Volume 3, Issue 6, June 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Credit Card Fraud Detection Using Hidden Markov Model and Its Performance

Avinash Ingole, Dr. R. C. Thool
Department of IT
S.G.G.S IE & T Nanded.
India

Abstract— Today's world is Internet world. Now-a-day popularity of E-commerce is increasing tremendously. Using E-commerce people do their financial transaction online like online shopping etc. Most popular mode for online and offline payment is using credit card, use of credit card has dramatically increased. So as credit card is becoming popular mode for online financial transactions, at the same time fraud associated with it are also rising. In this paper Hidden Markov Model (HMM) is used to model the sequence of operation in credit card transaction processing. HMM is trained using Baum-Welch algorithm with normal behaviour of 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 last we will see the performance of this system with the help of True Positive (TP) and False Positive (FP) parameters.

Keywords— Hidden Markov Model, credit card, fraud detection, online shopping, e-commerce (HMM)

I. INTRODUCTION

Popularity of online shopping is growing day by day. Credit card is the easiest way to do online shopping. According to an ACNielsen study conducted in 2005 one-tenth of the world's population is shopping online in same study it is also mentioned that credit cards are most popular mode of online payment[1]. In US it is found that total number of credit cards from the four credit card network(VISA, Master Card, American Express, and Discover) is 609 million and 1.28 billion credit cards from above four primary credit card networks plus some other networks (Store, Oil Company and other)[2]. If we consider the statistics of credit cards in India , it is found that total number of credit cards In India at the end of December-31-2012 is about 18 to 18.9 million [3][4]. In case of multinational banks, the average balance, or usage, per borrower for credit card holder has gone up from Rs 61,758 in 2011 to Rs 82,455 in 2012. During the same period, private bank customers' usage grew from Rs 39,368 to Rs 47,370[3]. 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. 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. Credit Card Fraud is defined as when an individual uses another individual s credit card for personal reasons while the owner of the card and the card issuer are not aware of the fact that the card is being used. And the persons using the card has not at all having the connection with the cardholder or the issuer and has no intention of making the repayments for the purchase they done [5]. Credit card based purchases can be categorized into two types:1) physical card and 2)virtual card. 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 the second kind of purchase, only some important information about a card (card number, expiration date, secure code) is required to make the payment. Such purchases are normally done on the Internet or over the telephone. To commit fraud in these types of purchases, a fraudster simply needs to know the card details.

II. CREDIT CARD FRAUD TECHNIQUES

There are many ways in which fraudsters execute a credit card fraud. As technology changes, so do the technology of fraudsters, and thus the way in which they go about carrying out fraudulent activities. Frauds can be broadly classified into three categories, i.e., traditional card related frauds, merchant related frauds and internet frauds. The different types of methods for committing credit card frauds are described below [6]:

A) Lost/ Stolen Cards

A card is lost / stolen when a legitimate account holder receives a card and loses it or someone steals the card for criminal purposes. This type of fraud is in essence the easiest way for a fraudster to get hold of other individual's credit cards without investment in technology. It is also perhaps the hardest form of traditional credit card fraud to tackle.

B) Account Takeover

This type of fraud occurs when a fraudster illegally obtains a valid customers' personal information. The fraudster takes control of (takeover) a legitimate account by either providing the customers' account number or the card number.

The fraudster then contacts the card issuer, masquerading as the genuine cardholder, to ask that mail be redirected to a new address. The fraudster reports card lost and asks for a replacement to be sent.

C) Fake and Counterfeit Cards

The creation of counterfeit cards, together with lost / stolen cards poses highest threat in credit card frauds. Fraudsters are constantly finding new and more innovative ways to create counterfeit cards. Some of the techniques used for creating false and counterfeit cards are listed below:

1) Erasing the magnetic strip: A fraudster can tamper an existing card that has been acquired illegally by erasing the metallic strip with a powerful electro-magnet. The fraudster then tampers with the details on the card so that they match the details of a valid card, which they may have attained, e.g., from a stolen till roll. When the fraudster begins to use the card, the cashier will swipe the card through the terminal several times, before realizing that the metallic strip does not work. The cashier will then proceed to manually input the card details into the terminal.

This form of fraud has high risk because the cashier will be looking at the card closely to read the numbers. Doctored cards are, as with many of the traditional methods of credit card fraud, becoming an outdated method of illicit accumulation of either funds or goods.

