Int. J. Nonlinear Anal. Appl. (11) No. 2, 437-445
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2020.4576

A Novel Method for Detection of Fraudulent Bank Transactions using Multi-Layer Neural Networks with Adaptive Learning Rate

Maryam Faridpour, Alireza Moradi*

Department of Electrical and Computer Engineering, Mahdishahr Branch, Islamic Azad University, Mahdishahr, Iran.

(Communicated by Ehsan Kozegar)

Abstract

Fraud refers to earn wealth including property, goods and services through immoral and non-legal channels. Fraud has always been in action and experiences an increasing trend worldwide. Fraud in financial transactions not only leads to losing huge financial resources, but also leads to reduction in trust of customers on using modern banking systems and hence, reduction in efficiency of the systems and optimal management of financial transactions. In recent years, by emerging new technologies of banking industry, new means of fraud are discovered. Although a new information system carry advantages and benefits, new opportunities are made for fraudsters. The applications of fraud detection methods encompasses detection of frauds in an organization, analysis of frauds and also user/customer behavior analytics in order to predict future behavior and reduce the fraud risks. In recent decades, employing new technologies in management of banking transactions has risen. Banks and financial institutions inevitably migrated from traditional banking to modern online banking to provide effective services. Although, the use of online banking systems improves the management of financial processes and speeds up services to customers of institutions, but some issues would also be carried. Financial frauds is one of the issues which organizations seek to prevent and reduce effects. In this paper, a novel machine learning based model is presented to detect fraud in electronic banking transactions using profile data of bank customers. In the proposed method, transactional data from banks are leveraged and a multi-layer perceptron neural network with adaptive learning rate is trained to prove the validity of a transaction and hence, improve the fraud detection in electronic banking. The proposed method shows promising results compared with logistic regression and support vector machines.

*Corresponding Author: Alireza Moradi
Email address: maryam.faridpour68@gmail.com, alireza.moradi@msh-iau.ac.ir (Maryam Faridpour, Alireza Moradi*)

Received: August 2019 *Revised:* August 2020

Keywords: Electronic Banking, Fraud Detection, MLP Neural Network, Adaptive Learning Rate.
2010 MSC: 26D15, 26D10.

1. Introduction

Banking Fraud has been an ever-growing issue with huge consequences to banks and customers alike, both in terms of financial losses, trust and credibility. As per the Nilson report, it is anticipated that card frauds alone would amount to a whopping $30 billion worldwide by 2020. Also, with the technology disruption in both banking and payments (due to a plethora of payment channels — credit/debit cards, smartphones, kiosks), the number of transactions has increased exponentially in recent years. Fraudsters have also become extremely smart, adopting innovatory fraudulent tactics. As a result it has compounded the problem.
Primarily, most banks employ Rule-based Systems with manual evaluation for detecting fraud. Although these systems were doing a pretty decent job, in the recent years, they have become more inconsistent. That's because new fraud patterns are evolving rapidly and these systems are unable to evolve accordingly, allowing frauds to go undetected, and resulting in huge financial losses. There are banks which also have systems built on RDBMS, but their performance is even worse when compared to Rule-based Systems.
Considering all these challenges and shortcomings, Machine Learning can play a vital role in effective and efficient fraud detection in the banking industry. In this paper, a multi-layer perceptron neural network with adaptive learning rate is proposed to detect fraudulent and non-fraudulent bank transactions.
The rest of the paper is organized as follows. In section 2, related works are described and discussed from two perspectives of supervised and unsupervised techniques. In section 3, a multi-layer perceptron neural network with adaptive learning rate is mathematically formulated. In section 4, results and discussions are described to compare the proposed method against logistic regression and support vector machines. Conclusion of the paper is made in section 5.

2. Related Works

The methods to detect fraudulent and non-fraudulent records are surveyed as two classifications of supervised and unsupervised techniques.

