HMM : AN EFFICIENT CREDIT CARD FRAUD DETECTION.

Akshaya Tupe, Tanuja Shinde , Dipali Taware.
S.V.P.M's COE Malegaon(Bk).
Department Of Computer Engg.

***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, cases of fraud associated with it are also rising. 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 behavior 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. At the same time we will alert to cardholder about Fraud via SMS and E-mail.**

Index Terms*—Internet, online shopping, credit card, e-commerce security, fraud detection, Hidden Markov Model.

I. Introduction

The popularity of online shopping is growing day by day. According to an ACNielsen study conducted in 2005, one-tenth of the worlds population is shopping online and credit card is the most popular mode of payment (59 percent). As the number of credit card users rises world-wide, the opportunities for attackers to steal credit card details and, subsequently, commit fraud are also increasing. 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. As the number of credit card users rises world-wide, the opportunities for attackers to steal credit card details and, subsequently, commit fraud are also increasing. 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.

II. Comparison Between Existing System and Proposed System

A. *Existing System*

1) Credit Card Transaction Fraud Detection by using Hidden Markov Model

   Nitin Mishra, Ranjit Kumarand and Shishir Kumar Shandilya proposed a HMM (Hidden Markov Model) based fraud detection system for credit card fraud detection. The method works on the statistical behavior of users transactions. Since the original transactions are not available due to privacy policies of bank we used here synthetically generated data for a credit card user, and then HMM model is trained .The system has been tested on a Pentium 4 PC with 2 GB of RAM; the test program is coded in MATLAB 7.5. In that paper, the HMM based method of detection outliers is used for set up a detection model, which could mine fraud transactions as outliers thereby provide decision support to prevent frauds and to control risks. Many outlier detection algorithms such as base on statistics and distance are gain good application. This shows the maximum accuracy up to 83

2) A Web Services-Based Collaborative Scheme for Credit Card Fraud Detection.

In the web services-based collaborative detection scheme, participant banks plays as service consumers, while Fraud Patterns Mining Service Center (FPMSC) serves as the service provider.To achieve data exchange across heterogeneous applications of banks, participant banks must obey uniform data formats for validating the exchanged data. FPMSC publishes a WSDL file that describes the implementation and interface specification of its provided service.Banks must obey the regulations in the WSDL file, so that they can know what data should be sent and what patterns will be replied, and understand how to access the service.In the WSDL file, the input message of provided service is defined as fraud transactions sent from banks, while the output message is defined as fraud patterns replied to banks. Participant banks must transform their individual fraud transactions stored in legacy formats to an XML document that must pass through the validation by Patterns-schema access the collaborative fraud patterns mining service provided by FPMSC.The valid XML document is enveloped in SOAP Envelope within SOAP Message.The SOAP Message can be sent to FPMSC via popular protocols such as HTTP, SMTP, and MIME. FPMSC extracts fraud patterns from the integrated fraud transactions using Fraud Patterns Mining (FPM) algorithm.When receiving the SOAP message sent from FPMSC, participant banks can interpret the contents of PMML document with SOAP message based on pattern schema and retrieve fraud patterns.Through those fraud patterns, banks can enhance their original fraud detection systems to avoid suffering fraud attacks.

3) Problem Reduction In Online Payment System Using Hybrid Model

To solve the problem of fixed fraud data here presents a new model which does not required fixed fraud data but can detect the fraud by cardholder spending behavior mobile activities. This model uses the combination of two techniques ,one is based on the spending profile i.e. HMM other is based on the mobile implicit authentication system.

4) Credit Card Fraud Detection Using Decision Tree for Tracing E-mail and IP.

Decision Tree have become one of the most powerful and popular approaches in knowledge discovery and Data Mining, the science and technology of exploring large and complex of data in order to discover useful patterns. Decision Tree was originally implemented in decision theory and statistic or highly effective tool in other areas such as Data Mining, Text Mining, Information extraction and Machine learning. Decision Tree is a method commonly used in Data Mining. Decision Tree is a classifier in the form of tree Structure.Fundamental algorithm used for it is Hunts Algorithm. Decision Tree algorithm is a data mining induction techniques that recursively partitions a data set of records using depth-first greedy approach (Hunts et al, 1966) or breadth-first approach (Shafer et al, 1996) until all the data items belong to a particular class. A decision tree structure is made of root, internal and leaf nodes. The tree Structure is used in classifying unknown data records. At each internal node of the tree, a decision of best split is made using impurity measures (Quinlan, 1993). The tree leaves are made up of the class labels which the data items have been group.

Decision Tree using Credit Card Fraud Detection: In credit card fraud detection using Decision tree the Transaction amount divided into two levels such as High and Low. We can find the location of the customer through IP address. IP address traces the transaction location of the customer/merchants. Email Tracing can be divided in to two types:

a) Merchant Tracing Customer Email Address
b) Customer Tracing Merchants Email Address.