2) Creating a fake card: A fraudster can create a fake card from scratch using sophisticated machines. This is the most common type of fraud though fake cards require a lot of effort and skill to produce. Modern cards have many security features all designed to make it difficult for fraudsters to make good quality forgeries. Holograms have been introduced in almost all credit cards and are very difficult to forge effectively. Embossing holograms onto the card itself is another problem for card forgers.

3) Altering card details: A fraudster can alter cards by either re-embossing them — by applying heat and pressure to the information originally embossed on the card by a legitimate card manufacturer or by re-encoding them using computer software that encodes the magnetic stripe data on the card.

4) Skimming: Most cases of counterfeit fraud involve skimming, a process where genuine data on a card's magnetic stripe is electronically copied onto another. Skimming is fast emerging as the most popular form of credit card fraud.

Employees/cashiers of business establishments have been found to carry pocket skimming devices, a battery-operated electronic magnetic stripe reader, with which they swipe customer's cards to get hold of customer's card details. The fraudster does this whilst the customer is waiting for the transaction to be validated through the card terminal. Skimming takes place unknown to the cardholder and is thus very difficult, if not impossible to trace. In other cases, the details obtained by skimming are used to carry out fraudulent card-not-present transactions by fraudsters. Often, the cardholder is unaware of the fraud until a statement arrives showing purchases they did not make.

5) White plastic: A white plastic is a card-size piece of plastic of any color that a fraudster creates and encodes with legitimate magnetic stripe data for illegal transactions. This card looks like a hotel room key but contains legitimate magnetic stripe data that fraudsters can use at POS terminals that do not require card validation or verification (for example, petrol pumps and ATMs).

D) Merchant Related Frauds

Merchant related frauds are initiated either by owners of the merchant establishment or their employees. The types of frauds initiated by merchants are described below:

1) Merchant Collusion This type of fraud occurs when merchant owners and/or their employees conspire to commit fraud using their customers' (cardholder) accounts and/or personal information. Merchant owners and/or their employees pass on the information about cardholders to fraudsters.

2) Triangulation The fraudster in this type of fraud operates from a web site. Goods are offered at heavily discounted rates and are also shipped before payment. The fraudulent site appears to be a legitimate auction or a traditional sales site. The customer while placing orders online provides information such as name, address and valid credit card details to the site. Once fraudsters receive these details, they order goods from a legitimate site using stolen credit card details. The fraudster then goes on to purchase other goods using the credit card numbers of the customer. This process is designed to cause a great deal of initial confusion, and the fraudulent internet company in this manner can operate long enough to accumulate vast amount of goods purchased with stolen credit card numbers.

E) Internet Related Frauds

The Internet has provided an ideal ground for fraudsters to commit credit card fraud in an easy manner. Fraudsters have recently begun to operate on a truly transnational Understanding Credit Card Frauds level. With the expansion of trans-border or 'global' social, economic and political spaces, the internet has become a New World market, capturing consumers from most countries around the world. The most commonly used techniques in internet fraud are described below:

1) Site cloning: Site cloning is where fraudsters clone an entire site or just the pages from which you place your order. Customers have no reason to believe they are not dealing with the company that they wished to purchase goods or services from because the pages that they are viewing are identical to those of the real site. The cloned or spoofed site will receive these details and send the customer a receipt of the transaction via email just as the real company would. The consumer suspects nothing, whilst the fraudsters have all the details they need to commit credit card fraud.

2) False merchant sites: These sites often offer the customer an extremely cheap service. The site requests a customer's complete credit card details such as name and address in return for access to the content of the site. Most of these sites claim to be free, but require a valid credit card number to verify an individual's age. These sites are set up to accumulate as many credit card numbers as possible. The sites themselves never charge individuals for the services they

provide. The sites are usually part of a larger criminal network that either uses the details it collects to raise revenues or sells valid credit card details to small fraudsters.

3) Credit card generators: Credit card number generators are computer programs that generate valid credit card numbers and expiry dates. These generators work by generating lists of credit card account numbers from a single account number. The software works by using the mathematical Luhn algorithm that card issuers use to generate other valid card number combinations. The generators allow users to illegally generate as many numbers as the user desires, in the form of any of the credit card formats, whether it be American Express, Visa or MasterCard.

