THE PERFORMANCE OF CLUSTERING TECHNIQUE AND ARTIFICIAL INTELLIGENCE IN POWER SYSTEM FAULT INVESTIGATION

AMALINA ABDULLAH

School of Electric and Electronic Engineering, University Science Malaysia, Penang, Malaysia
E-mail: eenora@usm.my

Abstract - This paper studied the performance of clustering technique and artificial intelligence in power system fault investigation. The study looking forward pertinent to fault classification and fault location. In terms of clustering technique, the focus is on Gustafson-Kessel clustering algorithm. Where as the Adaptive Neuro-Fuzzy Inference System (ANFIS) is one of the technique in artificial intellkigence that believed to give a great performance. Both techniques were combined due to their advantages and tested using a number of cases pertinent to a few types of fault at power transmission line. This idea has potential to analyze the fault location without the knowledge of line parameters. The results show that the scheme can perform well and has a high percentage of accuracy.

Keywords - Gustafson-Kessel Clustering Algorithm, Power Transmission Line, Adaptive Neuro-Fuzzy Inference System

I. INTRODUCTION

The reliability and uninterrupted of electric supplies for power transmission lines are very essential. Research of fault analysis pertinent to fault detection,classification and location is important and possess endure. The deterministic conventional approach has helped a lot in fault analysis. However, the approaches have a few disadvantages such as less accuracy due to wide variation in respective system and complex mathematical calculation that need the handling of an expert operator.Most of the techniques for identifying fault location depend on parameters of transmission line that require a complex mathematical solution. A parameter free algorithm is an option to overcome this problem by using an artificial intelligence (AI) technique.Each type of fault occurs in the transmission line systemhas their personal characteristics which will lead for better understanding on developing a proper system.

Protection systems for transmission lines are among the most important parts in power systems.Thus, the reliability and uninterrupted of electric supplies are very essential. Each type of fault occurs in the transmission line systemhas their personal characteristics which will lead for better understanding on developing a proper scheme.The increasing complexities of the modern power transmission systems required the significance of fault location research studies. Most of the techniques for identifying fault location depend on parameters of transmission line. Thus, a complex mathematical solution will be considered.The deterministic conventional approach has helped a lot on fault identification and diagnosis. However, they have a few disadvantages such as less accuracy due to wide variation in respective system and complex mathematical calculation which need the handling of an expert operator. Accuracy becomes more important when the physical dimensions and size of the transmission lines are being considered.

This study proposed an implementation of clustering algorithm in an intelligent technique known as Adaptive Neuro-Fuzzy Inference System (ANFIS) combined with the advantage of wavelet transform as a pattern recognition method. This paper presents the training procedures and performance for the proposed scheme. It is organized as follows:Section II describes about the Fuzzy Clustering Algorithms. Section III explained ATP/EMTP for power system simulation. Fault detection algorithm is explained in section IV. Section V provides the arcitechture of ANFIS. Finally, conclusions are given in section VI.

II. FUZZY CLUSTERING ALGORITHMS

Conventional clustering means the process of classifying a set of object to respective cluster. Nevertheless, such a process is insufficient to represent many actual situations. Thus, a fuzzy clustering method is presented to construct clusters with uncertain boundaries.Asa result, this method will allow that one object might belong to some overlapping clusters to some degree. The fundamental of of fuzzy clustering is to deliberatethe clusters and which degree do the objects belong to the clusters [4]. As for fuzzy clustering, it is defined as a partition based clustering scheme and the advantage is when there are no clearof a definite group in the data set. Every individual data object belongs to not one, but all of the clusters with varying degrees of membership. However, there are very few instances where a partitioning scheme has been used to analyze families of molecular conformation. Table 1 shows the general review of comparison on Fuzzy C Varieties (FCV), Adaptive Fuzzy Clustering (AFC), Fuzzy C Means (FCM) Algorithm, Gustafson Kessel (GK) Algorithm and Gath-Geva (GG) Algorithm.

Among these algorithms, FCM and GK will be further discussed.