2.1. Supervised techniques

In supervised learning, samples of both fraudulent and non-fraudulent records, associated with their labels are used to create models. These techniques are often used in fraud analysis approach. One of the most popular supervised neural networks is back propagation network (BPN). It minimizes the objective function using a multi-stage dynamic optimization method that is a generalization of the delta rule. The back propagation method is often useful for feed-forward network with no feedback. The BPN algorithm is usually time-consuming and parameters like the number of hidden neurons and learning rate of delta rules require extensive tuning and training to achieve the best performance [1]. In the domain of fraud detection, supervised neural networks like back-propagation are known as efficient tool that have numerous applications [2], [3], [4]. Raghavendra Patidar et al. [5] used a dataset to train a three layers backpropagation neural network in combination with genetic algorithms (GA) [6] for credit card fraud detection. In this work, genetic algorithms was responsible for making decision about the network architecture, dealing with the network topology, number of hidden layers and number of nodes in each layer.

Also, Aleskerov et al. [7] developed a neural network based data mining system for credit card fraud detection. The proposed system (CARDWATCH) had three layers auto associative architectures. They used a set of synthetized data for training and testing the system. The reported results show very successful fraud detection rates.
In [8], a P-RCE neural network was applied for credit card fraud detection. P-RCE is a type of radial-basis function networks [9, 10] that usually applied for pattern recognition tasks. Krenker et al. proposed a model for real time fraud detection based on bidirectional neural networks [11]. They used a large data set of cell phone transactions provided by a credit card company. It was claimed that the system outperforms the rule based algorithms in terms of false positive rate.
Again in [12] a parallel granular neural network (GNN) is proposed to speed up data mining and knowledge discovery process for credit card fraud detection. GNN is a kind of fuzzy neural network based on knowledge discovery (FNNKD). The underlying dataset was extracted from SQL server database containing sample Visa Card transactions and then preprocessed for applying in fraud detection. They obtained less average training errors in the presence of larger training dataset.

2.2. Unsupervised techniques

The unsupervised techniques do not need the previous knowledge of fraudulent and normal records. These methods raise alarm for those transactions that are most dissimilar from the normal ones. These techniques are often used in user behavior approach. ANNs can produce acceptable result for enough large transaction dataset. They need a long training dataset. Self-organizing map (SOM) is one of the most popular unsupervised neural networks learning which was introduced by [13]. SOM provides a clustering method, which is appropriate for constructing and analyzing customer profiles, in credit card fraud detection, as suggested in [14]. SOM operates in two phase: training and mapping. In the former phase, the map is built and weights of the neurons are updated iteratively, based on input samples [15], in latter, test data is classified automatically into normal and fraudulent classes through the procedure of mapping. As stated in [16], after training the SOM, new unseen transactions are compared to normal and fraud clusters, if it is similar to all normal records, it is classified as normal. New fraud transactions are also detected similarly.
One of the advantages of using unsupervised neural networks over similar techniques is that these methods can learn from data stream. The more data passed to a SOM model, the more adaptation and improvement on result is obtained. More specifically, the SOM adapts its model as time passes. Therefore it can be used and updated online in banks or other financial corporations. As a result, the fraudulent use of a card can be detected fast and effectively. However, neural networks has some drawbacks and difficulties which are mainly related to specifying suitable architecture in one hand and excessive training required for reaching to best performance in other hand.

3. Proposed method

The most commonly used algorithm to train neural networks is gradient descent. The gradient is a numeric calculation allowing us to adjust the parameters of a network in order to minimize its output deviation. However, the learning time is a challenge. Standard version of gradient descent learns slowly. There, an improvement is required for the gradient descent in real-world applications. Gradient descent algorithms including Batch Gradient Descent (BGD), Stochastic Gradient Descent (SGD) and mini-Batch Gradient Descent (mini-BGD, the mixture of BGD ans SGD) are the base state-of-the-art gradient descent algorithms. In essence, the methods seek to update the weights θ of the network, with the help of a learning rate $\eta\eta$, the objective function $\mathrm{J}(\theta)$ and the gradient of it, $\nabla \mathrm{J}(\theta)$. What all gradient descent algorithms and its improvements have in common,

is the goal of minimizing J(θ) in order to find the optimal weights $\theta\theta$. The simplest of the three is the BGD.

