ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 2, February 2014

Credit Card Fraud Detection System Using Hidden Markov Model and K-Clustering

MohdAvesh Zubair Khan[1], Jabir Daud Pathan[2], Ali Haider Ekbal Ahmed[3]

B.E, Computer, Jaihind College of Engineering (Kuran),Pune, India[1,2,3]

Abstract: Credit card frauds are increasing day by day regardless of the various techniques developed for its detection. Fraudsters are so expert that they engender new ways for committing fraudulent transactions each day which demands constant innovation for its detection techniques as well. Many techniques based on Artificial Intelligence, Data mining, Fuzzy logic, Machine learning, Sequence Alignment, decision tree, neural network, logistic regression, naïve Bayesian, Bayesian network, metalearning, Genetic Programming etc., has evolved in detecting various credit card fraudulent transactions. A steady indulgent on all these approaches will positively lead to an efficient credit card fraud detection system. This paper presents a survey of various techniques used in credit card fraud detection mechanisms and Hidden Markov Model (HMM) in detail. HMM categorizes card holder's profile as low, medium and high spending based on their spending behaviour in terms of amount. A set of probabilities for amount of transaction is being assigned to each cardholder. Amount of each incoming transaction is then matched with card owner's category, if it justifies a predefined threshold value then the transaction is decided to be legitimate else declared as fraudulent. Existing fraud detection system may not be so much capable to reduce fraud transaction rate. Improvement in fraud detection practices has become essential to maintain existence of payment system. In this paper Hidden Markov Model (HMM) is used to model the sequence of operation in credit card transaction processing. If an incoming credit card transaction is not accepted by the trained HMM with sufficiently high probability, it is considered to be fraudulent.

Keywords: Fraud detection, Credit card fraud, Various Techniques for Credit Card Frauds, HMM, K-Means Clustering Algorithm, Baum-Welch, OTP.

I. INTRODUCTION

In day to day life, online transactions are increased to purchase goods and services. According to Nielsen study conducted in 2007-2008, 28 of the world's total population has been using internet.85 of these people has used internet to make online shopping and the rate of making online purchasing has increased by 40 from 2005 to 2008. The most common method of payment for online purchase is credit card. Around 60 of total transaction were completed by using credit card. In developed countries and also in developing countries to some extent, credit card is most acceptable payment mode for online and offline transaction. As usage of credit card increases worldwide, chances of attacker to steal credit card details and then, make fraud transaction are also increasing. Credit card can be used to purchases goods and services using online and offline transaction mode. It can be divided into two types:

A. Physical Card
B. Virtual Card

In the physical card based purchase, card holder has to produce the card at the merchant counter and merchant will sweep the card in the EMV (Euro pay, MasterCard and Visa) machine. Fraud transaction can be happened in this mode, only after the card has been stolen. It will be difficult to detect fraud in this type of transaction. If the card holder does not realize loss of the card and does not report to police or card issuing company, it can give financial loses to issuing authorities. In the second method of purchasing i.e. online, these transactions generally happen on telephone or internet and to make this kind of transaction, the user will need some important information about a credit card (such as credit card number, validity, CVV number, name of card holder). To make fraud

transaction to purchase goods and services, fraudster will need to know all these details of card only then he/she will make transactions. Most of the time, the cardholder may or may not know that when or where any person will be seen or stolen card information. To detect this kind of fraud transaction, we have proposed a Hidden Markov Model which is studying spending profile of the card holder. An HMM is to analyse the spending profile of each card holder and to find out any discrepancy in the spending patterns. Fraud detection can be detected on analysing of previous transactions data which helps to form spending profile of the card holder. Every card holder having unique pattern contains information about amount of transactions, details of purchased items, merchant information, date of transaction etc. It will be the most effective method to counter fraud transaction through internet. If any deviation will be noticed from available patterns of the card holder, then it will generate an alarm to the system to stop the transaction.

II. LITERATURE SURVEY

Abhinav Srivastava et al describe the "Credit card fraud detection method by using Hidden Markov Model (HMM)", [8]. In this paper, they model the sequence of operations in credit card transaction processing using a Hidden Markov Model (HMM) and show how it can be used for the detection of frauds. An HMM is initially trained with the normal behaviour of a cardholder.

