^j - 2\beta M_{l+1/2}^j)}{4} \tag{13}$$

III. Method based on non-polynomial spline in compression approximations

We consider the one-space dimensional non linear partial differential equation

$$w_{tt} = w_{rr} + f(r, t, w, w_r), 0 < r < 1, t > 0 \tag{14}$$

with the initial conditions and boundary conditions prescribed by (2) and (3) respectively.Let w_l^j and W_l^j be the approximate and exact solutions of $w(r,t)$ at the grid point(r_l, t_j), respectively.

At the grid point (r_l, t_j), we may rewrite the differential equation (14) as

$$W_{tt_l}^j = W_{rr_l}^j + f(r_l, t_j, W_l^j, W_{r_l}^j) \equiv F_l^j \tag{15}$$

Similarly at grid point $(r_{l\pm1/2}, t_j)$

$$W_{tt_{l\pm1/2}}^j = W_{rr_{l\pm1/2}}^j + f(r_{l\pm1/2}, t_j, W_{l\pm1/2}^j, W_{r_{l\pm1/2}}^j) \equiv F_{l\pm1/2}^j \tag{16}$$

Using Taylor's expansion, we obtain

$$W_{l+1}^j - 2W_l^j + W_{l-1}^j = \frac{h^2}{12}(\overline{W}_{tt_{l+1}}^j + 10\overline{W}_{tt_l}^j + \overline{W}_{tt_{l-1}}^j) - \frac{h^2}{3}(F_{l+1/2}^j + F_l^j + F_{l-1/2}^j) + O(\tau^2h^2 + \tau^2h^4 + h^6) \tag{17}$$

For derivation of method , we need the following approximations:

$$\overline{W}_{l\pm1/2}^j = \frac{W_{l\pm1}^j + W_l^j}{2}, \tag{18.1}$$

$$\overline{W}_{tt_l}^j = \frac{W_l^{j+1} - 2W_l^j + W_l^{j-1}}{\tau^2}, \tag{18.2}$$

$$\overline{W}_{tt_{l\pm1}}^j = \frac{W_{l\pm1}^{j+1} - 2W_l^j + W_{l\pm1}^{j-1}}{\tau^2}, \tag{18.3}$$

$$\overline{W}_{tt_{l\pm1/2}}^j = \frac{\overline{W}_{tt_{l\pm1}}^j + \overline{W}_{tt_{l\pm1}}^j}{2}, \tag{18.4}$$

$$\overline{W}_{r_l}^j = \frac{W_{l+1}^j - W_{l-1}^j}{2h}, \tag{18.5}$$

$$\overline{W}_{r_{l\pm1/2}}^j = \pm\frac{(W_{l\pm1}^j - W_l^j)}{h}, \tag{18.6}$$

$$\overline{W}_{rr_l}^j = \frac{W_{l+1}^j - 2W_l^j + W_{l-1}^j}{h^2}. \tag{18.7}$$

We define the following approximations:

$$\overline{F}_l^j = f(r_l, t_j, W_l^j, \overline{W}_{r_l}^j) \tag{19.1}$$

$$\overline{F}_{l\pm1/2}^j = f(r_{l\pm1/2}, t_j, W_{l\pm1/2}^j, \overline{W}_{r_{l\pm1/2}}^j) \tag{19.2}$$

$$\overline{M}_l^j = (\overline{W}_{tt_l}^j - \overline{F}_l^j) \tag{19.3}$$

$$\overline{M}_{l\pm1/2}^j = (\overline{W}_{tt_{l\pm1/2}}^j - \overline{F}_{l\pm1/2}^j) \tag{19.4}$$

$$\widehat{W}_{r_{l+1/2}}^j = \frac{W_{l+1}^j - W_l^j}{h} + \frac{h(\alpha\overline{M}_l^j - 2\beta\overline{M}_{l+1/2}^j)}{4} \tag{19.5}$$

$$\widehat{W}_{r_{l-1/2}}^j = \frac{W_l^j - W_{l-1}^j}{h} + \frac{h(2\beta\overline{M}_{l-1/2}^j - \alpha\overline{M}_l^j)}{4} \tag{19.6}$$

$$\widehat{F}_{l\pm1/2}^j = f(r_{l\pm1/2}, t_j, \overline{W}_{l\pm1/2}^j, \widehat{W}_{r_{l\pm1/2}}^j) \tag{19.7}$$

$$\widehat{W}_l^j = W_l^j - \frac{h^2}{4}\overline{W}_{rr_l}^j \tag{19.8}$$

Furthur, let

$$\widehat{W}_{r_l}^j = \overline{W}_{r_l}^j - \frac{\alpha h}{2}(\overline{M}_{l+1/2}^j - \overline{M}_{l-1/2}^j) \tag{20}$$

