ACEEE *Proc. of Int. Conf.* on *Recent Trends in Information, Telecommunication and Computing, ITC*

Real-Time Credit-Card Fraud Detection using Artificial Neural Network Tuned by Simulated Annealing Algorithm

Azeem Ush Shan Khan, Nadeem Akhtar and Mohammad Naved Qureshi
Aligarh Muslim University, Department of Computer Engineering, Aligarh, India
Email: azeem5257@gmail.com
Aligarh Muslim University, Department of Computer Engineering, Aligarh, India
Email: {nadeemalakhtar, navedmohd786}@gmail.com

***Abstract*— Now-a-days, Internet has become an important part of human's life, a person can shop, invest, and perform all the banking task online. Almost, all the organizations have their own website, where customer can perform all the task like shopping, they only have to provide their credit card details. Online banking and e-commerce organizations have been experiencing the increase in credit card transaction and other modes of on-line transaction. Due to this credit card fraud becomes a very popular issue for credit card industry, it causes many financial losses for customer and also for the organization. Many techniques like Decision Tree, Neural Networks, Genetic Algorithm based on modern techniques like Artificial Intelligence, Machine Learning, and Fuzzy Logic have been already developed for credit card fraud detection. In this paper, an evolutionary Simulated Annealing algorithm is used to train the Neural Networks for Credit Card fraud detection in real-time scenario. This paper shows how this technique can be used for credit card fraud detection and present all the detailed experimental results found when using this technique on real world financial data (data are taken from UCI repository) to show the effectiveness of this technique. The algorithm used in this paper are likely beneficial for the organizations and for individual users in terms of cost and time efficiency. Still there are many cases which are misclassified i.e. A genuine customer is classified as fraud customer or vise-versa.**

***Index Terms*— Credit Card fraud detection, Simulated Annealing, Machine Learning, Training, Classification, Artificial Neural Network (ANN), Activation Function.**

I. INTRODUCTION

Credit card fraud is a kind of theft or unauthorized activity to make payment using credit card in an electronic payment system as a fake source of fund. The purpose of credit card fraud is to obtain money or make payment without owner permission. It involves illegal use of card or card information without the owner permission though it is a criminal deception and banned by laws. Because of the advancement in technology and software's, users can hide their identity and locations while committing any transaction over the web, which increases the fraud over the web. There are many methods of credit card fraud depicted in "figure 1", all are mentioned in details "see [1]", are as follows:

1. First fraud is that user using its own credit card knowing that they have no money in his credit card and

DOI: 02.ITC.2014.5.65

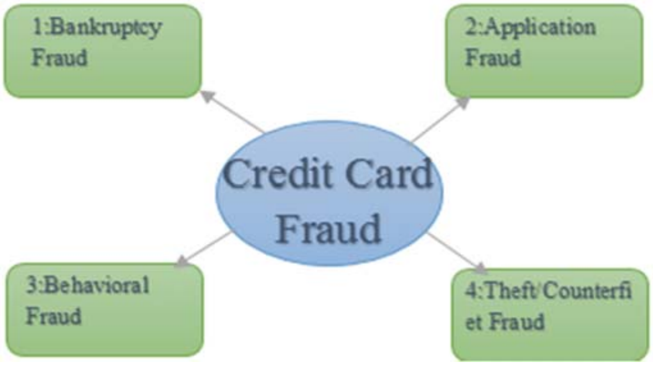

Fig 1: Types of Credit Card Fraud

bank has to pay by sending bill to the address.

2. Second fraud is done when submitting the application form to the bank for issuing the credit card with the fake information.
3. Third fraud is done online when purchasing any item by submitting the card information of any credit card without letting known to the owner.
4. Fourth fraud is the stealing of any credit card and using it by showing as an owner of this card until that card become blocked by bank.

