Study of Hidden Markov Model in Credit Card Fraudulent Detection

V. Bhusari
College of Computer Engineering,
Bharati Vidhyapeeth University,
Pune-411043, India

S. Patil
College of Computer Engineering,
Bharati Vidhyapeeth University,
Pune-411043, India

ABSTRACT
The most accepted payment mode is credit card for both online and offline in today's world, it provides cashless shopping at every shop in all countries. 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 business processing 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. Hidden Markov Model is the statistical tools for engineer and scientists to solve various problems. In this paper, it is shown that credit card fraud can be detected using Hidden Markov Model during transactions. Hidden Markov Model helps to obtain a high fraud coverage combined with a low false alarm rate.

Keywords: Hidden Markov Model, fraud transaction, credit card.

1. INTRODUCTION
In day to day life credit cards are used for purchasing goods and services by the help of virtual card for online transaction or physical card for offline transaction. In physical transaction, Credit cards will insert into payment machine at merchant shop to purchase goods. Tracing fraudulent transactions in this mode may not be possible because the attacker already steal the credit card. The credit card company may go in financial loss if loss of credit card is not realized by credit card holder. 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. Small transactions are generally undergo less verification, and are less likely to be checked by either the card issuer or the merchant. Card issuers must take more precaution against fraud detection and financial losses. Credit card fraud cases are increasing every year. In 2008, number of fraudulent through credit card had increased by 30 percent because of various ambiguities in issuing and managing credit cards. Credit card fraudulent is approximately 1.2% of the total transaction amount, although it is not small amount as compare to total transaction amount which is in trillions of dollars in 2007[1-3].

Hidden Markov Model will be helpful to find out the fraudulent transaction by using spending profiles of user. It works on the user spending profiles which can be divided into major three types such as 1) Lower profile; 2) Middle profile; and 3) Higher profile. For every credit card, the spending profile is different, so it can figure out an inconsistency of user profile and try to find fraudulent transaction. It keeps record of spending profile of the card holder by both way, either offline or online. Thus analysis of purchased commodities of cardholder will be a useful tool in fraud detection system and it is assuring way to check fraudulent transaction, although fraud detection system does not keep records of number of purchased goods and categories. Every user represented by specific patterns of set which containing information about last 10 transaction using credit card [4, 10]. The set of information contains spending profile of card holder, money spent in every transaction, the last purchase time, category of purchase etc. The potential threat for fraud detection will be a deviation from set of patterns.

2. 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; hence the name Hidden Markov Model [5-7].

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 [8].

In this prediction process, HMM consider mainly three price value ranges such as [14-15]

1) Low (l),

2) Medium (m) and,

3) 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.

3. CREDIT CARD FRAUD DETECTION USING HMM
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 [9]. 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 [13].

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 [12]. 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 [13].

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 [11]. 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 [15]. 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 behavioral profile of the cardholder, shipping address, and billing address and so on [10]. The probabilities of initial set have chosen based on the spending behavioral profile of card holder and construct a sequence for further 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 [12].

For the security purpose, the Security information module will get the information features and its store's in database [8].

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 [9]. 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 [11].

The system and tools for pre-authorizing business provided that a connections tool to a retailer and a credit card owner [14]. 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 [11].

The details are received as network data in the database only if an accurate individual recognition code is used with the communication [8]. 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 [12].

A. 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 V_1, V_2 . . . V_M, form the study symbols by the side of the issuing bank [14]. 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 V_k 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 [13].

In this prediction process it considers mainly three price value ranges such as 1) low (l) 2) Medium (m) and 3) High (h)[23]. 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 [16].

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.

4. 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

I) 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.

II) 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.

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 1.

