Hidden Markov Model Using Transaction Patterns for ATM Card Fraud Detection

Nkemnole E. B.[1] and Akinsete A. A.[2]
[1]Department of Mathematics, University of Lagos, Nigeria.
[2]Department of Mathematics, Marshall University, Huntington, WV, USA.

Abstract

Estimating the rate at which transactions happen in the social and business world is becoming imperative and has been attracting a lot of research attention. The increase in transactions using the ATM card has become bedeviled with increasing rate of attendant fraud associated with it. This study proposes a hidden Markov model (HMM) based on the Poisson distribution (HMM[Pois]), the generalized Poisson distribution (HMM[GenPois]), and the Gaussian distribution (HMM[Gauss]) for which optimal detection of patterns of anomalies is computed using the forward-backward algorithm. The proposed estimation procedure based upon the three distributions for the HMM model is used to construct a sequence of operations in ATM card transaction processing, and detect fraud by studying spending profile of the cardholder, followed by checking an incoming transaction against spending behavior of the cardholder. If the transaction satisfies a predefined threshold value, then the transaction is decided to be legitimate else, the transaction is declared as fraudulent. The evaluation statistics used shows that the HMM[Gauss] is the most appropriate model in detecting ATM card fraudulent transactions.

Keywords: Hidden Markov Model, Generalized Poisson, Optimal Detection, Forward-Backward Algorithm, Estimation

1. Introduction

Science and technology have really made human life less cumbersome through inventions of many useful devices. One of such devices is the ATM cards. This device contains confidential details such as the Card Number details, Card Member name, and other pieces of information related to the Card owner. These cards are used everywhere for deposits, withdrawals, account information, online shopping, regular purchasing and other forms of transactions. Due to its convenience, there is steady increase of its use. The convenience notwithstanding, the use of the cards is also susceptible to fraudsters who, if care is not taken, can cause enormous loss of money both for the card holder and the issuing banks. A quite number of techniques have been developed in the bid to detect and curb credit card fraudulent transactions. Some of these techniques are based on Artificial Intelligence, Data mining, Fuzzy logic, Machine learning, Sequence Alignment, decision tree, neural network, logistic regression, naïve Bayesian, Bayesian network, Genetic Programming etc.

This work strives to discover patterns which appear and reappear over a space of time as it concerns the pattern of commands someone uses in instructing a computer, sequence of words in sentences, and sequence of phonemes in spoken words. By finding these patterns, there is high probability of predicting the possible behavior or habits of ATM users, making it easy to spot cases that deviate from normal transaction pattern of the user. It is our reasoned opinion that such calculations can provide extra security on the ATM system.

Consequently, a Hidden Markov Model (HMM) based on the Poisson distribution (HMM[Pois]), the generalized Poisson distribution (HMM[GenPois]) and the Gaussian distribution (HMM[Gauss]) for which optimal detection of patterns of variances using the forward-backward algorithm is developed for the analysis of the spending profile of the card holder and to find out any inconsistency in the spending patterns. HMM model identifies transaction patterns of the user. Accordingly, it can assist in preserving and updating a database that defines the operational behavior of the identified user in the form of the pattern. The behavioural pattern of the user will be checked once there is a transaction of user to see if it aligns with previous patterns. Once there is nonconformity with the person's behavioural pattern, the transaction will be blocked. For the card to be unblocked, the user will have to prove ownership through a stipulated pattern. Various techniques proposed for the detection of credit card fraud transaction are briefly explained in section 2.

2. Related Literature

Credit card fraud detection has generated a handful of literature from scholars worldwide. These scholars, in one way or the other, strived to showcase a number of techniques that have been developed to detect fraudulent transactions using the credit/debit card.

Ghosh and Reilly (1994) developed a Neural Network for "Credit card fraud detection. In this study, using data from a credit card issuer, a neural network based fraud detection system was trained on a large sample of labelled credit card account transactions and tested on a holdout data set that contained all account activity over a two-month period of time. The neural network was trained on examples of fraud due to lost cards, stolen cards, application fraud, counterfeit fraud, mail-order fraud and non-received issue (NRI) fraud. The network detected significantly more fraud accounts with significantly fewer false positive over rule based fraud detection procedures.

