ISSN No. 0976-5697

Volume 3, No. 3, May-June 2012

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Credit Card Fraud Detection Using Hidden Markov Model

Shailesh S. Dhok
Department of Information Technology
P.RM.I.T.&R. Badnera,
Amravati, India
shaileshdhok@gmail.com

Dr. G. R. Bamnote
Department of Computer Science & Engineering,
P.RM.I.T.&R. Badnera
Amravati, India
grbamnote@rediffmail.com

Abstract: Due to a rapid advancement in the electronic commerce technology, the use of credit cards has dramatically increased. As credit card becomes the most popular mode of payment for both online as well as regular purchase, it provides cashless shopping. It will be the most convenient way to do online shopping, paying bills etc. Hence, risks of fraud transaction using credit card has also been increasing. In the existing credit card fraud detection system, fraudulent transaction will be detected after transaction is done. It is difficult to find out fraudulent and regarding loses will be barred by issuing authorities. In this paper, we 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. If an incoming credit card transaction is not accepted by the trained HMM with sufficiently high probability, it is considered to be fraudulent. Hidden Markov Model helps to obtain a high fraud coverage combined with a low false alarm rate.

Keywords: Internet, online shopping, credit card, e-commerce security, fraud detection, Hidden Markov Model.

I. INTRODUCTION

The online shopping growing day to day. Credit cards are used for purchasing goods and services with the help of virtual card and physical card where as virtual card for online transaction and physical card for offline transaction. In a physical-card based purchase, the cardholder presents his card physically to a merchant for making a payment. To carry out fraudulent transactions in this kind of purchase, an attacker has to steal the credit card. If the cardholder does not realize the loss of card, it can lead to a substantial financial loss to the credit card company. In online payment mode, attackers need only little information for doing fraudulent transaction (secure code, card number, expiration date etc.). In this purchase method, mainly transactions will be done through Internet or telephone. To commit fraud in these types of purchases, a fraudster simply needs to know the card details. Most of the time, the genuine cardholder is not aware that someone else has seen or stolen his card information.

The only way to detect this kind of fraud is to analyse the spending patterns on every card and to figure out any inconsistency with respect to the "usual" spending patterns. Fraud detection based on the analysis of existing purchase data of cardholder is a promising way to reduce the rate of successful credit card frauds. Since humans tend to exhibit specific behavioristic profiles, every cardholder can be represented by a set of patterns containing information about the typical purchase category, the time since the last purchase, the amount of money spent, etc. Deviation from such patterns is a potential threat to the system.

II.LITERATURE REVIEW

Credit card fraud detection has drawn a lot of research interest and a number of techniques, with special emphasis on neural networks, data mining and distributed data mining have been suggested.

Ghosh and Reilly [1] have proposed credit card fraud detection with a neural network. They have built a detection system, which is trained on a large sample of labeled credit card account transactions. These transactions contain example fraud cases due to lost cards, stolen cards, application fraud, counterfeit fraud, mail-order fraud, and nonreceived issue (NRI) fraud. Recently, Syeda et al. [2] have used parallel granular neural networks (PGNNs) for improving the speed of data mining and knowledge discovery process in credit card fraud detection. A complete system has been implemented for this purpose. Stolfo et al. [3] suggest a credit card fraud detection system (FDS) using metalearning techniques to learn models of fraudulent credit card transactions. Metalearning is a general strategy that provides a means for combining and integrating a number of separately built classifiers or models. A metaclassifier is thus trained on the correlation of the predictions of the base classifiers. The same group has also worked on a cost-based model for fraud and intrusion detection. They use Java agents for Metalearning (JAM), which is a distributed data mining system for credit card fraud detection. A number of important performance metrics like True Positive—False Positive (TP-FP) spread and accuracy have been defined by them.

