

Survey Paper on Credit Card Fraud Detection

Suman

Research Scholar,GJUS&T Hisar

HCE Sonapat

Mitali Bansal

Mtech.C.S.E,HCE Sonapat

Abstract- Banking Sector involves a lot of transactions for their day to day operation and they have now realized that their main disquietude is how to detect fraud as early as possible. Due to fulsome advancement in technology, it is imperative for bank to save its money from fraud. The primary motive of this paper is to represent technologies that can be redounding to detect credit card fraud. These technologies will help to diagnose the credit card fraud and give the acquiescent result. The use of these techniques will help to distinguish the credit card transactions generally into two types as legitimate and fraudulent transactions. These techniques are generally based on the Supervised and Unsupervised Learning. The approach on which we are working is based on Unsupervised, in which we train our network so that it is able to detect fraudulent transactions.

Keyword: *Fraud Detection, Data Mining, Unsupervised*

I. Introduction

Bank is a financial institution which accepts deposit from public. And it become great disquietude for bank if there happens any kind of fraud in deposit. It is mentioned by K.Chan&J Stolfo et.al that there are many kind of fraud and generally financial fraud much affects the bank fraud. Due to fastest growing online banking activities, it is found that nearly 44% of population of U.S using these online transactions. According to John T.S Quah it is projected loss of \$8.2 billion in the year 2006 with \$3 billion in U.S alone. Philip K.Chan and Wei Fan et.al defines the Data Mining is a new emerging technology that can detect credit card fraud very effectively. As according to them, by the help of Data Mining we can detect

hidden patterns and can find out the relationship between data set.

Fraud act as the wrongful/criminal deception intended to result in financial or personnel gain. So, Credit Card Fraud is an illegal or fulsome use of card or unusual transaction behavior. As shown in the figure 1 there are so many frauds detected that affect the bank, merchants as well as customers. Some of them are listed below:

- a) Inception of mails of newly issued cards.
- b) Copying or replicating of card information through cloned websites.
- c) Phishing in which credit card number and password is hacked like through emails etc.
- d) Triangulation In this type of fraud, fraudster make an authentic looking website and advertises to sell goods at highly lower prices. Unaware users attract to those sites and make online transactions. They submit their card information to buy those goods. And then fraudsters use these card information to make genuine transactions.

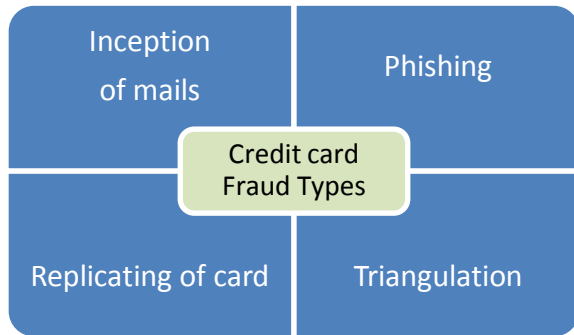


Fig: 1 General Types of Credit Card Fraud

It is interesting to note that credit card fraud affect owner the least and merchant the most .The existing legislation and card holder protection policies as well as insurance scheme affect most the merchant and customer the least. Card issuer bank also has to pay the administrative cost and infrastructure cost .Studies show that average time lag between the fraudulent transaction date and charge back notification can be high as 72 days, thereby giving fraudster sufficient time to cause severe damage.

In this paper we address some solutions to detect credit card fraud as early as possible. The following section introduce some approaches on the basis of supervised and Unsupervised learning.After that we explore the techniques that comes under the supervised and unsupervised learning techniques before going on to result and conclusion.

II. Literature Survey in fraud detection

Fraud act as the wrongful or criminal deception intended to result in financial or personal benefit. It is a deliberate act that is against the law, rule or policy with a aim to attain unauthorized financial benefit.Lokesh Sharma ans Raghavendra Patidar works on emerging technology Neural Network,that can be used in banking or financial areas to detect fraud. They have been successfully applied to detect legitimate or fraudulent transactions.Association Rules can be applied to detect fraud.Linda Delamaire and Pointon et.al use the association rules to extract knowledge so that normal behavior pattern may be obtained in unlawful transactions.This proposed Methodology has been applied on data about credit card fraud of the most important retail companies in Chile. In the area of fraud detection, neural network

like feed forward neural network with back propagation have found immense application.Usually such applications need to know previous data and on the behalf of this previous data they detect the fraud. Another statistically approach is feed forward network in which there is certain kind of relationship is found between user data and other parameters to get. the result. By the help of this approach or by using SOM, data can be filter out to analyze customer behavior (John T.S Kuah, M.Sriganesh). Another new emerging technology of Credit card fraud detection is based on the genetic algorithm and scatter search. Ekrem DuMan,M Hamdi Ozclik published an approach that was base on genetic algorithm and scattering search. In this approach, each transaction is scored and based on these score transactions are divided into fraudulent or legitimate transactions. They focused on a solution to minimize the wrongly classified transactions. They merge the Meta heuristic approaches scatter search and genetic algorithm.

