Review Paper on Credit Card Fraud Detection

[1]Suman
Research Scholar, GJUS&T Hisar
HCE Sonepat

[2] Nutan
Mtech.CSE ,HCE Sonepat

Abstract

Due to the theatrical increase of fraud which results in loss of dollars worldwide each year, several modern techniques in detecting fraud are persistently evolved and applied to many business fields. Fraud detection involves monitoring the activities of populations of users in order to estimate, perceive or avoid undesirable behavior. Undesirable behavior is a broad term including delinquency, fraud, intrusion, and account defaulting. This paper presents a survey of current techniques used in credit card fraud detection and telecommunication fraud. The goal of this paper is to provide a comprehensive review of different techniques to detect fraud.

Keywords: Fraud detection, data mining, support vector machine, anomalies.

Introduction

Credit card fraud can be defined as “Unauthorized account activity by a person for which the account was not intended. Operationally, this is an event for which action can be taken to stop the abuse in progress and incorporate risk management practices to protect against similar actions in the future”. In simple terms, Credit Card Fraud is defined as when an individual uses another individual’s credit card 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. And the persons using the card has not at all having the connection with the cardholder or the issuer and has no intention of making the repayments for the purchase they done. Fraud detection involves identifying Fraud as quickly as possible once it has been perpetrated. Fraud detection methods are continuously developed to defend criminals in adapting to their strategies. The development of new fraud detection methods is made more difficult due to the severe limitation of the exchange of ideas in fraud detection. Data sets are not made available and results are often not disclosed to the public. The fraud cases have to be detected from the available huge data sets such as the logged data and user behavior. At present, fraud detection has been implemented by a number of methods such as data mining, statistics, and artificial intelligence. Fraud is discovered from anomalies in data and patterns. The different types of methods for committing credit card frauds are described below.

Types of Frauds :Various types of frauds in this paper include credit card frauds, telecommunication frauds, and computer intrusions, Bankruptcy fraud, Theft fraud/counterfeit fraud, Application fraud, Behavioral fraud [2].

Credit Card Fraud*:* Credit card fraud has been divided into two types:

(1) Offline fraud and On-line fraud. Offline fraud is committed by using a stolen physical card at call center or any other place

(2) . On-line fraud is committed via internet, phone, shopping, web, or in absence of card holder.

Telecommunication Fraud*:* The use of telecommunication services to commit other forms of fraud. Consumers, businesses and communication service provider are the victims.

Computer Intrusion: Intrusion Is Defined As The act of entering without warrant or invitation; That means "potential possibility of unauthorized attempt to access Information, Manipulate Information Purposefully. Intruders may be from any environment, An outsider (Or Hacker) and an insider who knows the layout of the system [1].

Bankruptcy Fraud*:* This column focuses on bankruptcy fraud. Bankruptcy fraud means using a credit card while being absent. Bankruptcy fraud is one of the most complicated types of fraud to predict [1].

Theft Fraud/ Counterfeit Fraud: In this section, we focus on theft and counterfeit fraud, which are related to one other. Theft fraud refers using a card that is not yours. As soon as the owner give some feedback and contact the bank, the bank will take measures to check the thief as early as possible. Likewise, counterfeit fraud occurs when the credit card is used remotely; where only the credit card details are needed [2].

Application Fraud*:* When someone applies for a credit card with false information that is termed as application fraud. For detecting application fraud, two different situations have to be classified. When applications come from a same user with the same details, that is called duplicates, and when applications come from different individuals with similar details, that is termed as identity fraudsters. Phua et al. [3] describes application fraud as "demonstration of identity crime, occurs when application forms contain possible, and synthetic (identity fraud), or real but also stolen identity information (identity theft)."

Credit card fraud detection methods

On doing the literature survey of various methods for fraud detection we come to the conclusion that to detect credit card fraud there are multiple approaches like [11] [2].