III.IMPACT OF FRAUDS

*A) Impact of Fraud on Cardholders***:** It's interesting to note that cardholders are the least impacted party due to fraud in credit card transactions as consumer liability is limited for credit card transactions by the legislation prevailing in most countries. This is true for both *card-present* as well as *card-not- present* scenarios. Many banks even have their own standards that limit the consumer's liability to a greater extent. They also have a cardholder protection policy in place that covers for most losses of the cardholder. The cardholder has to just report suspicious charges to the issuing bank, which in turn investigates the issue with the acquirer and merchant, and processes chargeback for the disputed amount [6].

B) Impact Of Fraud On Merchants: Merchants are the most affected party in a credit card fraud, particularly more in the *card-not-present* transactions, as they have to accept full liability for losses due to fraud. Whenever a legitimate cardholder disputes a credit card charge, the card-issuing bank will send a chargeback to the merchant (through the acquirer), reversing the credit for the transaction. In case, the merchant does not have any physical evidence (e.g. delivery signature) available to challenge the cardholder's dispute, it is almost impossible to reverse the chargeback. Therefore, the merchant will have to completely absorb the cost of the fraudulent transaction [6].

C) Impact of Fraud on Banks (Issuer/Acquirer): Based on the scheme rules defined by both MasterCard and Visa, it is sometimes possible that the Issuer/Acquirer bears the costs of fraud. Even in cases when the Issuer/Acquirer is not bearing the direct cost of the fraud, there are some indirect costs that will finally be borne by them. Like in the case of charge backs issued to the merchant, there are administrative and manpower costs that the bank has to incur. The issuers and acquirers also have to make huge investments in preventing frauds by deploying sophisticated IT systems for detection of fraudulent transactions [6].

IV.MOTIVATION

So if we consider above frauds and their impacts it is necessary to have some detection systems. Several credit card fraud detection systems have been developed by many researchers. In this paper we use Hidden Markov Model to detect the credit card fraud. HMM use cardholder's spending behavior to detect fraud. We model a credit card transaction processing sequence by the stochastic process of an HMM.

V. HIDDEN MARKOV MODEL

An HMM is a double embedded stochastic process with two hierarchy levels. It can be used to model complicated stochastic processes as compared to a traditional Markov model. An HMM has a finite set of states governed by a set of transition probabilities. In a particular state, an outcome or observation can be generated according to an associated probability distribution. It is only the outcome and not the state that is visible to an external observer.

Mathematically an HMM can be defined as below [7]:

1. N is number states in the model and set of state is S={ $s_1, s_2, s_3, ...s_N$ }. Where $s_1, s_2, s_3, ...s_N$ are individual states. State at any time t is denoted by q_t
2. M is number of distinct observation symbols. Observation symbols correspond to physical output of system being modeled. We denote set of observation symbols V={ $v_1, v_2, v_3, ...v_M$ } and $v_1, v_2, v_3, ...v_M$ are individual observation symbols
3. State transition probability matrix A=[a_{ij}] .where a_{ij} is transition probability from state *i* to *j*.

$$a_{ij} = P\left(q_{t+1} = S_j \,|\, q_t = S_i\right),\ 1 \le i \le N, 1 \le j \le N; t = 1, 2, ... \quad (1)$$

4. The observation symbol probability matrix B= [$b_j(k)$]. Where $b_j(k)$ is the probability distribution of observation symbol *k* at state *j*.

$$b_j(k) = P\left(V_k \,|\, S_j\right),\ 1 \le j \le N, 1 \le k \le M \quad (2)$$

5. Initial state distribution $\pi = [\pi_i]$ where $\pi_i = P(q_1 = S_i),\ \ 1 \le i \le N$ (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.

It is evident that a complete specification of an HMM requires the estimation of two model parameters, N and M, and three probability distributions A, B, and π . We use the notation $\lambda = (A, B, \pi)$ to indicate the complete set of parameters of the model, where A, B implicitly include N and M.

VI. CREDIT CARD FRAUD DETECTION USING AN HMM

HMM uses cardholder's spending behavior to detect fraud. In our Implementation, three behavior of cardholder are taken into consideration.

1) Low spending behavior
2) Medium spending behavior
3) High spending behavior.

Different cardholders has their different spending behavior (low, medium, high).Low spending behavior of any cardholder means cardholder spend low amount, medium spending behavior of any cardholder means cardholder spend medium amount, high spending behavior of any cardholder means cardholder spend high amount. These profiles are observation symbols, therefore M=3 [8].

A) Generating Observation Symbols

For each cardholder, we train and maintain an HMM. To find one of the three observation symbols corresponding to individual cardholder's transactions, we run K-means clustering algorithm [9] on past transactions. We use random numbers as spending amounts in transactions. With clustering algorithm we get three clusters and clusters represent observation symbols. We then calculate clustering probability of each cluster, which is percentage of number of transaction in each cluster to total number of transactions.