Credit card fraud affect the organization by financial losses and individual user also affected if the information of credit card get steal. So it is important to find a solution which classifies a transaction into fraud or non-fraud. Many techniques have been developed for credit card fraud detection like Artificial Intelligence & Machine Learning and also based on locations [2]. In this paper, we focus on Machine Learning technique, basically it provides a system which is supposed to classify a current transaction into fraud or non-fraud.

In this paper, we are taking credit card fraud detection problem as a classification problem. Many classification algorithm have been developed [3], but the most popular one is Decision Tree. This algorithm is already been proposed for credit card fraud detection problem ("see [4]"). Basically there are two technique for credit card fraud detection:

1. Supervised
2. Un-supervised

These are the machine learning techniques, in which the first one uses training data, to build the model, which have all the attributes including class label i.e. it already contains the attribute which tells whether this previous transaction is fraud or not. And in the second technique, training data does not contain the class label i.e. this technique is class less. More study on these can be found in [5]. This paper propose a credit card fraud detection technique using Neural Networks and Simulated Annealing algorithm to adjust the weight of the neural network, Neural Network is a supervised machine learning technique. Many research papers has already been proposed for fraud detection using Neural Networks and many researchers have been uses Genetic Algorithm to adjust the weight of Neural Network in different fields [6]. This study of using Simulated Annealing to train Neural Network is one of the first to use for credit card fraud detection on real data set provided by UCI repository [7]. Basically, annealing is a process of heating and then cooling a solid to change the hardness of the solid and simulated annealing is to emulate this process.

The main aim of this process is to build a training model on the basis of previous transactions, called training data, for fraud detection. Once a learning of training model is complete, the model is capable of classifying the unseen online transaction as fraudulent or non-fraudulent in real time [8] & [9].

II. Artificial Neural Network

Artificial neural network works in the same way as a human brain does, human brain consist of number of neurons connected with each other, in the same way ANN consists of artificial neurons, called nodes in network, connected with each other. The idea of Artificial Neural Network was presented in late 1943 by Walter Pitts and Warren S.McCulloch as a data processing unit for classification or prediction problems [10]. For the first time, Dorronsoro "et al." in 1997 developed a system to detect credit card fraud by using

Neural Network. Now-a-days, ANN have been successfully applied in business failure prediction, stock price prediction, credit fraud detection and many more area.

ANN comes in many forms like Recurrent NN, Associative NN, etc. In this paper, we will discuss Feed-Forward Neural Network which will trained by Simulated Annealing method.

"Figure 2" depict a simple multi-layer feed forward neural network. It consist of an Input layer, an output layer and an Hidden Layer, hidden layer depends on the problem we are going to solve, it can be no or more than one hidden layer. The number of neurons in input layer corresponds to the number of input attributes in the training dataset which we will see later in this paper and the number of neurons in output layer is depend on the type of problem you are going to solve, in credit card fraud detection case we have two output one is fraud and the other one is non-fraud i.e., 0 and 1 respectively.

In the Feed-forward neural network, as we can see in "fig. 2", there is no feedback loop. Each of the neurons in each layer is connected with each other without making any loop and the link between these neurons has weights, represented by Wij. The connection between each neuron do not perform any calculation but is used to store the weights. These weights are initialized with some random values and changes at every iteration in training process. The simple neuron in each layer are often called perceptron is the simplest neuron network.

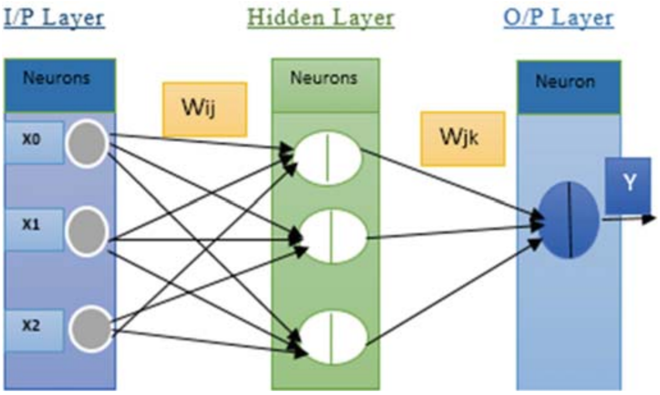

Fig 2: Simple Feed-Forward Neural Network

A feed-forward perceptron works by sending the input to the neurons and send to the output neuron after processing. This is a simple neuron, i.e. perceptron, "figure 3" which has three input and two output.

