GALGOTIAS UNIVERSITY

(Established under Galgotias University Uttar Pradesh Act No. 14 of 2011)

CREDIT CARD FRAUD DETECTION BY USING HIDDEN MARKOV MODEL

A Project Report of Capstone Project - 2

Submitted by

ABHISHEK CHATURVEDI

16SCSE101548

in partial fulfilment for the award of the degree

of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

Under the Supervision of

Dr. M. THIRUNAVUKKARASAN, M.TECH., Ph.D.,

ASSISTANT PROFESSOR

APRIL / MAY-2020

SCHOOL OF COMPUTING AND SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

Certified that this project report "CREDIT CARD FRAUD DETECTION MODEL BY USING HIDDEN MARKOV MODEL" is the bonafide work of "ABHISEK CHATURVEDI" who carried out the project work under my supervision.

SIGNATURE OF HEAD	**SIGNATURE OF SUPERVISOR**
Dr. MUNISH SABARWAL	Dr. M. THIRUNAVUKKARASAN
Ph.D. (Management), Ph.D. (CS)	B.E, M.E, Ph.D.,
Professor & Dean	**Assistant Professor**
School of Computer Science & Engineering	**School of Computer Science & Engineering**

ABSTRACT

Nowadays, credit card is the most accepted mode for the payment for both online and offline purpose, it offers cashless shopping, in petrol pump, in e-commerce and also for paying all types of bills at everywhere in all countries. It will be the more easiest way to do all type of transactions etc. Hence, the risk of fraudulent transactions using credit cards is also increasing. In the current credit card fraud detection business processing system, fraudulent transaction will recognize the transaction after completion. It's very difficult to discover fraud and loses by issuing authorities. Engineers and scientists use the Hidden Markov Model as a statistical tool to solve the various problems about credit cards. In this paper, it is showing how the Hidden Markov Model is a useful method to detect the credit card fraud during transactions. Hidden Markov Model helps to achieve high fraud scope combined with low false alarm rate.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

A, B = Probability Matrices.

Π = Initial State Distribution.

α= Probability

λ = Set of Matrices

R = Number of Observation in a Sequence.

V = Set of Symbols.

M = Number of Symbols in the Alphabets.

O = Observation.

1. INTRODUCTION

1.1 Motivation

In our daily routine life almost every person used credit cards for purchasing goods and services with the help of virtual card for online transactions and physical card for offline purpose. In physical-card based purchases, the cardholder physically hands over his card to a merchant to make a payment. To commit counterfeit transactions in such purchases, an attacker must steal a credit card. If the cardholder does not notice the loss of the card, it can cause considerable financial loss to the credit card company. In online payment mode, attackers need only small amount of information to conduct counterfeit transactions (card number, expiry date, cvv). In this procurement method, the transaction is mainly done through internet or telephone. In this types of purchases a fraudster has to know the details of the credit card to make fraud. Mostly, the actual cardholder isn't aware that someone else has stolen his credit card information. The only path to find out such fraud is to explore the spending habits on every card and to detect any discrepancies in relation to normal spending habits. Detection of fraud based on an analysis of the cardholder's current purchase data is a promising path to reduce successful credit card fraud rates. As humans display specific behaviour profiles, each cardholder can be deputized by a set of habits containing information about the specific purchase category, the time since the last purchase, the amount of money spent, and more. Deviation are a probable threat to the system from such patterns. And to avoid computational complexity and to provide better accuracy in fraud detection in proposed work.

1.2 Purpose

a) The objectives of credit card fraud detection are to reduce losses due to payment fraud for both merchants and issuing banks and increase revenue opportunities for merchants.

b) Adding more security layers to the buying process greatly reduces checkout velocity and, in turn, convenience for the buyer.

c) The other objective is to detect the fraudulent activity happened during credit transactions by strange person with the use of credit card details.

d) Credit card fraud detection confirms the exchange to be malicious, it raises an alarm, and the issuing bank declines the transaction.