Aleskerov et al. [4] present CARDWATCH, a database mining system used for credit card fraud detection. The system, based on a neural learning module, provides an interface to a variety of commercial databases. Kim and Kim [5] have identified skewed distribution of data and mix of legitimate and fraudulent transactions as the two main reasons for the complexity of credit card fraud detection. Based on this observation, they use fraud density of real transaction data as a confidence value and generate the weighted fraud score to reduce the number of misdetections.

CONFERENCE PAPER
National Conference on Information and Communication Technology for Development
Organized by PRMITR, Amravati (MS) India
http://mitra.ac.in/forthcoming.html

Fan et al. [6] suggest the application of distributed data mining in credit card fraud detection. Brause et al. [7] have developed an approach that involves advanced data mining techniques and neural network algorithms to obtain high fraud coverage. Chiu and Tsai [8] have proposed Web services and data mining techniques to establish a collaborative scheme for fraud detection in the banking industry. With this scheme, participating banks share knowledge about the fraud patterns in a heterogeneous and distributed environment. To establish a smooth channel of data exchange, Web services techniques such as XML, SOAP, and WSDL are used. Phua et al. [9] have done an extensive survey of existing data-mining-based FDSs and published a comprehensive report. Prodromidis and Stolfo [10] use an agent-based approach with distributed learning for detecting frauds in credit card transactions. It is based on artificial intelligence and combines inductive learning algorithms and metal earning methods for achieving higher accuracy.

Phua et al. [11] suggest the use of metaclassifier similar to in fraud detection problems. They consider naïve Bayesian, and Back Propagation neural networks as the base classifiers. A metaclassifier is used to determine which classifier should be considered based on skewness of data. Although they do not directly use credit card fraud detection as the target application, their approach is quite generic. Vatsa et al. [12] have recently proposed a game-theoretic approach to credit card fraud detection. They model the interaction between an attacker and an FDS as a multi stage game between two players, each trying to maximize his payoff.

HMM-based applications are common in various areas such as speech recognition, bioinformatics, and genomics. In recent years, Joshi and Phoba [13] have investigated the capabilities of HMM in anomaly detection. They classify TCP network traffic as an attack or normal using HMM. Cho and Park [14] suggest an HMM-based intrusion detection system that improves the modeling time and performance by considering only the privilege transition flows based on the domain knowledge of attacks. Ourston et al. [15] have proposed the application of HMM in detecting multistage network attacks. Hoang et al. [16] present a new method to process sequences of system calls for anomaly detection using HMM. The key idea is to build a multilayer model of program behaviors based on both HMMs and enumerating methods for anomaly detection. Lane [17] has used HMM to model human behavior. Once human behavior is correctly modeled, any detected deviation is a cause for concern since an attacker is not expected to have a behavior similar to the genuine user. Hence, an alarm is raised in case of any deviation.Ease of Use

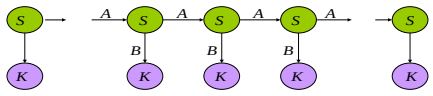
III. PROPOSED SYSTEM DESIGN

A. Hidden Markov Model:

A Hidden Markov Model is a finite set of states; each state is linked with a probability distribution. Transitions among these states are governed by a set of probabilities called transition probabilities. In a particular state a possible outcome or observation can be generated which is associated symbol of observation of probability distribution. It is only the outcome, not the state that is visible to an external observer and therefore states are ``hidden" to the outside observer; hence the name Hidden Markov Model (Shown in figure1). Hence, Hidden Markov Model is a perfect solution for addressing detection of fraud transaction through credit card. One more important benefit of the HMM-based approach is an extreme decrease in the number of False Positives transactions recognized as malicious by a fraud detection system even though they are really genuine. In this prediction process, HMM consider mainly three price value ranges such as.

a. Low (l),
b. Medium (m) and,
c. High (h).

First, it will be required to find out transaction amount belongs to a particular category either it will be in low, medium, or high ranges.



