

A New Novel Cluster based Analysis of Bank's Customer Data with Self-Organized Map Networks

Seyyed Reza Khaze
Department of Computer
Engineering, Dehdasht
Branch, Islamic Azad
University, Dehdasht,
Iran

Emita Davoudi
Takiyeh
Department of Computer
Engineering, Qazvin
Branch, Islamic Azad
University, Qazvin, Iran

Isa Maleki
Department of Computer
Engineering, Dehdasht
Branch, Islamic Azad
University, Dehdasht,
Iran

Farhad Soleimanian
Gharehchopogh
Department of Computer
Engineering, Hacettepe
University, Ankara,
Turkey

ABSTRACT

In the current global competing environment, creation of knowledge base and the use of it have been advantageous for the banks and the financial institutions and accounting and are being transformed to a strategic tool for competing among them and so data mining has been understood more and more in this field lately. In the today competing globe, banks and the financial institutions are trying to reach the advantage and be better than the others. Also, except execution of the business processes, the creation of the data knowledge and the use of it advantageous for the bank is being transformed to a strategic tool for competing. Taking into consideration this necessity, we have applied the Self-Organized Map (SOM) network in some cases of citizens in the banks of West Azerbaijan Province located at Republic Islamic of Iran. It is essential to cluster based solidarity analysis among of the specifications of customers to find common behavior points of them. However, it could be used to maintain the customers and find the new ones by the high responsible of programming of the banks. This approach leads to higher benefits and efficiencies in extracting and mining the likes and the wants of the customers. The results of the clustering analysis showed that the perspective of the customer about bank services and the effect of the electronic banking in banks selection, hold very similar junction patterns.

Keywords

Classification, Artificial Neural Network, Self-Organized Map, Learning, Algorithm.

1. INTRODUCTION

The grouping algorithms and the data mining prediction algorithms have proved that can have better correction rate in comparison to the other statistical methods for predicting, grouping and clustering operations [1]. And taking into consideration that grouping and prediction are of the important procedures in the financial and accounting problems, data mining tools have proved themselves in this field very well. Of the fields the data mining tools are used, are the financial status prediction, cheating diagnosis management and credit risk estimation [2]. In the today competing globe, banks and the financial institutions are trying to achieve the advantage and be better than the others. Also, except execution of the business processes, the creation of the data knowledge and the use of it advantageous for the bank is being transformed to a strategic tool for competing. It must be taken into consideration that the ability of production and absorbing the data have increased considerably lately and must be taken into consideration that the information of these data are very important, so the accessibility of the data and the

need for transforming them to the knowledge encourages the IT industry for using data mining.

As the banking industry has faced many changes, now it has understood the need for data mining which helps it in competing market. The banks and the financial institutions have stepped on the way of data mining knowledge to find new customers and diagnose the cheating processes and service according to the shopping patterns of the customers [3]. The banks have been informed of the advantages of data mining in creation of the data mining for decision making and surely will benefit this and this will be advantageous in future competing. Lately the financial problems have been the challenge for the companies and predicting the break of the factories and are also a stressful point for the ones in relation with exchange market.

As removing the regulations in the financial servicing industries and accepting the new wide technologies lead to the increase in competing in financial markets, and taking into consideration that the responsibility of the business strategies center of any company is to maintain the customers and also absorb the new ones, this knowledge could play important role in this struggle taking into consideration the abilities of data mining algorithms.

In this paper, we have studied the solidarity among the specifications of the customers to find the common behavior points of them by the SOM networks, and the analysis can be used for maintaining the customers and finding the new ones by the high responsible people of the programming center of the banks. So, we have studied the abilities of the SOM networks in clustering. In the first section, the ex-activities are studied and then the clustering of the bank customers using the SOM networks is analyzed.