- Gass algorithm
- Bayesian networks
- Hidden markov model
- Genetic algorithm
- A fusion approach using dempster-shafer theory and bayesian learning.
- Decision tree
- Neural network
- Logistic Regression

Gass algorithm

This algorithm is a combination of genetic algorithm and scatter search [11]. In this section, we first describe the basic operating principles of genetic algorithms and scatter search and then explain the steps of the suggested GASS algorithm. Genetic algorithms are inspired from natural evolution. The basic idea is that the survival chance of stronger members of a population is larger than that of the weaker members and as the generations evolve the average fitness of the population gets better. Normally the new generations will be produced by the crossover of two parent members. However, sometimes some random mutations can also occur on individuals which in turn increase the diversity in the population. 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-specified number of generations have passed, and the best solution found until then is selected. The SS is another evolutionary algorithm which shares some common characteristics with the GA. It operates on a set of solutions, the reference set, by combining these solutions to create new ones. The main mechanism for combining solutions is such that a new solution is created from the linear combination of two other solutions [21].In SS diversity in the reference set is very important and next time it will be determined again first a number of best solutions are selected and then these are coupled with a number of most diverse solutions to form the new reference set. Unlike the population of the GA, the reference set of SS is usually kept smaller as each solution in it is desired to be subjected to the recombination operator.

The suggested GASS algorithm basically follows the steps of GA but it has some components from SS. As compared to typical GA implementations we kept the size of the population smaller and we made sure some minimum level of diversity is attained at each generation. Also, for the reproduction we used a standard recombination operator rather than the classical cross-over operator of GA. We also used the mutation operator which is common to both GA and SS implementations. The steps of the GASS and the parameter values used are detailed below:

(a) Number of parent solutions (size of the reference set): Number of starting solutions which also equal to the number of parents selected for each generation is an important parameter which can influence the convergence speed of the procedure. The population size is determined according to the size of the problem, i.e. bigger population for larger problem. We have taken this to be 50 where three of them are determined

as to be the solutions which will generate the maximum number of alerts (MAX), the one that will generate the minimum number of alerts (MIN) and the one currently used in the production (PRD). The remaining 47 solutions are obtained by producing random numbers for all 43 parameters. Note that, MAX and MIN bring a certain level of diversity to the reference set.

(b) Number of children: For the ease of implementation we decided to recombine every possible pair of parents and this way we obtained 1225 children in each generation.

(c) Reproduction (recombination): We took the weighted average of the

parameter values of the two parent solutions and obtained the child solution. For each generation a random number between zero and one is determined and this number is used as the weight of the first parent in all recombination operators used in that generation. The weight of the second parent is equal to one minus the determined random number. This type of reproduction is not common in GA implementations but it can be regarded as a typical operator for SS.

(d) Mutation operator: One of the 43 parameters is picked up randomly and its value is changed randomly within its allowable range.

(e) Recombination and mutation probabilities: All children are generated by the recombination operator. Then, one of the children solutions is randomly picked up and mutation operator is applied to it.

(f) Fitness function: As described above, the fitness value of an individual solution is determined as the total amount of savings incurred from fraud losses.

(g) Selection: The best three members of the generation are automatically selected. To keep having diversity in all generations, the three named solutions, MAX, MIN and PRD are also automatically transferred to the next generation. The remaining 44 solutions are determined by the roulette selection procedure.

(h) Termination criterion: We decided to run the generations until no improvements are observed for at least 10 generations.

Bayesian networks

For the purpose of fraud detection, two Bayesian networks to describe the behavior of user are constructed. First, a Bayesian network is constructed to model behavior under the assumption that the user is fraudulent (F) and another model under the assumption the user is a legitimate (NF). The 'fraud net' is set up by using expert knowledge. The 'user net' is set up by using data from non fraudulent users. During operation user net is adapted to a specific user based on emerging data. By inserting evidence in these networks and propagating it through the network, the probability of the measurement x less than two above mentioned hypotheses is obtained. This means, it gives judgments to what degree observed user behavior meets typical fraudulent or non fraudulent behavior. These quantities we call p(X│NF) and p (X│F).