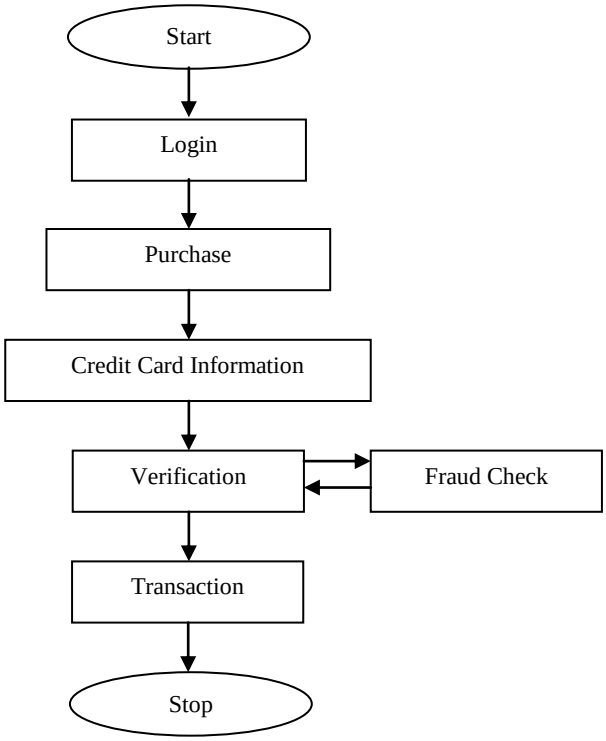

Fig.1: Flowchart of HMM module for credit card fraudulent detection.

5. RESULTS AND DISCUSSION

In this section, it is shown that fraud detection will be checked on last 10 transactions and also calculate percentage of each spending profile (low, medium and high) based on total number of transactions. In Table 1, list of all transactions are shown.

Table 1, list of all transactions happened till date.

No. of Transaction	Amount	No. of Transaction	Amount
1st	140	11th	210
2nd	125	12th	550
3rd	15	13th	800
4th	5	14th	110
5th	10	15th	35
6th	125	16th	118
7th	15	17th	20
8th	120	18th	148
9th	10	19th	141
10th	280	20th	6

The most recent transaction is placed at the first position and correspondingly first transaction is placed at the last position in the table.

The pattern of spending profile of the card holder is shown in Figure 2 based on all transactions done.

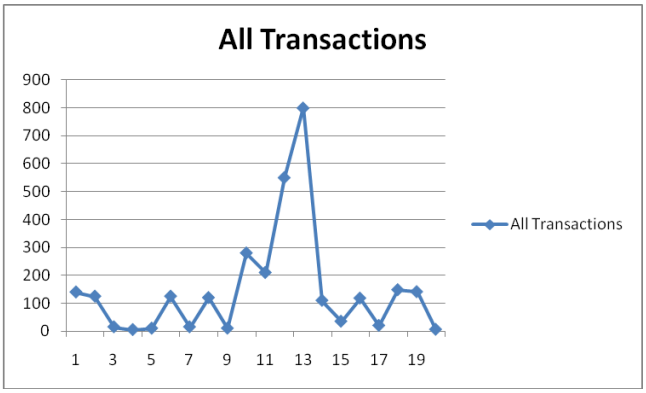

Fig. 2: Spending profile of all transactions.

The percentage calculation of each spending profile (low, medium and high) of the card holder based on price distribution range as mentioned earlier is shown in Figure 3.

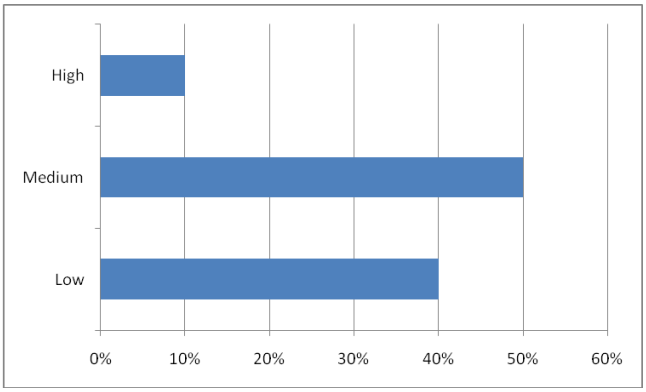

Fig. 3: Percentage of each spending profile.

It has been noticed that medium spending profile has maximum percentage of 50, followed by low profile 40% and then 10% of high spending profile as per details of transactions in Table 1.

Fraud detection mean distribution is shown in Figure 4, where probability of false transaction compared with that of genuine transaction.