S : are the values for the hidden states
K : are the values for the observations
A : are the state transition probabilities
B : are the observation state probabilities

Figure: 1 HMM finite set of states

B. Credit Card Fraud Detection Using Hidden Markov Model:

In this section, it is shown that system of credit card fraud detection based on Hidden Markov Model, which does not require fraud signatures and still it is capable to detect frauds just by bearing in mind a cardholder's spending habit. The particulars of purchased items in single transactions are generally unknown to any Credit card Fraud Detection System running either at the bank that issues credit cards to the cardholders or at the merchant site where goods is going to be purchased.

As business processing of credit card fraud detection system runs on a credit card issuing bank site or merchant site. Each arriving transaction is submitted to the fraud detection system for verification purpose. The fraud detection system accept the card details such as credit card number, cvv number, card type, expiry date and the amount of items purchase to validate, whether the transaction is genuine or not.

The implementation techniques of Hidden Markov Model in order to detect fraud transaction through credit cards, it create clusters of training set and identify the spending profile of cardholder. The number of items purchased, types of items that are bought in a particular transaction are not known to the Fraud Detection system, but it only concentrates on the amount of item purchased and use for further processing. It stores data of different amount of transactions in form of clusters depending on transaction amount which will be either in low, medium or high value ranges.

It tries to find out any variance in the transaction based on the spending behavioural profile of the cardholder, shipping address, and billing address and so on. The probabilities of initial set have chosen based on the spending behavioural profile of card holder and construct a sequence for further

CONFERENCE PAPER
National Conference on Information and Communication Technology for Development
Organized by PRMITR, Amravati (MS) India
http://mitra.ac.in/forthcoming.html

processing. If the fraud detection system makes sure that the transaction to be of fraudulent, it raises an alarm, and the issuing bank declines the transaction.

For the security purpose, the Security information module will get the information features and its store's in database.

If the card lost then the Security information module form arises to accept the security information. The security form has a number of security questions like account number, date of birth, mother name, other personal question and their answer, etc. where the user has to answer it correctly to move to the transaction section. All these information must be known by the card holder only. It has informational privacy and informational self determination that are addressed evenly by the innovation affording people and entities a trusted means to user, secure, search, process, and exchange personal and/or confidential information.

The system and tools for pre-authorizing business provided that a connections tool to a retailer and a credit card owner. The cardholder initiates a credit card transaction processing by communicating to a credit card number, card type with expiry date and storing it into database, a distinctive piece of information that characterizes a particular transaction to be made by an authoritative user of the credit card at a later time.

The details are received as network data in the database only if an accurate individual recognition code is used with the communication. The cardholder or other authoritative user can then only make that particular transaction with the credit card. Since the transaction is pre-authorized, the vendor does not need to see or transmit an accurate individual recognition code.

C. Techniques and 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 rang 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 M ¼ 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.

D. Model Description:

In existing models, the bank is verified credit card information, CVV number, Date of expiry etc., but all these information are available on the card itself. Nowadays, bank is also requesting to register your credit card for online secure password. In this new model, after feeding details of card at merchant site, then it will transfer to a secure gateway which is established at bank's own server. But, it is not verifying that the transaction is fraudulent or not. If hackers will get secure code of credit card by phishing sites or any other source, then it is very difficult to trace fraudulent transaction. In proposed model based on HMM will help to verify fraudulent of transaction during transaction will be going to happen. It includes two modules are as follow

a. ***Online Shopping:*** It comprises with many steps, first is to login into a particular site to purchase goods or services, then choose an item and next step is to go to payment mode where credit card information will be required. After filling all these information, now the page will be directed to proposed fraud detection system which will be installed at bank's server or merchant site.

b. ***Fraud Detection System:*** All the information about credit card (Like Credit card number, credit card CVV number, credit card Expiry month and year, name on credit card etc.) will be checked with credit card database. If User entered database is correct then it will ask Personal Identity number (PIN). After matching of Personal Identity number (PIN) with database and account balance of user's credit card is more than the purchase amount, the fraud checking module will be activated.

The verification of all data will be checked before the first page load of credit card fraud detection system. If user credit card has less than 10 transactions then it will directly ask to provide personal information to do the transaction. Once database of 10 transactions will be developed, then fraud detection system will start to work.

