A review of various credit card fraud detection techniques

[1]*Yogeshwari Dilip Hardas and* [2]*Prof. Ankush Pawar*
[1]*ME Computer Science, ARMIET, Sapgaon*
[2]*(H.O.D) Dept. of Computer Science, DRIEMS, Neral*
yohardas143@gmail.com ankushpawar1981@gmail.com

Abstract— **In present scenario when the term fraud comes into a discussion, credit card fraud clicks to mind so far. With the great increase in credit card transactions, credit card fraud has increasing excessively in recent years. Fraud detection includes monitoring of the spending behavior of users/ customers in order to determination, detection, or avoidance of undesirable behavior. As credit card becomes the most prevailing mode of payment for both online as well as regular purchase, fraud relate with it are also accelerating. Fraud detection is concerned with not only capturing the fraudulent events, but also capturing of such activities as quickly as possible. The use of credit cards is common in modern day society. Fraud is a millions dollar business and it is rising every year. Fraud presents significant cost to our economy worldwide. Modern techniques based on Data mining, Machine learning, Sequence Alignment, Fuzzy Logic, Genetic Programming, Artificial Intelligence etc., has been introduced for detecting credit card fraudulent transactions. This paper shows how data mining techniques can be combined successfully to obtain a high fraud coverage combined with a low or high false alarm rate. [6]**

Keywords—decision tree; fuzzy Darwinian system; genetic algorithm; artificial neural network;hidden markov model

I. INTRODUCTION

Financial fraud can be defined as an intentional act of deception. Intention could be anything like getting profit in terms of money, land etc. There are several frauds come under financial fraud like bank fraud, insurance fraud, securities and commodities fraud and other related financial fraud like mass marketing and corporate fraud [1]. Credit card fraud is a part of bank fraud. Credit card fraud can be defined as when someone uses credit card of another one for personal reasons while the owner of the card and the card issuer are not aware of the fact that the card is being used. Further, the individual using the card has no connection with the cardholder or issuer, and has no intention of either contacting the owner of the card or making repayments for the purchases made [6]. There are several ways by which credit card fraud can take place for example counterfeiting which is related to online transaction in which customer and

Merchant are in different physical location and hence verification of signature cannot be done so it may lead to an unusual transaction because merchant does not know that whether the customer providing the credit card information is indeed the authorized cardholder or a fraudster.

Another example is phishing which has been very prevalent method to steal information of a credit card holder that means fraudsters keep sending spam emails to credit card holders by putting a bank link to make their mails look like a real bank mail and ask them to provide their information by giving one simple reason like database crash so they need information for recovery purpose. After getting the information fraudsters use it for buying some stuff. Likewise so many other techniques are there by which credit card fraud takes place. [11]

The main aims are, firstly, to identify the different types of credit card fraud, and, secondly, to review alternative techniques that have been used in fraud detection. The focus here is in Europe, and so ethical issues arising from other cultures are not taken into account but for a discussion of these the reader is referred to Chepaitis (1997) and Gichure (2000). Indeed, transaction products, including credit cards, are the most vulnerable to fraud [1] [3].

In recent years so many countries have been affected by credit card frauds due to which not only common people but also financial institution and corporate sectors are getting affected by losing their money. A similar statistics of annual global fraud losses is given in Table I.

It can be clearly observed that credit card frauds have been increasing year by year. Hence credit card frauds need to be detected at reasonable time so that fraudsters could be prevented from committing illegal activities. An alternative approach could be trying to make use of general purpose Meta heuristic approaches like genetic algorithms.

Genetic algorithms are evolutionary algorithms which aim at obtaining better solutions as time progresses, they have been successfully applied to many problem domains from astronomy to sports, from optimization to computer science etc.

In this study, we try to solve our classification problem by using only a genetic algorithm solution.