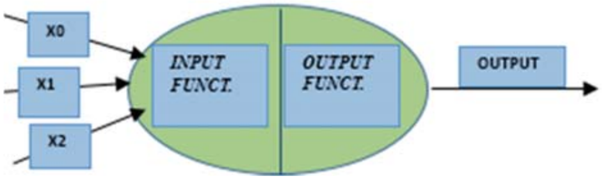

Fig 3: Simple Feed-Forward Perceptron

All the perceptron in the Neural Network have two functions i.e. Input and Activation Function. As the name suggest, the input function collects all the input and perform summation function on the input and then transfer the result to the activation function. An activation function perform some operation on the result after summation and then transfer to the next level. Let's take an example, in the above figure we have three inputs, let's say, I0, I1, I2 and two output Z1, Z2 and their corresponding weights. Now, input function will perform the summation on the inputs multiplied by their corresponding weights. Let's say the output of input function is S.

$$S = \sum_{2}^{x=0} Ix.Zx$$

The result of this summation function is then pass to activation function. Activation function scale the value of S in proper range. Common activation function are sigmoid activation function which works on threshold, if the value of S exceeded the threshold value then the node pass output.

There are two activation function which is commonly used in Neural Networks, Sigmoid and Hyperbolic Tangent Activation Function. It depends on the training dataset on which we are going to train the network that which activation function is good.

The “figure 4” shows sigmoid activation function graph, which refers to one of the case of logistic function. It works for real input values and it only returns positive value (“refer [11]”). The formula of sigmoid function is:

$$S(t) = \frac{1}{1+e^{-t}}$$

The Hyperbolic Tangent activation function (TANH) is the next version of sigmoid function because it produces both negative as well as positive values as shown in “figure 5”.

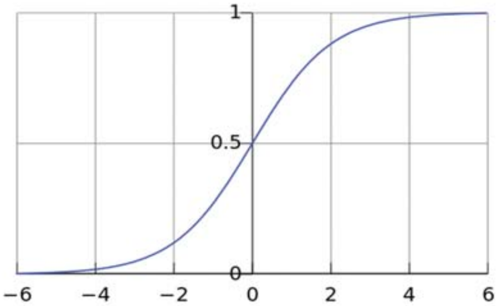

Fig 4: Sigmoid Activation Function Graph

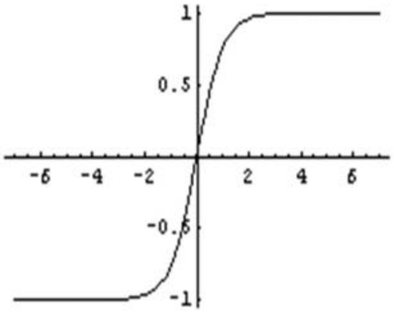

Fig 5: Hyperbolic Tangent Activation Function Graph

The equation of hyperbolic tangent function is given by:

$$F(x) = \frac{e^{2x}-1}{e^{2x}+1}$$

In this project, we have use both of the above activation function and the result of fraud detection is better with TANH.

III. Simulated Annealing

Basically Annealing is a thermodynamics process, it is a heat treatment process upon any metal to change the structure of the metal. It involves heating of any metal slightly above its critical temperature and then cooling it down slowly. It makes the metal harder or stronger and makes the structure of the metal homogenous. The emulation of the process of annealing is called Simulated Annealing.

