Improving Genetic Algorithm Operators for Analyzing Anomalous Behavior of Online Customers

Morteza Talebi

***Abstract*— Nowadays, banks and financial institutes have experienced a transition toward electronic banking leading to larger number of customers and huge amount of transactions. As a result of these changes new issues and challenges emerge which necessitate deep investigation of data. One of these challenges is examining customers' behavior and identifying anomalies so that fraud could be detected. On the other hand, exact investigation of huge transaction data is not possible using conventional methods. Therefore, financial institutes are seeking for novel solutions enabling them to detect anomalous behaviors rapidly. In this paper a novel solution is proposed based on data mining and artificial intelligence. For this purpose, fuzzy expert systems and genetic algorithm are utilized. First off, data collected from transactions are analyzed to extract behavioral features. Afterwards, these features are converted to fuzzy rule base of the proposed fuzzy expert system. Subsequently, the resulted fuzzy rules are optimized using genetic algorithm. In the proposed genetic algorithm, crossover and mutation operators are defined in fuzzy form which significantly improves their efficiency and convergence speed.**

***Index Terms*— Anomalous behavior, Fuzzy expert system, Fuzzy rule, Fuzzy inference system, Genetic algorithm.**

I. INTRODUCTION

Anomaly and fraud detection in financial transactions is a complicated and difficult task [10]. There is not any worldwide accepted definition of financial fraud [11]. Wang et al defined fraud as: "as a deliberate act or deliberate to act that is contrary to law, rule, or policy with intent to obtain unauthorized financial benefit". Fraud is as old as human life and meanwhile it is the most profitable business all over the world. The amount of financial transactions are considerably increasing. Recently, technologic progress has provided criminals with many methods to commit fraud.Creating a new information system may offer different pros and cons; however, it provides more opportunities for fraud commitment. The mentioned frauds inflict irrecoverable loss on investors damaging their competitiveness [1].

Results of a survey conducted by KPMG institute in 2003 revealed increasing fraud rate. It demonstrated that 75% of studied organizations have experienced examples of fraud. The statistics imply a 13% growth comparing to 1998 [2].

Intelligent techniques for fraud and anomaly detection are able to detect frauds in an organization and analyze them. These schemes identify customers' behavior and try to predict their behavior, which, in turn, decreases fraud risk.

II. RELATED WORK

A wide ranged research has been carried out on anomaly and fraud detection using data mining methods. Genetic algorithm is one of the popular methods. In [3] genetic algorithm is utilized for intrusion detection. In this paper some detectors are created using genetic algorithm and Minkowski distance function is exploited instead of Euclidean function. The obtained results demonstrate that improvements might be achieved in genetic algorithm when Minkoowski measure is employed.

Recently, some researchers have tried to combine different approaches to obtain better results. As an example in [4] a hybrid method combining genetic algorithm and Support Vector Machine is proposed to detect anomaly in a dataset. In this method, genetic algorithm is used to select the best subset of features which show anomaly. The obtained optimal data set is employed to train SVM and the results are improved.

To overcome fraud problem in financial institutes genetic algorithm and scatter search algorithm are combined. Merging fuzzy inference systems and artificial intelligence methods is a novel idea which has been used for some problems. Fraud detection in electricity grid is investigated in [6]. In this study fuzzy inference system and SVM method are utilized to face frauds.

Research works have demonstrated that data mining approaches are even successful in fraud detection in insurance industry. Exploiting these methods for fraud detection in financial institutes has risen significantly. In 1994 Gash and Reily proposed a method based on artificial neural networks to detect credit cards fraud [7]. Moreover, game theory has been utilized for fraud

Morteza Talebi, Department of Information, Faculty of Management (Electronic Branch), Young Research and Elite Clob (Electronic Branch), Islamic Azad University, Tehran, Iran.

detection. In this method interactions between attacker and fraud detection system are modeled as a multi-level game between two players each of which tries to maximize its benefits [8]. In [9] a comprehensive review on different data mining based method for fraud detection, proposed between 1997 and 2008 is provided.

In the paper at hand, a combination of fuzzy expert system and genetic algorithm is employed to solve anomaly detection problem. The proposed method is implemented and evaluated using a valid data set. To assess the proposed system data collected from online purchasing websites are utilized.

III. CUSTOMERS' ANOMALOUS BEHAVIOR

One of the main objectives of this study is detecting customers' anomaly. To achieve a clear definition of "suspicious user", exact definitions for anomalous behavior of internet banking users must be determined. These definitions are crucial for choosing the best approach to facing these behaviors. In order to extract anomalous behavior features, customers' behavior in online auction websites is investigated.

