ISSN:0975-9646

Genetic K-means Algorithm for Credit Card Fraud Detection

Pooja Chougule[#1], A.D. Thakare [*2], Prajakta Kale [#3], Madhura Gole[*4], Priyanka Nanekar[#5]

[#]*Department of Computer Engineering, Savitribai Phule Pune University, Pimpri Chinchwad College of Engineering, Pune, India*

[*]*Department of Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, India*

***Abstract*— Rapid growth in electronic commerce technology has led to a tremendous increase in the use of online credit card payment mode. With the usage of credit cards, the number of frauds associated with it also increases. In order to avoid credit card frauds, proper security measures need to be taken. This work reflects an attempt to detect fraudulent credit card transactions by using k-means along with genetic algorithm. Genetic Algorithm is a powerful optimization technique. The k-means algorithm groups the credit card transactions based on the independent attribute values. But, with the increase in input size, it results in outliers. Hence to provide optimized detection of frauds, we used genetic algorithm. The significant results by proposed model are observed over simple K-means and Simple Genetic Algorithm.**

***Keywords*— Fraud Detection, E-commerce technology, Credit Card, K-means, Genetic Algorithm (GA), Data Mining.**

I. INTRODUCTION

In today's electronically advanced world e-commerce plays crucial role in global business. Due to time and ease constraints customers prefer to use online payment mode. Along with the increase in online payment mode fraudulent transactions are also increased more. In 2005 total fraud loss in USA associated with credit cards was 2.7 billion out of which 1.6 billion losses was due to online shopping. In 2006 fraud loss, increased up to 3 billion, whereas losses due to online shopping increased up to 1.7 billion [11]. In consideration with e-commerce business main actors are e-commerce merchants and customers. After detection and giving report of fraudulent transactions, customers are paid back. For every fraud transaction e-commerce merchants needs to pay chargeback's. So here merchant is the one who incurs the loss.

To address this problem different business organizations use different tools for fraud detection. The fraudsters are very adaptive they adapt to new technologies very faster so it is better to use continuous enhancement in fraud detection systems [4]. Fraudulent transactions are scattered along with genuine transactions. So there is need to detect fraudulent transaction to reduce losses. It is convenient to use data mining approach for detecting credit card frauds [12]. In this paper we are using k-means algorithm to convert the transactions into clusters [3]. Then after clustering to optimize the results of clustering we are using genetic algorithm [9].

II. RELATED WORK

A. K-means Algorithm

In this scheme the input data is classified into specified number of groups. It is unsupervised learning approach used when there is no prior knowledge about particular class of observations in a dataset [3]. This scheme classifies n data points into pre-specified k clusters. The data will be grouped into k-clusters according similarities among the cluster. In first step k-clusters need to be defined. In Second step randomly centroid for each cluster will be choose. Centroid of particular cluster means mean value of that cluster. In third step distance of data from centroid of each cluster need to be computed. Data will be grouped into particular cluster, according to minimum distance of data from centroid of cluster. Next time again centroid will be recomputed for each cluster because different values come in cluster. These steps will be repeated until there is no change in the output.

B. Hidden Markov Model

HMM is finite set of states associated with some probabilities with it [2]. Every state generates outcome according to the certain probability associated with that particular state. The outcomes of state can be visible but the states are hidden, so named Hidden Markov Model.

Hidden Markov Model is used for detecting credit card frauds by analysing the spending profiles of credit card holder. Spending profiles of the user can be calculated according to user's past history of transaction in terms of attributes like transaction amount, IP address, shipping address & location of last transaction, etc. HMM model categories spending profiles of the user into 3 different categories such as high, medium, low [2]. HMM is carried out in two steps, in first step HMM model is been trained on basis of past transaction history and in second step HMM takes the input and check whether transaction details are accepted by trained HMM or not, otherwise it raise an alarm.

