ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

Review of Proficient Approaches for Determining Identity Frauds and New Data Mining Based Method Under Investigation

Vaibhav Sirasat, Prof. Sweta Kale

M.E. Dept. of I.T., R.M.D. Sinhgad School of Engineering, Savitribai Phule Pune University, Pune, Maharashtra, India

Professor, Dept. of I.T., R.M.D. Sinhgad School of Engineering, Savitribai Phule Pune University, Pune, Maharashtra, India

ABSTRACT: Recently during study it is observed that, there are many methods presented for crime detection and most of the methods are based on data mining concepts. The crime detection problems are modeled by data mining concepts. Basically crimes are nothing but the social nuisance as well as cost since from last many years in our society. Crime detection techniques evaluated based on its detection speed, faster the detection speed, efficient the detection technique. Based on existing purchase data analysis of cardholder (credit/debit card) is one of the challenging method for reducing the successful credit card frauds ratio. There are many existing methods presented those are non-data mining having suffered from some limitations. These methods basically were used business rules, scorecards and known fraud matching. To overcome such limitations recently new approaches presented for detecting the crime. In this paper we are first presenting the different types of frauds, after that, different kinds of fraud detection techniques and then data mining based techniques investigated here.

KEYWORDS: Crime Detection, credit card/debit card, clustering, data mining, identity crime, scorecards.

I. INTRODUCTION

Fraud refers to obtaining goods/services and money by illegal way. Fraud deals with events which involve criminal motives that, mostly, are difficult to identify. Credit cards are one of the most popular objective of fraud but not the only one. Credit card fraud, a wide-ranging term for theft and fraud committed or any similar payment mechanism as a fraudulent resource of funds in a transaction. Credit card fraud has been expanding issue in the credit card industry. Detecting credit card fraud is a difficult task when using normal process, so the development of the credit card fraud detection models has become of importance whether in the academic or business organizations currently. Furthermore, role of fraud has been changed suddenly during the last few decades along with advancement of technologies. In this paper we are discussing about identify crime in more details. At one extreme, synthetic identity fraud refers to the use of plausible but fictitious identities. These are effortless to create but more difficult to apply successfully. At the other extreme, real identity theft refers to illegal use of innocent people's complete identity details. These can be harder to obtain (although large volumes of some identity data are widely available) but easier to successfully apply. In reality, identity crime can be committed with a mix of both synthetic and real identity details. In the fight against fraud, actions fall under two broad categories: fraud prevention and fraud detection.

Fraud prevention describes measures to stop fraud occurring in the first place. These include PINs for bankcards, Internet security systems for credit card transactions and passwords on telephone bank accounts. In contrast, fraud detection involves identifying fraud as quickly as possible once it has been perpetrated. Apply fraud detection once fraud prevention has failed, using detection methods continuously, as we will usually be unaware that fraud prevention has failed. The US law requires offending organizations to notify consumers, so that consumers can mitigate the harm.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

As a result, these organizations incur economic damage, such as notification costs, fines, and lost business. Credit applications are Internet or paper-based forms with written requests by potential customers for credit cards, mortgage loans, and personal loans. Credit application fraud is a specific case of identity crime, involving synthetic identity fraud and real identity theft. As in identity crime, credit application fraud has reached a critical mass of fraudsters who are highly experienced, organized, and sophisticated. Their visible patterns can be different to each other and constantly change. They are persistent, due to the high financial rewards, and the risk and effort involved are minimal. Based on anecdotal observations of experienced credit application investigators, fraudsters can use software automation to manipulate particular values within an application and increase frequency of successful values. Duplicates (or matches) refer to applications which share common values.

There are two types of duplicates: exact (or identical) duplicates have the all same values; near (or approximate) duplicates have some same values (or characters), some similar values with slightly altered spellings, or both. This paper argues that each successful credit application fraud pattern is represented by a sudden and sharp spike in duplicates within a short time, relative to the established baseline level. Duplicates are hard to avoid from fraudsters point-of-view because duplicates increase their' success rate. The synthetic identity fraudster has low success rate, and is likely to reuse fictitious identities which have been successful before.

The identity thief has limited time because innocent people can discover the fraud early and take action, and will quickly use the same real identities at different places. In this paper we are discussing the recently presented efficient methods for identity crime detection which is based on data mining concepts. In section II we are discussing the different types of frauds. In below section III we are presenting the different methods those are used for detecting the credit card frauds and crimes.