According to the investigations, in online auction websites k-core feature properly matchesfrauds' behavior. Using k-core feature one may deeply understand density of a trading network. High k-core corresponds to high density of trading network. This feature is capable of detecting collusion behavior (when users rank or rate each other high). Wang et al used k-core feature to find inflated reputation in trading networks [13-14]. The results of their research demonstrated that k-core feature is sufficient for detecting inflated reputation. So, in this study it is tried to combine other effective features with k-core.

Another prominent feature is seller density detection which is used for extracting online customers' reputation [15]. The seller density is defined as follows. Two sellers (Si and Sj) are linked if minimum number of sellers in the last step of auction, which includes both sellers, is a definite value (min_buyers) and final price of each auction is at least equal to MIN_VALUE.

The number of buyers (Nij) who buy from these sellers is considered as the link power denoted by link(si,sj). The neighborhood of Si seller is denoted by N(Si). According to min_buyers and MIN_VALUE thresholds (defined by user) this neighborhood is introduced as a set of sellers {Sj} linked with Si. Min_buyers threshold is utilized to detect sellers with considerable sells. Besides, MIN_VALUE threshold is used to protect the system against fraudswhich useaggregation to counterfeit identity of credible sellers. Thus, another measure called score (showing seller density) is defined as follows [16].

$$score(s_i) = \sum_{sj \in N(si)} density(s_i). \log_{\min\_buyers} |link(s_i, s)j| \qquad (1)$$

Popular crime economy analytical viewmight be applied to this system. According to this view the first motivation for a crime includes three factors costs, benefits and risks [9]. The criminal tends to obtain the best possible benefits while the costs and risks are minimized. Based on this view, auctionfraudulent participants try to generate several accounts at low expenses and try to rateeach other. Furthermore, they commit fraudin a short period so that their risk is minimized. Hence, such accounts have low positive rate (minimum cost for building a few fraud accounts), short tender period (minimum risk) and rapid increase in negative rate (the most essential achievement of this is large number of complaints and negative credits). In the following, a feature called Negative/Positive Ratio is introduced to detect such features.

$$NP\ Ratio = \frac{Negative\_Rating\_Score}{Positive\_Rating\_Score} \qquad (2)$$

Real auctions information imply that positive rating score of fraud accounts are about 10 while negative rating score is almost 3. N/P ratio of an actual auction seller is much less than 0.3. So, N/P ratio may help us with detecting fraud accounts.

In addition to above mentioned features, main credit scores of auction website are utilized to detect fraud accounts. To sum up, this study uses a combination of social network analysis, crime economy view and main credit of the auction website to detect anomaly in auctions. The detection features of this study consist of four factors; k-core, score (sellers density), N/P Ratio and credit score (total negative rating score).

IV. THE PROPOSED APPROACH

The proposed system comprises a fuzzy expert system whose rule weights are optimized using genetic algorithm. Thus, genetic algorithm is implemented to optimize detection rules of expert system. Figure 1, depicts architecture of fuzzy expert system. Characteristics of input and output variables are presented in table 1.

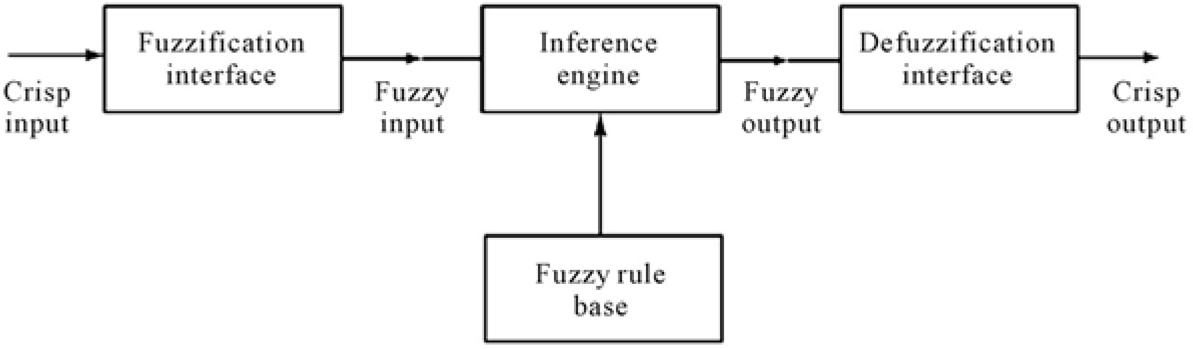


Figure 1.Expert fuzzy system architecture

Table 1.Input and output variable

Fuzzy values of membership functions	Number of membership functions	Variable type	Variable name
NS,PS	2	input	Kcore
NS,PS	2	input	Score
NB,NS,ZO,PS,PB	5	input	Negative-Positive-Ratio
NB,NS,ZO,PS,PB	5	input	negative-rating-score
N3,N2,N1,N,O,P,P1,P2,P3	9	input	fuzzyoutput