Fuzzy C-means (FCM) is known as a data clustering technique which each data point belongs to a cluster to some degree that is specified by a membership grade. This technique was introduced by Jim Bezdek in 1981 [1] as an improvement on earlier clustering methods [5]. In FCM, data are assured to each cluster by means of a membership function, as the fuzzy behavior of this algorithm. For this purpose, we simply have to build an appropriate matrix named U that has factors of numbers between 0 and 1. The next step is building up the membership's degree between data and the centers of clusters. However, FCM clustering is sensitive to the selection of distance metric. In [6], the author has claimed that the Euclidean distance possible for improvement when all clusters are *spheroids* with same size or when all clusters are well separated. It is based on minimization of the following objective function:

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}^m \left\|x_i - c_j\right\|^2, 1 \le m < \infty$$

Where m is any real number greater than 1, u_{ij} is the degree of membership of x_i in the cluster j, x_i is the ith of d-dimensional measured data, c_j is the dimension center of the cluster, and ||*|| is any norm expressing the similarity between any measured data and the center.Fuzzy partitioning is executed through an iterative optimization of the objective function shown above, with the update of membership u_{ij} and the cluster centers c_j by:

$$u_{ij} = \frac{1}{\sum_{k=1}^{C}\left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|}\right)^{\frac{2}{m-1}}}$$

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^m \cdot x_i}{\sum_{i=1}^{N} u_{ij}^m}$$

The iteration will stop if a termination criterion between 0 and 1, whereas k is the iteration steps. This practice related to a local minimum or a saddle point of J_m.

The proper algorithm is :

1. Initialize U=$[u_{ij}]$ matrix, $U^{(0)}$

2. At k-step: calculate the centers vectors $C^{(k)}=[c_j]$ with $U^{(k)}$

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^m \cdot x_i}{\sum_{i=1}^{N} u_{ij}^m}$$

3. Update $U^{(k)}$, $U^{(k+1)}$

$$u_{ij} = \frac{1}{\sum_{k=1}^{C}\left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|}\right)^{\frac{2}{m-1}}}$$

4. If $\| U^{(k+1)} - U^{(k)}\| < \varepsilon$ then STOP; else return to step 2.

The Gustafson-Kessel (GK) algorithm [7] is a great clustering technique that has been implemented in various domainsincluding classification, image processing, and system identification [8]. In [6, 8], the authors proposed a GK algorithm which uses the well-known Mahalanobis distance as the metric in FCM. They found that the GK algorithm is better than Euclidean distance based algorithms when data shape is being considered. In [10], the authors proposed a new toughformulation, which is well-known from the Euclidean distance, to increase the performance of FCM. It has a benefit which the local adaptation of the distance metric to the shape of the cluster is done by estimating the cluster covariance matrix [9]. The FCM-GK algorithm is based on iterative optimization of an objective functional of the c-means type:

$$J_m(U, V, \{A_i\}) = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^m d_{ikA_i}^2$$

The distance norm d_{ikA_i} as in the above equation.The metric of each cluster is defined by a local norm-inducing matrix Ai, which is used as optimization variables in the functional. This condition will allow the distance norm to adapt to the local topological arrangement of the data. The minimization in the aim of GK functional is achieved by using the alternating optimization (AO) method according to the well-known algorithm [1].For l = 1; 2; _ _ _ s where l represent every step in the iteration process and s is the maximum number of iterations allowed (stop criteria).These steps explained the implemented modified GK algorithm;

Step 1: Compute cluster prototypes (means)

$$V_i^{(l)} = \frac{\sum_{k=1}^{N}(\mu_{ik}^{(l-1)})^m Z_k}{\sum_{k=1}^{N}(\mu_{ik}^{(l-1)})^m}, \quad 1 \le i \le K \ (1)$$

Step 2: Compute the cluster covariance

$$F_i = \frac{\sum_{k=1}^{N}\left(\mu_{ik}^{(l-1)}\right)^m (z_k - V_i^{(l)})(z_k - V_i^{(l)})^T}{\sum_{k=1}^{N}\left(\mu_{ik}^{(l-1)}\right)^m}, \quad 1 \le i \le K \ (2)$$

Add a scaled identity matrix:

$$F_i = (1-\gamma)F_i + \gamma \det(F_0)^{\frac{1}{n}} I \ ,(3)$$

Extract eigenvalues λ_{ij} and eigenvectors ϕ_{ij} from F_i.