1.3 Overall Description

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. A Hidden Markov Model has a finite set of states governed by a set of transition probabilities. In a particular state, observation or an outcome can be generated according to an associated probability distribution. So it is only the outcome and not the state that is visible to an external observer. HMM uses cardholder's spending behaviour to detect fraud. In implementation, three behaviour of cardholder are taken into consideration.

a) Low spending behaviour (L)
b) Medium spending behaviour (M)
c) High spending behaviour (H)

Different cardholders have their different spending behaviour (low, medium, high). Low spending behaviour of any cardholder means cardholder spend low amount (L), medium spending behaviour of any cardholder means cardholder spend
medium amount(M), high spending behaviour of any cardholder spending high amount(H).

ALGORITHM STEPS:

Training Phase: Cluster creation

STEP 1: To Identify the profile of cardholder from their purchasing

STEP 2: The probability calculation depends on the amount of time that has clapsed since entry into the current state.

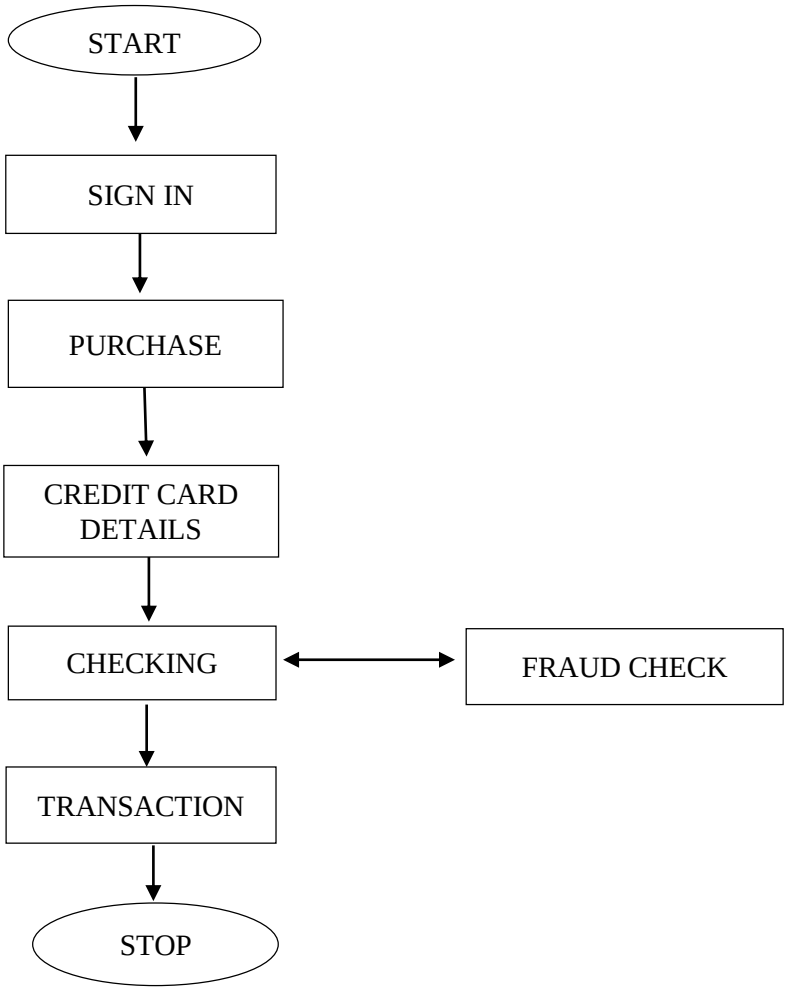
STEP 3: To construct the training sequence for training model

Detection Phase: Fraud detection

STEP 1: To Generate the observation symbol.

STEP 2: To form new sequence by adding in existing sequence

STEP 3: To Calculate the probability difference and test the result with training phase.



In this Technique Clustering algorithm are used for creating three clusters and clusters represent observation symbols. Then calculate clustering probability of each cluster, which is percentage of number of transaction in each cluster to total number of transactions. Then calculate fraudulent transaction. But in this proposed system no need to check the original user as we maintain a log. We can find the most accurate detection using this technique. This reduces the tedious work of an employee in the hank. A one-time password(OTP) is a password that is valid for only one login session or transaction.