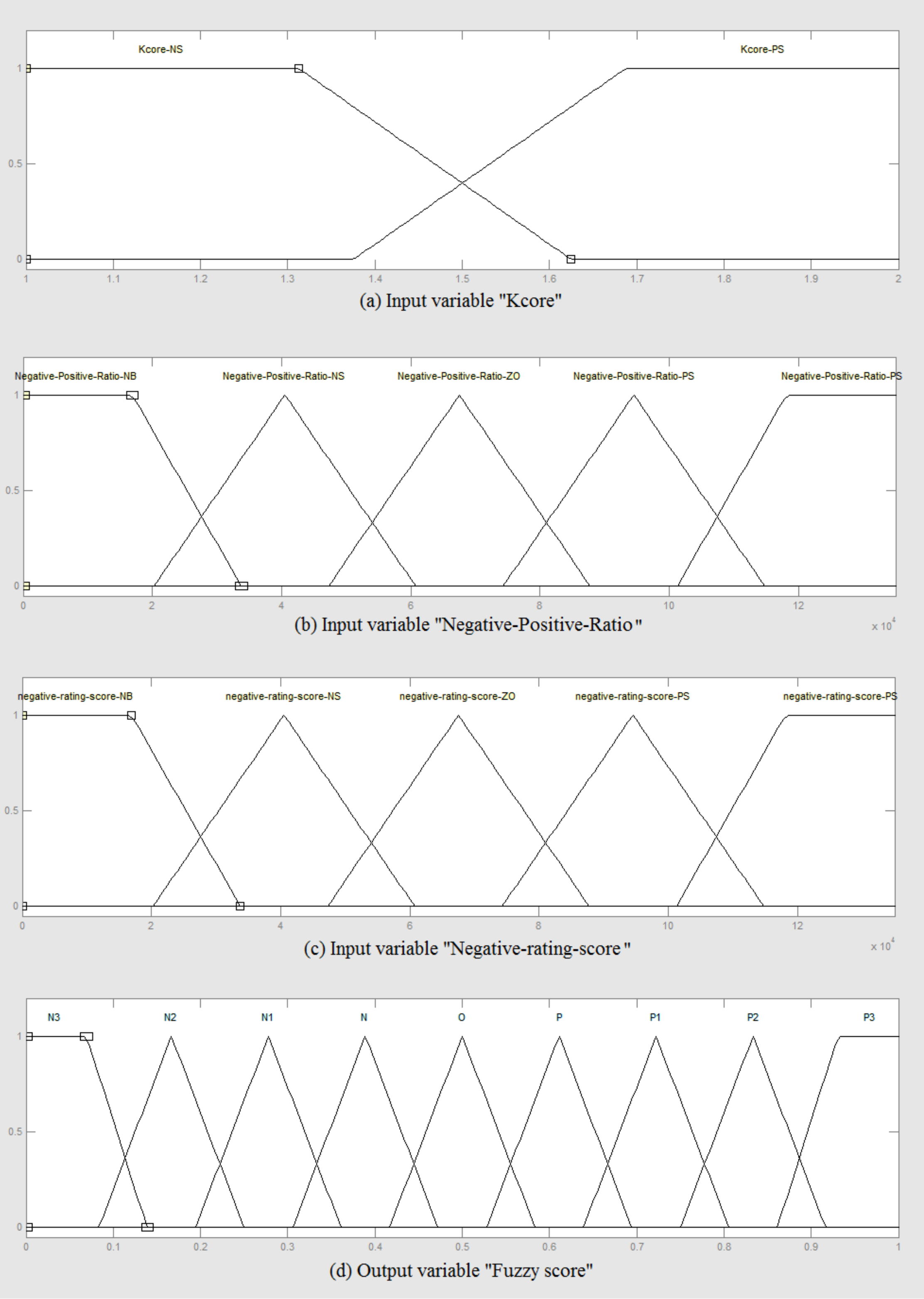
Membership functions of all variables are triangular or trapezoidal for sake of simplicity of computations. Among density inputs (k-core or Score) the most suitable one is selected according to experiments. Thus, inputs consist of density (k-core or Score), Negative/Positive Ratio and negative rating score. Density is only divided into two levels; whereas, Negative-Positive-Ratio and negative rating score are divided into 5 levels. Figures 2-5 illustrate inputs and output of fuzzy rule base membership functions. Table 2 includes rules of fuzzy rule base. To obtain acceptable results using the proposed control method, it is simulated and tested using existing data. As a result priority of each fuzzy rule could be derived using genetic algorithm.