Peer group analysis made by David Weston and Whitrow is a good solution regarding credit card fraud detection. Peer group analysis is a good approach that is based on unsupervised learning and it monitors the behavior over time as well. This peer group technique can be used to find anomalous transaction and help to detect the fraud in time.

All these technologies have their pros and cons as well. As Linda Delmaire works on association rule is a simple method that initially need large data set in which it can find frequent item set. As work done by Lokeh et.al is on Neural Network that can be applied in Supervised as well as Unsupervised Approach. As Unsupervised approach is little bit more complex but give more optimized results. John T.S Quah and M.Sriganesh works on Real time Credit card Fraud detection using computational intelligence that works on Self Organising Map. Ekrem et.al combined the genetic algorithm and Scatter search approach that is really helpful to find anomalous transcations.David Weston provides a good solution to find credit card fraud detection using Peer Group analysis method. So, the main motive of our paper is to represent all important technologies that can detect the fraud as early as possible and to avoid the loss as much as possible.

III. Various Techniques to Detect Credit Card Fraud

There are many emerging technologies that are able to detect credit card fraud detection. Some of condign technologies that will work on some parameters and able to detect fraud earlier as well are listed below:

3.1 Biometric Approach

Kenneth Aguilar & Cesar Ponce et al. defines that all human have particular characteristics in their behavior as well as in their physiological characteristics as depicted in Figure 2. Here behavioral characteristics mean any human’s voice, signature, keystroke etc. And physiological characteristics means fingerprints, face image or hand geometry. Biometric Data mining is an application of knowledge discovery techniques in which we provide biometric information with the motive to identify patterns.

3.1.1) Behavioral Characteristics: According to Revett Henrique Santos, we can have following characteristics that are able to identify patterns.

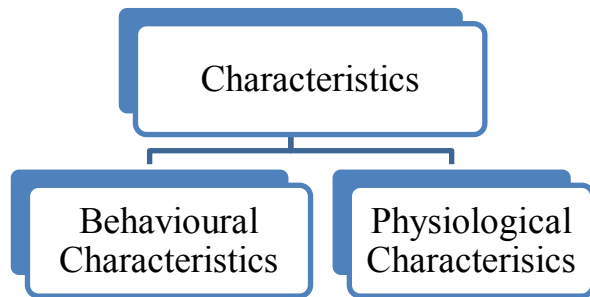


Fig: 2 Characteristics that can be recognized by Biometric System.

a)KeyStroke Patterns: Revett & Henrique Santos defined keystroke pattern in the term of keyboard duration or latency. According to them, pattern of Striking of keys of every person is unique. So it will help to identify the legitimate or fraudulent persons.

b)Mouse Movement: User identification during mouse movement is done by measuring the temperature and humidity of user’s palm and his/her

intensity of pressing. These parameters can be helpful to recognize suspicious behavior.

c)Online Behavior: According to Revett & Santos, this characteristic can be done by observing and collecting data of user’s behavior over long period of time. It is observed that user’s behavior is not random rather than it is centric.

3.1.2) Physiological Characteristics: In physiological characteristics, Hernandez & Diaz defined the following characteristics that are able to detect unique patterns.

a) Fingerprints: As this is a biological physiological characteristic in which every user has unique fingerprints that able to identify legitimate or fraudulent user.

b) Face recognition: According to (Lovell & Chemn, 2010), there are several application of data mining for face recognition: 1) Person recognition and Location Services on a planetary wide sensor, 2) Searching of video in a multimedia database. In Credit card fraud detection it can be used if a registered user uses its credit card for performing transactions.

3.2) Learning Learning is generally done with or without the help of teacher. Generally division of learning take place as shown in Figure3.

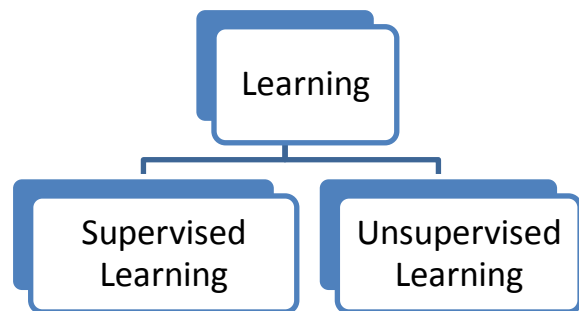


Figure: 3

The learning that take place under the supervision of teacher is termed as supervised learning. But In which there is no guidance of teacher is termed as Unsupervised learning. These Learning are explored as:

