AN IMPROVING CREDIT CARD FRAUD DETECTION USING A NOVEL DATA MINING TECHNIQUES

S.K.Saravanan[1], Dr.G.N.K. Suresh Babu[2]
[1]Research Scholar, Research and Development Centre, Bharathiar University, Coimbatore, India.
[2]Associate Professor, Department of Computer Applications,
Acharya Institute of Technology, Bangalore, India.

Abstract

Credit Card Fraud is an one of the major ethical issues in the credit card industry. 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. Due to fast growth of E-Commerce, use of credit cards for online purchases has dramatically increased and it caused an increase in the credit card fraud. As credit card has became the most popular mode of payment for online and regular purchase, frauds associated with it are rising. The sub-aim is to present, compare and analyze recently published findings in credit card fraud detection. This article defines common terms in credit card fraud and highlights key statistics and figures in this field. Depending on the type of fraud faced by banks or credit card companies, various measures can be adopted and implemented. The proposals made in this paper are likely to have beneficial attributes in terms of cost savings and time efficiency. The significance of the application of the techniques reviewed here is in the minimization of credit card fraud. Yet there are still ethical issues when genuine credit card customers are misclassified as fraudulent.

Keywords: Credit card, Fraud Detection, Security, Genetic Algorithms

I. INTRODUCTION

In the last decade, credit card fraud has started to pose a great threat to the businesses all over the worldwide and it seems to make an impact on the economy. It has become very important for business organisations to counter these credit card frauds effectively, for which understanding the credit cards is considered to be equally important.

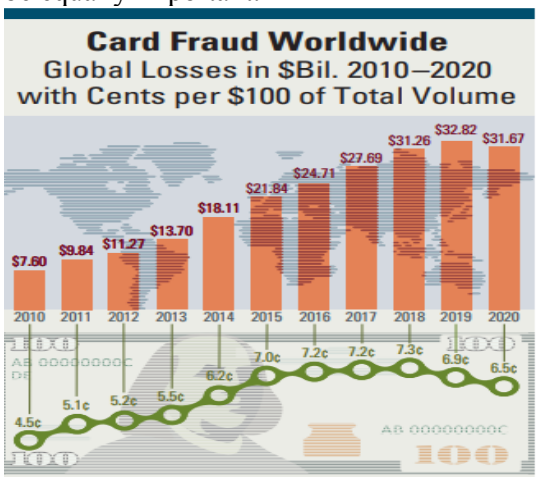

Fig.1 Credit Card Frauds Worldwide
(The Nelson Report, 2016)

Credit card frauds make a greater impact on the merchants when compared to the consumers; merchants are considered to face more risks in the credit card transactions. While consumers may face trouble trying to get a fraudulent charge reversed, merchants lose the cost of the product sold, pay chargeback fees, and fear from the risk of having their merchant account closed. Increasingly, the card not present scenario, such as shopping on the internet poses a greater threat as the merchant (the web site) is no longer protected with advantages of physical verification such as signature check, photo identification, etc. In fact, it is almost impossible to perform any of the 'physical

world' checks necessary to detect who is at the other end of the transaction. This makes the internet extremely attractive to fraud perpetrators. According to a recent survey, the rate at which internet fraud occurs is 12 to 15 times higher than 'physical world' fraud. However, recent technical developments are showing some promise to check fraud in the card not present scenario.

II. CREDIT CARD FRAUD

Joshi (2006) defines credit card fraud 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" (Baesens, Vlasselaer & Verbeke, 2015). 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 (Namdev, Kumar & Bansal, 2015). Maes, et.al. (2002) and Ogwueleka (2008) defines credit card fraud detection as the process of identifying the legitimate transactions and fraudulent transactions.

Fig.2 Credit Card Transaction Processing Steps

III. DIFFERENT TYPES OF CREDIT CARD FRAUDS

Bhatla, Prabhu & Dua (2003) classifies credit card frauds into 3 major categories such as traditional card related frauds, merchant related frauds, and internet frauds. Frauds related to traditional cards like counterfeit, application, application etc, internet frauds line generation of credit cards, fake merchant sites etc and frauds related to merchants like triangulation, merchant collusion etc are the three categories involved in credit card frauds.