and

$$\widehat{F}_l^j = f(r_l, t_j, \widehat{W}_l^j, \widehat{W}_{r_l}^j) \tag{21}$$

Then the proposed non-polynomial spline in compression method for the differential equation (1)

$$W_{l+1}^j - 2W_l^j + W_{l-1}^j = \frac{h^2}{12}(\overline{W}_{tt_{l+1}}^j + 10\overline{W}_{tt_l}^j + \overline{W}_{tt_{l-1}}^j) - \frac{h^2}{3}(\widehat{F}_{l+1/2}^j + \widehat{F}_l^j + \widehat{F}_{l-1/2}^j) + \widehat{T}_l^j. \quad (22)$$

where $\widehat{T}_l^j = O(\tau^2h^2 + \tau^2h^4 + h^6)$
Note that, the initial and Dirichlet boundary conditions are given by (2) and (3) respectively. Incorporating the initial and boundary conditions, we can write the spline in compression method in a tridiagonal form. If the differential equation(1) is linear system we use the Gauss elimination (tridiagonal solver)method; in the non-linear case, we can use the Newton-Raphson iterative method.(see[9-10])

IV. Application of method to wave equation in polar coordinates

We consider the one dimensional wave equation in polar coordinates

$$w_{tt} = w_{rr} + a(r)w_r + f(r,t), 0 < r < 1, t > 0 \quad (23)$$

with the following initial conditions

$$w(r,0) = a_0(r), w_t(r,0) = a_1(r), 0 \le r \le 1 \quad (24)$$

and boundary conditions

$$w(0,t) = a_3(t), w(1,t) = a_4(t), t \ge 0 \quad (25)$$

where $a(r) = \frac{\gamma}{r}$ and $\gamma = 1, 2$.
Applying the method (22) to the differential equation (23)we obtain

$$W_{l+1}^j - 2W_l^j + W_{l-1}^j = \frac{h^2}{12}(\overline{W}_{tt_{l+1}}^j + 10\overline{W}_{tt_l}^j + \overline{W}_{tt_{l-1}}^j) - \frac{h^2}{3}(a_{l+1/2}\widehat{W}_{r_{l+1/2}}^j + a_l\widehat{W}_{r_l}^j + a_{l-1/2}\widehat{W}_{r_{l-1/2}}^j) - \frac{h^2}{3}(f_{l+1/2}^j + f_l^j + f_{l-1/2}^j) + \widehat{T}_l^j. \quad (26)$$

where $\widehat{T}_l^j = O(\tau^2h^2 + \tau^2h^4 + h^6)$

and

$$a_s = a(r_s), f_s^j = f(r_s, t_j), s = l, l \pm 1/2 \quad (27.1)$$

$$\overline{F}_l^j = f(r_l, t_j) + a(r_l)\overline{W}_{r_l}^j \quad (27.2)$$

$$\overline{F}_{l\pm1/2}^j = f(r_{l\pm1/2}, t_j) + a(r_{l\pm1/2})\overline{W}_{r_{l\pm1/2}}^j \quad (27.3)$$

$$\overline{M}_l^j = \overline{W}_{tt_l}^j - \overline{F}_l^j \quad (27.4)$$

$$\overline{M}_{l\pm1/2}^j = \overline{W}_{tt_{l\pm1/2}}^j - \overline{F}_{l\pm1/2}^j \quad (27.5)$$

$$\widehat{W}_{r_{l+1/2}}^j = \frac{W_{l+1}^j - W_l^j}{h} + \frac{h(\alpha\overline{M}_l^j - 2\beta\overline{M}_{l+1/2}^j)}{4} \quad (27.6)$$

$$\widehat{W}_{r_{l-1/2}}^j = \frac{W_l^j - W_{l-1}^j}{h} + \frac{h(2\beta\overline{M}_{l-1/2}^j - \alpha\overline{M}_l^j)}{4} \quad (27.7)$$

V. Numerical Illustrations

In this section, we have computed problem (23)-(25) using the proposed scheme described by equation (22) and compared our results with second order numerical scheme. The exact solutions are provided. The right hand side homogeneous function, initial and boundary conditions may be obtained by using the exact solution as a test procedure. The proposed scheme is a three level scheme. The value of w at t=0 is known from the initial condition. To begin the computation, we need the numerical value of w of required accuracy at $t = \tau$, so we discuss an explicit method of $O(\tau^2)$ for calculating the value of w at first time level in order to solve the differential equation (23) using the proposed scheme (22)

Since the values of w and w_t are known explicitly at $t = 0$, so the values of successive tangential derivatives $w, w_r, w_{rr}, ..., w_{tr}, w_{trr}, ...$etc. are known at $t = 0$. An approximation for w may be written as

$$w_l^1 = w_l^0 + \tau(w_t)_l^0 + \frac{\tau^2}{2}(w_{tt})_l^0 + O(\tau^3) \quad (28)$$

From equation (23), we have

$$(w_{tt})_l^0 = [w_{rr} + a(r)w_r + f(r,t)]_l^0 \quad (29)$$

Then using the initial values and their successive tangential derivative values from (29) we obtain the value of w_{tt} at $t = 0$ and then subsequently from (28), we can compute the value of w at first time level, i.e. at $t = \tau$.