$$\theta = \theta - \eta \cdot \nabla_\theta J(\theta) \tag{3.1}$$

It tries to reach the minimum of J(θ), by subtracting from θ the gradient of J(θ) (refere to Figure 3 for a visualization). The algorithm always computes over the whole set of data, for each update. This makes the BGD the slowest and causes it to be unable to update online. Additionally, it performs redundant operates for big sets of data, computing similar examples at each update and it converges to the closeset minimum depending on the given data, resulting in potential suboptimal results. An often used algorithm is the SGD.

$$\theta = \theta - \eta \cdot \nabla_\theta J\left(\theta; x^{(i)}; y^{(i)}\right) \tag{3.2}$$

Contrary to BGD, SGD updates for each training example ($x^{(i)}$;$y^{(i)}$), thus updating according to a single example step. Furthermore, this fluctuation enables the SGD to jump to minima farther away, potentially reaching a better minimum. But thanks to this fluctuation, SGD is also able to overshoot. This can be counteracted by slowly decreasing the learning rate. In the exemplary code shown in Figure 2, a shuffle function is additionally used in the SGD and mini-BGD algorithm, compared to the BGD. This is done, as it is often preferable to avoid meaningful order of the data and thereby avoid bias of optimization algorithm, although sometimes better results can be achieved with data in order. In this case the shuffle operation is to be removed. Lastly, there is the mini-BGD.

$$\theta = \theta - \eta \cdot \nabla_\theta J\left(\theta; x^{(i:+n)}; y^{(i;+n)}\right) \tag{3.3}$$

The mini-BGD updates for every mini-batch of n training examples. This leads to a more stable convergence, by reducing the variance of the parameters. When people talk abput a SGD algorithm, they often refer to this version.

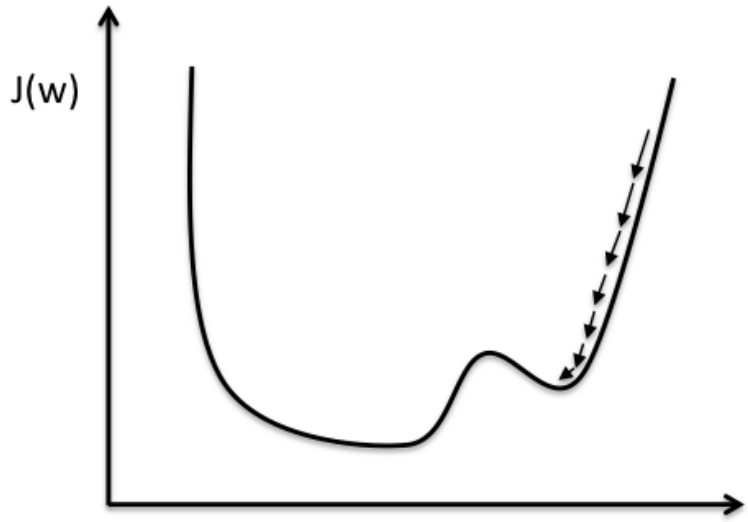

Figure 1: Local minima may occur in J(θ) (here J(w)), which may result in suboptimal solution for some gradient descent methods.

3.1. Adaptive Learning Rate Method

As an improvement to traditional gradient descent algorithms, the adaptive gradient descent optimization algorithms or adaptive learning rate methods can be utilized. Several versions of these algorithms are described below.

Momentum can be seen as an evolution of the SGD.

$$\begin{aligned} \mathrm{vt} &= \mathrm{yvt} - 1 + \eta \nabla_\theta \mathrm{J}(\theta) \\ \theta &= \theta - \mathrm{vt} \end{aligned} \tag{3.4}$$

While SGD has problems with data having steep curves in one direction of the gradient, Momentum circumvents that by adding the update vector of the time step before multiplying it with a γ, usually around 0.9. As an analogy, one can think of a ball rolling down the gradient, gathering momentum (hence the name), while still being affected by the wind resistance ($0< \gamma < 1$).