This method was developed by adapting some changes in Metropolis-Hastings algorithm, also known as Monte Carlo method, invented by M.N. Rosenbluth and published in a paper in 1953 [12]. This method was developed by Scott Kirkpatrick, C. Daniel Gelatt and Mario P. Vecchi in 1983 [13], and later on by Vlado Cerny in 1985 [14]. Corana “et al.” (1987) and Goffe (1994) had proposed some changes which was suitable to train discrete-valued weights. In this study, the implementation of simulated annealing is based on these algorithms, which is adjusted to find the best configuration of weights in artificial neural network. The basic procedure are as follows:

1. Heat the system at high temperature T and generate a random solution.
2. As the algorithm progress, T decreases at each iteration and each iteration forms a nearby model.
3. Then cool the system slowly until the minimum value of T is reached and generate a model at each iteration, which takes the system towards global minima.

In each iteration, a solution is generated which is compared with current solution, by using acceptance function, if it is better than current solution than it get replaced by this solution. The terminology and definitions used in Simulated Annealing is defined in [15] and [16]. These definitions are used in this paper to train the neural network for fraud detection. The main definitions which is needed for this algorithm are: (1) a method is to generate initial solution, by generating worst solution at the beginning helps to avoid converging to local minimum, (2) a Perturbation Function to find a next solution with whom the current solution is compared, (3) an Objective Function is to be defined to evaluate and rate the current solution on the basis of performance, (4) an Acceptance Function, which is used to check whether the current solution is good or not in comparison with the current one, a very basic one is exp((currentSol-nextSol)/currentTemp), (5) and the last one is stopping criteria, there are many stopping criteria's, in this paper we have used an threshold value of objective function as an stopping criteria.

IV. Training Of Ann

ANN is made up of connection between neurons in each layer and links connecting these neurons has some weights on it, so the adjustment of weights to learn the relationship between the input and the given output i.e. label, is called learning or training of neural network. The most popular training algorithm is BackPropogation which was given by Salchenberger "et al." in 1992. The main problem of this algorithm is that it gets stuck in local minima and the error still remains the same. An evolutionary algorithm, Simulated Annealing and Genetic Algorithm, was given to solve this problem of local minima, among these algorithm simulated annealing is preferred because it takes less time in comparison with genetic algorithm.

As we know, to perform training of ANN we should have millions of data but to train neural network for credit card fraud detection we don't have much amount of data of previous transaction to perform training upon. In this paper, we have used Simulated Annealing for training and it gives very good result in comparison with genetic and backpropogation, which we will see later.

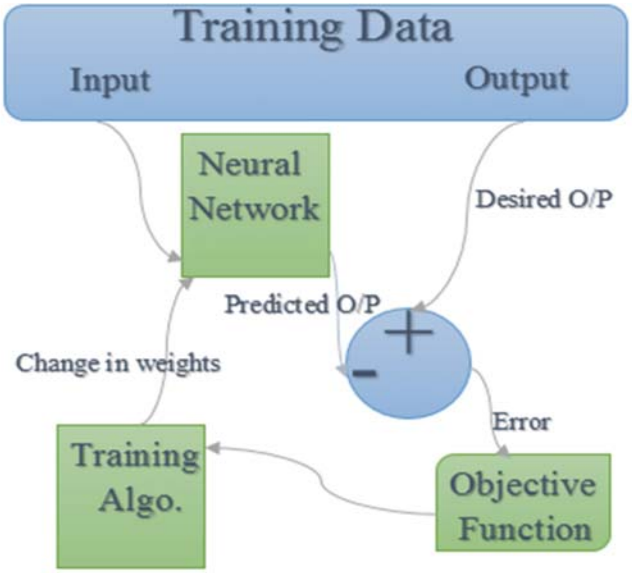

Fig 6: ANN training model

"Figure 6" depicts a basic model for the training process of ANN. In this paper we have used supervised learning [17], so our data consist of both input and desired output. A random weight is generated for each connection and output is calculated based on current weight & input. Obviously, in the initial states the desired output is different from current output which can be calculate by using any error function like Mean Squared Error (MSE) or Sum Squared Error (SSE). Now, according to the training algorithm the weights are adjusted and repeat these steps again until some threshold value for error function will reached.