Merchant Related Frauds

Merchant related frauds are initiated either by owners of the merchant establishment or their employees. The types of frauds initiated by merchants are described below:

i. Merchant Collusion: This type of fraud occurs when merchant owners or their employees conspire to commit fraud using the cardholder accounts or by using the personal information. They pass on the information about cardholders to fraudsters.

ii. Triangulation: Triangulation is a type of fraud which is done and operates from a web site. The products or goods are offered at heavily discounted rates and are also shipped before payment. The customer while browse the site and if he likes the product he place the online information such as name, address and valid credit card details to the site. When the fraudsters receive these details, they order goods from a legitimate site using stolen credit card details. The fraudsters then by using the credit card information purchase the products.

INTERNET RELATED FRAUDS

The internet is the base for the fraudsters to make the frauds in the simply and the easiest way. Fraudsters have recently begun to operate on a truly transnational level. With the expansion of trans-border, economic and political spaces, the internet has become a new worlds market, capturing consumers from most countries around the world. The below described are most commonly used techniques in Internet fraud: i. Site cloning: Site cloning is where fraudsters close an entire site or just the pages from which the customer made a purchase. Customers have no reason to believe they are not dealing with the company that they

wished to purchase goods or services from because the pages that they are viewing are identical to those of the real site. The cloned site will receive these details and send the customer a receipt of the transaction through the email just as the real company would do. The consumer suspects nothing, while the fraudsters have all the details they need to commit credit card fraud.

FALSE MERCHANT SITES

Some sites often offer a cheap service for the customers. That site requests the customer to fill his complete details such as name and address to access the webpage where the customer gets his required products. Many of these sites claim to be free, but require a valid credit card number to verify an individual s age. These kinds of sites in this way collect as many as credit card details. The sites themselves never charge individuals for the services they provide. The sites are usually part of a larger criminal network that either uses the details it collects to raise revenues or sells valid credit card details to small fraudsters.

CREDIT CARD GENERATORS

These are the computer programs that generate valid credit card numbers and expiry dates. These generators work by generating lists of credit card account numbers from a single account number. The software works by using the mathematical Luhn algorithm that card issuers use to generate other valid card number combinations. This makes the user to allow to illegally generating as many numbers as he desires, in the form of any of the credit card formats.

ERASING THE MAGNETIC STRIP

This is the type of the fraud where the fraudsters erase the magnetic stripe by using the powerful electro-magnet. The fraudster then tampers with the details on the card so that they match the details of a valid card, which they may have attained, for example, when the fraudster begins to use the card, the cashier will swipe the card through the terminal several times, before realizing that the metallic strip does not work. The cashier will then proceed to manually input the card details into the terminal. This kind of fraud is having high risk because the cashier will be looking at the card closely to read the numbers.

CREATING A FAKE CARD

Today we have sophisticated machines where one can create a fake card from using the scratch. This is the common fraud though fake cards require a lot of effort and skill to produce it. Modern cards are having many security features, all designed to make it difficult for fraudsters to make good quality fraudulent. After introducing the Holograms in the credit cards it makes very difficult to forge them effectively.

IV. LITERATURE REVIEW

Since last two decades, research on the data mining techniques for credit card fraud detection has been started; Chan, et.al. (1999) addressed the growing credit card transactions in the US payment system that is considered to be leading to greater stolen credit card accounts. In the early years of credit card usage, banks faced a huge problem in analysing massive amounts of transaction data that efficiently compute fraud detectors in a timely manner. There are also several problems associated with the skewed distributions of training data and non-uniform cost per error. Chan, et.al (1999) conducted a study to address the three most important problems associated with the credit card transactions especially in e-commerce such as scalability, efficiency and technical issues. Chan, et.al (1999) proposed a fraud detection model that has the combination of multiple fraud detectors referred as distributed data mining of models demonstrates a significant reduction in credit card frauds.