C. Dempster-Shafer Theory

Dempster-Shafer theory consists of four main components as rule-based engine, Dempster-Shafer theory, transaction history database and Bayesian learning method. Transaction history database is data repository of both

fraudulent and genuine transactions [4]. Rule-based filter have certain rules such as address mismatch and outlier detection to determine the level of transaction deviation from normal behaviour. The rule-based filter is scalable in terms of rules. In second phase of fusion approach Dempster-Shafer theory will act as adder. This theory combine address mismatch, outlier detection rules and classifies transaction into different groups.

Bayesian learning gives optimal decision. The main drawback of this approach is there is high possibility of conflict in evidences which will decreases accuracy of modelling [4].

D. Biometrics

In this technique we can accurately verify if the credit card transaction is fraud. This is implemented by the IRPV (Iris Recognition and Palm Vein) technique [5]. It provides the 2 step mechanism for the security Iris Recognition and Palm Vein Authentication. This method provides 99.9% accuracy for fraud detection as no two persons have same attributes for Iris and Palm. The biggest disadvantage of the system is the cost for the palm and vein recognition device is more. Some disaster may happen with the person due to which the person can lose identity; in this case he can't access his credit card. So this method is not used nowadays.

E. Genetic Algorithm

Genetic Algorithm is heuristic search algorithm which follows survival of fittest rule of natural selection [10]. Basically there are three steps in GA that are selection, crossover and mutation. Selection calculates fitness of individuals in each generation. Crossover will perform combination of individuals to generate new individual. Finally mutation will do random modification on newly generated individual by crossover step. This procedure will be repeated until best solution is found after producing certain number of generation [1].

This approach use fitness score and inherited good properties of parents to produce new generation due to which GA provide better results. Also it gives optimization search method by using better region of search space.

III. PROPOSED SYSTEM

As shown in Fig. 1 first the dataset loaded. In second step on each transaction rules will be applied from rule engine module. The rule engine content following rules:

Average daily spending, CC Usage Frequency, CC Usage Location, Proxy Port check, IP Address Check, Wrong Password Attempt Check, Authentication Type Check, CC Balance, CC Overdraft.

For every transaction summation of all critical values by each rule is computed and then k-means clustering algorithm applied on summation of critical values for each transaction.

In order to overcome execution time of credit card fraud risk assessment model k-means algorithm is used which will form three clusters low risk, medium risk, and high risk. Genetic algorithm select fittest individuals from medium risk and high risk cluster then perform single point crossover and one bit flip mutation to produce new population until best result found and give results in terms of critical, monitorable and ordinary records.

Genetic algorithm is robust algorithm and gives optimization solution but due to more number of iteration to produce new population increases execution time.

Hence in credit card fraud risk assessment model we have used Genetic k-means algorithm to use features of simple GA and to overcome execution time.