1.4 Problem Statement

a) Online payment doesn't require physical card.
b) Anyone who knows the details of card can make fraud transactions.
c) Currently, card holder comes to know only after the fraud transaction is carried out.
d) Currently, the mechanism to track the fraud transaction is not very efficient and don't give accuracy in detection.

2. LITERATURE SURVEY

Credit card extortion recognition has pulled in a ton of research intrigue and numerous techniques: -

Aleskerov [1]. have proposed card watch, a database digging framework approach utilized for credit-card misrepresentation discovery. The framework, in view of a neural learning module, gives an interface to a different business database. Monika, Ishu Trivedi and Mrigya Mridushi [2] have proposed Credit card misrepresentation location with a Genetic Algorithm Technique which involves methods for finding ideal answer for the issue and verifiably producing the aftereffect of the deceitful transaction.

The fundamental intention is to get the fake transaction and to build up a strategy to producing test information.

This algorithm is a heuristic methodology used to tackle high intricacy computational issues. It is an advancement method and developmental hunt dependent on the hereditary and regular determination. The execution of a skilled fraud catching framework is compulsory for all credit card giving banks and their clients to contract their misfortunes. And also concluded that, this strategy demonstrates exact in discovering the deceitful transaction and limiting the quantity of bogus alarm. Hereditary Algorithm is appropriate for this situation like application fields. The utilization of this algorithm in credit card misrepresentation identification framework is probably going to brings about extortion discovery or expectation after the transactions is distinguished in a brief time frame. This will ultimately save banks and clients from a lot of losses and also will minimize risks. Stolfo and Prodromidis [3]. have proposed an operator based strategy with disseminated learning for getting false during card transactions. It's based on artificial intelligence and consolidates inductive learning algorithm and metal acquiring techniques to accomplish an extraordinary exactness. Syeda [4]. presents a parallel granular neural systems (PGNNs) to improve the pace of information mining and information revelation process in Credit-card misrepresentation recognition. A total framework has been actualized for this reason. [5] Peer bunch investigation made by David Weston and Whitrow. is a decent arrangement about credit card false identification. Peer bunch investigation is a decent technique that depends on unsupervised learning and it watched the conduct after some time too. This peer bunch strategy can be utilized to discover peculiar transaction and help to get the extortion in time. Linda Delmaire and Pointon et al. [6]. working on association rule is a basic strategy that at first need big dataset that can find frequent item set and also to take out knowledge so that ordinary personal conduct might be acquired in unlawful transactions.

This approach presents here has been applied on data about credit card fraud of the most important retail companies in Chile.

John T.S Quah and M. Sriganesh [7]. have proposed a Real-Time Credit Card Fraud recognition utilizing computational knowledge that takes a shot at Self Organizing Map. Brause et al. [8]. have built a strategy that includes propelled information mining methods and neural system calculations to discover high misrepresentation inclusion. Ekrem et.al [9]. consolidated the genetic algorithm and Scatter search approach that is extremely useful to discover odd transactions. Chiu and Tsai [10] presents Web Services and Data Mining technique to manufactures a community oriented plan to recognize misrepresentation in the financial area. With this plan, those banks are taking an interest share their insight about the extortion designs in a heterogeneous and conveyed condition. To set up a smooth channel of information trade, Web administrations strategies like XML, SOAP, and WSDL are utilized.

3. PROPOSED SYSTEM

3.1 Techniques and Algorithm Used

a) A mechanism is developed to determine whether the given transaction is fraud or not.
b) The mechanism uses Hidden Markov Model to detect fraud transaction.
c) Hidden Markov Model works on the basis of spending habit of user.
d) Classifies users into low, medium or high category.
e) The details of items purchased in Individual transactions are usually not known to any Fraud Detection System(FDS) running at the bank that issues credit cards to the cardholders. Hence, we feel that HMM is an ideal choice for addressing this problem.