The equation (23)-(25) is solved, using the proposed scheme (22). The maximum absolute errors (MAE) are compared with Central Difference Scheme (CDS) in Table 1, for $\gamma = 1, 2, t = 1$ and $\sigma = \frac{\tau}{h^2} = 3.2$. Figure 1 gives the graph of exact vs.numerical solution at time level $t = 1$,$h = 1/32$,$\gamma = 1$

Here, exact solution $w(r,t) = \cosh r \sin t$
Initial conditions: $w(r,0) = 0, w_t(r,0) = \cosh r$
Boundary conditions: $w(0,t) = \sin t, w(1,t) = (\frac{e+e^{-1}}{2})\cos t$
Table 1: Maximum Absolute Errors :

t=1,$\sigma = 3.2$				
h	$Proposed Scheme$ (22)		CDS	
	γ=1	γ=2	$\gamma = 1$	γ=2
1/8	$3.2633(-5)$	$3.3839(-5)$	$5.2011(-4)$	$1.0247(-3)$
1/16	$2.0948(-6)$	$2.1810(-6)$	$1.3916(-4)$	$2.7007(-4)$
1/32	$1.3356(-7)$	$1.3914(-7)$	$3.5865(-5)$	$6.8675(-5)$
1/64	$8.5274(-9)$	$8.8328(-9)$	$9.1102(-6)$	$1.7267(-5)$

VI. Conclusion

In this paper using three grid points and three evaluations of function f, we have derived a new half step spline in compression method of $O(\tau^2 + h^4)$ accuracy for the solution of one-dimensional wave equation.Our method is directly applicable to wave equation in poolar coordinates which is the

main attraction of our work. For a fixed parameter $\sigma = \frac{\tau}{h^2}$, the proposed method behaves like a fourth order method. The accuracy and efficiency of the proposed method are exhibited from the numerical computations.

REFERENCES

[1] MK Jain, SRK Iyengar, and ACR Pillai. Difference schemes based on splines in compression for the solution of conservation laws. *Computer Methods in Applied Mechanics and Engineering*, 38(2):137–151, 1983.

[2] CT Kelley. Iterative methods for linear and nonlinear equations siam publications, 1995.

[3] A Khan and T Aziz. Parametric cubic spline approach to the solution of a system of second-order boundary-value problems. *Journal of optimization theory and applications*, 118(1):45–54, 2003.

[4] Wei-Dong Li, Zhi-Zhong Sun, and Lei Zhao. An analysis for a high-order difference scheme for numerical solution to utt= a (x, t) uxx+ f (x, t, u, ut, ux). *Numerical Methods for Partial Differential Equations*, 23(2):484–498, 2007.

[5] RK Mohanty, David J Evans, and Urvashi Arora. Convergent spline in tension methods for singularly perturbed two-point singular boundary value problems. *International Journal of Computer Mathematics*, 82(1):55–66, 2005.

[6] RK Mohanty and Navnit Jha. A class of variable mesh spline in compression methods for singularly perturbed two point singular boundary value problems. *Applied Mathematics and Computation*, 168(1):704–716, 2005.

[7] RK Mohanty, Navnit Jha, and David J Evans. Spline in compression method for the numerical solution of singularly perturbed two-point singular boundary-value problems. *International Journal of Computer Mathematics*, 81(5):615–627, 2004.

[8] N Papamichael and JR Whiteman. A cubic spline technique for the one dimensional heat conduction equation. *IMA Journal of Applied Mathematics*, 11(1):111–113, 1973.

[9] GF Raggett and PD Wilson. A fully implicit finite difference approximation to the one-dimensional wave equation using a cubic spline technique. *IMA Journal of Applied Mathematics*, 14(1):75–78, 1974.

[10] Richard S Varga. *Matrix iterative analysis*, volume 27. Springer Science & Business Media, 2009.