2. PREVIOUS WORK

Data mining has played vast role in financial services, banks and business institutions lately which we study some here. Many of the data mining algorithms produce the lists of the regulations in which finding the beneficiary regulations depends on the user. Introducing the law selection mechanism, it is possible to use them for bank loans. Also production of law is possible by utilizing the decision tree C4.5 [4]. The results have showed that these methods of data mining are very effective in finding the favored laws.

Also in Thailand, electronic banking is presented by different financial institutions which include business and governmental banks. On this perspective, electronic banking is very limited because of the limited use of the private data. The use of data mining techniques for analyzing the use of the

historical data of the electronic banking of the business banks in Thailand which includes SOM algorithms and RFM marketing techniques, has led to advantageous results. Utilizing the results of diagnosing the relation between electronic banking services, it is possible to reach the production of the new service packets for any part of the users of the electronic banking. A Priori algorithm which is a data mining algorithm could be used in this diagnosis. In addition to it, using the diagnosis algorithms of Outlier which is based on the k th vice versa close neighbor, it is possible to decrease the credit card cheating processes. Other researchers [5], have presented another diagnosis algorithm which not only decreases the mining repetitions in one unit, but also has showed on the real and unreal data sets that is more effective and efficient in comparison to the other methods.

The data mining solution is also usable for the direct bank marketing risks. Also the results show that the best comprehensive model is presented by support vector machine which is more efficient in predicting. In addition to other data mining algorithms, the Artificial Neural Networks (ANN)s are also one of the best methods of data mining which are used for designing the architecture of the ANNs for the cheating diagnosis system related to Credit Card, which could be used with no supervisor according to the learning methods. The research findings have shown that the ANNs have diagnosed more than 95% cases of cheating [6]. Also the competing advantage of the maintenance and avoiding the end of the old companies in comparison to the new ones is the accurate diagnosis of the data mining methods. Data mining methods for better model diagnosis are important for predicting.

Other researchers [7] have utilized the clementine software and a data base of 300 records of the customers of Iran Insurance Company in Anzali city (in Iran) and proved the effectiveness of the data mining tools in diagnosis of the reliability of the costumers. At the late years, the common economic models like decomposition, analysis and logistic regression have been used for predicting the customer selection. Although there have been interest in ANNs for analyzing the behavior of the customer and modeling the customer decision making process, the ANNs have been considered as a branch of artificial intelligence. ANNs has been used in the fields related to the business lately and most of the studied have selected feed forward multi-layer ANNs for analyzing the customer selection problems. Researches in [8] have compared the experimental power of predicting the Artificial Probable ANN (APNN) which is a specific state of the ANNs to the MLFN of logistic model on the customer selection in electronic and non-electronic banking. The data of this research for analysis have been achieved via sending email to 1960 families. The experimental results show that both ANN models (APNN and MLFN) have showed the customer selection results accurately but the logistic models did not. On the other hand the APNN holds the best accuracy total percent and the least error rate as a prediction model. By the transformation of the banking operations to online processes, the need for predicting the banking cheats and security failing has been highlighted. Such violations exist all over the economic system. The information security system predicts these challenges as a problem and tries to find the insecure transactions in the bank operations section. Some researchers [9] have utilized the ANNs techniques for predicting the cheating problem in banking sector in a study. These researchers have utilized the Nigeria Banks as a reference and have designed a model based on the ANNs in

which the feed forward multi-layer ANNs are used on the database system to collect the educational data.

3. APPLICATION OF SOM IN CLUSTERING

The goal of extracting the information from the data, from different methods of relational learning to statistics and ANNs, is to find knowledge of the complex data of high dimensions and volume which are not analyzable without computer [10]. Clustering analysis is one of the methods of extracting the information from data, and also is a method of learning with no supervision which is used for grouping the data according to the similarity or closeness insertion and could classify the data into the separate groups. Clustering does not lean on the pre-identified classes or the goal specifications in contrast to grouping [12] and is able to classify the large and complex data into the groups of similar specifications [13]. The SOM networks are of the ANNs with no supervision learning which hold powerful abilities in analyzing the complex spaces [14]. This model of the ANNs was first introduced by Kohonen in 1981 and inspiring the eye retinal neurons and was first used experimentally in 1984 for voice diagnosis and transforming it to text and it is able to present the large data in two dimensional manner[15].