B. Demerits Of Existing System

In existing system of credit card fraud detection system fraudulent transaction is detected after transaction is done.It is difficult to find out fraudulent and regarding losses will be barred by issuing authorities.

III. Proposed Model

HMM (Hidden Markov Model) A HMM is a finite set of states; each state is linked with a probability distribution. Transaction among these states is governed by a set of probabilities called transition probabilities. In this prediction process, HMM considers mainly three price value ranges such as,

1) Low (l),
2) Medium(m),
3) High(h).

Model Description

The whole system is divided into three different models, these models are

1) Registration And Training
   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 banks server or merchant site.k-means is an unsupervised learning algorithm for finding out the centroids of generated clusters.The Baum-Welch algorithm is used for calculating the HMM transition state and probability.Forward Backward procedure is used in training phase of HMM. i.e this two procedures are used for training the HMM.
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 users 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.HMM receives cardholder transaction sequence and decide that transaction is fraud or real. The process flow of proposed module is shown in following Figure.
3) Alerting User Regarding Fraud Transaction
   If HMM concluded that the transaction is fraud then the SMS is send to the credit card holder about fraud detected. The SMS is send to mobile phone via SMTP gateway server on the internet.

Merits of Proposed System:

1) It gives better accuracy than other system.
2) The system can scalable for handling large volumes of transaction.

HMM Can be Mathematically defined as follows:

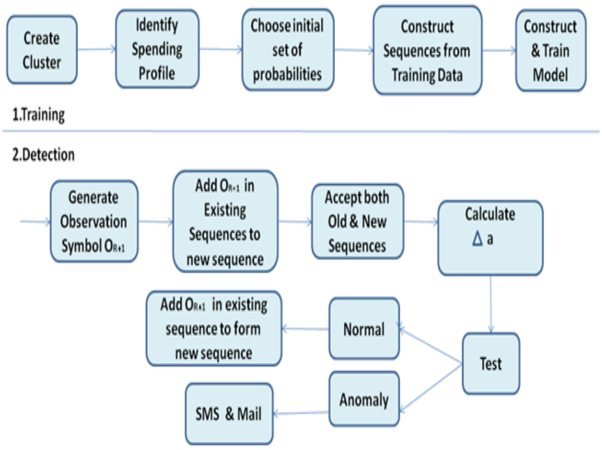
Create Cluster
Identify Spending Profile
Choose initial set of probabilities
Construct Sequences from Training Data
Construct & Train Model
1.Training
2.Detection
Generate Observation Symbol O_{R+1}
Add O_{R+1} in Existing Sequences to new sequence
Accept both Old & New Sequences
Calculate Δ a
Add O_{R+1} in existing sequence to form new sequence
Normal
Test
Anomaly
SMS & Mail

Fig. 1. Proposed Model Of Credit Card Fraud Detection

1) N is number of states in the model set of states is $s = s_1, s_2, s_3 ... s_N$, Where $s_1, s_2, s_3, ..., s_N$ are individual states at any time t is q_t.
2) M is number of distinct observation symbols. Observation symbols corresponding to physical output of system being modeled. We denote set of observation symbols V=$v1, v2, v3, ..., vM$. Where $v_1, v_2, v_3 ... v_M$ are individual observation symbols.
3) State transition probability matrix A=$[a_{ij}]$. Where a_{ij} is transition probability from state *i* and *j*.
   a$_{ij} = P\left(q_{t+1} = S_j | q_t = S_i\right),$
$$1 \leq i \leq N, 1 \leq j \leq N; t = 1, 2, ... \quad (1)$$
4) The observation symbol probability matrix B=$[b_j\left(k\right)]$. Where $b_j\left(k\right)$ is the probability distribution of observation symbol k at state j.
$$b_j\left(k\right) = P\left(V_k | S_j\right), 1 \leq j \leq N, 1 \leq k \leq M \quad (2)$$
5) Initial state distribution $\pi = [\pi_i]$ where
$$\pi_i = P\left(q_1 = S_i\right), 1 \leq i \leq N \quad (3)$$
6) The observation sequence $O = O_1, O_2, O_3, ..., O_R$, where each observation sequence O_t one of the observation symbols from V, and R is the number of observations in the sequence.

1) **Generating Observation Symbols**
   Equation which calculates Spending profile using K-means clustering algorithm, $SP = MAX\left(P_i\right)$ Where P_i percentage of number of transaction those belongs to cluster $i, 1 \leq i \leq M$.