By postulating the probability of fraud P (F) and P (NF) =1-P(F) in general and by applying Bayes' rule, it gives the probability of fraud, given the measurement x,

$$P(F|X) = \frac{P(F)P(X|F)}{P(F|X)} \qquad \text{----- - (1)}$$

Where the denominator p(x) can be calculated as

$$P(x) = P\ (F)\ p\ (X│F) + P\ (NF)\ p\ (X│NF) \text{ --------- (2)}$$

The fraud probability P (F│X) given the observed user behavior x can be used as an alarm level. On the one hand, Bayesian networks allow the integration of expert knowledge, which we used to initially set up the models [4]. On the other hand, the user model is retrained in an unsupervised way using data. Thus our Bayesian approach incorporates both, expert knowledge and learning.

Hidden markov model

A Hidden Markov Model is a double embedded stochastic process with 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. HMM[5], Baum Welch algorithm is used for training purpose and K-means algorithm for clustering.HMM sores data in the form of clusters depending on three price value ranges low, medium and high[6].

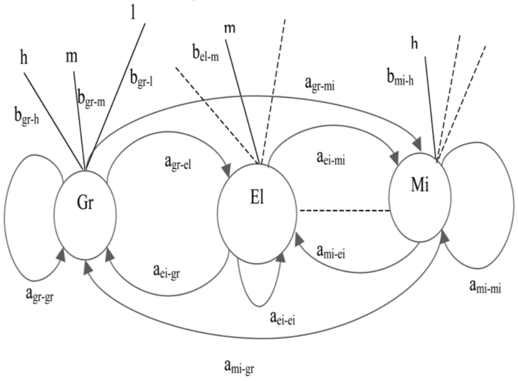

Fig.2: HMM for credit card fraud detection

The probabilities of initial set of transaction have chosen and FDS checks whether transaction is genuine or fraudulent. Since HMM maintains a log for transactions it reduces tedious work of employee but produces high false alarm as well as high false positive[7]. The initial choice of parameters affects the performance of this algorithm and, hence, they should be chosen carefully. We consider the special case of fully connected HMM in which every state of the model can be reached in a single step from every other state, as shown in Fig. 2. Gr, El, Mi, etc., are names given to the states to denote purchase types like Groceries, Electronic items, and miscellaneous purchases. Spending profiles of the individual cardholders are used to obtain an initial estimate for probability matrix B

Genetic algorithm

Genetic algorithms, inspired from natural evolution were first introduced by Holland (1975). Genetic algorithms are evolutionary algorithms which aim at obtaining better solutions as time progresses**.** Fraud detection problem is classification problem, in which some of statistical methods many data mining algorithms have proposed to solve it. Among decision trees are more popular. Fraud detection has been usually in domain of E-commerce, data mining [8].

GA is used in data mining mainly for variable selection [9] and is mostly coupled with other DM algorithms [10]. And their combination with other techniques has a very good performance. GA has been used in credit card fraud detection for minimizing the wrongly classified number of transactions [10]. 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.

A fusion approach using Dempster-Shafer theory and Bayesian learning.

As mentioned in [19] First approach i.e. Dempster-Shafer Theory basically proposes Fraud Detection System using information fusion and Bayesian learning in which evidences from current as well as past behavior are combined together and depending on certain type shopping behavior establishes an activity profile for every cardholder. It has advantages like: - high accuracy, processing speed, reduces false alarm, improves detection rate, applicable in E-commerce. But one disadvantage of this approach is that it is highly expensive. Dempster–Shafer theory and Bayesian learning is a hybrid approach for credit card fraud detection [18][11] which combines evidences from current as well as past behavior. Every cardholder has a certain type of shopping behavior, which establishes an activity profile for them. This approach proposes a fraud detection system using information fusion and Bayesian learning of so as to counter credit card fraud.

The FDS system consists of four components, namely, rule-based filter, Dempster–Shafer adder, transaction history database and Bayesian learner. In the rule-based component, the suspicion level of each incoming transaction based on the extent of its deviation from good pattern is determined. Dempster–Shafer’s theory is used to combine multiple such evidences and an initial belief is computed [20]. Then the initial belief values are combined to obtain an overall belief by applying Dempster–Shafer theory. The transaction is classified as suspicious or suspicious depending on this initial belief. Once a transaction is found to be suspicious, belief is further strengthened or weakened according to its similarity with fraudulent or genuine transaction

Decision tree

Decision trees are statistical data mining technique that express independent attributes and a dependent attributes logically AND in a tree shaped structure. Classification rules, extracted from decision trees, are IF-THEN expressions and all the tests have to succeed if each rule is to be generated [11]. Decision tree usually separates the complex problem into many simple ones and resolves the sub problems through repeatedly using [11][12]. Decision trees are predictive decision support tools that create mapping from observations to possible consequences. There are number of popular classifiers construct decision trees to generate class models.