Find $\lambda_{i\,max} = max_j \lambda_{ij}$ and set:

$$\lambda_{ij} = \lambda_{i\,max} / \beta \quad \forall j \text{ for which } \lambda_{i\,max} / \lambda_{ij} > \beta (4)$$

Reconstruct F_i by

$$= [\phi_{i1} \dots \phi_{in}]\,\mathrm{diag}(\lambda_{i1}, \dots, \lambda_{in})[\phi_{i1} \dots \phi_{in}]^{-1}$$

Step 3: Compute the distances

$$D^2_{ikA_i} = (z_k - V_i^{(l)})^T [\rho_i \det(F_i)^{\frac{1}{n}} F_i^{-1}] (z_k - V_i^{(l)}) ,$$

$1 \le i \le K$, $1 \le k \le N$.(5)

Step 4: Update the partition matrix

for $1 \le k \le N$, if $D_{ikA_i} > 0$ for $1 \le i \le K$,

$$\mu_{ik}^{(l)} = \frac{1}{\sum_{j=1}^{K}(D_{ikA_i}/D_{jkA_i})^{2/(m-1)}} \quad \text{otherwise}$$

$\mu_{ik}^{(l)} = 0$ if $D_{ikA_i} > 0$, and $\mu_{ik}^{(l)} \in [0,1]$ (6)

with $\sum_{i=1}^{K} \mu_{ik}^{(l)} = 1$ otherwise until $\|U^{(l)} - U^{(l-1)}\| < \epsilon$

This algorithm was implemented using the improved covariance estimation proposed by [1]. Another main feature is the ability on adapting the distance-inducing matrix correspondingly, when the GK algorithm is implemented in the extraction of Takagi–Sugeno fuzzy model from data. Overfitting is able to be reduced when the number of training samples is low in comparison to the amount of clusters. This is attained by including a scaled unity matrix to the calculated covariance matrix. In implementing the GK clustering algorithm, it is applied with Takagi Sugeno models [7). A drawback of this algorithm is that only clusters with equal volumes can be found and the resulted clusters cannot be directly used to form membership functions. The GK algorithm allows calculating an adaptive distance as a way to detect clusters with different geometries within a data set. Using a modified GK algorithm [11] has improved the performance of training stage and able to find a partitioning of the data.

III. ATP/EMTP FOR POWER SYSTEM SIMULATION

A test system of transmission system is developed by using the ATP/ EMTP program. The purpose is to obtain a large amount of data through the simulation. The power system model is 100km double circuit transmission line, has two generators of115 kV each located on either end of the transmission line along with a three phase fault simulator used to simulate faults at various positions on the transmission line. Figure 1 shows the one-line diagram of the system. The robustness of double circuit/ parallel line is most significant since the faults are commonly happened in this type of system voltage [10]. The parameters setting for generating training patterns are as follows:

- Fault types consist of single line to ground (SLG), double Line (DL), double line to ground (DLG) and three phase (ABC)
- Fault location on the each transmission line were at every 2km measured from the sending end.
- The fault resistances; 0.5Ω, 5 Ω, 25 Ω, 100 Ω, 200 Ω.
- The loading conditions over the protected transmission line is assumed to be nearly constant.

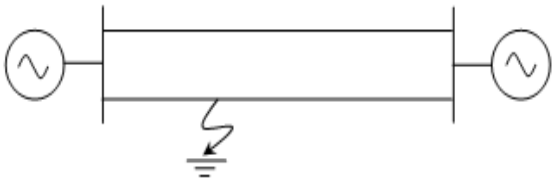


Figure1. One-line diagram

IV. FAULT DETECTION ALGORITHM

Fault detection is the first phase of the propose algorithm as illustrated in figure 2. For a fault waveforms, RMS value varies as the amplitude of the waveform varies.. If a fault occurs, the voltage rms value of the phase in fault decreases while corresponding current rms value increases.

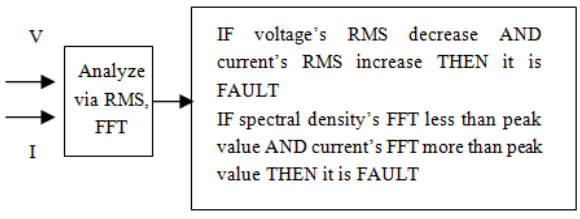


Figure 2. Rules for fault detection algorithm

FFT is a tool which effectively analyzes different frequency components present in a signal. For no fault, waveforms have a strong peak at 50 Hz for all 3 phases with almost same amplitude. For fault condition, the magnitude of peak at 50 Hz of FFT for the voltage in fault will decrease significantly.