2) **HMM Training**
   Baum-Welch algorithm is as follows: particular observation sequence is $O_1, O_2, O_3....O_T$.
   initialization: set $\lambda = (A, B, \pi)$ with random initial conditions. The algorithm updates the parameter of λ iteratively until convergence, following the procedure below

The forward procedure: We define:$\alpha_i\left(t\right) = P\left(O_1, O_2, O_3....O_t, S_t = i | \lambda\right)$, which is the probability of seeing the partial sequence $O_1, O_2, O_3....O_t$ and ending up in state i at time t. We can efficiently calculate $\alpha_i\left(t\right)$ recursively as:

$$\alpha_i\left(t\right) = \pi_i b_i(O_1) \quad (4)$$

$$\alpha_j\left(t+1\right) = b_i\left(O_t + 1\right) \sum_{i=1}^{N} \alpha_i\left(t\right).a_{ij} \quad (5)$$

The backward procedure: This is the probability of the ending partial sequence $O_1, O_2, O_3....O_T$ given that we started at state i, at time t.We can efficiently calculate $\beta_i\left(t\right)$ as:

$$\beta_i\left(t\right) = 1 \quad (6)$$

$$\beta_i\left(t\right) = \sum_{j=1}^{N} \beta_j\left(t+1\right) a_{ij} b_j\left(O_{t+1}\right) \quad (7)$$

Using α and β, We can calculate the following variables:

$$\gamma_i\left(t\right) = P\left(S_t = i | O, \lambda\right) = \frac{\alpha_i\left(t\right)\beta_i\left(t\right)}{\sum_{j=1}^{N} \alpha_j\left(t\right)\beta_j\left(t\right)} \quad (8)$$

$$\xi_{ij}\left(t\right) = P\left(S_t = i, S_{t+1} = j | O,\right.$$

$$\lambda = \frac{\alpha_i\left(t\right) a_{ij} \beta_j\left(t+1\right) b_j\left(O_{t+1}\right)}{\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i\left(t\right) a_{ij} \beta_j\left(t+1\right) b_j\left(O_{t+1}\right)} \quad (9)$$

having γ and ϵ, one can define update rules as follows:

$$\overline{\pi_i} = \gamma_i\left(1\right) \quad (10)$$

$$\overline{a_{ij}} = \frac{\sum_{t=1}^{T-1} \epsilon_{ij}\left(t\right)}{\sum_{t=1}^{T-1} \gamma_i\left(t\right)} \tag{11}$$

$$\overline{b_i}\left(t\right) = \frac{\sum_t = 1^T \delta_{O_t..o_k} \gamma_i\left(t\right)}{\sum_{t=1}^{T} \gamma_i\left(t\right)} \tag{12}$$

Using the updated values of A, B and π, a new iteration is performed until convergence.

3) **Fraud Detection**
let initial sequence of observation symbol of length R up to time t is $O_1, O_2, O_3, ..., O_R$. We calculate the probability of acceptance of this sequence by HMM, let α_1 be the probability of acceptance of this sequence by HMM, let α_1 be probability of acceptance.
$\alpha_1 = P\left(O_1, O_2, O_3, ..., O_R|\lambda\right)$.
At time $t+1$ sequence is $O_2, O_3, O_4, ..., O_{R+1}$, let α_2 be the probability of acceptance of this sequence
$\alpha_2 = P\left(O_2, O_3, O_4, ..., O_{R+1}|\lambda\right)$.
Let $\triangle\alpha = \alpha_1 - \alpha_2$
If $\triangle\alpha > 0$, it means new sequence is accepted by an HMM with low probability, and it could be a fraud. The new added transaction is determined to be fraudulent if percentage change in probability is above threshold, that is
$Threshold \leq \triangle\alpha|\alpha_1$
If O_{R+1} is malicious, the issuing bank does not approved the transaction, and the FDS discards the symbol. Otherwise,O_{R+1} is added in the sequence permanently, and the new sequence is used as the base sequence for determining the validity of the next transaction

IV. Conclusion

We have proposed an application of HMM in Credit card fraud detection in different steps . 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.where as the type of item have been considered to be states of the HMM. we have suggested a method for finding the spending profile of card holder,as well as deciding the values of observation symbol. HMM can detect whether an incoming transaction is fraudulent or not. The system is also scalable for handing large volume of transaction.

References

[1] Siva Parvati.Nelluri, Shaikh. Nagul , Dr .M.Kishorekumar *"Credit Card Fraud Detection Using Hidden Markov Model"*.

[2] SHAILESH S. DHOK. *"Credit Card Fraud Detection Using Hidden Markov Model."*

[3] Avinash Ingole, Dr. R. C. Thool. *"Credit Card Fraud Detection Using Hidden Markov Model and Its Performance."*

[4] Dr R.DHANAPAL , GAYATHIRI.P *"Credit Card Fraud Detection Using Decision Tree and for Tracing E-mail And IP.* Abhinav Srivastava, Amlan Kundu, Shamik Sural, Senior Member, IEEE, and Arun K. Majumdar, Senior Member, IEEE *IEEE Transaction for "Credit Card Fraud Detection Using Hidden Markov Model"*

[5] Sonali N.Jadhav,Kiran Bhandari. *"Anomaly Detection Using Hidden Markov Model"*

[6] P.Amarnath Raddy,K.Srinivas. *"Credit card Fraud Detection and alerting Using Hidden Mark over Model SMS Gateway."*