By using this observation, determine users spending profile. The purchase amount will be checked with spending profile of user. By transition probabilistic calculation based on HMM, it concludes whether the transaction is real or fraud. If transaction may be concluded as fraudulent transaction then user must enter security information. This information is related with credit card (like account number, security question and answer which are provided at the time of registration). If transaction will not be fraudulent then it will direct to give permission for transaction.

CONFERENCE PAPER
National Conference on Information and Communication Technology for Development
Organized by PRMITR, Amravati (MS) India
http://mitra.ac.in/forthcoming.html

If the detected transaction is fraudulent then the Security information form will arise. It has a set of question where the user has to answer them correctly to do the transaction. These forms have information such as personal, professional, address; dates of birth, etc are available in the database. If user entered information will be matched with database information, then transaction will be done securely. And else user transaction will be terminated and transferred to online shopping website. The flowchart of proposed module is shown in Figure 2.

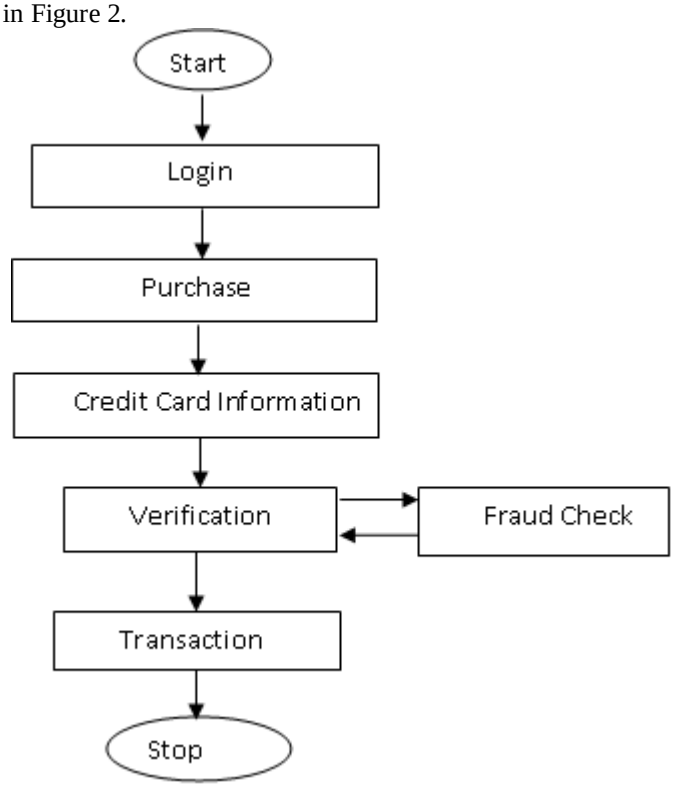


Figure 2: Flowchart of HMM module for credit card fraudulent detection

IV. CONCLUSION

We have discussed an application of HMM in credit card fraud detection. The different steps in credit card transaction processing are represented as the underlying stochastic process of an HMM. We have study the ranges of transaction amount as the observation symbols, whereas the types of item have been considered to be states of the HMM. We have study the suggested a method for finding the spending profile of cardholders, as well as application of this knowledge in deciding the value of observation symbols and initial estimate of the model parameters. It has also been explained how the HMM can detect whether an incoming transaction is fraudulent or not. The system is also scalable for handling large volumes of transactions.

V. ACKNOWLEDGMENT

First & foremost, i would like to express my sincere gratitude towards my project guide prof. Dr. G. R. Bamnote for his valuable guidance, encouragement, and optimism. I feel proud of his eminence and vast knowledge, which will guide me throughout my life.

VI. REFERENCES

[1]. Ghosh, S., and Reilly, D.L., 1994. Credit Card Fraud Detection with a Neural-Network, 27th Hawaii International l Conference on Information Systems, vol. 3 (2003), pp. 621-630.