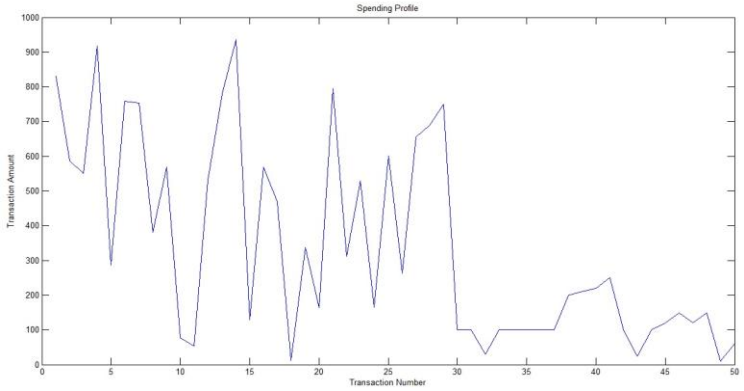

Fig.1 Spending Distribution of amounts in transactions

Fig.1 is graph between transaction number and transaction amount. It shows spending distribution of 50 transactions. It indicate that how cardholder spent amounts in 50 transactions. Fig.2 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 3 indicates that clustering probability of each observation symbol. In this fig.3 clustering probability of high spending is highest among three. It can be said that spending profile of given cardholder is high spending. Following equation calculates spending profile.

$SP = \underset{i}{MAX}\left(P_i\right)$ Where P_i percentage of number of transaction those belongs to cluster *i*, $1 \le i \le M$

B) An HMM Training

Training of an HMM is an offline process. We use Baum-Welch algorithm to train an HMM. Baum-Welch algorithm uses observation symbols generated at the end of k-means clustering. At the end of training phase we get an HMM corresponding to each cardholder. Baum-Welch algorithm is as follow [10]:

Particular observation sequence is $O_1, O_2, O_3 \ldots O_T$.

Initialization: set $\lambda = (A, B, \pi)$ with random initial conditions. The algorithm updates the parameters of λ iteratively until convergence, following the procedure below:

The forward procedure: We define: $\alpha_i(t) = P(O_1, O_2, O_3, \ldots O_t, S_t = i \mid \lambda)$, which is the probability of seeing the partial sequence $O_1, O_2, O_3, \ldots O_t$ and ending up in state *i* at time *t*.

We can efficiently calculate $\alpha_i(t)$ recursively as:

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

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

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

$$\beta_i(T) = 1 \tag{6}$$

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

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

$$\gamma_i(t) \equiv P(S_t = i \mid O, \lambda) = \frac{\alpha_i(t)\beta_i(t)}{\sum_{j=1}^{N} \alpha_j(t)\beta_j(t)} \tag{8}$$

$$\xi_{ij}(t) \equiv P(S_t = i, S_{t+1} = j \mid O, \lambda) = \frac{\alpha_i(t) a_{ij} \beta_j(t+1) b_j(O_{t+1})}{\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i(t) a_{ij} \beta_j(t+1) b_j(O_{t+1})} \tag{9}$$

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

$$\overline{\pi_i} = \gamma_i(1) \tag{10}$$

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

$$\overline{b_i}(k) = \frac{\sum_{t=1}^{T} \delta_{O_t, O_k} \gamma_i(t)}{\sum_{t=1}^{T} \gamma_i(t)} \tag{13}$$

(Note that summation in nominator of $\overline{b_i}(k)$ is only over the observed symbols equal to O_k). Using the updated values of A, B and π, a new iteration is performed until convergence.