II. LITERATURE SURVEY

Genetic algorithms: In [2], For predictive purposes, genetic algorithms are often acclaimed as a means of detecting fraud. In order to establish logic rules which is capable of classifying credit card transactions into suspicious and non-suspicious classes, genetic algorithm has been suggested that is based on genetic programming. However, this method follows the scoring process. In the experiment as described in their study, the database was made of 4,000 transactions along with 62 fields. As for the similarity, tree, training and testing samples were employed. For this purpose, different types of rules were tested with the different fields. The best rule among these is with the highest predictability. Their method has proven results for real home insurance data and could be one best method against credit card fraud. Chan et al. (1999) has developed an algorithm for prediction of suspect behaviour. Origin of their research is that cost model evaluated and rated b whereas other studies use evaluation based on their prediction rate/the True Positive Rate (TPR) and the error rate/the False NegativE Rate (FNR). Wheeler & Aitken (2000) formed the idea of combining different algorithms to maximize the power of prediction. Article by, Wheeler & Aitken, presents different algorithms: diagnostic algorithms, diagnostic resolution strategies, best match algorithms, density selection algorithms, probabilistic curve algorithms and negative selection algorithms. As a conclusion from their investigation that probabilistic algorithms and neighbourhood-based algorithms have been taken to be appropriate techniques for classification, and further it may be improved using additional diagnostic algorithms for decision-making in borderlines cases as well as for calculation of confidence measures and relative risk measures.

The inspiration for GANN, by combining genetic algorithms with neural networks comes from nature. In GANN, the genetic algorithm is used to find some parameters. Main query is how exactly Genetic Algorithm and Neural Network can be combined. Neural Network has been encoded in the genome of the Genetic Algorithm. In GANN the procedure involves generation of number of random individuals. Designing of neural network is according to the genome information which helps in evaluation of parameter strings. Performance can be easily determined after back-propagation training. To find an optimal network, few GANN strategies rely only on the GA. In this case no training set takes place which are further evaluated and ranked according to parameter performance. Genetic Algorithm (GA) is a search heuristic that copies the process of natural evolution and is used to generate useful and appropriate solutions for optimization problems and search problems. Genetic algorithms (GA) belongs to the larger class of Evolutionary Algorithms (EA), generate solutions to optimization problems using some techniques such as mutation, inheritance, selection, and crossover.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

Neural networks: In [3], 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. From the given cases, nonlinear mapping relations from the input space to output space. Neural networks [3] can learn and summarizes the internal assumptions of data even without knowledge of the potential data principles in advance. 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:

- Difficulty to confirm the structure
- Excessive training
- Efficiency of training and so on.

Decision Tree: In [4], 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. There are many advantages of Decision tree method [4]. At first the high flexibility that it is a non-parameter method without any notion for the data distribution. Good haleness on the other side. Nearby, it is explainable, which is also the reason of its varied utilization. After that, the conception of a similarity tree using decision tree logic has been developed. A similarity tree refers to edges are labelled with values of attributes and pertaining nodes that are labelled with attribute names, that satisfy some condition and "leaves", an intensity factor which implies as the ratio of the number of transactions that satisfy these condition(s) over the total number of legitimate transaction in the particular behaviour The advantage of the similarity tree method is that it is suggested that it is easy to implement, to display and to understand. Still, system has some disadvantages that, the requirements to check each transaction one by one. Similarity trees have given proven results that worked on decision trees and especially on another type of fraud, inductive decision tree in order to establish an intrusion detection system.

Clustering techniques: In [5], Two clustering techniques have been suggested for behavioural fraud by Bolton & Hand. Peer group analysis is a system that allows identifying accounts which are behaving differently from others at one moment in time whereas previously, they were behaving the same. These certain accounts are then flagged as suspicious. Then fraud analysts have been used to uncover those cases. Hypothesis behind peer group analysis is that if accounts that were behaving the same for a certain period of time and then one account, still behaving significantly differently, then this account has to be notified. Another approach, Breakpoint analysis [5] uses a different hypothesis which states that if a change of card usage is notified on an individual basis, the account must be investigated. Or we can say that based on the transactions of a single card, the break-point analysis can identify suspicious behaviour/pattern. Signals of suspicious behaviour are a sudden transaction for a high amount, and a high frequency of usage without any knowledge to cardholder(s).