[2]. Syeda, M., Zhang, Y. Q., and Pan, Y., 2002 Parallel Granular Networks for Fast Credit Card Fraud Detection, Proceedings of IEEE International Conference on Fuzzy Systems, pp. 572-577 (2002).

[3]. Stolfo, S. J., Fan, D. W., Lee, W., Prodromidis, A., and Chan, P. K., 2000. Cost-Based Modeling for Fraud and Intrusion Detection: Results from the JAM Project, Proceedings of DARPA Information Survivability Conference and Exposition, vol. 2 (2000), pp. 130-144.

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

[5]. M.J. Kim and T.S. Kim, "A Neural Classifier with Fraud Density Map for Effective Credit Card Fraud Detection," Proc. Int'l Conf. Intelligent Data Eng. and Automated Learning, pp. 378-383, 2002.

[6]. W. Fan, A.L. Prodromidis, and S.J. Stolfo, "Distributed Data Mining in Credit Card Fraud Detection," IEEE Intelligent Systems, vol. 14, no. 6, pp. 67-74, 1999.

[7]. R. Brause, T. Langsdorf, and M. Hepp, "Neural Data Mining for Credit Card Fraud Detection," Proc. IEEE Int'l Conf. Tools with Artificial Intelligence, pp. 103-106, 1999.

[8]. C. Chiu and C. Tsai, "A Web Services-Based Collaborative Scheme for Credit Card Fraud Detection," Proc. IEEE Int'l Conf. e-Technology, e-Commerce and e Service, pp. 177-181, 2004.

[9]. C. Phua, V. Lee, K. Smith, and R. Gayler, "A Comprehensive Survey of Data Mining-Based Fraud Detection Research," http:// www.bsys.monash.edu.au/people/cphua/. Mar. 2007.

[10]. S. Stolfo and A.L. Prodromidis, "Agent-Based Distributed Learning Applied to Fraud Detection," Technical Report CUCS-014-99, Columbia Univ., 1999.

[11]. C. Phua, D. Alahakoon, and V. Lee, "Minority Report in Fraud Detection: Classification of Skewed Data," ACM SIGKDD Explorations Newsletter, vol. 6, no. 1, pp. 50-59, 2004.

[12]. V. Vatsa, S. Sural, and A.K. Majumdar, "A Game-theoretic Approach to Credit Card Fraud Detection," Proc. First Int'l Conf. Information Systems Security, pp. 263-276, 2005

CONFERENCE PAPER
National Conference on Information and Communication Technology for Development
Organized by PRMITR, Amravati (MS) India
http://mitra.ac.in/forthcoming.html

[13]. S.S. Joshi and V.V. Phoha, "Investigating Hidden Markov Models Capabilities in Anomaly Detection," Proc. 43rd ACM Ann. Southeast Regional Conf., vol. 1, pp. 98-103, 2005,

[14]. S.B. Cho and H.J. Park, "Efficient Anomaly Detection by Modeling Privilege Flows Using Hidden Markov Model," Computer and Security, vol. 22, no. 1, pp. 45-55, 2003.

[15]. D. Ourston, S. Matzner, W. Stump, and B. Hopkins, "Applications of Hidden Markov Models to Detecting Multi-Stage Network Attacks," Proc. 36th Ann. Hawaii Int'l Conf. System Sciences, vol. 9, pp. 334-344, 2003.

[16]. X.D. Hoang, J. Hu, and P. Bertok, "A Multi-Layer Model for Anomaly Intrusion Detection Using Program Sequences of System Calls," Proc. 11th IEEE Int'l Conf. Networks, pp. 531-536, 2003.

[17]. T. Lane, "Hidden Markov Models for Human/Computer Interface Modeling," Proc. Int'l Joint Conf. Artificial Intelligence, Workshop Learning about Users, pp. 35-44, 199

CONFERENCE PAPER
National Conference on Information and Communication Technology for Development
Organized by PRMITR, Amravati (MS) India
http://mitra.ac.in/forthcoming.html