In this presentation the SOM networks maintain the large data topology in two dimensions [16]. Totally if the data of the entrance space are close to each other, they will remain close in the SOM network plane [17]. So, the SOM plans can include the entrance space of large dimensions in a two dimensional plan maintaining the entrance data topologic structure. Till now the SOM networks are used vastly in data mining, presenting the complex spaces and clustering the spaces of high dimensions [18]. As this network must be educated like other ANNs for extraction of data, it is possible to use the no supervision methods according to the competing learning algorithm to educate it. First the weight vector of any neuron is produced randomly and the primary structure of the network is formed and then by the network education process the weight vector is set in a way that could cover a part of the information of the under analysis space.

4. PROPOSED METHOD, RESULT AND DISCUSSION

SOM network capabilities in clustering the data [19] could contribute considerably to clustering analysis of the banks customers. In this paper we have extracted the information using the clustering analysis using SOM networks in some cases among the customers of the banks of West Azerbaijan province in Islamic Republic of Iran. The collected information includes the 50 people.

Any record of data base includes the account type, the view of the account holder to the bank services of their account, the advertisement level of the bank, the behavior of the clerks and the managers of the different branches of the bank, support of the latest modules of electronic and internet banking, beneficiary, the loan wage, and also the decision to continue working with the bank as the selected bank for any banking process for any customer.

The clustering analysis of these information not only classifies the customers in different programming groups based on the specifications and views of the customers of each cluster and with respect to their differences in comparison to other clusters, but also considers the similarities of each cluster and creates a theoretical link among the members of any cluster of the customers which leads to better benefits and efficiency in

extracting and mining the wants and interests of the customers. First we created a SOM network and in the first step we selected the parameters like dimension and primary weight vector of any SOM neuron for clustering the bank customers. The used network of us for clustering the bank customers is single layered in which the neurons are organized in a mesh manner in it. Creating this network, the number of the rows and the columns of the mesh must be defined. We have considered the mesh rows and columns as 10. The total scheme of the network we have designed could be seen in Figure 1.

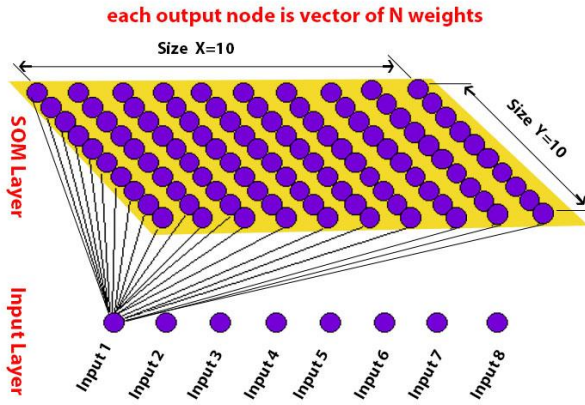


Fig 1: The Total Scheme of The Som Network Architecture For Clustering The Bank Customers

In the next stage the information related to the bank customers must be presented and it is important to begin finding the best neuron for any customer. The customer information could enter the system all at once or could be presented one by one for any customer and the education process would start then. It must be considered that the SOM network uses the SOM algorithm for training as default. In the educational procedure, the weight vectors of any neuron moves toward the center of a cluster of the bank customers. The neurons next to each other in SOM topology affect each other and get close to each other. If any customer like X, is considered as it includes n quantity of the specification, then it is possible to present that customer as follows [18].

$$X = [X_1, X_2, \dots, X_n] \in \mathbb{R}^n$$

If the tth neuron weight vector is introduced as follows:

$$m_1 = [m_{t1}, m_{t2}, \dots, m_{tn}] \in \mathbb{R}^n$$

Then the related neuron to any of the bank customers or named the BUM or winner neuron is defined by the following equation.