Table 1.fuzzy rules

1. If (Kcore is NS) and (Negative-Positive-Ratio is NB) and (negative-rating-score is NB) then (Fuzzy_Score is N3)
2. If (Kcore is NS) and (Negative-Positive-Ratio is NS) and (negative-rating-score is NB) then (Fuzzy_Score is N2)
3. If (Kcore is NS) and (Negative-Positive-Ratio is ZO) and (negative-rating-score is NB) then (Fuzzy_Score is N1)
4. If (Kcore is NS) and (Negative-Positive-Ratio is PS) and (negative-rating-score is NB) then (Fuzzy_Score is N)
5. If (Kcore is NS) and (Negative-Positive-Ratio is PB) and (negative-rating-score is NB) then (Fuzzy_Score is O)
6. If (Kcore is NS) and (Negative-Positive-Ratio is NB) and (negative-rating-score is NS) then (Fuzzy_Score is N2)
7. If (Kcore is NS) and (Negative-Positive-Ratio is NS) and (negative-rating-score is NS) then (Fuzzy_Score is N2)
8. If (Kcore is NS) and (Negative-Positive-Ratio is ZO) and (negative-rating-score is NS) then (Fuzzy_Score is N1)
9. If (Kcore is NS) and (Negative-Positive-Ratio is PS) and (negative-rating-score is NS) then (Fuzzy_Score is N)
10. If (Kcore is NS) and (Negative-Positive-Ratio is PB) and (negative-rating-score is NS) then (Fuzzy_Score is O)
11. If (Kcore is NS) and (Negative-Positive-Ratio is NB) and (negative-rating-score is ZO) then (Fuzzy_Score is N1)
12. If (Kcore is NS) and (Negative-Positive-Ratio is NS) and (negative-rating-score is ZO) then (Fuzzy_Score is N1)
13. If (Kcore is NS) and (Negative-Positive-Ratio is ZO) and (negative-rating-score is ZO) then (Fuzzy_Score is N)
14. If (Kcore is NS) and (Negative-Positive-Ratio is PS) and (negative-rating-score is ZO) then (Fuzzy_Score is O)
15. If (Kcore is NS) and (Negative-Positive-Ratio is PB) and (negative-rating-score is ZO) then (Fuzzy_Score is O)
16. If (Kcore is NS) and (Negative-Positive-Ratio is NB) and (negative-rating-score is PS) then (Fuzzy_Score is N)
17. If (Kcore is NS) and (Negative-Positive-Ratio is NS) and (negative-rating-score is PS) then (Fuzzy_Score is N)
18. If (Kcore is NS) and (Negative-Positive-Ratio is ZO) and (negative-rating-score is PS) then (Fuzzy_Score is O)
19. If (Kcore is NS) and (Negative-Positive-Ratio is PS) and (negative-rating-score is PS) then (Fuzzy_Score is P)
20. If (Kcore is NS) and (Negative-Positive-Ratio is PB) and (negative-rating-score is PS) then (Fuzzy_Score is P)
21. If (Kcore is NS) and (Negative-Positive-Ratio is NB) and (negative-rating-score is PS) then (Fuzzy_Score is O)
22. If (Kcore is NS) and (Negative-Positive-Ratio is NS) and (negative-rating-score is PS) then (Fuzzy_Score is O)
23. If (Kcore is NS) and (Negative-Positive-Ratio is ZO) and (negative-rating-score is PS) then (Fuzzy_Score is O)
24. If (Kcore is NS) and (Negative-Positive-Ratio is PS) and (negative-rating-score is PS) then (Fuzzy_Score is P)
25. If (Kcore is NS) and (Negative-Positive-Ratio is PB) and (negative-rating-score is PS) then (Fuzzy_Score is P1)
26. If (Kcore is PS) and (Negative-Positive-Ratio is NB) and (negative-rating-score is NB) then (Fuzzy_Score is O)
27. If (Kcore is PS) and (Negative-Positive-Ratio is NS) and (negative-rating-score is NB) then (Fuzzy_Score is O)
28. If (Kcore is PS) and (Negative-Positive-Ratio is ZO) and (negative-rating-score is NB) then (Fuzzy_Score is P)
29. If (Kcore is PS) and (Negative-Positive-Ratio is PS) and (negative-rating-score is NB) then (Fuzzy_Score is P1)
30. If (Kcore is PS) and (Negative-Positive-Ratio is PB) and (negative-rating-score is NB) then (Fuzzy_Score is P1)
31. If (Kcore is PS) and (Negative-Positive-Ratio is NB) and (negative-rating-score is NS) then (Fuzzy_Score is O)
32. If (Kcore is PS) and (Negative-Positive-Ratio is NS) and (negative-rating-score is NS) then (Fuzzy_Score is O)
33. If (Kcore is PS) and (Negative-Positive-Ratio is ZO) and

(negative-rating-score is NS) then (Fuzzy_Score is P)