TABLE I: Annual Global Fraud Losses (Credit & Debit Cards)

Year	Amount (in Billions)
2000	$2.7
2001	$3.1
2002	$3.1
2003	$3.6
2004	$4.2
2005	$4.3
2006	$4.8
2007	$5.5
2008	$6.4
2009	$6.9
2010	$7.6
2011	$9.8
2012	$11.2

II. Literature Study Of Credit Card Fraud

Credit card fraud is an important and interesting work of research technology. Several techniques has been develop for credit card fraud detecting system in online transaction, such techniques base on artificial intelligence, fuzzy logic, data mining, machine learning, genetic algorithm, decision tree, Bayesian network, neural network, clustering algorithm, etc., that evaluate of various credit card fraud transaction.

Ghosh and Reilly [2] have proposed a neural network Method to detect credit card fraud transaction. They have built a detection system, which is trained on a large sample of labeled credit card account transactions. These sample

Contain example fraud cases due to lost cards, stealing cards, application fraud, stolen card details, counterfeit fraud etc.

Kokkinaki and other have proposed the technics of decision tree. This technics of decision tree are simple and easy to the implementation, decision trees is reduces misclassification of incoming transaction of data, but this is not for use dynamically adaptive of online transaction. [13]

Meas, Suggest of fraud detection technics using the bayesian network, in this technics, improving the fraud detecting by removing highly correlated attribute, ANN was found the credit card fraud predication faster of the testing phase, at using transaction profile. Bayesian algorithm is performed better result of fraud detection only on neural network. [6][9]

Chan and Stalfo, have proposed the a technics of multiclassifier meta learning issues of credit card transaction, it detecting the fraud detection 46 % improving of overall fraud.

Kim, method improving number fraud detection classifier and compare only on the neural network by using the unsupervised algorithm of data mining.

Bolton and Hand et al. [12] it has proposed credit card Detection using unsupervised method by frequency of Transactions and observing abnormal spending behavior.

Hidden Markov Model is one of the best methods for observation spending profile generate at the state transaction. HMM is statically model for best engineering practice. Hidden markov model is best for using the FDS (fraud Detection system).

III. Various Techniques Of Credit Card Fraud In Detailed

a. *Neural networks*:

Neural network is defined as a set of interconnected nodes designed to represent functioning of the human brain. Each node has a weighted connection to several other linked nodes in adjacent layers. Single node take input received from linked nodes and use the weights of the connected nodes together with easy function for computation of output values. Neural networks can be created for supervised and/or unsupervised learning. The user specifies the number of hidden layers along with the number of nodes within a specific hidden layer. The output layer of the neural network may contain one or several nodes depending upon the application. Recently, neural network researchers have several associated methods from statistics and numerical analysis into their networks. [9][10]

Neural networks can learn and summarizes the internal assumptions of data even without knowledge of the potential data principles in advance. According to **Rumelhart,** (1986), Neural networks topologies, or architectures, formed by organizing nodes into layers and attach layers of neurons with modified weighted interconnections And it can match its own behavior to the new environment along with the results of formation of evolution capability from present environment to the new possible situation. Statistical methods are sometime unusual

in the practice research even though the common advantages of the neural networks in application of credit card fraud detection. On the other side, there are still many disadvantages for the neural networks, such as

(1) Difficulty to confirm the structure,

(2) Excessive training,

(3) Efficiency of training and so on.

b. Decision Tree:

After introducing the concept of learning system, decision tree method has been developed, that can deals with continuous data. The decision tree is a table of tree shape with connecting lines to available nodes. Each node is either a branch node followed with more nodes or only one leaf node assigned by classification. With this strategic approach of separating and resolving, decision tree usually detach the complex problem into many simple ones and resolves the sub-problems through repeatedly using, data mining method to discover training various kinds of classifying knowledge by constructing decision tree. The basis of decision tree model is how to construct a decision tree with high precision and small scale.[13]

There are many advantages of Decision tree method.