Aleskerov, et *al*. (1997) advanced a fraud detection system known as "Card watch" which is built upon the neural network learning algorithm. This detection system is designed for commercial implementation and so can take care of large datasets, and parameters of an analysis can be easily adjusted within a graphical user interface. It makes use of three main neural network learning techniques, namely conjugate gradient, back propagation, and batch back propagation. Practically, it is convenient for large financial institutions due to its ease of operation with commercial databases. But, the shortcoming of this system lies in the fact that one has to build a separate neural network for each customer, which will surely result in a very large overall network that requires relatively higher amounts of resources to maintain.

Dorronsoro et *al*. (1997) developed a neural network based fraud detection system called "Minerva". It entrenches itself deep in credit card transaction servers to detect fraud in real time. It uses a novel nonlinear discriminant analysis technique that combines the multilayer perceptron architecture of a neural network with Fisher's discriminant analysis method. Here, there is no need for a large set of historical data because Minerva acts solely on immediate previous history, and is able to classify a transaction in 60ms. But, the shortcoming of this system is the difficulty in determining a meaningful set of detection variables and the difficulty in obtaining effective datasets to train with.

Kokkinaki (1997) proposed the establishment of a user profile for every credit card account and to assess incoming transactions against the corresponding user's profile. The features used in creating these profiles include credit card numbers, transaction dates, type of business, place, amount spent, credit limit and expiration time. In order to get a user's habits, Kokkinaki employed the use of a Similarity Tree algorithm, a variation of Decision Trees. The study showed that the method has a very small probability for false negative errors. However, there is need for constant updates as the user profiles are not adaptive when user habits and fraud patterns change. Stolfo et *al*. (1997) research on the class distribution of a training set and its effects on the performance of multi-classifiers on the credit card fraud domain showed that increasing the number of minority instances in the training process results in fewer losses due to fraudulent transactions. Besides, the fraud distribution for training was varied from 10% to 90% and it was discovered that maximum savings were realized when the fraud percentage used in training was 50%.

By combining a rule-based classification approach with a neural network algorithm, Brause et *al*. (1999) identified fraud cases through their study of credit card payment. In this approach the rule-base classifier first checked to see if a transaction was fraudulent, and then the transaction classification was verified by a neural network. This procedure increases the probability for the correct analysis of fraud, and therefore able to reduce the number of false alarms while increasing the confidence level.

Ehramikar (2000) revealed that the most predictive Boosted Decision Tree classifier is one that is trained on a 50:50 class distribution of fraudulent and legitimate credit card transactions. The study also shows that training decision tree classifiers on datasets with a high distribution of legitimate transactions leads to high fraudulent cases classified as legitimate (a high false negative rate). This means that predictive model over fitting occurs when the training dataset has a majority of legitimate transactions.

To reduce the number of fraud investigations in the credit approval process, Wheeler and Aitken (2000) came up with a case-based reasoning system which consists of two parts, a retrieval component and a decision component. The retrieval component uses a weighting matrix and nearest-neighbor strategy to ascertain and extract the right cases to be used in the final diagnosis for fraud, while the decision component utilizes a multi-algorithm strategy to evaluate the retrieved cases. The nearest-neighbour and Bayesian algorithms were used in the multi algorithm strategy. Initial results of 80% non-fraud and 52% fraud recognition from Wheeler and Aitken indicate that their multi-algorithmic case-based reasoning system is capable of high accuracy rates.

Through the observation of uncharacteristic spending behavior and occurrence of transactions, Bolton and Hand (2001) suggested an unsupervised credit card detection system. The mean amount spent over a specified time window was used as the comparison statistic. The study recommended the Peer Group Analysis (PGA) and the Break Point Analysis (BPA) techniques as unsupervised outlier detection tools. The results of the study indicated that the PGA technique was able to effectively identify local anomalies in the data, and the BPA technique can efficiently determine fraudulent behavior by comparing transactions at the beginning and end of a time window.