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

Relay Selection and Beamwidth Adaptation to Overcome Blockage in mm-Wave WPAN

Anindita Saha*, Farah Ashrafi*, Sadia Islam*, Risala Tahsin Khan*, M. S. Kaiser†
*†Institute of Information Technology
Jahangirnagar University, Dhaka-1342, Bangladesh
†Anglia Ruskin IT Research Institute, Anglia Ruskin University, UK
Email:{sahaanindita73,farahashrafi1971,sadiaislam0192}@gmail.com,risala,mskaiser}@juniv.edu

***Abstract*—This work presents relay selection and beamwidth adaptation algorithms for millimeter-wave communication network by selecting best relay. Due to the high path loss, the mm-Wave links suffer from shadowling. Thus improving link availability is a challenge for the system designer. In this work, the link availability is further improved by optimal power allocation and beamwidth adaptation. The channel model considered the path loss and shadowing. The throughput and link availability of the proposed algorithms have been evaluated. The results have been compared with the existing state of art methods. It has been found that the proposed algorithms for the relay selection and beamwidth adaptation outperformed the existing results.**

***Keywords*—mm-Wave; Beamwidth; Power allocation; Relay selection; Blocking.**

I. Introduction

The 60-GHz mill-meter wave (mm Wave) band is very felicitous for Wireless Personal Area Network (WPAN). Mill-meter-wave, which is also known as extremely high frequency by the International Telecommunications Union (ITU), can be used for high-speed indoor wireless broadband communications. Relay devices can significantly overcome signal blockage and propagation loss characteristics of mm Wave.

It needs higher costs in manufacturing of greater precision hardware due to component with smaller size. At extremely high frequencies, there is significant attenuation. Moreover mm Waves can hardly be used for long distance applications.Mm Wave communications suffer from huge propagation loss. Consequently, the penetration power becomes low through concrete walls or other objects. Due to its short range, mm Wave travels by line of sight, so its high-frequency wavelengths can be blocked by physical objects like human bodies , furniture etc. There are interferences with oxygen and rain at higher frequencies therefore further research is going on to reduce this.

A large number of research works have been reported for blockage of mmWave. Previously there were many cross layer anti-blocking scheme used, most of them including multi-hop, deflection routing, scheduling, spatial reuse.Some papers focused on inter and intra-group communication scenarios and proposed two hops relay selection schemes for each case[1]. Some paper proposed relay selection strategies based on the criterion of outage probability minimum or on the criterion of overall network throughput capability Maximum. Another paper proposed a fast relay selection algorithm based on the geographical location information, access point (AP) would simulate[2] and select best relay in preliminary sectors to save the time consumption[3]. Moreover other relay selection strategy establised on the feedback beamforming (BF) information through designed sector sweep (SSW) report frame for mmWave WPANs with the aim of minimizing the outage probability and maximizing the overall network throughput capacity[4].There were also work on relay selection and source and relay power allocation under mixed line-of-sight (LOS) and non-line-of-sight (NLOS) path scenarios for both power saving and robustness enhancement.

This paper describes, the best way of relay selection and the design of beamwidth adaptation for power consumption to cooperative multicast transmission strategy for 60 GHz mm Wave WPANs. After comparing SINR and degree of mobility, from the proposed relay selection algorithm we get the number of best relays. In this paper we introduce a core factor degree of mobility. When the AP transmits packets on carrier frequency, it can be distorted. We consider the capability of moving or of being moved of relays, readily from place to place. We also proposed an algorithm for beamwidth adaptation. Beamwidth adaptation is important to save power. Saving the power is the key to improve system performance for 60 GHz mm Wave.

The remainder of the paper is organized as follows. The system model is described in Section II. Timing diagram is also explained in this paper. A relay selection algorithm is proposed in Section III. An algorithm for beamwidth adaptation is analyzed in Section IV and evaluated by intensive simulations in Section V. Section VI is the conclusion where Appendix A is the proof of the path loss equation and related works are presented in Appendix B.

II. System Model

Figure 1 shows a 60 GHz mm Wave WPANs with multiple relays. The base station (BS) /access point (AP) has information to be transmitted to Wireless User Equipment (WUE). Here AP is a sender and WUE is a destination. If there is a blocking between AP and WUE link. The information can not be transmitted due to shadowing. THus a number of randomly distributed relays assist Access Point (AP) to transmit information to WUE. Our target is to establish a non-line-of-sight (NLOS) connection avoiding the obstacles.