3.1.1 Application of HMM in Credit Card Fraud Detection

In this section, we present credit card fraud detection system based on Hidden Markov Model, which does not require fraud signatures and still is able to detect frauds just by bearing in mind a cardholder's spending habit. The 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 fraud detection system, we consider three different spending profiles of the card holder which is depending upon price range, named high (h), medium (m) and low (l). In this set of symbols, we define V = {l, m, h} and M =3. The price range of proposed symbols has taken as low (0, $101], medium ($101, $501] and high ($501, up to credit card limit]. After finalizing the state and symbol representations, the next step is to determine different components of the HMM, i.e. the probability matrices A, B, and Π so that all parameters required for the HMM is known. These three model parameters are determined in a training phase using the forward-backward algorithm. The initial choice of parameters affects the performance of this algorithm and, hence, it is necessary to choose all these parameters carefully. We consider the special case of fully connected HMM in which every state of the model can be reached to every other state just in a single step, as shown in Fig. 1. 1, 2, 3 etc., are names given to the states to denote different purchase types such as bill payment, restaurant, electronics items etc.

In the figure 1, it has been shown that probability of transition from one state to another (for example from 1 to 2 and vice versa, represented as a1-2 and a2-1, respectively) and also probabilities of transition from a particular state (1, 2, or 3) to different spending habits h, m, or l (for example, b1-h, b1-m, etc.).

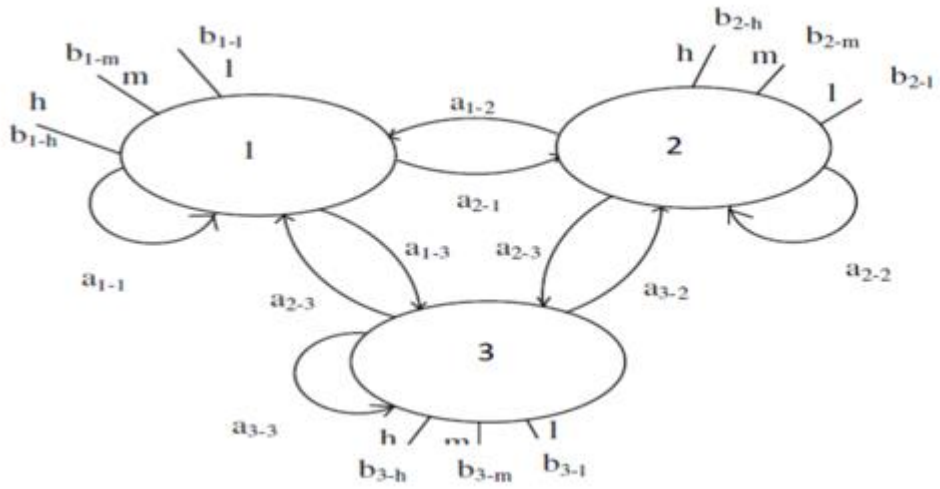


Fig 2. Transition of Different States

The most important thing is to estimate HMM parameters for each card holder. The forward backward algorithm starts with initial HMM parameters and converges to the nearest likelihood values. After deciding HMM parameters, we will consider to form an initial sequence of the existing spending behaviour of the card holder. Let O_1, O_2, O_R be consisting of R symbols to form a sequence. This sequence is recorded from cardholder's transaction till time t. We put this sequence in HMM model to compute the probability of acceptance. Let us assume be this probability is $\alpha 1$, which can be calculated as,

$$\alpha 1 = P (O_1, O_2, O_3, ...O_R \mid \lambda),$$

Let O_{R+1} be new generated sequence at time t+1, when a transaction is going to process. The total number of sequences is R+1. To consider R sequences only, we will drop O_1 sequence and we will have R sequences from O2 to OR+1.

Let the probability of new R sequences be $\alpha 2$,

$$\alpha 2 = P (O_2, O_3, O_4 \ldots\ldots O_{R+1} \mid \lambda),$$

Hence, we will find, $\Delta\alpha = \alpha_1 - \alpha_2$.

If $\Delta\alpha > 0$, it means that HMM consider new sequence i.e. O_{R+1} with low probability and therefore, this transaction will be considered as fraud transaction if

and only if percentage change in probability is greater than a predefined threshold value.

$$\Delta\alpha \mid \alpha 1 \geq \text{threshold value},$$

The threshold value can be calculated empirically. This Fraud detection system if finds that the present transaction is a malicious, then credit card issuing bank will regret the transaction and FDS discard to add O_{R+1} symbol to available sequence. If it will be a genuine transaction, FDS will add this symbol in the sequence and will consider in future for fraud detection.