To improve the learning proficiency of a neural network, Kim and Kim (2002) came up with a fraud density map technique. The fraud density map (FDM) looks at the inconsistent distributions of legitimate and fraudulent transactions between the training data and real data. It modifies the bias found in the training data by reflecting the distribution of the real data onto the training data through the changing of a weighted fraud score.

Through the application of artificial neural networks (ANN) and Bayesian belief networks (BBN) to a real world dataset, Maes et *al*. (2014) discovered that by performing a correlation analysis on the features and removing the feature that was strongly correlated with many of the other features clear improvements to the results were obtained. In addition, the results of the study revealed that BBNs yield superior fraud detection results and their training period is shorter. On the other hand, ANN was found to be able to compute fraud predictions faster in the testing stage.

To address the credit card fraud problem, Chen et *al*. (2004) developed a questionnaire-responded transaction (QRT) data of users. To develop the QRT models, the study applied the support vector machine algorithm to the data, which were then used to decide if new transactions were fraudulent or genuine. The research results showed that even with very little transaction data the QRT model has a high accuracy in identifying fraud. Chiu and Tsai (2004) identified the problem of credit card transaction data having a natural skewness towards legitimate transactions. The ratio of fraud transactions to normal transactions is extremely low for an individual financial institution (FI), and this makes it difficult for FIs to maintain updated fraud patterns. The study proposed web service techniques for FIs to share their individual fraud transactions to a centralized data center and applied a rule-based data mining algorithm to the combined dataset to detect credit card fraud. Foster and Stine (2004) used a fully automated stepwise regression model to predict personal bankruptcy. The results from this thesis indicate that standard statistical models are competitive with decision trees. The benefit of this model is that it can easily understand the procedures in the prediction process. But the disadvantage lies in the fact it is difficult to follow the process from input to the output prediction.

Joshi and Phoba (2005) have examined the capacities of HMM in detection of irregularities. They classify transmission control protocol (TCP) network traffic as an attack or normal using HMM. Cho and Park (2003) proposed an HMM-based intrusion detection system which strives to improve the modeling time and performance through an exclusive consideration of the privilege transition flows based on the domain knowledge of attacks. Ourston et *al*. (2003) studied the detection of multistage network attacks through the application of HMM. Hoang and Hu (2004), tackle the issue of irregularity detection using HMM via a new method to process sequences of system calls. The idea is to develop a multilayer model of program behaviors based on both HMMs and enumerating methods for anomaly detection. Lane (1999) has used HMM to model human behavior. Once human behavior is appropriately modeled, any identified anomaly is a cause for concern since we do not anticipate an invader to have the same behavioral pattern as the genuine user. Hence, an alarm is raised in case of any deviation. Lately, Ashphak et *al*. (2013), Bhusari and Patil (2011), Mohdavesh et *al*. (2014), Jadhav and Bhandari (2013), Singh and Singh (2015) have investigated the capabilities of HMM in anomaly detection.

3.0 Methodology

3.1 Hidden Markov Model

The HMM is a stochastic finite *state,* a web of relations associated with a probability distribution. Transitions among the states are ruled by a set of probabilities known as *transition probabilities.* In a specific state, according to the associated probability distribution, an *observation* can be generated. The outcome is visible, but not the state. In other words, the states are ``hidden" to the outside; hence the name Hidden Markov Model. It is a statistical model which comprises a set of observations which are produced by an unobservable set of states (Elliot et *al*. (1995), Rabiner (1989), and MacDonald &

Zucchini (1997)). For Rabiner and Juang (1986), a hidden Markov model can be thought as a doubly-embedded stochastic process with an underlying state sequence $\{X_k\}_{k\geq 0}$ that is not observable or hidden. The sequence of state is Markovian and hidden because it can only be seen through another set of stochastic processes $\{Y_k\}_{k\geq 0}$ that produce the sequence of observation where each observation is a probabilistic function of the corresponding state. In HMM, the true state $\{X_k\}_{k\geq 0}$ is hidden, but leads to observable consequences $\{Y_k\}_{k\geq 0}$ as shown in figure 1.