V. ANFIS ARCHITECTURE

After the process in figure 2, the fault type and estimation of distance will be analyze. Wavelet transforms is used to give information about the frequency components present in the signal [3]. It has received great attention due to its advantage on analysis of transient waveform better than other types of transform approach [4]. In most of the reference research papers [5-9], wavelet coefficients are used for data input to the neural network. These coefficients are obtained by applying Digital Wavelet Transformation (DWT) on the data under study [10]. This paper implemented the wavelet transform for data reduction pertinent to detailed coefficient (Cd). DWT employs two sets of functions, called scaling functions and wavelet functions, which are associated with low pass and high pass filters, respectively. Align with Nyquist's rule, the signal decomposed into different frequency bands as high pass and low pass filtering of the time domain signal**.** After passing the signal through a half band low pass filter, half of the samples can be eliminated since the signal now has the highest frequency of p/2 radians instead of p radians.

Next step is implementation of ANFIS architecture which contains a 5-layer feedforward neural network as shown in Figure 3. ANFIS has several properties where the output is Takagi Sugeno-type (TS) of zero-th order. ANFIS has been used in fault detection and fault classification in power system during the last 2 decades [10]. Recently, the researchers have spent their interest in its capability for estimation of fault location in transmission line and distribution network [10, 14-15].

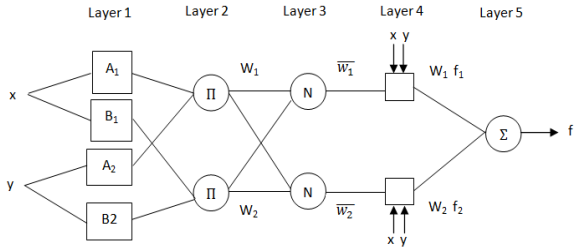


Figure 3. 5 layer architectures in ANFIS

TS Fuzzy which implemented as the main model in ANFIS requires two main stages; structure identification and parameter optimization [11-12]. Gustafson-Kessel (GK) clustering algorithm that has a lot of advantages [12-14] is employed to detect clusters lies in linear subspaces.Itemploys an adaptive distance norm to detect clusters of different geometrical shapes in one data set and possible to reduce the overfitting when the number oftraining samples is low in comparison to the number of clusters. Figure 4 shows the proposed ANFIS procedure. It ivolves a few important steps or stages in signal processing and pattern recognition. These are importance pre process for the input data.