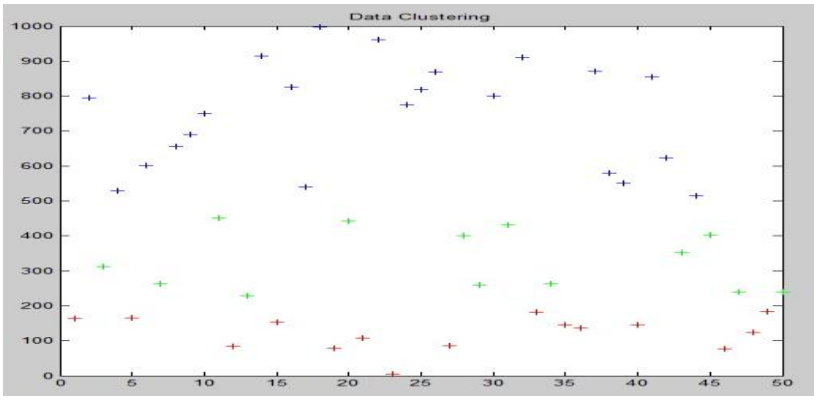

Fig.2 data clustering

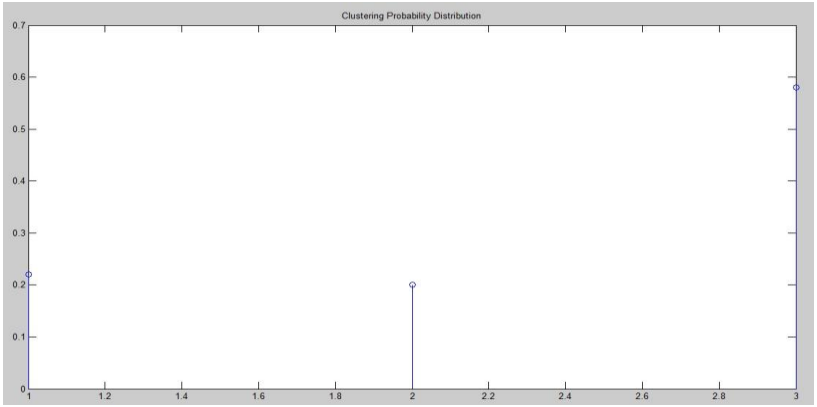

Fig.3 Clustering Probability

C) Fraud Detection

Let initial sequence of observation symbols of length *R* up to time *t* is $O_1, O_2, O_3, \ldots O_R$. In our implementation we have taken 50 as length of sequence. We calculate the probability of acceptance of this sequence by HMM, let α_1 be the probability of acceptance

$$\alpha_1 = P(O_1, O_2, O_3, \ldots O_R \mid \lambda).$$

At time *t+1* sequence is $O_2, O_3, O_4, \ldots O_{R+1}$, let α_2 be the probability of acceptance of this sequence

$$\alpha_2 = P(O_2, O_3, O_4, \ldots O_{R+1} \mid \lambda).$$

$$\text{Let } \Delta\alpha = \alpha_1 - \alpha_2$$

If $\Delta\alpha > 0$, 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

$$Threshold \leq \Delta\alpha / \alpha_1$$

The threshold value can be learned empirically and Baum-Welch algorithm calculates it automatically. If O_{R+1} is malicious, the issuing bank does not approve 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 [8]. The reason for including new non-malicious symbols in the sequence is to capture the changing spending behavior of a cardholder. Fig.4 [8] shows complete process flow of our system. Fig.4 indicates that the system is divided into two parts-one is generating observation symbol and training as shown in fig.4a and other is detection as shown in fig.4b. Training part is performed offline, whereas detection part is an online process.