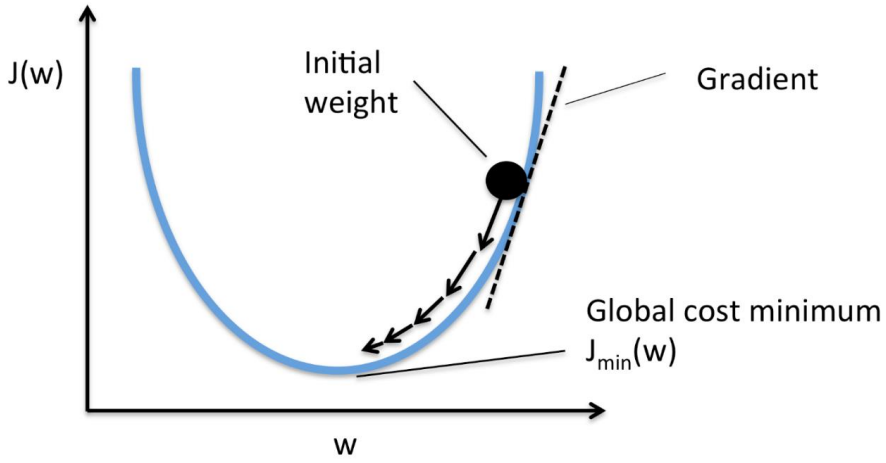

Figure 2: Visualization of the analogy for Momentum using a $\gamma = 0.9$.

4. Results

In this paper, a perceptron multi-layer neural network with two hidden layers is designed such that the input layer contains 30 nodes taking bias aside which is taken from the 30 independent variables. The middle layer of the network is designed to contain 10 nodes and a bias. The output layer of the network contains two nodes taken from the number of dependent variables. All the nodes of each layer is fully connected to the nodes of previous layer. The connections carry the weights of the network elements. Fig. 3 shows the discriminability of the dataset features and Fig. 4 shows the classification of the features into appropriate and inappropriate features.

The network is evaluated with different number of network layers and different number of hidden layers; the reported error rate is drawn in Fig. 5 and Fig. 6, respectively.

Since the identification of fraud transaction is based on classification. Classification measures are used to evaluate the performance of the proposed method against other methods. The proposed method which is a multi-layer perceptron neural network is evaluated against logistic regression and support vector machine.