Decision tree methods C5.0,C&RT and CHAID. The work demonstrates the advantages of applying the data mining techniques including decision trees and SVMs to the credit card fraud detection problem for the purpose of reducing the bank's risk. The results show that the proposed classifiers of C&RT and other decision tree approaches outperform SVM approaches in solving the problem under investigation.

Neural network

Fraud detection methods based on neural network are the most popular ones. An artificial neural network [13][14] consists of an interconnected group of artificial neurons .The principle of neural network is motivated by the functions of the brain especially pattern recognition and associative memory [15]. The neural network recognizes similar patterns, predicts future values or events based upon the associative memory of the patterns it was learned. It is widely applied in classification and clustering. The advantages of neural networks over other techniques are that these models are able to learn from the past and thus, improve results as time passes. They can also extract rules and predict future activity based on the current situation. By employing neural networks, effectively, banks can detect fraudulent use of a card, faster and more efficiently. Among the reported credit card fraud studies most have focused on using neural networks. In more practical terms neural networks are non-linear statistical data modeling tools. They can be used to model complex relationships between inputs and outputs or to find patterns in data.

There are two phases in neural network [16] training and recognition. Learning in a neural network is called training. There are two types of NN training methods supervised and unsupervised. In supervised training, samples of both fraudulent and non fraudulent records are used to create models. In contrast, unsupervised training simply seeks those transactions, which are most dissimilar from the norm. On other hand, the unsupervised techniques do not need the previous knowledge of fraudulent and non fraudulent transactions in database. NNs can produce best result for only large transaction dataset. And they need a long training dataset.

Logistic regression

Two advanced data mining approaches, support vector machines and random forests, together with the well known logistic regression [18], as part of an attempt to better detect (and thus control and prosecute) credit card fraud. The study is based on real-life data of transactions from an international credit card operation. It is well-understood, easy to use, and remains one of the most commonly used for data-mining in practice. It thus provides a useful baseline for comparing performance of newer methods. Supervised learning methods for fraud detection face two challenges. The first is of unbalanced class sizes of legitimate and fraudulent transactions, with legitimate transactions far outnumbering fraudulent ones. For model development, some form of sampling among the two classes is typically used to obtain training data with reasonable class distributions. Various sampling approaches have been proposed in the literature, with random oversampling of minority class cases and random under sampling of majority class cases being the simplest and most common in use; others include directed sampling The second problem in developing supervised models for fraud can arise from potentially undetected fraud transactions, leading to mislabeled cases in the data to be used for building the model. For the purpose of this study, fraudulent transactions are those specifically identified by the institutional auditors as those that caused an unlawful transfer of funds from the bank sponsoring the credit cards. These transactions were observed to be fraudulent ex post. Our study is based on real-life data of transactions from an international credit card operation. The transaction data is aggregated to create various derived attributes.

Support vector machine

The basic idea of SVM classification algorithm is to construct a hyper plane as the decision plane which making the distance between the positive and negative mode maximum [17]. The strength of SVMs comes from two important properties they possess - kernel representation and margin optimization. Kernels, such as radial basis function (RBF) kernel, can be used to learn complex regions. A kernel function represents the dot product of projections of two data points in a high dimensional feature space. In SVMs, the classification function is a hyper-plane separating the different classes of data. The basic technique finds the smallest hyper sphere in the kernel space that contains all training instances, and then determines on which side of hyper sphere a test instance lies. If a test instance lies outside the hyper sphere, it is confirmed to be suspicion. SVM can have better prediction performance than BPN(Back propagation network) in predicting the future data.