Rather than using single algorithm techniques, a second group of research studies focused on applying multiple algorithm techniques in credit card fraud detection. The most quoted research is the meta-learning technique proposed by Chan and Stolfo (1998). In their research they utilized naïve Bayesian, C4.5, CART, and RIPPER as base classifiers and combined them by implementing a stacking method. It was found that their multi-classifier meta-learning approach can significantly reduce the loss amount due to fraudulent transactions by using a 50:50 fraudulent to legitimate distribution in the datasets for training. Brause, Langsdorf, &

Hepp (1999) combined a rule-based technique with a neural network to identify fraudulent credit card transactions. It was found that this combined technique increases the probability for the diagnosis of fraud to be correct and therefore is able to decrease the number of false alarms while increasing the confidence level. Phua, Alahakoon, & Lee, (2004) proposed the use of backpropogation neural networks, naïve Bayesian, and C4.5 algorithms as base classifiers, and to combine the base classifiers' predictions using a metaclassifier technique to detect fraudulent automobile insurance claims. Duman and Ozcelik (20110 used a novel combination of the genetic algorithm and the scatter search algorithm to detect credit card fraud in a large Turkish bank. By combining these two algorithms Duman and Ozcelik were able to improve the bank's existing fraud detection strategy by 200%. The abovementioned studies show that the neural network technique is still the most widely used method in fraud detection and that multiple algorithm techniques often improve upon single algorithm techniques. However, none of these studies have looked into applying a multiple algorithm technique to a post-neural network dataset. Furthermore, previous credit card fraud detection methods lack integration with existing commercial fraud detection systems.

V. 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

- 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 (Benson, Raj, & Portia, 2011). Tthe 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 (Hung, et.al, 2002).

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 $p(X \mid NF)$ and $p(X \mid F)$.

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 (Bhusari & Patil, 2011).

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 Ecommerce, data mining (Kalyani & Devi, 2012). GA is used in data mining mainly for variable selection (Bidgoli, et.al., 2003) and is mostly coupled with other DM algorithms (Duman & Ozceli, 2011). 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. 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

Singh, et.al., (2011) addresses the first approach i.e. DempsterShafer 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.

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 (Kundu, et.al., 2009). Decision tree usually separates the complex problem into many simple ones and resolves the sub problems through repeatedly using (Sahin & Duman, 2011). 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 (Chang, et.al., 2006) 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 (Patidar & Sharma, 2011). The neural network recognizes similar patterns, predicts future values or events based upon the associative memory of the patterns it was learned.

Logistic regression

Two advanced data mining approaches, support vector machines and random forests, together with the well known logistic regression (Bhattacharyya, et.al., 2011), 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.

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 (Pun, 2011). 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.

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.

VI CONCLUSION

Data mining is a process that uses a variety of data analysis tools to discover patterns and relationships in data that may be used to make a valid prediction (Edelstien, 1999). The six basic steps of data mining process are defining the problem, preparing data, exploring data, building models, exploring and validating models, deploying and updating models. Fraud detection solutions are used in the fields like credit card, e-commerce, telecommunication, insurance etc in order to protect personal information of customers. The two main disadvantages involved in the research process of fraud detection in data mining are lack of personal information for conducting experiments and lack of well developed techniques and research methods. In order to overcome these issues proper categorization and comparison has been done using similar literature in this paper. Few innovative methods are also suggested in this paper. Every banking industry should be alerted immediately on any fraudulent activity that affects their process. Huge database is maintained in every bank. Possibilities are high for misuse of information of highly profitable organisation. Data mining is a process that uses a variety of data analysis tools to discover patterns and relationships in data that may be used to make a valid prediction (Edelstien, 1999). The six basic steps of data mining process are defining the problem, preparing data, exploring data, building models, exploring and validating models, deploying and updating models. Genetic algorithm is a novel one in this literature in terms of application domain. If this algorithms applied into bank credit card fraud detection system, the probability of fraud transaction can be predicted soon after credit card transactions. And a series of anti-fraud strategies can be adopted to prevent banks from great losses and reduce risks. The objective of the study was taken differently than the typical classification problems in that we had a variable misclassification cost. As the standard data mining algorithms does not fit well with this situation. We decided to use multi population genetic algorithm to obtain an optimized parameter.