In 1988, Jonathan Engel publish a paper [18] in which he had explained the training process for feed-forward NN using simulated annealing, we have used this paper to implement the simulated annealing algorithm for training purpose, while a brief description of algorithm is given below in this paper(for pseudo code refer [19]). There are the series of different steps which simulated annealing algorithm has to follow at each cycle. A cycle is completed when it follows all the steps shown in the "fig. 7" and randomized the weights at each cycle. The 'n' number of cycle is fixed by the programmer, at each iteration it will perform n cycles and after one iteration is completed, the current temperature gets lowered and checked against the minimum allowed lowest temperature, if it is not less than the threshold value then again a cycle of randomization is repeated. The method used in this paper for temperature reduction is based on start and stop temperatures. Its equation is given by:

NewTemp= Ratio * currentTemp

The ratio causes the new temperature lies in between start and stop temperature, ratio is calculated at each cycle and it decreases the temperature logarithmically. Its equation is given by:

$$Ratio = \frac{e^{log(stopTemp/startTemp))}}{cycles - 1}$$

The values of start and stop temperature is decided by hit and trial method, you have to check the result by putting different values and compare to find the best one. The high temperature will cause more randomization in weights.

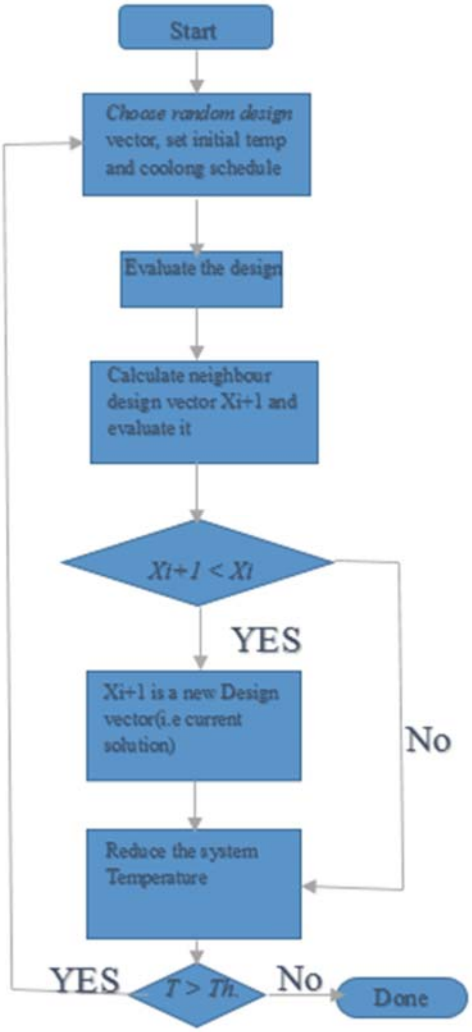
Start
Choose random design vector, set initial temp and coolong schedule
Evaluate the design
Calculate neighbour design vector Xi+1 and evaluate it
Xi+1 < Xi
YES
No
Xi+1 is a new Design vector(i.e current solution)
Reduce the system Temperature
YES
T > Th.
No
Done

Fig 7: Simulated Annealing method for training ANN

The main part in the training process of an ANN is the randomization of weights, simulated annealing uses previous input values and current temperature to randomize the weights, it depends on the type of problem we are going to solve with trained neural network. In this paper, we have used TANH as an activation function so we have to normalize the input, output and their corresponding weights in between (-1,1) and a weight matrix is created which acts as a linear array of double data type. The randomization of weights is not a complex task, in this paper we have generated a random number and multiply it with current temperature.