4. RESULTS AND DISCUSSION

It is very difficult to do simulation on real time data set which is not providing from any credit card bank on security reasons. In Table 1, it is shown that a random data set of all transactions happened is categorized according to their types of purchase. With the help of this, we calculate probability of each spending profile (h, l and m). Fraud detection of incoming transaction will be checked on last 10 transactions.

No. Of Transaction	Amount	No. Of Transaction	Amount
1st	141	11th	211
2nd	126	12th	551
3rd	16	13th	801
4th	6	14th	111
5th	11	15th	36
6th	126	16th	119
7th	16	17th	21
8th	121	18th	149
9th	11	19th	142
10th	281	20th	7

The latest transaction is set in any case and comparably first transaction is set in the last spot in the table.

The example of spending profile of the cardholder is appeared in Figure-3 dependent on all transaction performed.

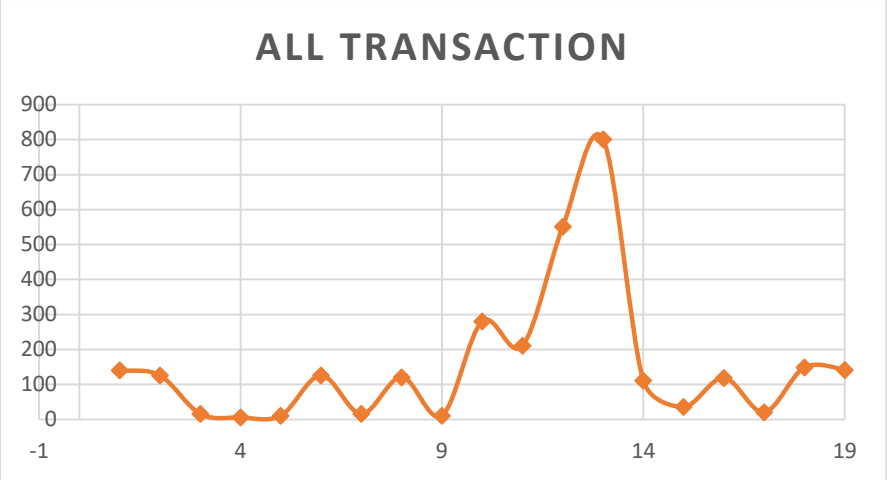


Fig. 3: Expense Profile of All Transaction.

The percentage estimation of every single expense profile (low, medium and high) of the cardholder dependent on the value conveyance limit is portrayed before is appeared in Figure 4.

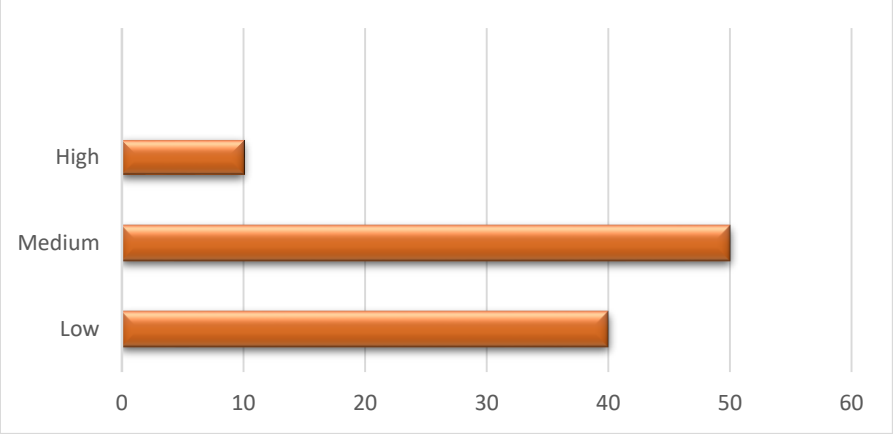


Fig. 4: Percentage of each expense profile

It is seen that the most extreme percentage in the medium consumption profile is 50, trailed by low profile 40 % and afterward 10% of high use profile as indicated by the transactions subtleties Table 1.