REFERENCES

[1] Aleskerov, E.; Freisleben, B.; and Rao, B. (1997). CARDWATCH: A neural network-based database mining system for credit card fraud detection. Proceeding of the IEEE/IAFE on Computational Intelligence for Financial Engineering, 220-226.

[2]Al-Khatib, A.M. (2012) Electronic payment fraud detection techniques. World of Computer Science and Information Technology Journal. Vol.2, No.4, 137–141

[3] Baesens, B., Vlasselaer, V., & Verbeke, W. (2015) Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques: A Guide to Data Science for Fraud Detection, New Jersey: John Wiley & Sons

[4] Bhatla, T.P.; Prabhu, V.; and Dua, A. (2003). Understanding credit card frauds. Crads Business Review# 2003-1, Tata Consultancy Services.

[5] Bhattacharyya, S., Jha, S., Tharakunnel, K. & Westland, C. (2011). "Data mining for credit card fraud: A comparative study", Decision Support Systems, Vol. 50 pp. 602–613.

[6] Bhusari, V. & Patil, S. (2011) "Study of Hidden Markov Model in Credit Card Fraudulent Detection", International Journal of Computer Applications, Vol. 20, No.5.

[7] Chan, P.K. & Stolfo, S.J. (1998). Toward scalable learning with nonuniform class and cost distributions: A case study in credit card fraud detection. Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining, 164-168.

[8] Chan, P.K., Fan, W., Prodromidis, A.L., & Stolfo, S.J. (1999) Distributed data mining in credit card fraud detection, IEEE Intelligent Systems and their Applications. Vol. 14, No. 16, pp. 67 – 74

[9] Doronsoro, J.R., Ginel, F., Sgnchez, C. & Cruz, C.S. (1997) Neural fraud detection in credit card operations, IEEE Transactions on Neural Networks, Vol.8, No. 4, pp. 827 – 834

[10] Edelstien, H.A. (1999). Introduction to data mining and knowledge discovery. (2nd Ed.), Two Crows Corporation

[11] Haimowitz, I.J.; and Schwarz, H. (1997). Clustering and prediction for credit line optimization. Proceedings of AAAI-97 Workshop on AI Approaches to Fraud Detection and Risk Management, 29-33.

[12] Hanagandi, V.; Dhar, A.; and Buescher, K. (1996). Density-based clustering and radial basis function modeling to generate credit card fraud scores. Proceedings of the IEEE/IAFE 1996 Conference on Computational Intelligence for Financial Engineering (CIFEr), 247-251

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

[14] Ibrahim, L.M. (2010). Anomaly network intrusion detection system based on distributed time-delay neural network (DTDNN). Journal of Engineering Science and Technology (JESTEC), 5(4), 457-471.

[15] Joshi, M. (2006) Black Cards Forensics: Classification of the ATM and Credit Card Fraud Schemes, Pune: India Forensic Research Foundation

[16] Maes, S.; Tuyls, K.; Vanschoenwinkel, B.; and Manderick, B. (2002). Credit card fraud detection using Bayesian and neural networks. Proceeding International NAISO Congress on Neuro Fuzzy Technologies.

[17] Ogwueleka, F. N. (2008). Credit card fraud detection using data mining techniques. Ph.D. Dissertation. Department of Computer Science. Nnamdi Azikiwe University, Awka, Nigeria.

[18] Stolfo, S.J.; Fan D.W.; Lee, W.; Prodromidis, A.; and Chan, P.K. (1997). Credit card fraud detection using meta-learning: Issues and initial results. Proceedings of AAAI-97 Workshop on AI Approaches to Fraud Detection and Risk Management, 83-90.

[19] Ramakalyani, K. & Umadevi, D. (2012) "Fraud Detection of Credit Card Payment System by Genetic Algorithm", International Journal of Scientific & Engineering Research Vol. 3, No. 7.

[20] Patidar, R. & Sharma, L. (2011) "Credit Card Fraud Detection Using Neural Network". International Journal of Soft Computing and Engineering, Vol. 1, No. 32.

[21] Guo, T. & Li, G. (2008) "Neural Data Mining For Credit Card Fraud Detection". IEEE, Proceedings of the Seventh International Conference on Machine Learning and Cybernetics.