$$c = \arg \min_t \{d(X, m_1)\}$$

In which the C is the winner neuron and d (X, m1) is the Euclid distance between any of the bank customers and the tth weight vector. In the next stage the weigh vector related to each neuron must be updated. The operation takes place by the following relation.

$$m_i(t+1) = m_i(t) + \alpha(t)h_{ci}(t)[X(t) - m_i(t)]$$

In which $0 < \alpha < 1$ is the learning rate, and $h_{ci}(t)$ is the neighboring rate of the ith and cth neuron (winner neuron).

The neighboring rate of the winner neuron to the ith neuron is calculated via the following relation.

$$h_{ct} = e^{-\frac{\|r_c - r_i\|^2}{2\sigma^2(t)}}$$

In which σ is the controlling factor of the function amplitude and is reduced in the learning process. Also, r_i and r_c are the position of the ith and cth (winner) neurons in SOM plan. At last we study the ending condition of the algorithm. If it is not true, the algorithm continues execution from the second step. So it must be considered that as the algorithm of training the SOM networks is based on Euclid distance, the data of any dimension of the under study space must be normalized independently. When the training phase of the SOM networks end, a plan of the neurons is achieved which is in fact an abstract of the under analysis space of the network. The default topology is hexagonal. The figure 2 shows the SOM topology of the bank customers clustering.

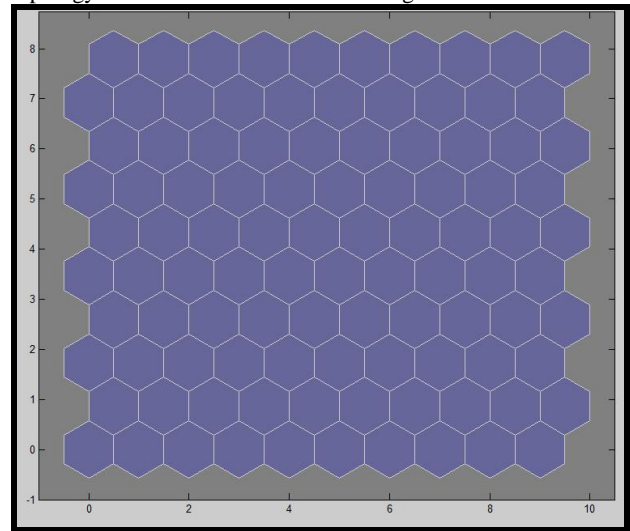


Fig 2: SOM topology for clustering the bank customers

One of the outputs of SOM is the clustering matrix and the related clustering plan. Elements of this matrix show the distance of the neighbor neurons from each other. It is clear that if the specifications of two parts of the under analysis space are similar to each other, then the distance of the related weight vectors would not be so much, and in other words both neurons belong to the cluster of the under study space. In contrast the more the distance of the neighbor neurons, the more the differences related to their space, and so they could be placed in two separate clusters. Each hexagonal shape presents a neuron in this topology. As it is clear this network contains 100 neurons. As each of the bank customers holds 8 specifications, the entrance space is four dimensional. SOM topology is two dimensional and to show the eight dimensional relations in two dimensional SOM format, the weight distance matrix called U matrix is used. In the figure U matrix which is the distance between neighbors in SOM is

showed.