1. High flexibility
2. Good haleness
3. It is explainable, which is also the reason of its varied utilization.

Its disadvantages is that, it requires to check each transaction one by one.

c. Fuzzy Darwinian System :

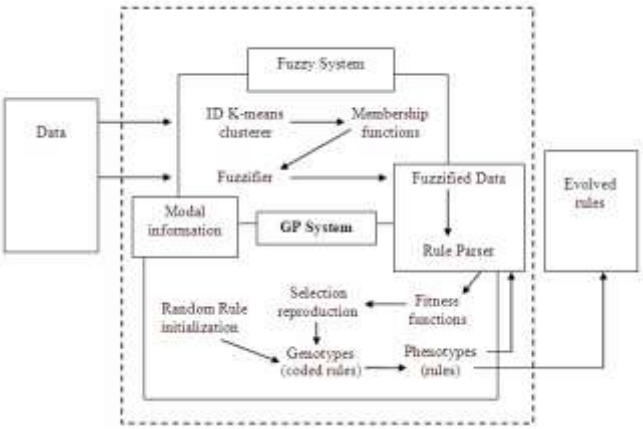
This technique [12] uses genetic programming to evolve fuzzy logic rules capable of classifying credit card transactions into "suspicious" and non-suspicious classes. It describes the use of an evolutionary-fuzzy system capable of classifying suspicious and non-suspicious credit card transactions. The system developed comprises two main elements: a Genetic Programming (GP) search algorithm and a fuzzy expert system.

When the data is provided to the FDS system, the system first clusters the data into three groups namely low, medium and high (fuzzy clustering). The genotypes and phenotypes of the GP System consist of rules which match the incoming sequence with the past sequence. Genetic Programming is used to evolve a series of variable-length fuzzy rules which characterize the differences between classes of data held in a database. The system is being developed with the specific aim of insurance-fraud detection which involves the challenging task of classifying data into the categories: safe and suspicious.

For classification of transactions, when the customer's payment is not overdue or the number of overdue payment is less than three months, the transaction is considered as "non suspicious, otherwise it is considered as suspicious.

The Fuzzy Darwinian detects suspicious and non - suspicious data and it easily detects stolen credit card Frauds. This system has very high accuracy and produces a low false alarm in comparison with other techniques, but it is highly expensive [1]. The speed of the system is low.

The complete system is capable of attaining good accuracy and intelligibility levels for real data. It has very high accuracy and produces a low false alarm, but it is not applicable in online transactions and it is highly expensive. The processing speed of the system is low.



Fiigure1: Fuzzy Darwinian System[4]

d. Hidden Markov Model :

A Hidden Markov Model is a double embedded stochastic process which is used to model much more complicated stochastic processes as compared to a traditional Markov model. If an incoming credit card transaction is not accepted by the trained Hidden Markov Model with sufficiently high probability, it is considered to be fraudulent transactions. A Hidden Markov Model [8] is initially trained with the normal behavior of a cardholder. It works on the user spending profiles which can be divided into three types such as

1. Lower profile;
2. Middle profile; and
3. Higher profile

For every credit card, the spending profile is different, so it can figure out an inconsistency of user profile and try to find fraudulent transaction. It tries to find any anomaly in the transaction based on the spending profile of the cardholder, shipping address, and billing address, etc. Every user is represented by specific patterns of set which containing information about last 10 transaction using credit card [1][8]. The set of information contains spending profile of card holder, money spent in every transaction, the last purchase time, category of purchase etc. The potential threat for fraud detection will be a deviation from set of patterns.

In the process of HMM each incoming transaction is submitted to the FDS for verification. FDS receives the card details and the value of purchase to verify whether the transaction is genuine or not. If the FDS confirms the transaction to be malicious, it raises an alarm and the issuing bank declines the transaction. The concerned cardholder may then be contacted and alerted about the possibility that the card is misused. HMM never check the original user as it maintains a log. The log which is maintained will also be a proof for the bank for the transaction made. HMM reduces the tedious work of an employee in bank since it maintains a log. HMM produces high false alarm as well as high false positive.

e. Genetic Algorithm (GA) :

Genetic algorithms, inspired from natural evolution were first introduced by Holland (1975). Genetic algorithm (GA) is a search technique used in computing to find exact or approximate solutions to optimization and search problems.