978-1-5090-5769--6/16/$31.00 2016 IEEE

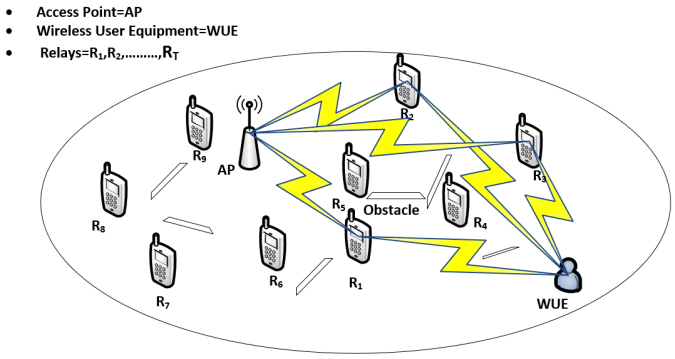


Fig. 1. System Model

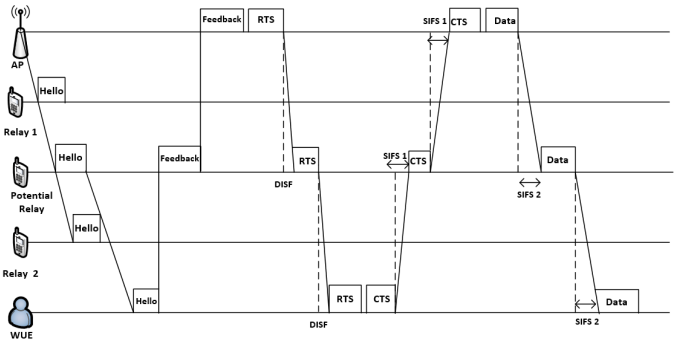


Fig. 2. Timing Diagram for finding relays

In figure 1, there are R_1 to R_T relays, AP can establish a NLOS link with WUE through R_1, R_2, R_3indefinitely these three relays are potential relays.

Figure 2 shows a timing diagram consists of an AP, relays, an WUE. AP is an omnidirectional antenna which sends hello packets in every 3° angle which includes the address of AP and WUE. Whenever there is a relay on the exact angle as well as there is no obstacle between them , it will receive and re-send this packet to the WUE, that relay will be termed potential relay.

Because of the clearance between relay and WUE, WUE will conveniently receive the hello packets. WUE will send a feedback via potential relay with the information of distance, capacity, SINR , degree of mobility. AP subsequently sends a RTS after DIFS duration, WUE will respond with CTS after SIFS1 duration. RTS (Ready to Transmit) frame contains five fields which are frame control, duration, RA (Received Address), TA (Transmitter Address), FCS and DIFS (Distributed Inter-Frame Space). FCS is the time delay to transmit RTS from AP to Potential relay and DIFS is the time delay to transmit RTS from potential relay to WUE. CTS (Clear to Send) frame consists of frame control, duration, RA (Receiver Address), FCS and SIFS (Short Inter-Frame Space). FCS is the time delay to transmit CTS from AP to Potential relay and SIFS is the time delay to transmit CTS from potential relay to WUE. After obtaining CTS , WUE will send data which will be received after SIFS1 by AP , following that AP will send data to WUE which will be received after SIFS2 duration. RTS frame is RTS and CTS frames are used to enhance the virtual carrier sense process. All these are happened via potential relay.

III. Problem Formulation

WUE which is regarded as the destination will decode the signals coming from the AP successfully. Reliability must be considered for choosing appropriate relay. Reliability is denoted as $\mathbb{R}_{\mathrm{i}}$,

$$\mathbb{R}_{\mathrm{i}} = \frac{\psi_{out_{\mathbb{R}_{\mathrm{i}}}}}{\psi_{in_{\mathbb{R}_{\mathrm{i}}}}} \approx 1 \tag{1}$$

where $\psi_{in_{\mathbb{R}_{\mathrm{i}}}}$ out is the packet which is sent by AP and $\psi_{out_{\mathbb{R}_{\mathrm{i}}}}$ is the packet which is received by WUE. If $\mathbb{R}_{\mathrm{i}} < 1$ packet will drop in the channel, $\mathbb{R}_{\mathrm{i}} > 1$ extra packets will be are added with the main packets. Both scenarios are not admissible to select the relay.

Assuming that channel's power is P and channel gain is G. We can get EIRP , $EIRP = P \times G$

There is also a relation between gain, power and SINR as γ,

$$\gamma = \frac{GP\,|h|^2}{\sigma^2} \tag{2}$$

Where h is channel gain and σ is data rate.

Power equation will be,

$$P(d,f)(dB) = P_t(dB) + 10\alpha log_{10} d - 10\beta log_{10}(f') + X_s \tag{3}$$

Where P_t is transmit power, f' is degree of mobility, α and β are constants. We also consider shadowing X_s. X_s is a normal variable with zero mean, reflecting the attenuation (in dB) caused by shadowing.In case of no shadowing, this variable is 0. In case of only shadow fading this random variable may have Gaussian distribution with standard deviation in dB, resulting in log-normal distribution of the received power in Watt. Here, from Doppler shift we get the degree of mobility [10]

$$f' = f(\frac{v - v_0}{v - v_s}) \tag{4}$$

SINR has also relation with capacity C. The capacity of an additive white Gaussian noise (AWGN) channel assumed as Gaussian distribution is given by [11], [12]:

$$C = \frac{1}{2}\log_2(1+\gamma) \tag{5}$$

From this equation (3) we get,

$$\gamma = 2^{2C} - 1 \tag{6}$$

From equations (2) and (6) we can get EIRP.

$$EIRP = \frac{N(2^{2C}-1)}{|h|^2} \tag{7}$$

Let consider the transmitted signal from AP to potential relay is y_1 and another transmitted signal from potential relay to WUE is y_2.