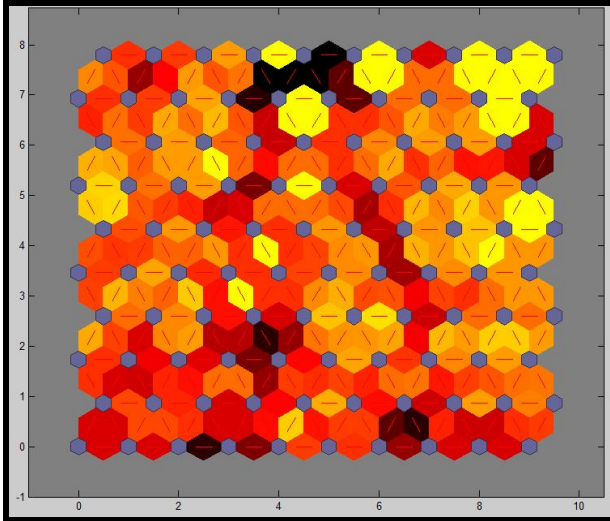


Fig3. SOM matrix of the distance between the neighbors (U matrix) for clustering the bank customers

In figure 3, the red lines between the neurons, show the junctions of the neighbor neurons. The color of the junction areas of the neurons shows the distance of the neurons. The dark colors show the more distance and the light ones show the lower distance. Also we can show the number of the bank customers related to each neuron which is seen in figure 4.

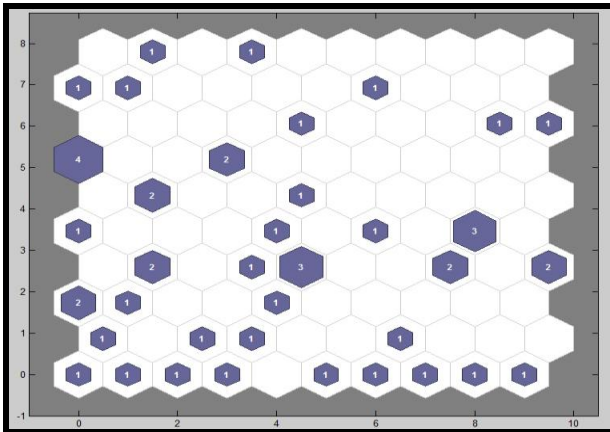


Fig4.SOM SAMLE HITS diagram of bank customers.

In figure 4 the highest value relates to the neuron 4. This means that 4 customers exist in the related cluster. We consider the information of the customers to any entrance and in the Table 1 the related index for any input is showed.

Table1. Mapping of input to the index

Input	index of input
Type of Account	Input 1
Customer opinions about bank services	Input 2
The impact of advertising in select banks	Input 3
Influence the behavior of employees and managers to choose the bank	Input 4
The impact of electronic banking in select banks	Input 5
The impact of benefits of the account to chosen the bank	Input 6
Impact loan fees on select bank	Input 7
Customer's decision to continue cooperation with banks	Input 8