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 2, February 2014

S.Ghosh and Douglas L.Reilly et al describes the "Credit card fraud detection With Neural Network", [11].In this paper they using data from a credit card issuer, a neural network based fraud detection system was trained on a large sample of labelled credit card account transactions and tested on a holdout data set that consisted of all account activity over a subsequent two-month period of time. The neural network was trained on examples of fraud due to lost cards, stolen cards, application fraud, counterfeit fraud, mail-order fraud and NRI (non-received issue) fraud. The network detected significantly more fraud accounts (an order of magnitude more) with significantly fewer false positives (reduced by a factor of 20) over rule based fraud detection procedures.

Sunil S Mhamane et al describes the "Use of Hidden Markov Model as Internet Banking Fraud Detection", [1]. In this paper they explained about how Fraud is detected using Hidden Markov Model also care has been taken to prevent genuine Transaction should not be rejected by making use of one time password which is generated by server and sent to Personal Mobile of Customer.

Pankaj Richhariya describes "A Survey on Financial Fraud Detection Methodologies",[2].The paper details as follows. Owing to levitate and rapid escalation of E-Commerce, cases of financial fraud allied with it are also intensifying and which results in trouncing of billions of dollars worldwide each year.

III. VARIOUS TECHNIQUES FOR CREDIT CARD FRAUD

A. Neural Networks

Neural network is the need 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. On the other side, there are still many disadvantages for the neural networks, such as:

1. Difficulty to confirm the structure.
2. Excessive training
3. Efficiency of training and so on.

B. Genetic Algorithms

For predictive purposes, 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, one algorithm that has been suggested by Bentley etal? (2000) 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.

C. Decision Tree

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. There are many advantages of Decision tree method. At first the high exibility that it is a non-parameter method without any notion for the data distribution. Good haleness on the other side.

IV. SYSTEM ARCHITECTURE

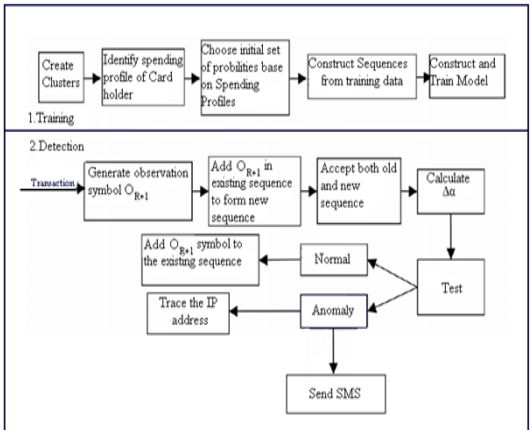

Fig 1: Flow Model of Credit Card FDS

V. WORKING OF SYSTEM

In Fig 2, if an authorized user performs an online transaction then his spending profile is matched into our database and if it matches then the transaction is performed successfully and then user is notified that transaction is done successfully.

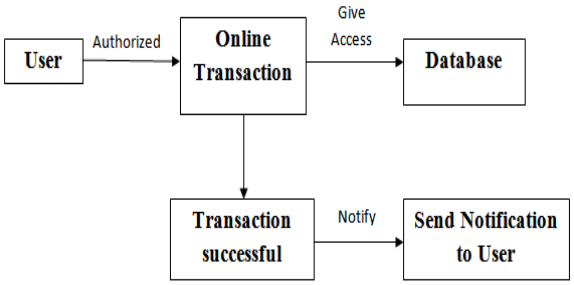

Fig 2: Authorized Users Access

In Fig 3, if an unauthorized user tries to perform an online transaction and if the spending profile doesn't matches into the database then access is blocked to that user and system failure occurs. HMM traces the IP address of the organization from where unauthorized user was trying to

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 2, February 2014

gain transaction and it also sends notification on authorized user's mobile number and raises the alarm to Admin System.

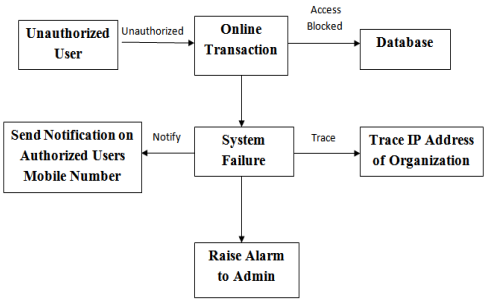

Fig 3: Unauthorized Users Access

VI. USE OF HMM, K-CLUSTERING, OTP FOR CREDIT CARD FRAUD DETECTION

A. HMM MODEL BACKGROUND

An HMM is a double embedded stochastic process much more complicated stochastic processes as compared to a traditional Markov model in fig 4.