Now,

$$y_1 = \sqrt{P}h_1 + n_1 \tag{8}$$

$$y_2 = (\sqrt{P}h_1 + n_1)Gh_2 + n_2 \tag{9}$$

Here, $\sqrt{P}$ is the power, h_1 is the path loss and shadowing when signal is transmitted from AP to Potential relay, h_2 is the path loss and shadowing when signal is transmitted from potential relay to WUE, n_1 is the noise for the first situation, n_2 is the noise for second situation. G is the channel gain.

Considering these two situations we can get an equation for SINR as,

$$\frac{\sqrt{p}h_1Gh_2}{p\left|h_1\right|^2\left|h_2\right|^2G^2} + \frac{h_1Gh_2 + n_2}{\sigma_1^2 + \sigma_2^2} \tag{10}$$

Due to the doppler shift,a delay of τ may introduce. By considering the effect of γ the(10) no equation would be,

$$\gamma = \frac{P(t-\tau)\left|h_1(t-\tau)\right|^2\left|h_2(t-\tau)\right|^2G^2}{\sigma_1^2 + \sigma_2^2} \tag{11}$$

Let, n number of relays are selected and h_{11} is the path loss and shadowing of AP to potential relay of the first relay,h_{12} is the path loss and shadowing of potential relay to WUE of the first relay. In this way if $n = 2, 3, 4, ..., \infty$ then considering MRC at the relay we can estimate the relay signal as $\bar{S}$

$$\bar{S} = \frac{h_{11}h_{12}y_1 + h_{21}h_{22}y_2 + ... + h_{(n-1)1}h_{(n-1)2}y_{n-1}}{\left|h_{11}\right|^2\left|h_{12}\right|^2 + \left|h_{21}\right|^2\left|h_{22}\right|^2 + ... + \left|h_{(n-1)1}\right|^2\left|h_{(n-1)2}\right|^2} \tag{12}$$

IV. Proposed Algorithm

A. Relay Selection Algorithm

In mm Wave links require high data rate, too many hops can seriously affect the individual flows. The algorithm tries to find maximum relays considering there are one Access Point (AP) and one WUE. Assume that, there is a set ϕ_R of all distributed relays,

$$\phi_R = \{R_1, R_2, ..., R_T\}$$

. Here, T is the total number of relays then there is an another set P_R of all potential relays,

$$\phi_{PR} = \{R_{P1}, R_{P2}, ..., R_{PT}\}$$

.

When the AP transmits packets on carrier frequency, it can be distorted by the motions of the potential relay. We consider the capability of moving or of being moved of relays as degree of mobility. This is a core factor which we are considering in this algorithm.

There is an another factor reliability. If the same transmitted data is not received by WUE then the channel can not be reliable, those relay will not be counted. Weights are assigned on the basis of SINR and degree of mobility, and the capacity C_{W_l} of specified weight is compared with the threshold capacity C_{Th} of WUE. If C_{W_l} is less or equal than C_{Th} then that specific weight which is included in a priority queue (maximum to minimum) (Algorithm 1 lines 6-8). To utilize the power (Algorithm 1 lines 11-16) based on the WUE's capacity, number of hops or number of relays are counted .The best relay would be the channel links with the maximum weights.

Algorithm 1 Relay Selection Algorithm

Input: Input : link length, traffic load, degree of mobility, reliability.
Output: One or Multiple best relays (count).

procedure PRIORITY QUEUE(considering a pair(AP, WUE))
  for $l = 1$ to size $\{\phi_{PR}\}$ **do**
    if relay is reliable
    $W_l(AP, WUE)$ = $\frac{\gamma_i}{\gamma_m} + \frac{f}{f_i'}$
    if $Cw_l <= C_{Th}$ then
    Enqueue $(Queue, W_l)$
  end for
  $count \leftarrow 0$, $totalcapacity \leftarrow 0$
  while (queue is not empty and $totalcapacity <= C_{Th}$) **do**
    value.C= $ExtractMax(Queue)$
    $totalcapacity = totalcapacity + value.C$
    $count = count + 1$
  end while

B. Beamwidth Adaptation Algorithm

Beamwidth Adaptation [Algorithm 2] is important to save power. To get the optimal beamwidth, we propose a algorithm where EIRP is calculated for every 3° angle in between minimum 3° to maximum 45°.Every sector has an angle of 45° so beamwidth cannot be bigger than that.