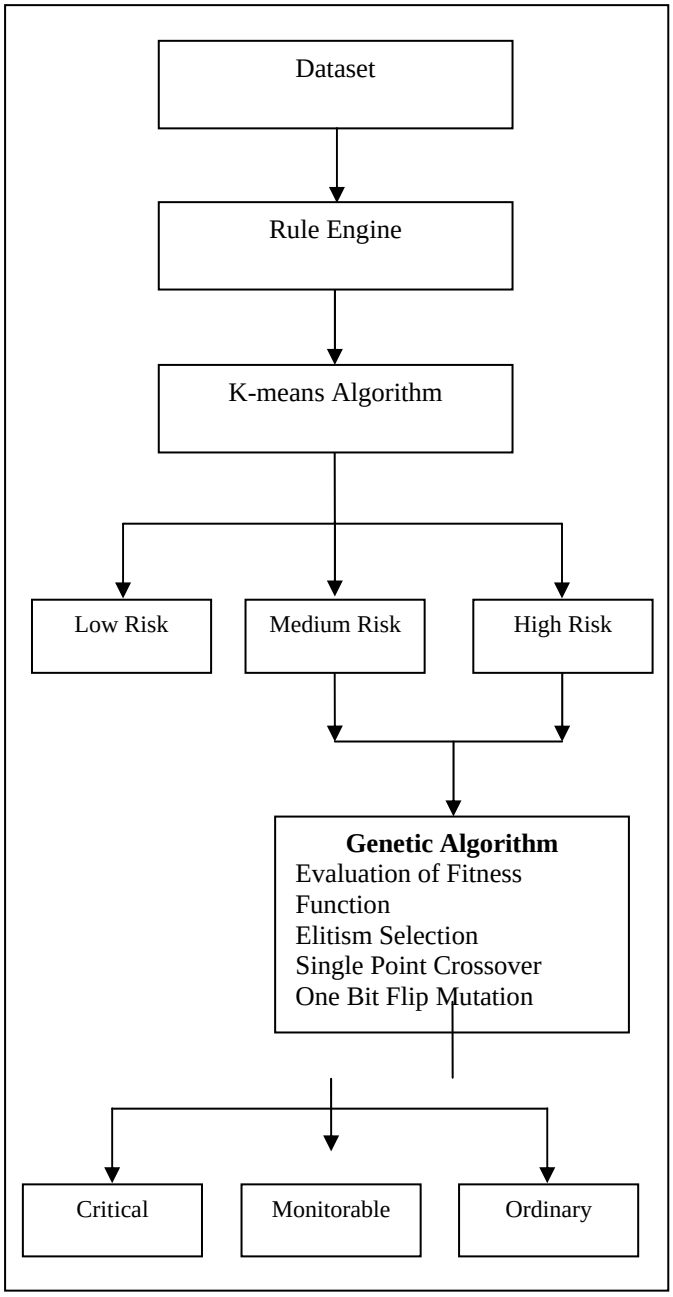

Fig. 1 Credit Card fraud risk assessment model

Genetic algorithm fitness function:

$$\text{Fitness Function} = \frac{\sum_{i=1}^{n} Vi}{\sum_{i=1}^{n} BiVi}$$

Where,
n = number of rules
Bi = Boolean vector

If critical value more than threshold set to 1
else set to 0
Vi = threshold value

Genetic K-means Algorithm for Credit Card Fraud Detection

Steps:

1. Load the dataset.
2. According to the rule engine calculate the critical values for each transaction in dataset.
3. Apply k-means clustering algorithm to generate 3 different clusters of records low risk, high risk and medium risk as per their critical values.
4. Apply genetic algorithm to medium and high risk cluster.
5. Until the obtained results are repeated.
6. Evaluate the fitness of each transaction.
7. Perform elitism selection to generate new population.
8. Perform single point crossover.
9. Perform one bit flip mutation.

Fig. 1 Genetic K-means Algorithm

The Fig. 2 describes the steps used in the Genetic K-means algorithm used in the implementation of the proposed system.

IV. EXPERIMENTAL RESULT

TABLE I Experimental Result

Parameters	Simple Genetic Algorithm	Genetic K-means for Fraud Detection
Average Daily Spending	13.3%	13.3%
CC Usage Frequency	2.3%	2.3%
CC Usage Location	6.6%	6.6%
Proxy Port check	1.6%	1.6%
IP Address Check	1.6%	1.6%
Wrong Password Attempt Check	4.3%	4.3%
Authentication Type Check	2.5%	2.5%
CC Balance	4.6%	4.6%
CC Overdraft	9.53%	9.53%
Execution Time	2 min 04 sec	1 min 53 sec

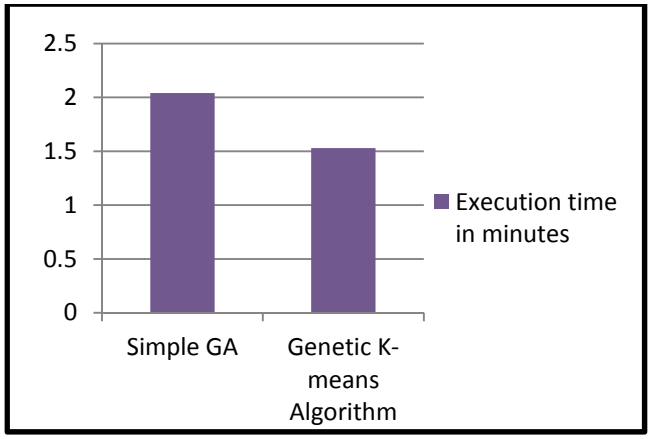

Fig. 2 Performance of Proposed System in terms of execution time

We have tested our proposed system and Simple GA on database of 1500 records according to our rules that is Average daily spending [1], CC Usage Frequency [1], CC Usage Location [1], Proxy Port check, IP Address Check, Wrong Password Attempt Check, Authentication Type Check, CC Balance [1], CC Overdraft [1] we calculated the percentage of critical records found by each rule.