Q = currentTemp * Random (N)

Then this number is multiplied by each value in the weight matrix Wij*Q and updates all the values. This task is performed in each 'n' cycles and the updated matrix is compared with the previous one, if it is better than previous then the weight matrix gets updated.

V. RESULTS

In this section, we present the result which we get by using this algorithm on real world dataset. The dataset is taken from UCI repository consist of 1000 instance. This dataset contains useful information about transaction. This dataset contains 20 attributes and values corresponding to each rows is converted into symbolic or numerical form because of some privacy agreement. But the types of attributes is mentioned, shown in the table I.

TABLE I: ATTRIBUTES USED FOR TRAINING AND EVALUATING THE NEURAL NETWORK

S. No	Attribute Name
1	Status of existing checking Account
2	Duration in month
3	History of credit taken
4	Purpose of transaction
5	Credit amount
6	Saving account/bonds
7	Present employment since
8	Instalment rate in %
9	Personal status & sex
10	Guarantors
11	Present address since
12	Property
13	Age
14	Other instalment plans
15	Housing
16	Existing credits at this bank
17	Job
18	Number of people liable to provide maintenance for
19	Telephone
20	Foreign worker(Yes or No)

In this paper, we have divided the dataset into two parts: 75% data is taken from dataset for the training purpose and 25% data is taken for the evaluation of trained neural network. So, the evaluation dataset is new for the trained neural network, by performing the evaluation task we can classify the unseen data as a fraud or non-fraud. In this paper we present the result by showing the percentage of correct & incorrect classified data by comparing the predicted label with the existing label.

The configuration of Neural Network and the list of parameter used in Simulated Annealing to train neural network and their respective values is shown in table 2 & 3 respectively.

TABLE II. PARAMETERS OF AN ARTIFICIAL NEURAL NETWORK

Input layer Neurons	Hidden layer Neurons	Output layer neurons	Activation Function
20	50	2	Hyperbolic Tangent Activation Function

TABLE III. PARAMETERS OF SIMULATED ANNEALING ALGORITHM TO TRAIN NEURAL NETWORK

Start Temperature	Stop temperature	Number of Cycles/Iteration
100	3	100

After running this program it takes almost two days and stop training after reaching 1% error. After the process of training is, it goes for the evaluation process. This project is based on real-time i.e. at the time when performing any transaction at any online portal, so the time taken to classify any unseen data should be less. We have implemented this algorithm in java and perform an evaluation on the 25% of the dataset. It classifies all the data i.e. 250 instances, within 5-10 seconds, which is good in comparison with different configuration of neural network.

TABLE IV. RESULT OF TRAINED NEURAL NETWORK WHEN EVALUATION DATASET IS APPLIED

Observed	Total	Correct	Percentage Correct
Total Case	250	224	89.6%
Fraud Case	173	159	92%
Non-Fraud Case	77	65	85%

Table IV display the result of the trained neural network using simulated annealing. As we can see, in the first row the total case is only 250 instances from which 224 data is correctly classified i.e. based on the pre-defined label, the label predicted by the trained network is correct.

VI. CONCLUSION

In this paper we showed that better result is achieved with ANN when trained with simulated annealing algorithm. As the result shows that the training time is high but the fraud detection in real time is considerably low and the probability of predicting the fraud case correctly in online transaction is high, which is a main measure to evaluate any ANN. In the table 3 we can see that 65% of total fraud case is correctly classified which is a very high percentage in comparison with genetic, resilient backpropogation and any other training algorithm.

The main problem in credit card fraud detection is the availability of real world data for the experiment. This approach can also be used in other applications which require classification task [20] e.g. software failure prediction, etc.

A. Future Work in this project

There will be a lot of work to be done for fraud detection because the activity of user is different in each transaction which causes the training of any ANN to be difficult. In this project the main task is to find the best configuration for neural network, we can use Genetic Algorithm for this task, it would find a better configuration by applying different combinations. So if we combine Simulated Annealing and Genetic Algorithm to create a best model, it will gives better result than any other.