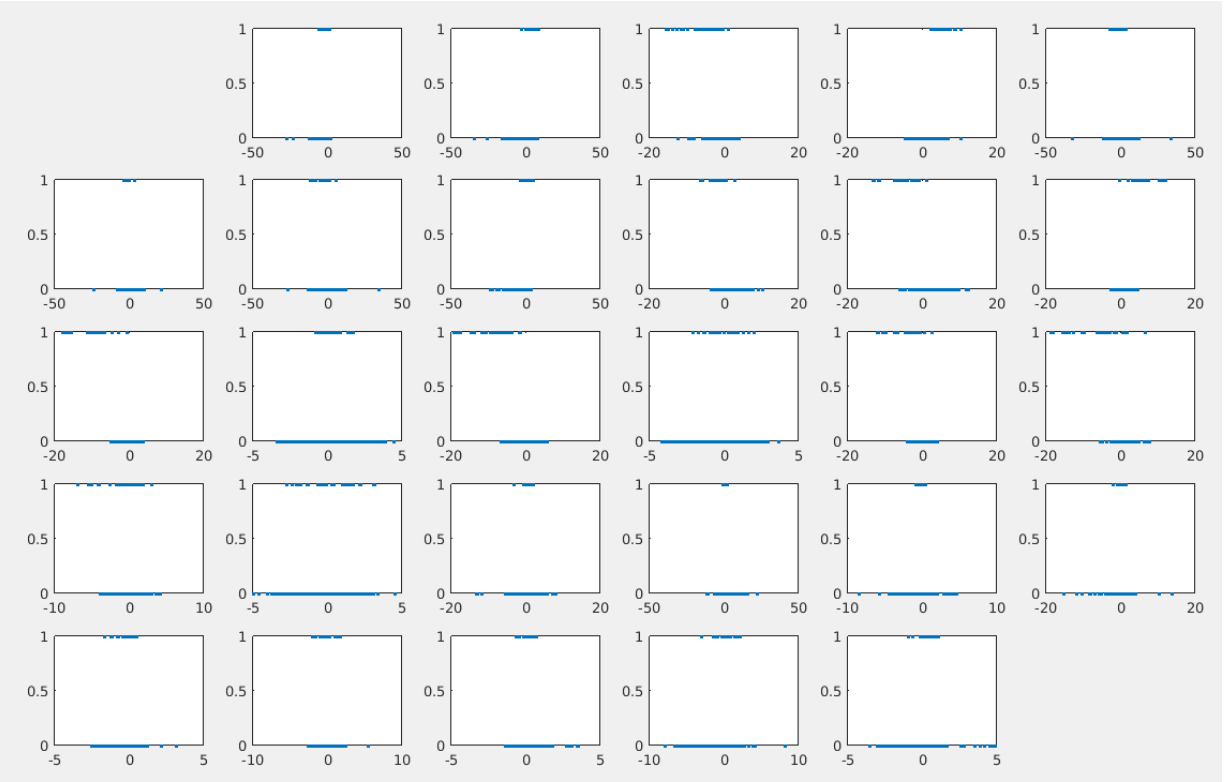

Figure 3: Discriminability of dataset feature

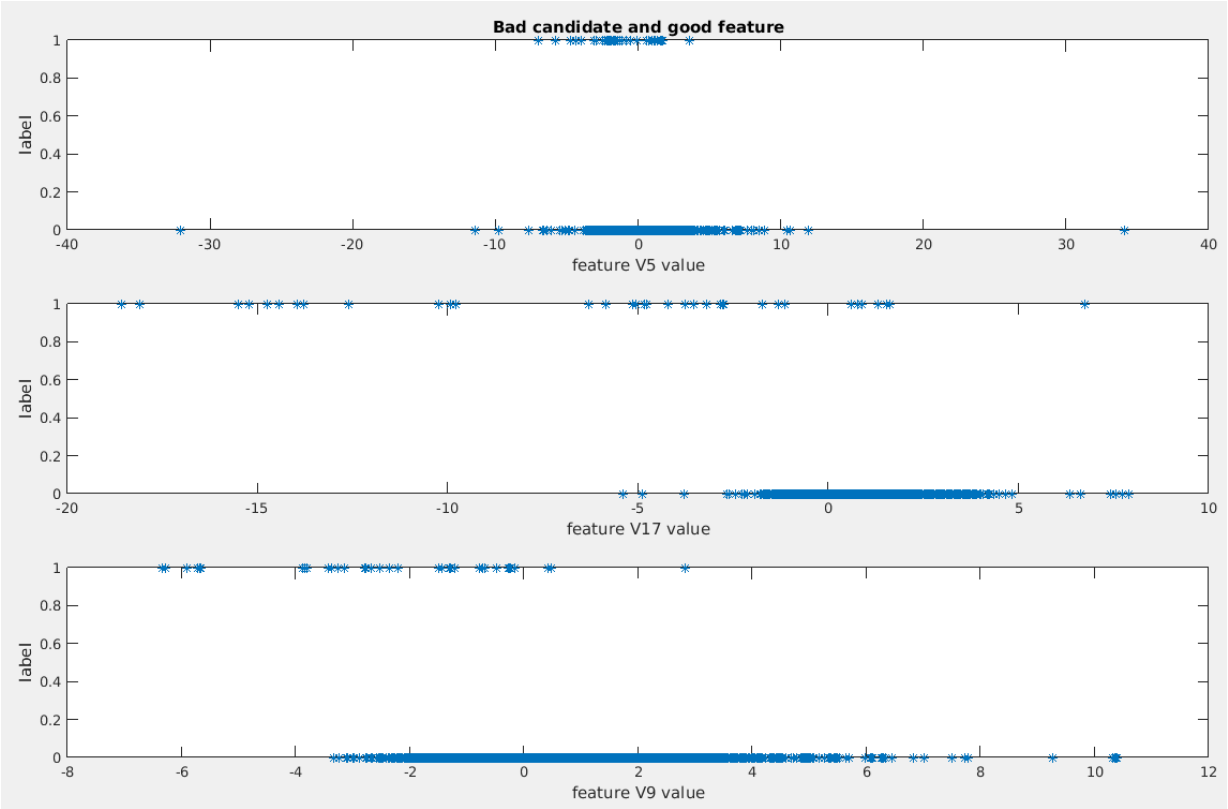

Figure 4: Separation of features into appropriate and inappropriate candidate feature

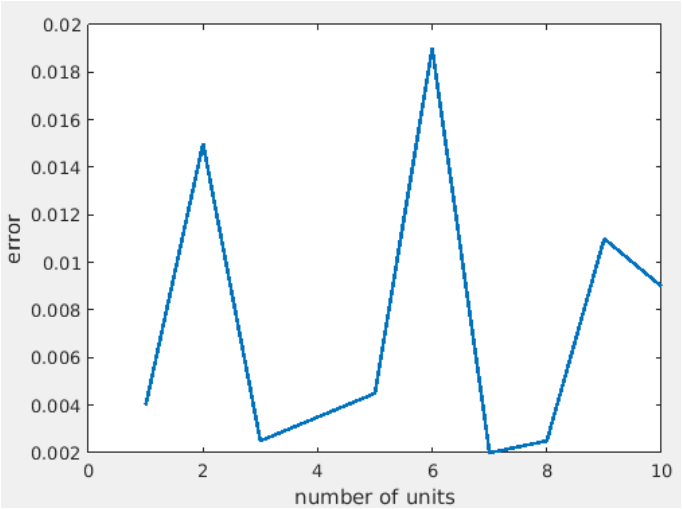

Figure 5: The comparison of different number network layers

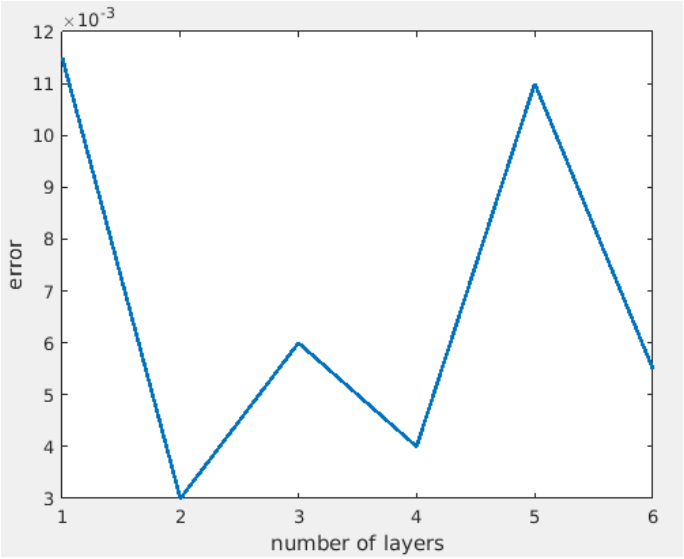

Figure 6: The comparison of different number of hidden layers of the network

Table 1: Results of the proposed method against logistic regression and support vector machine.

Learning Model	Accuracy on training data	Accuracy on test data
MLP Neural Network with Adaptive Learning Rate	0.9999	0.9990
Logistic Regression	0.9236	0.9723
Support Vector Machine	0.9154	0.9345

Table 2: Confusion matrix.