34. If (Kcore is PS) and (Negative-Positive-Ratio is PS) and (negative-rating-score is NS) then (Fuzzy_Score is P1)
35. If (Kcore is PS) and (Negative-Positive-Ratio is PB) and (negative-rating-score is NS) then (Fuzzy_Score is P1)
36. If (Kcore is PS) and (Negative-Positive-Ratio is NB) and (negative-rating-score is ZO) then (Fuzzy_Score is P)
37. If (Kcore is PS) and (Negative-Positive-Ratio is NS) and (negative-rating-score is ZO) then (Fuzzy_Score is P)
38. If (Kcore is PS) and (Negative-Positive-Ratio is ZO) and (negative-rating-score is ZO) then (Fuzzy_Score is P1)
39. If (Kcore is PS) and (Negative-Positive-Ratio is PS) and (negative-rating-score is ZO) then (Fuzzy_Score is P2)
40. If (Kcore is PS) and (Negative-Positive-Ratio is PB) and (negative-rating-score is ZO) then (Fuzzy_Score is P2)
41. If (Kcore is PS) and (Negative-Positive-Ratio is NB) and (negative-rating-score is PS) then (Fuzzy_Score is P1)
42. If (Kcore is PS) and (Negative-Positive-Ratio is NS) and (negative-rating-score is PS) then (Fuzzy_Score is P1)
43. If (Kcore is PS) and (Negative-Positive-Ratio is ZO) and (negative-rating-score is PS) then (Fuzzy_Score is P2)
44. If (Kcore is PS) and (Negative-Positive-Ratio is PS) and (negative-rating-score is PS) then (Fuzzy_Score is P2)
45. If (Kcore is PS) and (Negative-Positive-Ratio is PB) and (negative-rating-score is PS) then (Fuzzy_Score is P2)
46. If (Kcore is PS) and (Negative-Positive-Ratio is NB) and (negative-rating-score is PS) then (Fuzzy_Score is P1)
47. If (Kcore is PS) and (Negative-Positive-Ratio is NS) and (negative-rating-score is PS) then (Fuzzy_Score is P1)
48. If (Kcore is PS) and (Negative-Positive-Ratio is ZO) and (negative-rating-score is PS) then (Fuzzy_Score is P2)
49. If (Kcore is PS) and (Negative-Positive-Ratio is PS) and (negative-rating-score is PS) then (Fuzzy_Score is P2)
50. If (Kcore is PS) and (Negative-Positive-Ratio is PB) and (negative-rating-score is PS) then (Fuzzy_Score is P3)



(a) Input variable "Kcore"

(b) Input variable "Negative-Positive-Ratio"

(c) Input variable "Negative-rating-score"

(d) Output variable "Fuzzy score"

Figure 1. Fuzzy input and output variables, a) fuzzy input variable: Kcore or Score, b) fuzzy input variable: N/PRatio, c) fuzzy input variable: credit score (total negative rating score), d) fuzzy output variable

N/P Ratio	Density (K-core or Score) = NS					Density (K-core or Score) = PS				
	Negative rating score					Negative rating score				
	NB	NS	ZO	PS	PB	NB	NS	ZO	PS	PB
NB	G_0	G_1	G_2	G_3	G_4	G_{25}	G_{26}	G_{27}	G_{28}	G_{29}
NS	G_5	G_6	G_7	G_8	G_9	G_{30}	G_{31}	G_{32}	G_{33}	G_{34}
ZO	G_{10}	G_{11}	G_{12}	G_{13}	G_{14}	G_{35}	G_{36}	G_{37}	G_{38}	G_{39}
PS	G_{15}	G_{16}	G_{17}	G_{18}	G_{19}	G_{40}	G_{41}	G_{42}	G_{43}	G_{44}
PB	G_{20}	G_{21}	G_{22}	G_{23}	G_{24}	G_{45}	G_{46}	G_{47}	G_{48}	G_{49}

Table 2.Genes cryptography [17]

V. THE PROPOSED GENETIC ALGORITHM

Figure 3 demonstrates flowchart of the proposed genetic algorithm which is tasked with optimizing weights of fuzzy rules. Different steps are explicated in the following.