REFERENCES

[1] Linda Delamaire (UK), Hussein Abdou (UK), John Pointon (UK),"Credit card fraud and detection techniques: a review", Banks and Bank Systems, Volume 4, Issue 2, 2009.

[2] Nadeem Akhtar, Farid ul Haq, " Real Time Online Banking Fraud Detection Using Loaction Information",International Conference on Computational Intelligence and Information Technology – CIIT 2011, Pune, India.

[3] K. Cios, W. Pedrycs, and R. Swiniarski, Data Mining Methods for Knowledge Discovery. Boston: Kluwer Academic Publishers, 1998.

[4] Y. Sahin and E. Duman,"Detecting Credit Card Fraud by Decision Trees and Support Vector Machines", International conference of Engineers & computer Scientists 2011 Vol I, March 16 2011, Hong Kong.

[5] Bolton, R. J. and Hand, D. J.,"Statistical fraud Detection: A review". Statistical Science 28(3):235-255, 2002.

[6] Karl BlomStorm," Benchmarking an artificial neural network tuned by a genetic algorithm", VT 2012.

[7] UCI Machine Learning Repository,"http://archive.ics.uci.edu/ml/datasets.html", last accessed at 22/11/2013.

[8] Wai-cgiu Wong, Ada Wai-chee Fu,"Incremental Document Clustering for Web Page Classification", Department of Computer Science and Engineering, The Chinese University of Hong Kong, Shatin, Springer Japan 2002.

[9] W. Wong and A. Fu, "Incremental Document Clustering for Web Page Classification," Proc. 2000 Int'l Conf. Information Soc. in the 21st Century: Emerging Technologies and New Challenges (IS2000), 2000.

[10] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain", Psychological review, 65(6):386, 1958.

[11] Han, Jun; Morag, Claudio," The influence of the sigmoid function parameters on the speed of Backpropagation learning", In Mira, José, Sandoval, Francisco, From Natural to Artificial Neural Computation. pp. 195–201, 1995.

[12] Metropolis, Nicholas, Rosenbluth, Arianna W., Rosenbluth, Marshall N., Teller, Augusta H., Teller, Edward, "Equation of State Calculations by Fast Computing Machines", The Journal of Chemical Physics 21 (6): 1087, 1953.

[13] Kirkpatrick, S., Gelatt Jr, C. D., Vecchi, M. P., "Optimization by Simulated Annealing". Science 220 (4598): 671–680, 1983.

[14] Cerny, V., "Thermo dynamical approach to the traveling salesman problem: An efficient simulation algorithm", Journal of Optimization Theory and Applications 45: 41–51, 1985.

[15] P J van Laarhoven and E H Aarts,"Simulated Annealing: Theory and Applications", Kluwer Academic Publishers, 1987.

[16] R H Otten and L P Ginneken,"The Annealing Algoritm", Kluwer Academic Publishers, 1989.

[17] Y. Yang, J. Carbonell, R. Brown, T. Pierce, B. Archibald, and X. Liu, "Learning Approaches for Detecting and Tracking News Events," IEEE Intelligent Systems, vol. 14, no. 4, pp. 32-43, 1999.

[18] Jonathan Engel," Teaching Feed-Forward Neural Networks by Simulated Annealing", Norman Bridge Laboratory of Pllysics 161-33, California Institute of Technology, Pasadena, CA 91125, USA Complex, Systems 2, 1988.

[19] Mohamed Benaddy and Mohamed Wakrim,"Simulated Annealing Neural Network for Software Failure Prediction", International Journal of Software Engineering and Its Applications Vol.6, No. 4, October, 2012.

[20] Li, Y.H., Jain, A.K.: Classification of Text Documents. The Computer Journal. vol. 41, pp. 537--546 (1998).