SVMs are set of related supervised learning methods used for classification and regression they belong to a family of generalized linear classification. A special property of SVM is, SVM Simultaneously minimize the empirical classification error and maximize the geometric margin. So SVM called Maximum Margin Classifiers. SVM is based on the Structural risk Minimization (SRM). SVM map input vector to a higher dimensional space where a maximal separating hyper plane is constructed. Two parallel hyper planes are constructed on each side of the hyper plane that separate the data. The separating hyper plane is the hyper planes that maximize the distance between the two parallel hyper planes. An assumption is made that the larger the margin or distance between these parallel hyper planes the better the generalization error of the classifier will be .We consider data points of the form $\{(X_1, Y_1), (X_2, Y_2), (X_3\ Y_3), (X_4,Y_4)\ldots\ldots\ldots, (X_n, Y_n)\}$.

Where Y_n=1 / -1, a constant denoting the class to which that point Xn belongs.
n = number of sample. Each X_n is P -dimensional real vector. The scaling is important to guard against variable (attributes) with larger variances. We can view this training data, by means of the dividing hyper plane, which takes
W. X + b = O ----- (1) Where b is scalar and W is p-dimensional Vector. The vector W points perpendicular to the separating hyper plane. Adding the offset parameter b allows us to increase the margin. Absent of b, the hyper plane is forced to pass through the origin, restricting the solution. As we are interesting in the maximum margin, we are interested SVM and the parallel hyper planes [11]. Parallel hyper planes can be described by equation

$$W.X + b = 1$$

$$W.X + b = -1$$

If the training data are linearly separable, we can select these hyper planes so that there are no points between them and then try to maximize their distance.

Random forests

The popularity of decision tree models in data mining arises from their ease of use, flexibility in terms of handling various data attribute types, and interpretability. Single tree models, however, can be unstable and overly sensitive to specific training data. Ensemble methods seek to address this problem by developing a set of models and aggregating their predictions in determining the class label for a data point. A random forest model is an ensemble of classification (or regression) trees. Ensembles perform well when individual members are dissimilar, and random forests obtain variation among individual trees using two sources for randomness: first, each tree is built on separate bootstrapped samples of the training data; secondly, only a randomly selected subset of data attributes is considered at each node in building the individual trees. Random forests thus combine the concepts of bagging, where individual models in an ensemble are developed through sampling with replacement from the training data, and the random subspace method, where each tree in an ensemble is built from a random subset of attributes. Given a training data set of N cases described by B attributes, each tree in the ensemble is developed as follows:

- Obtain a bootstrap sample of N cases
- At each node, randomly select a subset of bbB attributes. Determine the best split at the node from this reduced set of b attributes
- Grow the full tree without pruning

Random forests are computationally efficient since each tree is built independently of the others. With large number of trees in the ensemble, they are also noted to be robust to over fitting and noise in the data.

Conclusion

Credit card fraud has become more and more rampant in recent years. To improve merchants' risk management level in an automatic and effective way, building an accurate and easy handling credit card risk monitoring system is one of the key tasks for the merchant banks. One aim of this study is to identify the user model that best identifies fraud cases. There are many ways of detection of credit card fraud. If one of these or combination of algorithm is applied into bank credit card fraud detection system, the probability of fraud transactions can be predicted soon after credit card transactions by the banks. And a series of anti-fraud strategies can be adopted to prevent banks from great losses before and reduce risks. This paper gives contribution towards the effective ways of credit card fraudulent detection.

References

[1] Linda Delamaire (UK), Hussein Abdou (UK), John Pointon (UK), "Credit card fraud and detection techniques: a review", Banks and Bank Systems, Volume 4, Issue 2, 2009 .

[2] Khyati Chaudhary, Jyoti Yadav, Bhawna Mallick, "A review of Fraud Detection Techniques: Credit Card", International Journal of Computer Applications (0975 – 8887) Volume 45– No.1, May 2012 .

[3] Vladimir Zaslavsky and Anna Strizhak," credit card fraud detection using selforganizing maps", information & security. An International Journal, Vol.18,2006.