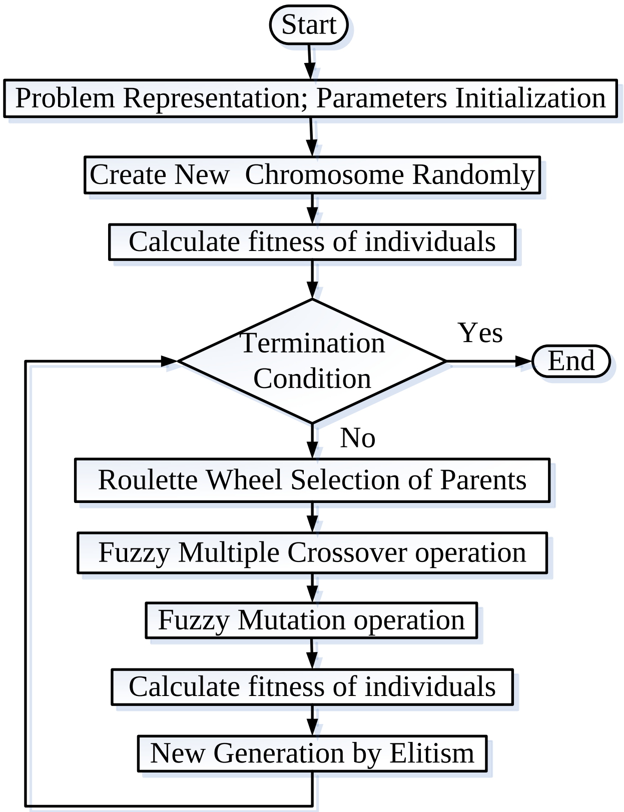

Start
Problem Representation; Parameters Initialization
Create New Chromosome Randomly
Calculate fitness of individuals
Termination Condition
Yes
End
No
Roulette Wheel Selection of Parents
Fuzzy Multiple Crossover operation
Fuzzy Mutation operation
Calculate fitness of individuals
New Generation by Elitism

Figure 2.the proposed genetic algorithm

VI. CHROMOSOME DEFINITION

Genes of chromosome are coded as a table of fuzzy control rules (G0 to G49). Each cell denotes the weight of corresponding rule. In figure 4, G0 is the output when negative rating score and N/P ratio are NB and density (k-core, score) is NS.

A. Fitness function

Objective function is designed such that the gene with the best capability of detecting anomaly is effectively identified. Additionally, the desired gene may be used for comparing features. More important gene corresponds tohigher threshold and vice versa. Fitness function is defined as follows.

$$E_i = Fuzzy\_Score \tag{3}$$

$$Threshold = \frac{Normal\_Max\_Fuzzy\_Score + Fraud\_Min\_Fuzzy\_Score}{2} \tag{4}$$

$$Recall = \frac{Count(Fraud\_Ei > Threshold)}{Count(Fraud\_Account)} \tag{5}$$

$$Precision = \frac{Count(Fraud\_Ei > \ Threshold)}{Count(Fraud\_Ei > Threshold) + Count(Normal\_Ei > Threshold)} \tag{6}$$

$$Goal = Max(F\text{-}measure) \tag{7}$$

$$F\text{-}measure = \frac{2 \cdot precision \cdot recall}{precision + recall} \tag{8}$$

In equation 3, $\mathbf{E_i}$ is fuzzy output of a chromosome of the population. In equation 5, **Recall** demonstrates how many times detection function is recalled. **Count (Fraud_E$_i$>Threshold)** determines the number of fraud accounts when threshold is E_i.

Threshold is calculated using equation 4 and it is depicted in figure 5. Threshold is set to be the average of minimum fuzzy score of fraud accounts list and maximum fuzzy score of normal accounts lists. **Count (Fraud_Account)** is the total number of accounts in fraud accounts list. **Precision** in equation 6 denotes precision of detection rule. **Count (Normal_E$_i$> Threshold)** is the number of accounts in normal accounts list while threshold is E_i. **Goal** (equation 7) is the objective of genetic algorithm which is desired to be maximized**. F-measure** is general and is a balanced value between recall and precision. Using this fitness function the optimal detection rule would be able to detect all fraud accounts and derive maximum **Count (Fraud_Ei > Threshold)**. This optimal value is equal to **Count (Fraud_Account)**. Therefore, the best value for **Recall** is almost 1. Moreover, optimal detection rules avoid detecting a normal account as a fraud account. Thus, minimum value of **Count (Normal_Ei > Threshold)** and optimal value would be zero. The best values for **Precision** and **F-measure** are close to 1.

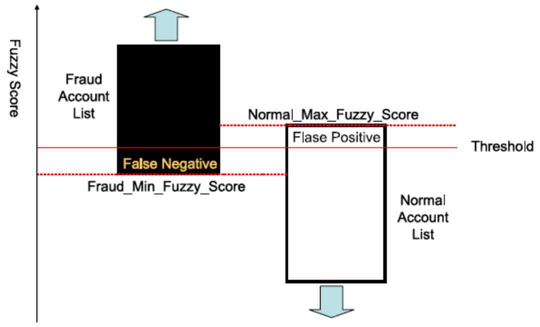


Figure 3. Threshold design [17]

B. Fuzzy Crossover Operator

In this study multiple crossover method is exploited. As a consequence, chromosomes are involved more effectively in offspring generation so that the search space is explored more properly and efficiency of search increases. In this paper, the number of crossover points is determined using a fuzzy inference system. The number of current iteration and best fitness changes in previous iterations are inputs of this fuzzy inference system. The output of the system is a fuzzy number specifying crossover point number according to a predetermined range. Figure 6 demonstrates generic model for the proposed fuzzy inference system. Membership functions of inputs and output are presented in figure 7. Max point number is assumed to be 5 in this paper.