The H.M.M. uses the Price range:

1. High.
2. Medium.
3. Low as Prediction.

The diagram below shows the general architecture of an instantiated HMM with two hierarchy levels.

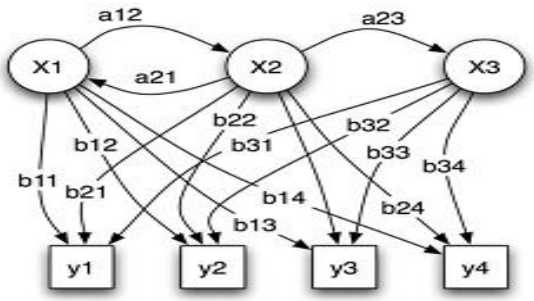

Fig 4.Architecture Of HMM

Technique & Algorithm Used:

To record the credit card transaction dispensation process in conditions of a Hidden Markov Model (HMM), it creates through original deciding the inspection symbols in our representation. We quantize the purchase values x into M price ranges V1, V2 . . . VM, form the study symbols by the side of the issuing bank. The genuine price variety for each symbol is configurable based on the expenditure routine of personal cardholders. HMM determine these prices range. Dynamically by using clustering algorithms (like K clustering algorithm) on the price values of every card holder transactions. It uses cluster Vk for clustering algorithm as k ¼ 1, 2. . . . M, which can be represented both observations on price value symbols as well as on price value range.

In this prediction process it considers mainly three price value ranges such as 1) low (l) 2) Medium (m) and 3) High (h). So set of this model prediction symbols is V { l, m,h}, so V ¼ f as l (low), m (medium), h (high) which makes M¼ 3.

E.g. If card holder perform a transaction as $ 250 and card holders profile groups as l (low) = (0, $ 100], m (medium) = ($ 200, $ 500], and h (high) = ($ 500, up to credit card limit], then transaction which card holder want to do will come in medium profile group. So the corresponding profile group or symbol is M and V (2) will be used. In various period of time, purchase of various types with the different amount would make by credit card holder. It uses the deviation in a purchasing amount of latest 10 transaction sequence (and adding one new transaction in that sequence) which is one of the possibilities related to the probability calculation.

In initial stage, model does not have data of last 10 transactions, in that case, model will ask to the cardholder to feed basic information during transaction about the cardholder such as mother name, place of birth, mailing address, email id etc. Due to feeding of information, HMM model acquired relative data of transaction for further verification of transaction on spending profile of cardholder.

B. K-Means Clustering Technique

K-Clustering algorithm used to determine the clusters. K-means is an unsupervised learning algorithm for grouping a given set of data based on the similarity in their attribute values.

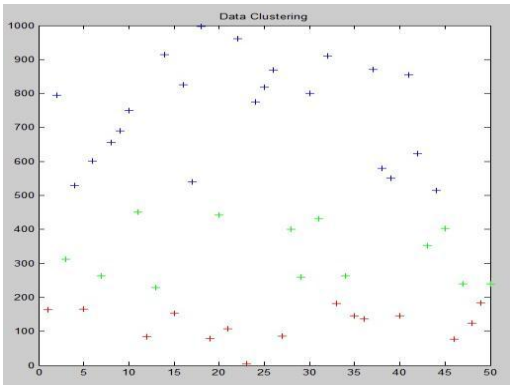

Fig 5.Data Clustering

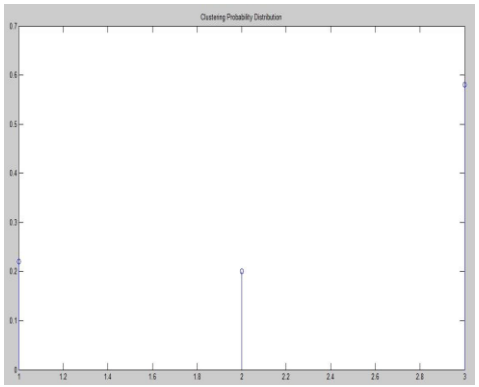

Fig 6. Clustering Probability

Fig.5 shows three clusters. Transactions in red forms low spending group, transactions in green form medium

spending group, and transactions in blue form high spending group. These groups are observation symbols in our implementation. Fig 6 indicates that clustering probability of each observation symbol. In this Fig.6 clustering probability of high spending is highest among three. It can be said that spending profile of given cardholder is high spending

C. One Time Password (OTP)

A one-time password (OTP) is a password that is valid for only one login session or transaction. OTPs avoid a number of shortcomings that are associated with traditional passwords

VII. ADVANTAGES

1. The detection of the fraud use of the card is found much faster that the existing system.
2. In case of the existing system even the original card holder is also checked for fraud detection. But in this system no need to check the original user as we maintain a log.
3. We can find the most accurate detection using this technique.
4. This reduces the tedious work of an employee in the bank.