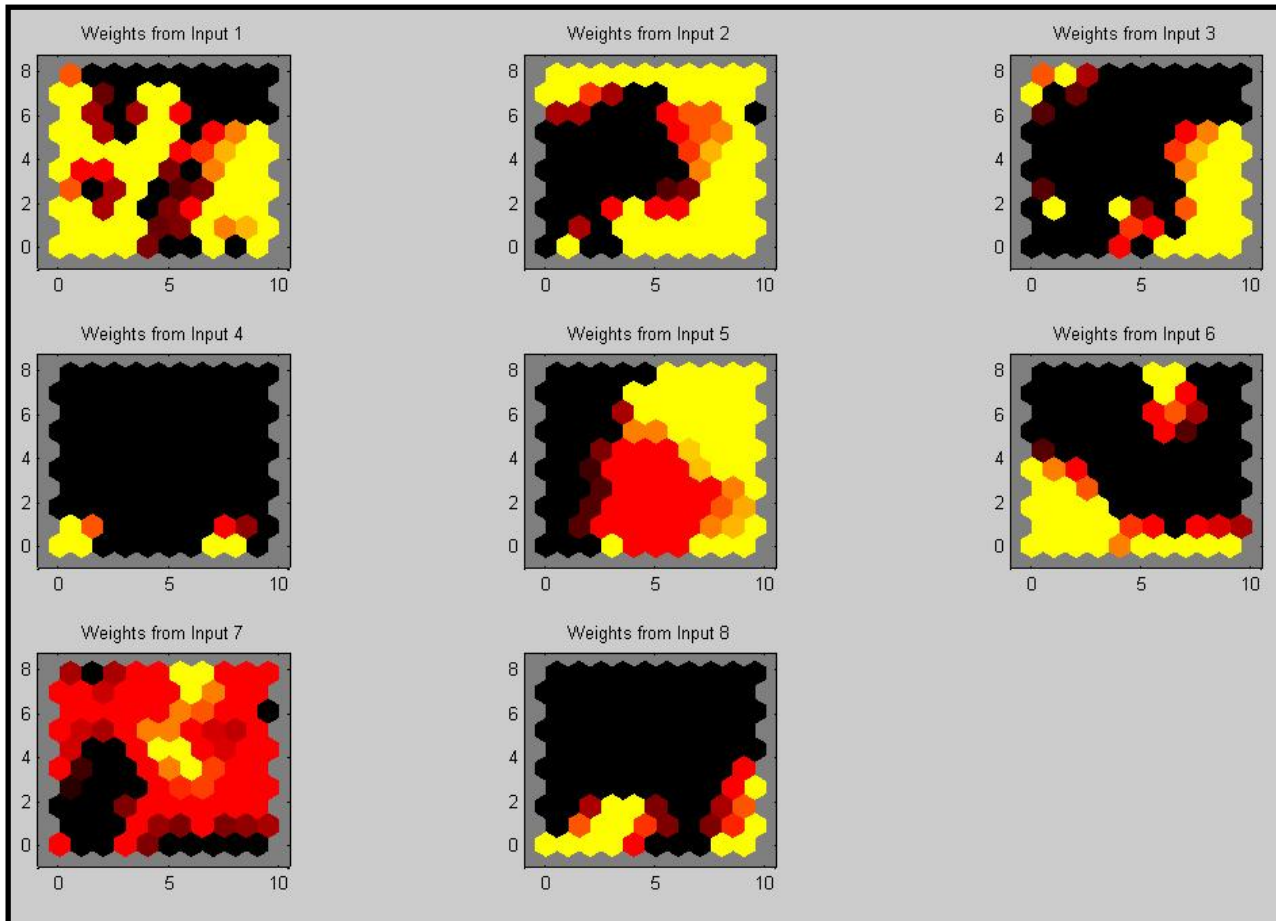


Fig5. The diagram of the weights of the plates related to bank customers clustering.

Figure 5 shows the diagram of the weights of the plates for any element of the input vectors.

These diagrams show the junction weight of the neurons and the darker colors show the larger weights. If the patterns of the junctions between two inputs are similar, it is possible to conclude that the inputs hold high rate of solidarity. It could be seen that the perspective of the customer about the bank services and effect of the electronic banking on selection of a bank and also the effect of the behavior of the clerks and the managers in selection of the bank and the decision of the customer for continuing working with bank hold similar junction patterns and it is possible to conclude that they all hold high rate of solidarity.

5. CONCLUSIONS

In the competing globe of today, creation of the knowledge base and the use of it are advantageous for the banks and the financial institutions and accounting and are being transformed to a strategic tool for their competition. At the late years the ability of production and absorption of the data has increased considerably and as the banking industry has faced many changes all over the world, this industry has felt the need for data mining. The clustering analysis is one of the methods of extracting the information from the data and data mining which is used for grouping the data according to the similarity and closeness insertion.

In this paper, we have extracted the information using the clustering analysis using SOM networks in some cases among the customers of the banks of West Azerbaijan province in Islamic Republic of Iran. The clustering analysis of these information not only classifies the customers in different programming groups based on the specifications and views of the customers of each cluster and with respect to their differences in comparison to other clusters, but also considers the similarities of each cluster and creates a theoretical link among the members of any cluster of the customers which leads to better benefits and efficiency in extracting and mining the wants and interests of the customers.