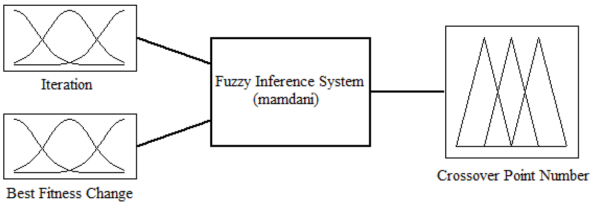


Figure 4.generic model of the proposed fuzzy inference system for crossover

C. Fuzzy Mutation Operator

Genetic algorithm may encounter difficulties due to local optimal solutions. To overcome this problem mutation operator is employed which randomly changes genes and leads them to other areas of search space avoiding local optimal solutions to dominate the algorithm. Mutation speed or the number of mutations is inversely related to fitness changes of the best solution in previous iterations and the current iteration. It is calculated using the proposed fuzzy inference system. The number of current iteration and fitness changes of the best solution in previous iterations are fed, as inputs, to this inference system while its output is a fuzzy number determining the number of mutations. Membership functions for inputs and outputs are depicted in figure 8. Mutation number is considered to be 4 in this paper.

VII. SIMULATION RESULTS

The proposed algorithm is simulated in MATLAB environment. In this section the proposed scheme is simulated and its results are compared to previous methods in literature.