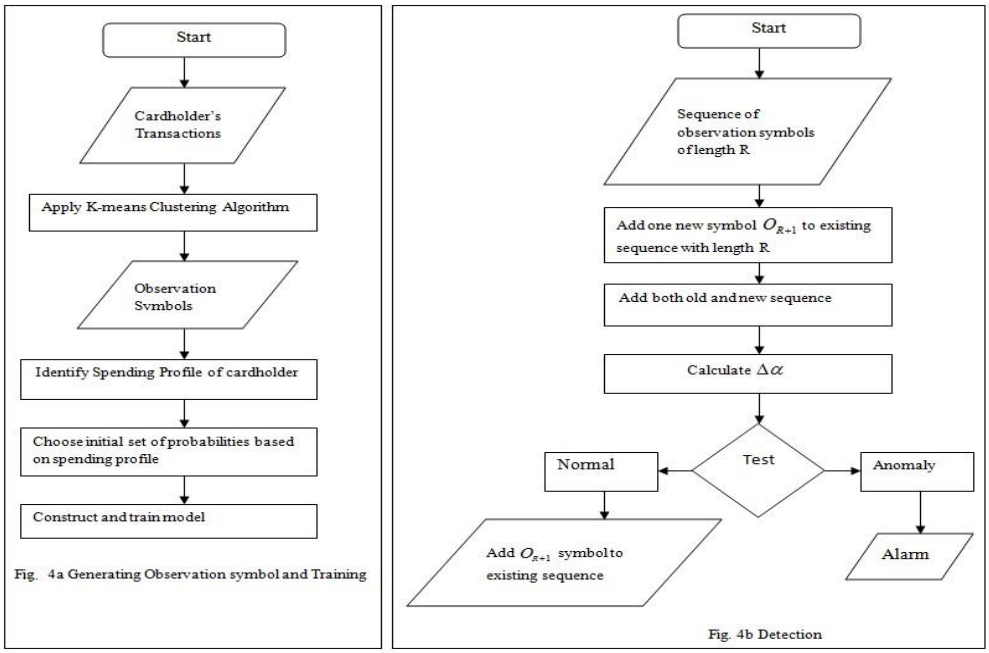

Fig.4 Flow diagram

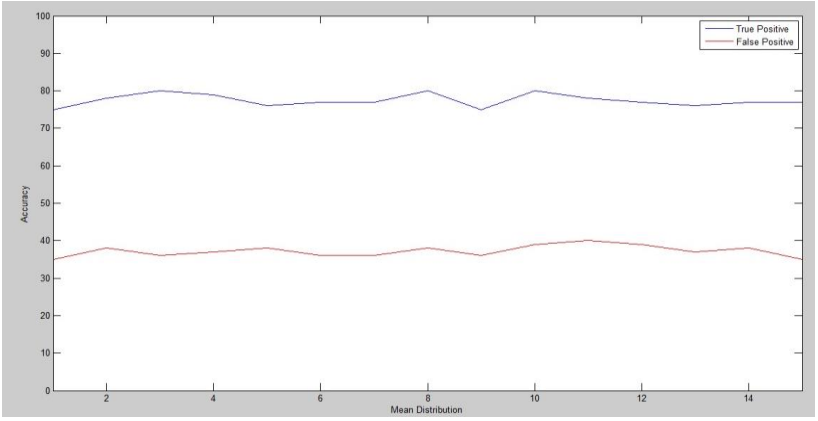

Fig. 5 Performance of system

VII. PERFORMANCE

The performance graph (fig. 5) denotes change in accuracy of system with respect to change in mean of the distribution function. For True Positive, we take fraud transactions as the baseline and for False Positive; we took genuine transactions as the baseline. In fig 5, it is noted that when accuracy of the true positive is going up corresponding accuracy of false positive is going down. We use standard metrics [11]-True Positive (TP) represents the fraction of fraudulent transactions correctly identified as fraudulent, whereas False Positive (FP) is the fraction of genuine transactions identified as fraudulent.

VIII. CONCLUSION

In this paper we used an HMM in detection of credit card fraud. We modeled 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, where as 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. Finally we calculate d the performance of system using TP and FP metrics and it is observed that accuracy of system is near to 75%.

REFERENCES

[1] "Global Consumer Attitude Towards On-Line Shopping," http://www2.acnielsen.com/reports/documents/2005_cc_onlineshopping.pdf, Mar. 2007.

[2] http://www.cardhub.com/edu/number-of-credit-cards/

[3] The Economic Times "http://articles.economictimes.indiatimes.com/2012-08-10/news/33137593_1_number-of-active-credit-credit-card-cibil"

[4] http://www.cardbhai.com/credit-debit-card-data/india-credit-card-holders-information-bank-wise-2012

[5] Raghavendra Patidar, Lokesh Sharma "Credit Card Fraud Detection using Neural Network"

[6] Tej Paul Bhatla, Vikram Prabhu & Amit Dua "Understanding Credit Card Frauds"

[7] L.R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," Proc. IEEE, vol. 77, no. 2, pp. 257-286, 1989.

[8] Credit card fraud detection using Hidden Markov Model", Abhinav Srivastava, Amlan Kundu, Shamik Sural, Senior Member, IEEE, and Arun K. Majumdar, Senior Member, IEEE

[9] A. K. Jain, M. N. Murty, and P.J Flynn. Data clustering:a review. ACM comput. Surv.,31(3):263-323 September 1999

[10] "A gentle tutorial of the EM algorithm and its application to Parameter Estimation for gaussian mixture and Hidden Markov Models", J. A..

[11] Bilmes S.J. Stolfo, D.W. Fan, W. Lee, A. Prodromidis, and P.K. Chan, "Cost-Based Modeling for Fraud and Intrusion Detection: Results from the JAM Project," Proc. DARPA Information Survivability Conf. and Exposition, vol. 2, pp. 130-144, 2000