The results of the clustering analysis showed that the perspective of the customer about bank services and the effect of the electronic banking in banks selection and also the effect of the behavior of the clerks and the managers in bank selection and the decision of the customer for continuing working with the bank, hold very similar junction patterns and it could be concluded that any of them hold high rate of solidarity.

6. REFERENCES

- [1] Sun, J., & Li, H. (2008). Data mining method for listed companies' financial distress prediction. *Knowledge-Based Systems*, 21(1), 1-5.
- [2] Chan, P. K., Fan, W., Prodromidis, A. L., & Stolfo, S. J. (1999). Distributed data mining in credit card fraud detection. *Intelligent Systems and their Applications*, IEEE, 14(6), 67-74.

- [3] Moin, K. I., & Ahmed, Q. B.(2012) .Use of Data Mining in Banking. *International Journal of Engineering Research and Applications*.2 (2,738-742)
- [4] Ikizler, N., & Guvenir, H. A. (2001). Mining interesting rules in bank loans data. In *Proceedings of the Tenth Turkish Symposium on Artificial Intelligence and Neural Networks*.
- [5] Ganji, V. R., & Mannem, S. N. P. (2012). Credit card fraud detection using anti-k nearest neighbor algorithm. *International Journal on Computer Science and Engineering (IJCSSE)*, 4(06).
- [6] Ogwueleka, F. N. (2011). Data mining application in credit card fraud detection system. *Journal of Engineering Science and Technology*, 6(3), 311-322.
- [7] Soeini, R. A., & Rodpysh, K. V. (2012). Evaluations of Data Mining Methods in Order to Provide the Optimum Method for Customer Churn Prediction: Case Study Insurance Industry. In *2012 International Conference on Information and Computer Applications (ICICA 2012) IPCSIT (Vol. 24)*.
- [8] Gan, C., Limsombunchai, V., Clemes, M., & Weng, A. (2005). Consumer choice prediction: Artificial neural networks versus logistic models. *Lincoln University. Commerce Division*..
- [9] Adeyiga, J. A., Ezike, J. O. J., Omotosho, A., & Amakulor, W. (2011). A neural network based model for detecting irregularities in e-Banking transactions. *African Journal of Computer and ICTs*, 4(2), 7-14.
- [10] Gharehchopogh, F.S., & Khaze, S.R. (2012), "Data Mining Application for Cyber Space Users Tendency in Blog Writing: A Case Study", *International Journal of Computer Applications*, 47(18), 40-46.
- [11] Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, 17(3), 37.
- [12] Cios, K. J., Pedrycz, W., & Swiniarsk, R. M. (1998). Data mining methods for knowledge discovery. *Neural Networks, IEEE Transactions on*, 9(6), 1533-1534.
- [13] Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, 33(1), 135-146.
- [14] Stuart Kauffman. (1993). *the origins of order: Self organization and selection in evolution*. Oxford University Press.
- [15] Kangas, J. A., Kohonen, T. K., & Laaksonen, J. T. (1990). Variants of self-organizing maps. *Neural Networks, IEEE Transactions on*, 1(1), 93-99.
- [16] Kaski, S., Kangas, J., & Kohonen, T. (1998). Bibliography of self-organizing map (SOM) papers: 1981-1997. *Neural computing surveys*, 1(3&4), 1-176.
- [17] Kiang, M. Y. (2001). Extending the Kohonen self-organizing map networks for clustering analysis. *Computational Statistics & Data Analysis*, 38(2), 161-180.
- [18] Fritzke, B. (1994). Growing cell structures—a self-organizing network for unsupervised and supervised learning. *Neural networks*, 7(9), 1441-1460.
- [19] Kaski, S., Nikkila, J., & Kohonen, T. (1998). Methods for interpreting a self-organized map in data analysis. In *In Proc. 6th European Symposium on Artificial Neural Networks (ESANN98)*. D-Facto, Brugfes.