[4] L. Mukhanov, "Using bayesian belief networks for credit card fraud detection," in Proc. of the IASTED International conference on Artificial Intelligence and Applications, Insbruck, Austria, Feb. 2008, pp. 221– 225.

[5] Abhinav Srivastava, Amlan Kundu, Shamik Sural and Arun K. Majumdar, "CreditCard Fraud Detection Using Hidden Markov Model" IEEE, Transactions On Dependable And Secure Computing, Vol. 5, No 1. , January-March 2008

[6] V. Bhusari, and S. Patil, "Study of Hidden Markov Model in Credit Card Fraudulent Detection", International Journal of Computer Applications (0975 – 8887) Volume 20– No.5, April 2011

[7] V.Bhusari ,S.Patil ," Study of Hidden Markov Model in Credit Card Fraudulent Detection ",International Journal of Computer Applications (0975 - 8887) Volume 20- No.5, April 2011

[8] K.RamaKalyani, D.UmaDevi, "Fraud Detection of Credit Card Payment System by Genetic Algorithm", International Journal of Scientific & Engineering Research Volume 3, Issue 7, July-2012

[9] Bidgoli, B. M., Kashy, D., Kortemeyer, G. & Punch, W. F "Predicting student performance: An Application of data mining methods with the educational web-based system LON-CAPA". In Proceedings of ASEE/IEEE frontiers in education conference. . (2003).

[10] Ekrem Duman, M. Hamdi Ozcelik "Detecting credit card fraud by genetic algorithm and scatter search". Elsevier, Expert Systems with Applications, (2011). 38; (13057–13063).

[11] S. Benson Edwin Raj, A. Annie Portia, "Analysis on Credit Card Fraud Detection Methods", International Conference on Computer, Communication and Electrical Technology – ICCCET2011, 18th & 19th March, 2011

[12] Y. Sahin and E. Duman, "Detecting Credit Card Fraud by Decision Trees and Support Vector Machines", International Multiconference of Engineers and computer scientists March, 2011.

[13] S. Benson Edwin Raj, A. Annie Portia "Analysis on Credit Card Fraud Detection Methods". IEEE-International Conference on Computer, Communication and Electrical Technology; (2011). (152-156).

[14] Ray-I Chang, Liang-Bin Lai, Wen-De Su, Jen-Chieh Wang, Jen-Shiang Kouh "Intrusion Detection by Backpropagation Neural Networks with Sample-Query and Attribute-Query". Research India Publications; (2006). (6-10).

[15] Raghavendra Patidar, Lokesh Sharma "Credit Card Fraud Detection Using Neural Network". International Journal of Soft Computing and Engineering (IJSCE), (2011). Volume-1, Issue; (32-38).

[16] Tao Guo, Gui-Yang Li "Neural Data Mining For Credit Card Fraud Detection". IEEE, Proceedings of the Seventh International Conference on Machine Learning and Cybernetics; (2008). (3630-3634).

[17] Joseph King-Fung Pun,"Improving Credit Card Fraud Detection using a Meta-Learning Strategy", Chemical Engineering and Applied Chemistry University of Toronto 2011

[18] Siddhartha Bhattacharyya, Sanjeev Jha, Kurian Tharakunnel, J. Christopher Westland, "Data mining for credit card fraud: A comparative study", Decision Support Systems 50 pp. 602–613,2011.

[19] Sandeep Pratap Singh, Shiv Shankar P.Shukla,Nitin Rakesh and Vipin Tyagi "Problem Reduction In Online Payment System Using Hybrid Model" International Journal of Managing Information Technology (IJMIT) Vol.3, No.3, August 2011

[20] Amlan Kundu, Suvasini Panigrahi, Shamik Sural and Arun K. Majumdar, "Credit card fraud detection: A fusion approach using Dempster-Shafer theory and Bayesian learning," Special Issue on Information Fusion in Computer Security, Vol. 10, Issue no 4, pp.354- 363, October 2009 .

[21] Hung, W. N. N., Song, X., Aboulhamid, E. M., & Driscoll, M. A. (2002). BDD minimization by scatter search. IEEE Transactions on Computer-Aided Design on Integrated Circuits and Systems, 21(8), 974–979.