III. DIFFERENT 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, Behavioural fraud.

Credit Card Fraud: Credit card fraud has been divided into two types: Offline fraud and On-line fraud.
Offline fraud: Is committed by using a stolen physical card at call centre or any other place.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

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. Hansen, McDonald, Messier, and Bell (1996) used a powerful generalized response model to predict management fraud. Model includes the "probit and logit" techniques. At first, this paper introduces Credit Card, its various types, then related work and possible techniques, models for detecting fraudulent/legal transactions.
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. Computer intrusion can be classified into three categories: misuse intrusions, network intrusions and host intrusions.
Misuse intrusions: analyse the information gather and compare it to large databases of attack signatures.
Network intrusions, individual packets flowing through a network are analysed. Passive intrusions, detects a potential security breach, logs the information and signals an alert.
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. Some methods or techniques may help in fraud prevention. The bank will send its users/customers an order to pay. However, the users will be recognized as being in a state of personal bankruptcy and not able to recover their unwanted loans. The bank will have to cover the losses itself. One of the possible ways to prevent bankruptcy fraud is by doing a pre-check with credit bureau in order to be informed about the past banking history of its customers. Foster & Stine (2004) presented a model to forecast personal bankruptcy among users of credit card.
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. Firstly, use of your copied card number and codes via various web-sites, where no signature or physical cards are required. Pago Report issues (2005), although in European E-commerce seems to be quite low, at only 0.83 percent along with the average charge-back ratio, significant concerns are notified in detailed analysis. For the listed credit card, the customers are contacted and if they do not react within certain time limit than the card is blocked.
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. (2006) describes application fraud as "demonstration of identity crime, occurs when application form(s) contain possible, and synthetic (identity fraud), or real but also stolen identity information (identity theft)." In most of the banks, eligibility for a credit card, applicants need to complete an application form. Application form is mandatory except for social fields. The bank would also ask for certain details as contact details, such as e-mail address, mobile phone number and land-line number. Confidential information will be the password.
Behavioural Fraud: Behavioural fraud occurs when sales are made on a "cardholder present" basis and details of legitimate cards have been obtained fraudulent basis.

IV. INVESTIGATED METHOD FOR IDENTITY FRAUD DETECTION

In the literature survey, we have studied different methods for crime detection, however most methods are suffered from different kinds of limitations. Thus to overcome this methods, recently new methods presented which is based on data mining concepts. Recently new method presented in [1]. The main objective of [1] is to achieve resilience by adding two new, real times, data mining-based layers to complement the two existing nondata mining layers discussed in the section. These new layers will improve detection of fraudulent applications because the detection system can detect more types of attacks, better account for changing legal behavior, and remove the redundant attributes. These new layers are not human resource intensive. They represent patterns in a score where the higher the score for an application, the higher the suspicion of fraud (or anomaly). In this way, only the highest scores require human intervention. These two new layers, communal and spike detection do not use external databases, but only the credit application database per se. And crucially, these two layers are unsupervised algorithms which are not completely

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

dependent on known frauds but use them only for evaluation. This investigated method is multilayered detection system which composed with two additional layers: communal detection (CD) and spike detection (SD). CD finds real social relationships to reduce the suspicion score, and is tamper resistant to synthetic social relationships. It is the white list-oriented approach on a fixed set of attributes. SD finds spikes in duplicates to increase the suspicion score, and is probe resistant for attributes. It is the attribute-oriented approach on a variable-size set of attributes. From the practical observations, this investigated approach overcomes limitations of existing methods.

V. CONCLUSION AND FUTURE WORK

Now days for banks and financial organizations, building the accessible, simple and precise risk monitoring system for credit cards is major challenge for improving the level of risk management in efficient way. In this paper we first discussed the different types of frauds. We have presented the review of different techniques those are used for detecting the frauds and crimes. We have demonstrated the limitations of existing methods as well. On the basis of that we have discussed the currently new method presented which is based on concepts of data mining. Practically this method is outperforming the existing methods. For the future work we will suggest to improve this currently presented method in different directions.

REFERENCES

1. Clifton Phua, Member, IEEE, Kate Smith-Miles, Senior Member, IEEE, Vincent Cheng-Siong Lee, and Ross Gayler, "Resilient Identity Crime Detection", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 24, NO. 3, MARCH 2012.
2. 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
3. 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.
4. Y. Sahin and E. Duman , "Detecting Credit Card Fraud by Decision Trees and Support Vector Machines " International MultiConference of Engineering and Computer Scientists 2011 Vol I,IMECS2011,Hong Kong.
5. R. Bolton and D. Hand, "Unsupervised Profiling Methods for Fraud Detection," Statistical Science, vol. 17, no. 3, pp. 235-255, 2001.

BIOGRAPHY

Vaibhav Sirasat is a M.E. Student in the Department of Information Technology, R.M.D Sinhgad School of Engineering, Savitribai Phule Pune University, Pune, MH, India.

Prof. Sweta Kale is a Professor in the Department of Information Technology, R.M.D Sinhgad School of Engineering,Savitribai Phule Pune University, Pune, MH, India.