Transaction mean distribution is appeared in Figure 5, where likelihood of bogus transaction contrasted and that of genuine transactions.

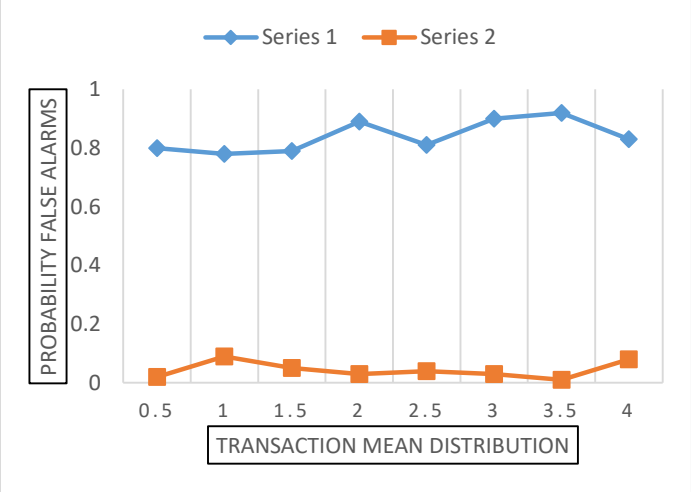


Fig 5. Transaction Mean Distribution

"Series 1 denotes Genuine Transaction and Series 2 denotes False Transaction".

In this figure, it is engaged that when likelihood of genuine transaction is declining and likelihood of bogus transaction are expanding and the other way around. It helps in recognizing the bogus alerts to distinguish deceitful transaction. In this manner, when the likelihood of bogus alert is more prominent than edge likelihood, at that point it will create a caution for counterfeit and furthermore dismisses the transaction.

5. CONCLUSION

We have structured a utilization of HMM to distinguish credit card counterfeit. The different phases of Credit-card transaction preparing are spoken to as the crucial stochastic procedure of an HMM. We have utilized the components of transactions amount as the throughout identification, while the sorts of things have been concentrated to be the states of the Hidden Markov Model. We have communicated a strategy for find the spending profile of cardholders, just as the capacity of capacity in choosing the estimation of transaction and starting evaluation of the model system. It is likewise depicted how a HMM can get whether an approaching transaction is fake or not. Provisional examinations admit that the precision of the system is close to 81% over a wide variety in the input information. The framework or system is also gaugeable to handle the huge amount of transactions.

6. REFERENCES

1. 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.
2. Andrea Dal Pozzolo, Giacomo Boracchi, Olivier Caelen, Cesare Alippi, and Gianluca Bontempi, "Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy," IEEE Trans. Neural Networks Learn. Syst., vol. 29, no. 8, pp. 3784–3797, Aug. 2018.
3. David J. Wetson, David J. Hand, M Adams, Whitrow and Piotr Jusczak "Plastic Card Fraud Detection using Peer Group Analysis" Springer, Issue 2008.
4. John T.S Quah, M Sriganesh "Real time Credit Card Fraud Detection using Computational Intelligence" ELSEVIER Science Direct,35 (2008).
5. Linda Delamaire, Hussein Abdou and John Pointon, "Credit Card Fraud and Detection technique", Bank and Bank System, Volume 4, 2009.
6. Raghavendra Patidar, Lokesh Sharma "Credit Card Fraud Detection using Neural Network" International journal of Soft Computing(IJCSE), Volume 32,38, Issue 2011.
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. Rinky D Patel and Dheeraj Kumar Singh “Credit Card Fraud Detection and Prevention of Fraud Using Genetic Algorithm” published by International
9. Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-6, January 2013.
10. S. Stolfo and A.L. Prodromidis, “Agent-Based Distributed Learning Applied to Fraud Detection,” Technical Report CUCS-014-99, Colombia University...,1999.
11. 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).
12. T. Lane, “Hidden Markov Models for Human/ Computer Interface Modeling” Proc. Int’l Joint Conf. Artificial Intelligence, Workshop Learning about Users, pp. 35-44,1999.