Actual\Predicted	Predicted non-fraud transaction	Predicted fraud transaction
Actual non-fraud transaction	85261	33
Actual fraud transaction	28	120

5. Conclusion

In this paper, a novel machine learning method is proposed which is able to identify suspicious transactions as frauds and report the results to the bank switch. Multi-layer perceptron neural network with adaptive learning rate is employed as a learning model to identify frauds in bank card transactions. A significant requirement for a method of fraud detection is the ability of real-time detection. Since banks always handle huge amounts of transactional data, the fraud detection model is expected to produce real-time results. A post-detection action after the fraud detection can be blockage of the suspicious account. Therefore, the costs of false detection of non-fraud transaction increases. On the other hand, risk of irreversibility of a suspicious transaction is tied to the risk of false detection of non-fraud transaction. Therefore, the risk factor level can be determined according to the bank strategy requirements. In future works, a dependent variable with numerous classifications can be used. For example, classes such as normal transaction, low-risk transaction, high-risk transaction, fraud transaction are defined based on the dependent variable. The classification scheme block the accounts associated to fraud transactions and more investigations will be conducted for low-risk transactions.

References

[1] Masoumeh Zareapoor, Seeja K.R, M. Afshar.Alam, "Analysis of Credit Card Fraud Detection Techniques: based on Certain Design Criteria", International Journal of Computer Applications (0975 – 8887) Volume 52– No.3, 2012.
[2] Fraud Brief – AVS and CVM, Clear Commerce Corporation, 2003, http://www.clearcommerce.com.
[3] All points protection: One sure strategy to control fraud, Fair Isaac, http://www.fairisaac.com, 2007.
[4] Clear Commerce fraud prevention guide, Clear Commerce Corporation, 2002, http://www.clearcommerce.c
[5] Raghavendra Patidar, Lokesh Sharma, "Credit Card Fraud Detection Using Neural Network", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-1, 2011.
[6] Holland, J. H. "Adaptation in natural and artificial systems." Ann Arbor: The University of Michigan Press. (1975).
[7] E. Aleskerov, B. Freisleben, B. Rao, CARDWATCH: A Neural Network-Based Database Mining System for Credit Card Fraud Detection", the International Conference on Computational Intelligence for Financial Engineering, pp. 220-226, 1997.
[8] Sushmito Ghosh, Douglas L. Reilly, Nestor, "Credit Card Fraud Detection with a Neural-Network", Proceedings of 27^{th} Annual Hawaii International Conference on System Sciences, 1994.
[9] Moody and C. Darken, "Learning with localized receptive fields." in Proc. of the 1988 Connectionist Models Summer School, D.S. Touretzky, G.E. Hinton and T.J. Sejnowski, eds., Morgan Kaufmann Publishers, San Mateo, CA, 1989, pp. 133-143.
[10] S.J. Nowlan, "Max likelihood competition in RBP networks," Technical Report CRG-TR-90- 2, Dept. of Computer Science, University of Toronto, Canada, 1990.
[11] Firouzian, Iman, Morteza Zahedi, and Hamid Hassanpour. "Cycle Time Optimization of Processes Using an Entropy-Based Learning for Task Allocation." International Journal of Engineering 32, no. 8 (2019): 1090-1100.
[12] Mubeena Syeda, Yan-Qing Zbang and Yi Pan," Parallel granular neural networks for fast credit card fraud detection", international conference on e-commerce application, 2002.
[13] Kohonen, T. "The self-organizing maps". In Proceedings of the IEEE (1990) 78 (9), pp 1464–1480.

[14] Vladimir Zaslavsky and Anna Strizhak "Credit card fraud detection using self organizing maps". Information & Security. An International Journal, (2006). Vol.18; (48-63).
[15] Vesanto, J., &Alhoniemi, E. (2000). "Clustering of the self-organizing map".IEEE Transactions on Neural Networks, (2009). 11; (586–600).
[16] Serrano-Cinca, C "Self-organizing neural networks for financial diagnosis". Decision Support Systems, (1996). 17; (227–238).