3.2.1)Supervised Learning According to Patdar and Lokesh Sharma following technologies are based on Supervised Learning approach in which we have an external teacher to check our output.

a)Bayesian Network: Baye’s theorem is derived by Thomas Bayes. These are statistical classifier that predict class membership probabilities such that whether a particular given tuple belongs to a particular class.In this X is considered as “evidence” and H will be some hypothesis such that X belongs to particular class C.In this,we have two kind of probability:

In this $P(Fr/X)$ and $P(X/Fr)$ are posterior probability conditioned on Hypothesis. And $P(Fr)$ and $P(X)$ are prior probability of Hypothesis. We calculate the posterior probability , $P(Fr/X)$,from $P(Fr)$, $P(X/Fr)$ and $P(X)$ are given

$$\text{Baye's theorem is: } P(Fr/X) = \frac{P(X/Fr)P(Fr)}{P(X)}$$

$P(Fr/X)$ is the fraud probability given the observed behavior user X.This Network can model the behavior based on the assumption that whether the user is fraudulent or legitimate.

b)Decision Tree Induction Dipti Thakur & Bhatia defined this as a type of supervised learning in which we make a decision tree to reach at a particular solution.As shown in figure4 they defined that in decision tree we have some internal nodes and each node represent a test on a particular attribute and each branch in decision tree represent an outcome of test and each leaf node will represent class label means output. Decision trees are used for classification in which we give a new transaction for which class label is unknown(means it is unknown whether it is fraudulent or legitimate) and the transaction value is tested against the decision tree. A path is traced from root node to output/class label for that transaction.

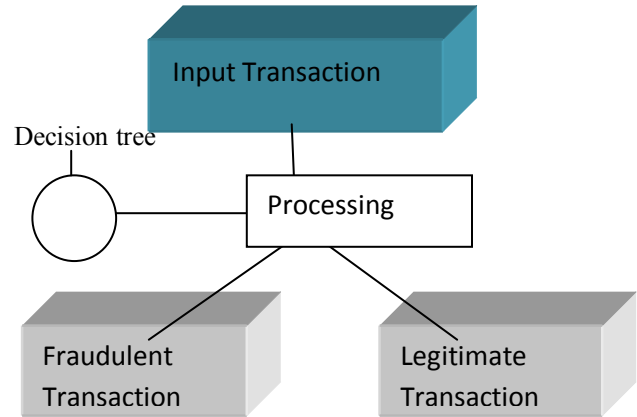


Fig: 4 Decision Tree

c)Support Vector Machine: In supervised learning Vapnik come with an idea of support Vector Machine.Joseph King-Fun Pun approached that in this classification algorithm we can construct a hyper plane as a decision plane which can make distinction between fraudulent and legitimate transaction. This Hyper plane Separate the different class of data. Support Vector Machine can maximize the geometric margin and simultaneously it can minimize the empirical classification. So, it is also called Maximum Margin classifier..The separating Hyperplane is a plane that exploit the distance between the two equivalent hyper plane.

3.2.2) Unsupervised Learning: In contrast to supervised Learning, unsupervised or self organized learning does not require an external teacher.Quah & Sriganesh defined that in this during training session the neural network receive a number of different input patterns, discovered significant features in these pattern and learn how to classify input pattern into different categories.

a)Peer Group Analysis Approach

In 2008 Watson and Whitrow et.al work on Peer Group Analysis in Plastic Card Fraud Detection. They defined it as an unsupervised method that analyzes the behavior over time by monitoring it. This approach can be used to identify credit card fraud detection by analyzing the fraudulent transactions. In this those transactions deviate from their peer group termed as anomalous/fraudulent transactions. They defined that there are generally two type of approach to detect fraud. One is, in which

Form of the fraud is known this can be detected by pattern matching. And when the type of fraud is not known then we approach anomaly detection techniques. Peer group analysis is an anomaly detection technique. Suppose we have $a_1, a_2, \dots, a_{n-1}, a_n$ time series representing the weekly amount spent on a credit card for a particular account and a_n is the target spend transaction a_n is fraudulent or not at time $t=n$. In this, in order to detect outlier transaction we can use the Mahalanobis distance of a_n from the centroid of its peer group. As it shown in figure 5 that by the help of the Peer group technique anomaly input are separated.