VIII. APPLICATION

1. Provide easy and well security to Online Shopping
2. Detect Frauds and trace the Location from where the transaction has been made

IX. CONCLUSION

In our paper we used an HMM in detection of credit card fraud. We modelled the sequence of transactions in credit card processing using an HMM. We have used clusters that are generated by using k-means clustering algorithm as our observation symbols. In our implementation we took three observation symbol which are spending ranges of cardholder that are low, medium, and high, whereas the type of item have been considered to be states of an HMM. An HMM is trained with Baum-Welch algorithm for each cardholder. It has been also explained that how an HMM can detect whether the incoming transaction is fraudulent or not.

REFERENCES

[1] Sunil S Mhamane and L.M.R.J Lobo "Use of Hidden Markov Model as Internet Banking Fraud Detection" International Journal of Computer Applications (0975 – 8887) Volume 45– No.21, May 2012

[2] Pankaj Richhariya et al "A Survey on Financial Fraud Detection Methodologies" BITS,Bhopal," International Journal of Computer Applications (0975 – 8887) Volume 45 No.22, May 2012

[3] "Credit Card Fraud Detection Using Hidden Markov Model Shailesh S. Dhok (2012).

[4] A Survey on Hidden Markov Model for Credit Card Fraud Detection Anshul Singh,Devesh Narayan.

[5] Raghavendra Patidar, Lokesh Sharma Credit Card Fraud Detection using Neural Network(2011).

[6] Fraud Detection of Credit Card Payment System by Genetic Algorithm K.RamaKalyani, D.UmaDevi (2012).

[7] The Economic Times http://articles.economictimes.indiatimes.com/2012-08-10/news/33137593_1_number-ofactive-credit-credit-card-cibil.

[8] Srivastava, Abhinav, Kundu, Amlan, Sural, Shamik and Majumdar, Arun K., (2008) "Credit Card Fraud Detection Using Hidden Markov Model", IEEE Transactions on Dependable and Secure Computing, Vol. 5, No. 1, pp. 37-48..

[9] Abhinav Srivastava, Amlan Kundu, Shamik Sural and Arun K. Majumdar, "Credit Card Fraud Detection Using Hidden Markov Model", IEEE transactions on dependable and secure computing, vol. 5, no. 1, January-march 2008.

[10] Maes, Sam, Tuyls Karl, Vanschoenwinkel Bram & Manderick, Bernard, (2002) "Credit Card Fraud Detection Using Bayesian and Neural Networks", Proc. of 1st NAISO Congress on Neuro Fuzzy Technologies. Hawana.

[11] S. Ghosh and D.L. Reilly, "Credit Card Fraud Detection with a Neural-Network," Proc. 27th Hawaii Int'l Conf. System Sciences:Information Systems: Decision Support and Knowledge-Based Systems,vol. 3, pp. 621-630, 1994.

[12] Sung-Bae Cho and and Hyuk-Jang Park ,"Efficient anomaly detection by modeling privilege flows using hidden Markov model" Department of Computer Science, Yonsei University,134 Shinchon-dong,Sudaemoon-ku,Seoul 120-749, Korea.

[13] E. Aleskerov, B. Freisleben, and B. Rao, "CARDWATCH: A Neural Network Based Database Mining System for Credit Card Fraud Detection," Proc. IEEE/IAFE: Computational Intelligence for Financial Eng., pp. 220-226, 1997.

BIOGRAPHIES

Mr. MohdAvesh Zubair Khan is currently pursuing B.E.degree in Computer Engineering from Jaihind College of Engineering, Kuran (University of Pune).

Mr. Jabir Daud Pathan is currently pursuing B.E. degree in Computer Engineering from Jaihind College of Engineering, Kuran (University of Pune).

Mr. Ali Haider Ekbal Ahmed is currentlypursuing B.E. degree in Computer Engineering from Jaihind College of Engineering, Kuran (University of Pune).