Credit Card Fraud Detection Using Hidden Markov Model

SHAILESH S. DHOK

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

I. INTRODUCTION

In day to day life credit cards are used for purchasing goods and services with the help of virtual card for online transaction or 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 analyze 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 behavioral 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.

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...
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; hence the name Hidden Markov Model. 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:

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.

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 behavioral profile of the cardholder, shipping address, and billing address and so on. 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.

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.

**B.1. 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 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 k¼ 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 holder 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 transactions 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.

**C. 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

**C.2 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.
Credit Card Fraud Detection Using Hidden Markov Model

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

**D Transaction Mean Distribution:**

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.**

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

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.

**IV. CONCLUSION**

We have proposed 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 used 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 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. Comparative studies reveal that the Accuracy of the system is close to 80 percent over a wide variation in the input data. The system is also scalable for handling large volumes of transactions.

**REFERENCES**


AUTHORS PROFILE

Shailesh S. Dhok received his B.E (Information Technology) degree in 2009 from Sant Gadge Baba Amravati University, Amravati , Maharashtra, INDIA. Now I am pursuing M.E.(Information Technology) degree from Sant Gadge Baba Amravati University, Amravati, Maharashtra, INDIA. Life time member of ISTE.