We calculate the beamwidth for which EIRP is maximum. Here Capacity of beam is also considered. Maximum EIRP with the capacity which is lower than threshold capacity will not be counted. Only that beamwidth will be selected which's EIRP is maximum with greater capacity than receiver's.

Algorithm 2 Beamwidth Adaptation Algorithm

Input: Input : EIRP
Output: Beamwidth θ

1: **procedure** PRIORITY QUEUE(max $EIRP$)
2: **for** ($\theta = 3, \theta <= 45$,θ= $\theta 3++$) **do**
3: calculate C_{EIRP_θ}
4: if$C_{EIRP_\theta} <= C_{Th}$ then
5: Enqueue($Queue, C_{EIRP_\theta}$)
6: **end for**
7: value.θ = $Extract - max(Queue)$
8: **return**

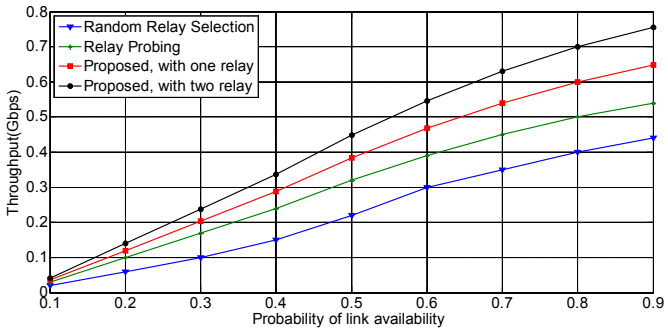


Fig. 3. Comparison of throughput performance under our algorithm with relay probing, and random selection of relay

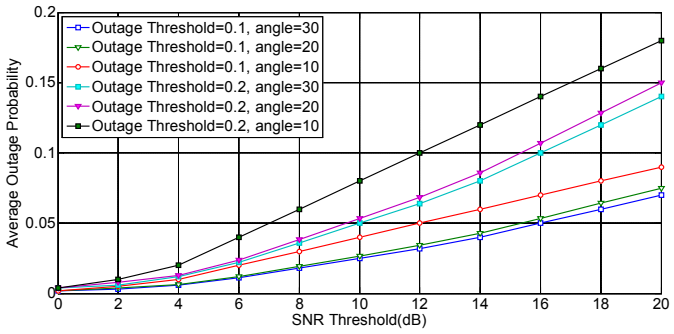


Fig. 4. Effect of SNR threshold on average outage probability

V. Performance Evolution

In this section, we provide performance evolution to generate more insights into the derived theoretical results. As we want to select optimal relay in mm wave the throughput must to be higher effect of the link availability. The Fig. 3 shows the effect of link availability on the throughput of the mm wave communication network. We compare our algorithm with relay probing, and random selection of relay. It has been found that the proposed relay selection technique gives more throughput compared to others. The proposed relay selection uses greedy approach and does not maintain the fairness. The throughput rises if the more relays are considered during selection.

Figure 4 shows the average outage probability raises with the rise of SNR threshold as well as outage threshold of the link. Moreover the less beamwidth increases the link failure probability as node mobility is considered. But it increases the throughout and outage probability.

VI. Conclusion

In this paper, we proposed two algorithms, one is relay selection algorithm and another one is beamwidth adaptation algorithm for mm Wave WPAN. The blocking probability of mm Wave communication network is very high and avoiding that probability is the main aim of this paper. A new factor degree of mobility is introduced here to calculate weights. Path loss and shadowing are also considered by the channel model. Our proposed relay selection algorithm gives us the best one or multiple relays and beam width adaptation algorithm is for saving power. We evaluate the performance of our proposed model with existing ones. Simulation result shows that our proposed model performs better than the other models.

Appendix A
Solution of the Path Loss Equation

The received power equation can be expressed as

$$P_L(f,d)[dB] = 10\alpha log_{10}(d) + 10\beta log_{10}(f) + X_s \quad (13)$$

where β is the shadowing factor.