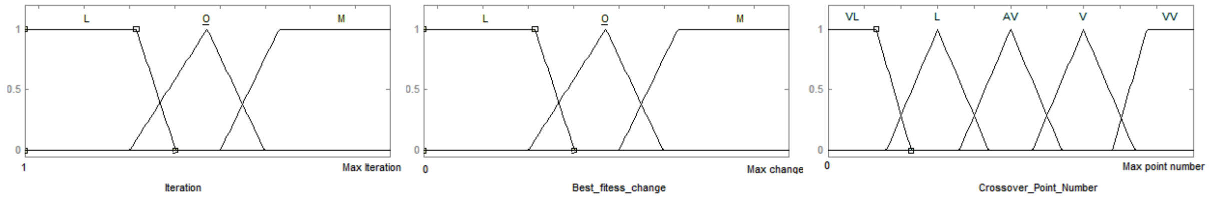


Figure 7.membership functions for crossover fuzzy inference system

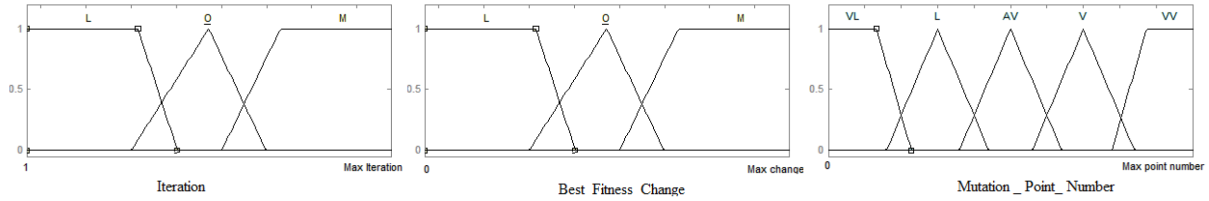


Figure 5.membership functions of the proposed fuzzy inference systems

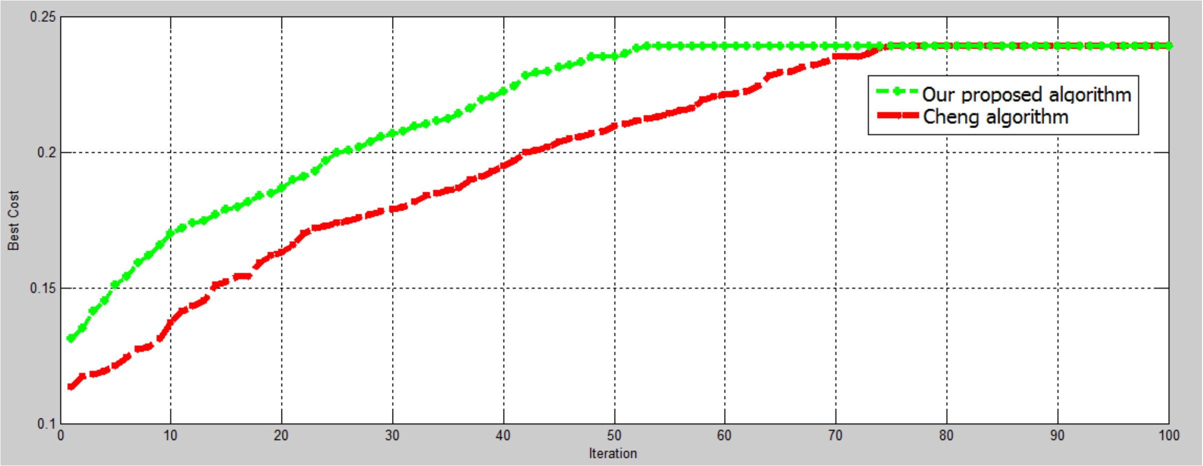


Figure 6.Comparison between the proposed algorithm and Cheng algorithm

In the first test K-core and Score features are assessed. For this purpose, genetic algorithm is executed 10 times for each of these features. The initial population and number of generations are assumed to be 200 and 50, respectively. Contrary to score feature,which reaches the desired solution in 31st generation, the k-score does not lead to any solutions even after 50 generations. The results demonstrate superiority of Score feature to K-core. Subsequently, another test is designed which is performed considering a population of 100 chromosomes with 100 iterations. It is designed to comparer performance of the proposed method to the algorithm introduced by Cheng [17]. Diagrams shown in figure 9 are average of executing the algorithm for 20 times. According to simulation results our proposed method significantly increases convergence speed. As can be seen in the figure, for similar recall number of fitness function, the proposed system is capable of improving algorithm performance and convergence speed.

CONCLUSION

Development of banks and financial institutes has led to rise in the number of customers and transactions. The challenges and issues resulted from this increase requires deep investigation of data. However, it is not possible to examine large amount of data using conventional methods. One of the most crucial challenges is anomaly and fraud detection. Thus, financial institutes are seeking for methods through which they could detect crimes rapidly.

In this paper, a novel method was proposed for fraud and anomaly detection based on data mining. The proposed scheme is a combination of fuzzy expert system and genetic algorithm. The proposed method was evaluated using valid datasets. The simulation results revealed that performance and convergence speed improve considerably. This is obtained using fuzzy crossover and fuzzy mutation operators.

REFERENCES

[1] Albrecht, W.S., Albrecht,C.C. & Albrecht, G.O. (2008). Current Trends in Fraud and Its Detection. Information Security Journal, 17.

[2] Institute, KPMG.(2003). Fraud Survey of 2003, From http://www.kpmg.com.

[3] Aziz, Amira Sayed A., et al. "Artificial Immune System Inspired Intrusion Detection System Using Genetic Algorithm." Informatica (Slovenia) 36.4 (2012): 347-357.

[4] Anil, S., and R. Remya. "A hybrid method based on genetic algorithm, self-organised feature map, and support vector machine for better network anomaly detection." 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT). IEEE, 2013.

[5] Duman, Ekrem, and M. Hamdi Ozcelik. "Detecting credit card fraud by genetic algorithm and scatter search." Expert Systems with Applications 38.10 (2011): 13057-13063.

[6] Nagi, Jawad, et al. "Improving SVM-Based Nontechnical Loss Detection in Power Utility Using the Fuzzy Inference System." Power Delivery, IEEE Transactions on 26.2 (2011): 1284-1285.

[7] Phua, C.; Alahakoon, D.; Lee, V. . 2004. "Minority Report in Fraud Detection: Classification of Skew Data", ACM SIGKDD Exploration Newsletter, 6, 1, 50-59.

[8] Vatsa, V.; Sural, S.; Majumadar, A.K.; 2005. "A Game Theoretic Approach to Credit Card Fraud Detection". Proc. First Int’l Conf. Information Systems Security, 263-276.

[9] Yeh, I.; Lien, C.; 2008. "the comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients", Expert Systems with Applications, 36 (2), 2473-2480.

[10] Yue D., X. Wu, Y. Wang, Y. Li and C. Chu, A Review of Data Mining-based Financial Fraud Detection Research, International Conference on Wireless Communications, Networking and Mobile Computing, 2007, pp. 5519–5522

[11] Ngai E.W.T., Yong Hu, Y.H. Wong, Yijun Chen, Xin Sun, The Application of Data Mining Techniques in Financial Fraud Detection: A Classification Framework and an Academic Review of literature; Decision Support Systems, vol. 50(3), 2010, pp. 559-569

[12] Wang J., Y. Liao, T. Tsai, G. Hung, Technology-based Financial Frauds in Taiwan: Issue and Approaches, IEEE Conference on: Systems, Man and Cyberspace, 2006, pp. 1120–1124

[13] Wang, J. C., & Chiu, C. C. (2008). Recommending trusted online auction sellers using social network analysis. Expert Systems with Applications, 34(3), 1666–1679.

[14] Wang, J. C., & Chiu, C. (2005). Detecting online auction inflated-reputation behaviors using social network analysis. In Annual conference of the North American Association for Computational Social and Organizational Science.

[15] Internet Crime Complaint Center (2009), 2008 Internet crime report.

[16] Morzy, M. (2008). New algorithms for mining the reputation of participants of online auctions. Algorithmica, 52(1), 95–112.

[17] Yu, Cheng-Hsien, and Shi-Jen Lin. "Fuzzy rule optimization for online auction frauds detection based on genetic algorithm." Electronic Commerce Research 13.2 (2013): 169-182.