GA is used in data mining mainly for variable selection [7] and is mostly coupled with other DM algorithms [1]. And their combination with other techniques has a very good performance.

They have been used in a number of applications in engineering and social science. Recently, they applied for optimization of the parameters of support vector machine for predicting bankruptcy [7], and hybrid with neural net work for detecting credit card fraud with high accuracy [1], and have been used along with Artificial Immune System for reducing a number of false alarm in credit card fraud detection. GA has been used in credit card fraud detection for minimizing the wrongly classified number of transactions. And is easy accessible for computer programming language implementation, thus, make it strong in credit card fraud detection. But this method has high performance and is quite expensive.

However, sometimes some random mutations can occur on individuals which in tum increase the diversity in the population. The GA as a solution procedure is depicted in Figure 1. It starts with a number of initial solutions which act as the parents of the current generation. New solutions are generated from these solutions by the cross-over and mutation operators. The less fit members of this generation are eliminated and the fitter members are selected as the parents for the next generation. This procedure is repeated until a pre-specifie number of generations have passed, and the best solution found until then is selected [6], [7].

IV. CONCLUSION AND FUTURE WORK

In this paper, we present a comparative study of fraud detection methods based on credit card. The main objective of this paper is to review methodology of different detection methods based on credit card. We have considered the most important parameter in different methods such as, accuracy, speed and cost. Comparison table was prepared in order to compare various credit card fraud detection mechanisms. All the techniques of credit card fraud detection described in the table1 have its own strengths and weaknesses. We found these result is mentioned in following table from the references that we have mentioned in end as the results show, the fraud detection systems based on Fuzzy Darwinian, has a very high accuracy with 100% true positive but with very low processing speed. In another view, HMM has a fast processing speed with low accuracy. At the same time, the processing speed in decision tree is very fast enough to enable detection of credit card fraud. [5]

All these techniques of credit card fraud detection discussed in this survey paper, have its own weaknesses as well as strengths. Thus, this survey enables us to build a hybrid approach for developing some effective algorithms which can perform well for the classification problem with variable misclassification costs and with higher accuracy.

References

[1] A. Shen, R. Tong, and Y. Deng, "Application of classification models on credit card fraud detection," June 2007.

[2] Ghosh and D.L. Reilly, "Credit Card Fraud Detection with a Neural-Network," Proc. 27th Hawaii International Conference on System Sciences: Information Systems: Decision Support and Knowledge-Based Systems, vol. 3, pp. 621-630, 1994.

[3] Md Delwar Hussain Mahdi, Karim Mohammed Rezaul, Muhammad Azizur Rahman "Credit Fraud Detection in the Banking Sector in UK: A Focus on E-Business**."** 2010.

[4] Amlan Kundu, Suvasini Panigrahi, Shamik Sural and Arun K. Majumdar, "BLAST-SSAHA Hybridization for Credit Card Fraud Detection," *IEEE Transactions On Dependable And Secure Computing*, vol. 6, Issue no. 4, pp.309-315, October-December 2009S.

[5] M.F. Gadi, X. Wang, and A.P. Lago, "Comparison with parametric optimization in credit card fraud detection, 2008.

[6] Tej Paul Bhatla, Vikram Prabhu & Amit Dua "Understanding Credit Card Frauds," 2003.

[7] Blickle, T., & Thiele, L. (1995). A Comparison of Selection Schemes used in Genetic Algorithms (Vol. 2). Zurich: Swiss Federal Institute.

[8] Abhinav Srivastava, Amlan Kundu, Shamik Sural, Arun K. Majumdar. "Credit Card Fraud Detection using Hidden Markov Model". IEEE Transactions on dependable and secure computing,Volume 5; (2008) (37-48).

[9] A. Vellidoa, P.J.G. Lisboaa, J. Vaughan "Neural networks in business: a survey of applications". Elsevier, Expert Systems with Applications, (1999).

[10] P.K. Chan, W. Fan, A.L. Prodromidis, S.J. Stolfo "Distributed Data Mining in Credit Card Fraud Detection". Data Mining; (1999).