Assuming, $B = P_L(f,d)[dB]$, $D = 10\log_1 0(d)[dB]$, and $F = 10\log_1 0(f)[dB]$ and consider $X_s = 0$ in equation (13), then equation (13) can be expressed as

$$\beta = B - \alpha D - \gamma F \quad (14)$$

The standard deviation of β is:

$$\beta = \sqrt{\frac{\sum(B - \alpha D - \gamma F)^2}{N}} \quad (15)$$

Minimizing the fitting error is equivalent to minimizing

$$\sum(B - \alpha D - \gamma F)^2$$

which means its partial derivatives with respect to α, γ should be zero, as shown by equation (13), and equation (14)

$$\frac{\delta\sum(B - \alpha D - \gamma F)^2)}{2\delta\alpha} = \alpha\sum D^2 + \gamma\sum DF - DB = 0 \quad (16)$$

$$\frac{\delta\sum(B - \alpha D - \gamma F)^2)}{2\delta\gamma} = \alpha\sum DF + \gamma\sum F^2 - \sum FB = 0 \quad (17)$$

It is found from equation (16) and equation (17) that

$$\alpha\sum D^2 + \gamma\sum DF - DB = 0 \quad (18)$$

$$\alpha\sum DF + \gamma\sum F^2 - \sum FB = 0 \quad (19)$$

Through calculation, we obtain the closed form solutions for α and γ using equation (18), and equation (19). Finally, the minimum standard deviation of β for path loss can be obtained by equation (15).

References

[1] W. u. Rehman, T. Salam and X. Tao, *Relay selection schemes in millimeter-wave WPANs*, 2014 International Symposium on Wireless Personal Multimedia Communications (WPMC), Sydney, NSW, 2014, pp. 192-197. doi: 10.1109/WPMC.2014.7014815.

[2] H. Chu, P. Xu, C. Yang, B. T. Oanh and S. Zhang, *Joint relay selection and power control for robust cooperative multicast in MmWave WPANs*, 2015 IEEE International Conference on Communication Workshop (ICCW), London, 2015, pp. 2787-2792. doi: 10.1109/ICCW.2015.7247601

[3] K. Song, R. Cai and D. Liu, *A fast relay selection algorithm over 60GHz mm-wave systems*, Communication Technology (ICCT), 2013 15th IEEE International Conference on, Guilin, 2013, pp. 676-680. doi: 10.1109/ICCT.2013.6820460

[4] H. Chu and P. Xu,*Relay Selection with Feedback Beamforming Information for NLoS 60GHz mm-Wave WLANs/WPANs*, in National Mobile Communications Research Laboratory, Southeast University, in IEEE ICC 2014 - Wireless Communications Symposium.

[5] G. Zheng, C. Hua, R. Zheng and Q. Wang, *A robust relay placement framework for 60GHz mmWave wireless personal area networks*, 2013 IEEE Global Communications Conference (GLOBECOM), Atlanta, GA, 2013, pp. 4816-4822. doi: 10.1109/GLOCOMW.2013.6855713

[6] M. S. Kaiser, M H Chaudary, R A Shah, K M Ahmed,*Neuro-fuzzy (NF) based relay selection and resource allocation for cooperative networks*, Electrical Engineering/Electronics Computer Telecommunications and Information Technology (ECTI-CON), 2010 International Conference on, Chaing Mai, 2010, pp. 244-248.

[7] M. A. Kabir and M. S. Kaiser, *Outage capacity analysis of MC-CDMA based on cognitive radio network*, Electrical Engineering and Information Communication Technology (ICEEICT), 2015 International Conference on, Dhaka, 2015, pp. 1-5. doi: 10.1109/ICEEICT.2015.7307463

[8] S. Kato, *R & D on Millimeter Wave (60 GHz) Systems and IEEE Standardization Updates*, 2007 IEEE International Conference on Ultra-Wideband, Singapore, 2007, pp. 514-517. doi: 10.1109/ICUWB.2007.4380999

[9] Ning Wei, Xingqin Lin, and Zhongpei Zhang, *Optimal Relay Probing in Millimeter Wave Cellular Systems with Device-to-Device Relaying*, in 8 oct 2015.

[10] Nusrat Jahan, Fariha Afsana, Mufti Mahmud, M. S. Kaiser, *An Adaptive Link Selection Algorithm for Cognitive Cooperative Network Using Modified Bat Algorithm*, In the Proceeding of 1st IEEE International Conference on Telecommunications and Photonics (ICTP), 2015, Dhaka.

[11] A. Sarker, N. Fatema, B. R. Binti and M. S. Kaiser, *Performance evaluation of energy efficient routing algorithm for ad-hoc network*, Electrical Engineering and Information & Communication Technology (ICEEICT), 2014 International Conference on, Dhaka, 2014, pp. 1-6. doi: 10.1109/ICEEICT.2014.6919094

[12] H, Chu, and P. Xu, W. Wang,C. Yang and W. Zhu, *BF-assisted joint relay selection and power control for cooperative multicast in MmWave networks*," 26th IEEE Annual International Symposium on Personal, Indoor, and Mobile Radio Communications, PIMRC 2015, Hong Kong, China, August 30 - September 2, 2015, pp. 2255–2259. doi:10.1109/PIMRC.2015.7343673