Fig: 5 Peer group Analysis

1. ● Population
2. ●● Peer Group Member
3. ○ Target

3.3) Neural Network: Neural Network is a set of connecting input and output units and these units are connected with their associated weights. It used connection between units that's why it is termed as connectionist learning. In this during the learning phase, the network learn by adjusting its weight to predict the class label of the input vector. The main advantage of neural network is its high tolerance of noisy data that's why it is more suitable for credit card fraud detection. They can also detect those patterns for which they have not been trained. Means, in any new transaction of Fraud it will be applicable to detect it as Legitimate or Fraudulent. According to Patdar and Lokesh Sharma There are generally two types of Learning take place in Neural Network:

1) Supervised Learning

2) Unsupervised Learning

3.3.1)Supervised Learning in Neural Network In Supervised Learning, Patdar and Lokesh Sharma analyze the Customer's past behavior and on the basis of that we conclude to our result. In neural network Backpropagation method works in Supervised mode. It iteratively process the data of training set and compare the value of each predicted value to the target value. For each training input the mean squared error is minimized between the network's prediction and actual target value. These all modifications are done in the backward direction by the help of each hidden layer. Hence it is termed as the Backpropagation method.

3.3.2)Unsupervised Learning in Neural Network

a)Self Organising Map In Research of Kuah & Sriganesh Self organizing Map configure its neuron according to topological structure of input data as shown in Figure 6. This process is called Self Organization because of iterative tuning weight of neurons. The result is clustering of input data. In Self Organizing map Zaslavsky and Strizhak[8] defined that we needn't any external teacher in this mode of learning. So, Verification of resultant matrix will be done on behalf of the presented past learning. As our approach is also based on the Unsupervised learning. There are mainly four step of processing take place in Self organizing Map that we applied on our inputs.

1) Initialization : First of all we have to selected small random value for synaptic weight in interval[0,1] and have to assign a small positive value of learning parameter a .

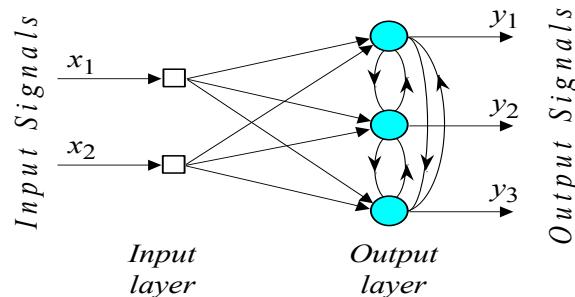


Fig: 6 Kohonen Network

2) Activation and Similarity matching: in this step we activated Kohonen Network for input vector X and find Winner.

3) Learning of Adaptive: Weight are trained by performing many steps.

4) Iteration: At that point we performed iterations,till we did not get a stability in our network.

IV. . Conclusion

Due to Fulsome advancement in technology, the use of credit card has increased and due to this, Fraud cases are affecting it directly. One of the main motives of this study is to explore as many techniques that can detect fraud effectively. If one of the above technologies is applied in bank then cases of credit card fraud will surely minimize. Here we represent the advanced technologies that can detect credit card fraud and save the bank from big loss.

V. References

[1] David J.Watson,David J.Hand,M Adams,Whitrow and Piotr Juszczak “Plastic Card Fraud Detection using Peer Group Analysis” Springer, Issue 2008.

[2] Dipti Thakur,Shalini Bhatia “Distribution Data Mining approach to Credit card Fraud Detection” SPIT IEEE Colloquium and International Conference,Volume 4,48, Issue 2002.

[3] Francisca Nonyelum Ogwueleka, “Data Mining application In Credit Card fraud Detection” Journal of Engineering Science and Technology, Vol 6 No 3,Issue 2011.

[4] Hernandez,Zavala,Diaz and Ponce “Biometric Data Mining Applied to On-Line Regonition Systems” CIICAp, Issue 2008.

[5] John T.S Quah,M Sriganesh “Real time Credit Card Fraud Detection using Computational Intelligence” ELSEVIER Science Direct,35 (2008) 1721-1732.

[6] Joseph King –Fung Pun, “Improving Credit Card Fraud Detection using a Meta Heuristic

Learning Strategy” Chemical Engineering and Applied Chemistry University of Tornto 2011.

[7] Kenneth Revett,Magalhaes and Hanrique Santos “Data Mining a Keystroke dynamic Based Biometric Databse Using Rough Set” IEEE

[8] Linda Delamaire ,Hussein Abdou and John Pointon, “Credit Card Fraud and Detection technique”, Bank and Bank System,Volume 4, 2009.

[9] Philip K Chan,Wei Fan,Andias Prodromidis,J.Stolfo, “Distributed Datan Mining For Credit Card Fraud Detection”, IEEE Intelligent System,Special Issue On Data Mining, 1999.

[10] Raghavendra Patidar,Lokesh Sharma “Credit Card Fraud Detection using Neural Network” International journal of Soft Computing(IJCSE),Volume 32,38,Issue 2011.

[11] Vladdimir Zasalavsky and Anna Strizhak, “Credit Card Fraud Detection using Self Organising Map”,Information and Security An International Journal,Vol 18,2006

[12] V.PriyaDarshini,G.Adiline Macriga, “An Efficient dta Mining For Credit Card Fraud detection Using finger Print Recognition”, IJACR, Vol 2, 2012.