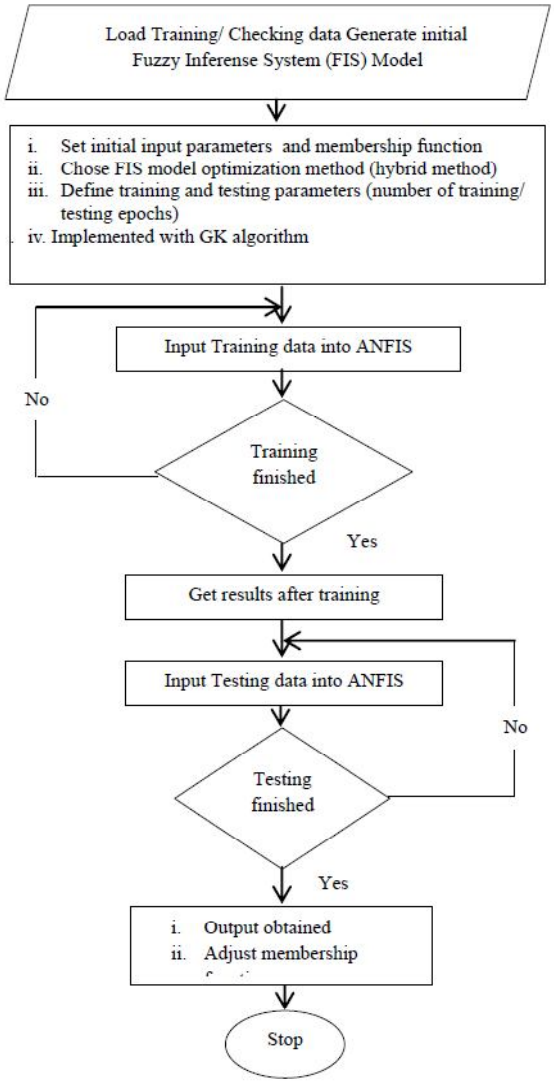


Figure 4. The proposed ANFIS structure

VI. RESULTS AND DISCUSSION

The first phase in identification is the fault classification. The inputs are 6 wavelet detail coefficients of the fault data; Cd_Va, Cd_Vb, Cd_Vc, Cd_Ia, Cd_Ib and Cd_Ic. Then it classed the data into two clusters using modified GK algorithm which is better than FCM. The membership of each data point is calculated for each defined class. The data point may belong to more than one class with a certain degree of membership. The membership to one class is assigned based on the highest value of the membership value. Figure 5 shows the analysis using confusion matrix. It illustrated the performance of fault classification pertinent to single line to ground. Class 1 stand for AG, class 2 for BG and class 3 for CG. It shows that the three types of fault were able to be classified between the actual and predicted.

The second phase in identification is fault location estimation. In this part, the same strategy has been used. We knew that our aim is to identify faults data pertinent to phase to ground (AG), phase B to ground (BG) and phase C to ground (CG). These are the class of fault data. The input is six wavelet coefficients. The same strategy on membership value and clustering are used in the program. The distance of fault location will be predicted based on the developed algorithm. Figure 6 shows the results of training and testing data set for respective distance and fault resistance. Actual and predicted values mean the error in calculating estimated distance from fault location. The plotting data of actual and predicted values are almost overlapping that shows the results are closely accurate. Table 1 presents the results of fault distance estimation pertinent to training phase.

Table 1. Results of Fault Distance Estimation

Fault Distance Estimation For Single Line to Ground					
Actual distance(km)	Estimation Error (Actual–Predicted)				
	R_F=0.5Ω	R_F = 5Ω	R_F = 25Ω	R_F = 100Ω	R_F = 200Ω
2.0	0.12	0.10	-0.096	-0.50	-0.53
10.0	-0.14	0.19	-0.16	-0.30	-0.26
20.0	0.29	0.26	-0.38	0.10	-0.22
30.0	-0.37	0.56	0.89	-0.15	-0.34
40.0	-0.27	-0.15	-0.55	0.10	0.25
50.0	0.12	0.19	-0.46	0.85	-0.13
60.0	0.13	0.29	-0.13	0.27	-0.32
70.0	-0.51	-0.26	0.12	0.11	-0.12
80.0	-0.14	-0.17	-0.12	0.59	0.32
90.0	0.54	0.13	0.79	0.28	0.12
100.0	-0.22	0.54	0.12	0.12	0.43