As shown in Table I result and also the Fig. 3 describes that the results generated by simple GA and Genetic k-means algorithm have same results but differs in the execution time which is of great effect when the size of the transactions increases accordingly.

V. CONCLUSION

Credit card fraud has been deeply rooted in the e-commerce industry. In this scenario more of the financial losses are associated with the e-commerce merchants. To save merchant from these losses we have proposed the Credit Card fraud risk assessment model. In order to improve fraud risk assessment we have used combination of two presented methods. In proposed model, genetic algorithm is applied on the clusters generated by k-means clustering algorithm. Genetic algorithm will optimize the output generated by k-means clustering.

The rule engine is used so that system is scalable in terms of rules. In future this model can be extended by adding various rules in rule engine to improve accuracy of the system.

ACKNOWLEDGMENT

We thank Prof. A. D. Thakare, Department of Computer Engineering, Pimpri Chinchwad College of Engineering, Pune for her guidance and support.

REFERENCES

[1] K. RamaKalyani and D. UmaDevi, "Fraud Detection Of Credit Card Payment System by Genetic Algorithm," International Journal Of Scientific Research, vol. 3, Issue 7, July 2012.

[2] A. Srivastava, A. Kundu, S. Sural and A. Majumdar, "Credit Card Fraud Detection Using Hidden Markov Model," IEEE Transactions On Dependable And Secure Computing, vol. 5, No. 1, January-March 2008.

[3] Vaishali, "Fraud Detection in Credit Card by Clustering Approach," International Journal of Computer Applications, vol. 98, No. 3, July 2014.

[4] S. Panigrahi, A. Kundu, S. Sural and A. Majumdar, "Credit Card Fraud Detection: A Fusion Approach Using Dempster-Shafer Theory and Bayesian Learning," ElseVeir.

[5] Prithika and P. Rajalakshmi, "Credit Card Duplication and Crime Prevention Using Biometrics," Computer Science and Engineering, R.M.K. Engineering College, Chennai, Tamil Nadu, India.

[6] V. Dheepa and R. Dhanapal, "Analysis of Credit Card Fraud Detection Methods," International Journal of Recent Trends in Engineering, vol. 2, No. 3, Nov. 2009.

[7] Benson and A. Portia A, "Analysis on Credit Card Fraud Detection Methods," IEEE International Conference on Computer, Communication and Electrical Technology, 18 th and 19 th, 152-156, 2011.

[8] C. phua, V. lee1, K. smith and R. gayler, "A Comprehensive Survey of Data Mining-based Fraud Detection Research," 2005.

[9] Blickle and Thiele, "A Comparison of Selection Schemes used in Genetic Algorithms," Zurich: Swiss Federal Institute of Technology, vol. 2, 1995.

[10] S. Vats, S. Dubey and N. Pandey, "Genetic algorithms for credit card fraud detection," Proceedings of the 2013 International Conference on Education and Educational Technologies.

[11] "Statistics for General and On-Line Card Fraud," www.epaynews.com/statistics/fraud.html, Mar. 2007.

[12] F. Ogwueleka, "Data mining application in credit card fraud detection system," Journal of Engineering Science and Technology, Vol. 6, No. 3, 2011.

[13] R. Patidar and L. Sharma, "Credit Card Fraud Detection Using Neural Network," International Journal of Soft Computing and Engineering (IJSCE), ISSN: 2231-2307, Volume-1, Issue-NCAI2011, June 2011.

[14] J. Dara and L. Gundemoni, "Credit Card Security and E-Payment," 2006.

[15] P. Chan, W. Fan, Prodromidis and Salvatore, "Distributed Data Mining in Credit Card Fraud Detection," IEEE December 1999.