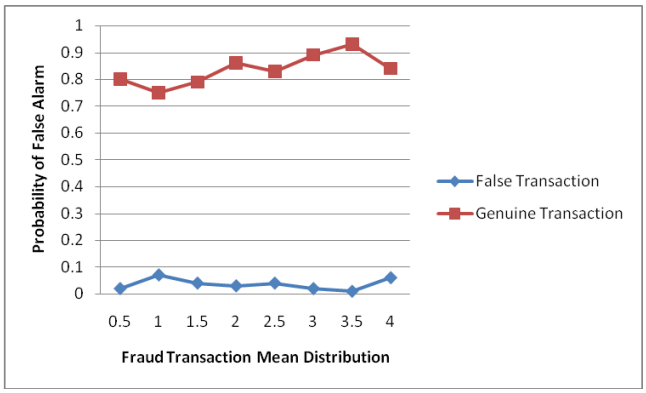

Fig 4: Probability of False Alarm compared with Fraud Transaction Mean Distribution.

In Figure 4, it is noted that when probability of genuine transaction is going down correspondingly probability of false transaction is going to increase and vise versa. It helps to find out the false alarm for the detection of fraud transaction. Hence, when the probability of false alarm will be more than threshold probability, then it will generate an alarm for fraudulent and also decline the transaction.

6. CONCLUSION

In this paper, it has been discussed that how Hidden Markov Model will facilitate to stop fraudulent online transaction through credit card. The Fraud Detection System is also scalable for handling vast volumes of transactions processing. The HMM-based credit card fraud detection system is not taking long time and having complex process to perform fraud check like the existing system and it gives better and fast result than existing system. The Hidden Markov Model makes the processing of detection very easy and tries to remove the complexity.

At the initial state HMM checks the upcoming transaction is fraudulent or not and it allow to accept the next transaction or not based on the probability result. The different ranges of transaction amount like low group, medium group, and high group as the observation symbols were considered. The types of item have been considered to be states of the Hidden Markov Model. It is recommended that a technique for finding the spending behavioral habit of cardholders, also the application of this knowledge in deciding the value of observation symbols and initial estimation of the model parameters

In our proposed model, we have found out more than 84% transactions are genuine and very low false alarm which is about 7 % of total number of transactions.

The relative studies and our results sure that the correctness and effectiveness of the proposed system is secure to 80 percent over a broad deviation in the input data.

7. REFERENCES

[1] Federal Trade Commission, 2009. Consumer sentinel network data book.

[2] Statistics for General and On-Line Card Fraud, March 2007.

[3] Global Consumer Attitude towards On-Line Shopping, March 2007.

[4] 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.

[5] 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).

[6] 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.

[7] 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.

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

[9] Brause, R., Langsdorf, T., and Hepp, M., 1999. Neural Data Mining for Credit Card Fraud Detection, Proceedings of IEEE International Conference Tools with Artificial Intelligence (1999), pp. 103-106.

[10] Chiu, C., and Tsai, C., 2004. A Web Services-Based Collaborative Scheme for Credit Card Fraud Detection, Proceedings of IEEE International Conference e-Technology, e-Commerce and e-Service (2004), pp. 177-181.

[11] Phua, C., Lee, V., Smith, K., and Gayler, R., 2007. A Comprehensive Survey of Data Mining-Based Fraud Detection Research (2007), March.

[12] Rabiner, L.R. 1989. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, Proceedings of IEEE, vol. 77, no. 2 (1989), pp.257-286.

[13] Ourston, D., Matzner, S., Stump, W., and Hopkins, B., 2003. Applications of Hidden Markov Models to Detecting Multi-Stage Network Attacks, Proceedings of 36th Annual Hawaii International Conference System Sciences, vol. 9 (2003), pp. 334-344.

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

[15] Kim, M.J., and Kim, T.S., 2002. A Neural Classifier with Fraud Density Map for Effective Credit Card Fraud Detection, Proceedings of International Conference on Intelligent Data Eng. and Automated Learning, (2002), pp. 378-383.

[16] Kaufman, L., and Rousseeuw, P.J., 1990. Finding Groups in Data: An Introduction to Cluster Analysis, Wiley Series in Probability and Math. Statistics, (1990).