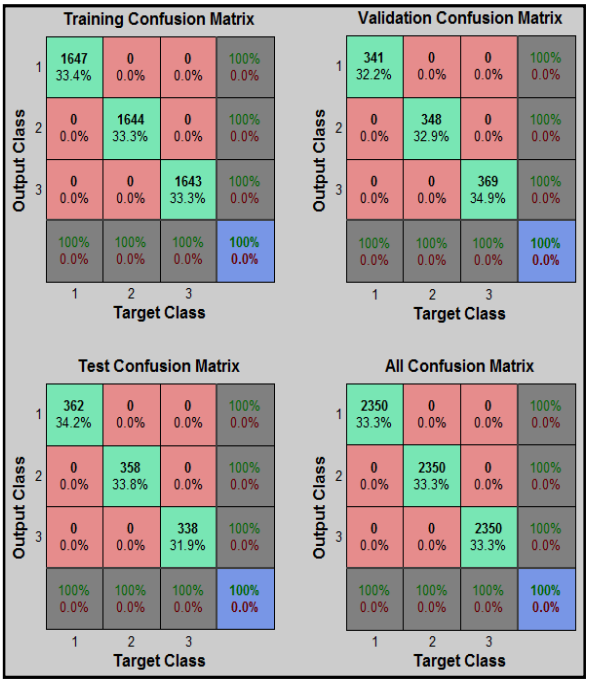


Figure 5. Confusion matrix on fault classification

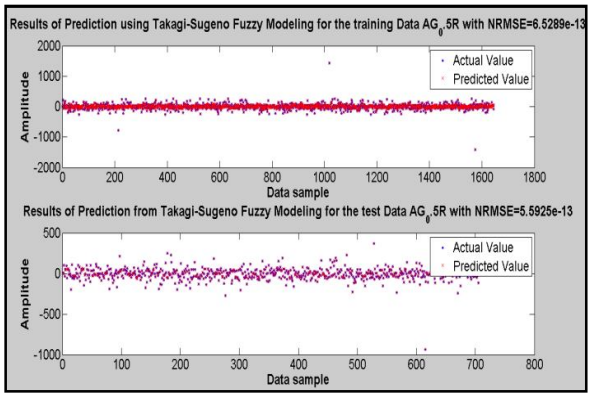


Figure 6. Illustrations of actual and predicted value

CONCLUSION

In this paper we have discussed a few fuzzy clustering algorithms on their advantages, disadvantages and their applications. A novel approach on GK algorithms in fault identification applied in fault analysis also presented. GK is possible to overcome clustering problem by using the distance matrice characteristics and capable to prevent overfitting that important in pattern recognition aspect. The proposed method also has very less involving of calculation and almost no affected by the fault resistance. The aspect of series impedance matrix per unit length is not considered to identify the fault distance. The results of fault distance are able to be estimated regardless the type of fault. Future work will focus on adapting the proposed method in real application where the actual fault data pertinent to type and distance are unknown.

REFERENCES

[1] Sato-Ilic, Mika; Jain, Lakhmi C., "Innovations in Fuzzy Clustering", Electronic Book, 2006

[2] J. C. Bezdek, "Pattern Recognition with Fuzzy Objective Function Algorithms", Plenum Press, New York, 1981

[3] F. Magnago and A. Abur, "Fault Location Using Wavelets", IEEE Transactions on Power Delivery, Vol. 13, Issue 4, October 1998.

[4] Zhao, S.Y., "Calculus and Clustering", China RenmingUniversity Press. 1987

[5] Gustafson, D.E., Kessel, W., "Fuzzy Clustering With aFuzzy Covariance Matrix", In: Proceedings of IEEE Conference on Decision Control, San Diego, CA, pp. 761–766, 1979

[6] Dave, R.N., "Boundary Detection Through Fuzzy Clustering", In IEEE International Conference on Fuzzy Systems, pages 127-134, San Diego, USA, 1992

[7] Babu˘ska, R., "Fuzzy Modeling for Control", Kluwer Academic Publishers, Boston, USA, 1998

[8] Raghu Krishnapuram and Jongwoo Kim, "A Note on the Gustafson–Kessel and Adaptive Fuzzy Clustering Algorithms", IEEE Transactions On Fuzzy Systems, VOL. 7, NO. 4, August 1999

[9] Wu, K.L., Yang, M.S., "Alternative C-Means ClusteringAlgorithms". Pattern Recog. 35, 2267–2278, 2002

[10] Zhang Qing-Chao et. Al, "Fault Location of Two-Parallel Transmission Line For Double Phase-to-Earth Fault Using One-Terminal Data", Journal of Zhejiang University, Vol. 4, No. 5, pp. 520 – 525, 2003

[11] H¨oppner, et. Al.: "Fuzzy Cluster Analysis", John Wiley & Sons, Chichester, England, 1999

[12] Paasche, et. Al, "Rapid Integration Of Large Airborne Geophysical Data Suites Using a Fuzzy Partitioning Cluster Algorithm" Exploration Geophysics 40, 277-287, 2009

[13] Dibya Jyoti Bora, Anil Kumar Gupta, "A Comparative study Between Fuzzy Clustering Algorithm and Hard Clustering Algorithm", International Journal of Computer Trends and Technology (IJCTT), vol. 10 no. 2, Apr 2014

[14] Ramadoni Syahputra, "A Neuro-fuzzy Approach For the Fault Location Estimation on Unsynchronized Two-Terminal Transmission Line", Int. Journal of Computer Science & Information Technology, Vol 5., No 1, February 2013

★★★