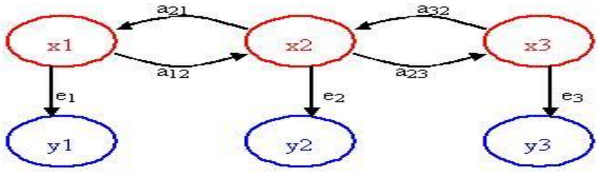

Figure 1. *Graphical representation of the dependence structure of a HMM, where* y_1, y_2, y_3 *are the observable states and* x_1, x_2, x_3 *are the hidden states*

- A set of states $(x^{'}s)$
- A set of possible output symbols $(y^{'}s)$
- A state transition matrix $(a^{'}s)$, probability of making transition from one state to another:

$$P = \{p_{ij}\}, \quad p_{ij} = p(X_{t+1} = j \mid X_t = i), \quad p_{ij} \geq 0, \quad 1 \leq i, j \leq N, \quad X_t,$$

denotes the current state.

- Output emission matrix $b_i(k)\}$, probability of emitting or observing a symbol at a particular state:

$B = [b_i(k)], \quad b_i(k) = p(y_t = v_k \mid x_t = i), \quad 1 \leq i \leq N, \quad 1 \leq k \leq M, v_k$, denotes the k^{th} observation symbol per state.

- Initial probability vector, probability of starting at a particular state: $\pi_i, i \in S, \ \pi_i = p(X_1 = s_i)$

The key point of figure 1 is that these *observable states,* y_1, y_2, y_3 , are directly dependent on some *hidden state,* x_1, x_2, x_3. These hidden states are what actually dictate the outcome of the observable states. The challenge is to figure out the hidden states, the emission probabilities and transition probabilities.

An application of HMM needs specification of two model parameters $(N \text{ and } M)$, and of the three probability measures $\{P, B, \pi\}$. For expediency, we use the compact notation

$$\lambda = \{P, B, \pi\}$$

to designate the complete parameter for HMMs.

Two assumptions can be detected in the model. Firstly, is the Markov assumption, which states that the current state is dependent only on the previous state,

(3.1) $$p(X_{t+1} \mid X_1^t) = p(X_{t+1} \mid x_t)$$

Secondly, the independence assumption states that the output observation at time t is dependent only on the current state; it is independent of previous observations and states: $p(Y_t \mid Y_1^{t-1}, X_1^t) = p(Y_t \mid X_t)$

3.1.1 Model Parameter Estimation

With the model and the observation sequence in consideration, the model parameter is estimated with the following estimation algorithm. The first two are pattern recognition problems: Finding the probability of an observed sequence given a HMM (evaluation); and finding the sequence of hidden states that most probably generated an observed sequence (decoding). The third problem is generating a HMM given a sequence of observations (learning). It deals with the training of the model which is of most significant interest.

Evaluation: Given a model $\lambda = (A, B, \pi)$, and a sequence of observations $O = (o_1, \cdots, o_t)$, q_t hidden states how do we compute $p(O \mid \lambda)$? We use the forward algorithm to calculate the probability of an observation sequence given a particular HMM.

The forward variable $\alpha_{(t)}(i)$ is defined as:

$$\alpha_t(i) = P(o_1 o_2 \cdots o_t, q_t = s_i \mid \lambda)$$

(3.2)

$\alpha(t)$ stores the total probability of ending up in states s_i at time t, given the observation sequence $o_1 o_2 \cdots o_t$ then the sum of $\alpha_t(i)$ gives the probability of the observation, given the HMM, λ.

$$P(O \mid \lambda) = \sum_{i=1}^{N} \alpha_T(i)$$

(3.3)

The forward variable at each time t is calculated inductively as follows:

1. Initialisation $\alpha_1(i) = \pi_i b_i(o_1), \; 1 \le i \le N$
2. Induction $\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i) a_{ij} \right] b_j(o_{t+1}), 1 \le t \le T-1, 1 \le j \le N$
3. Update time set $t = t + 1$; Return to step 2 if $t < T$; else terminate algorithm.

4. Termination $P(O \mid \lambda) = \sum_{i=1}^{N} \alpha_T(i)$

Full details of the procedure as well as the various implementation issues, are described in Bhar and Hamori (2004), and Rabiner (1989).

Decoding:

Similarly, a model estimate that finds the most probable sequence of hidden states given a sequence of observations is the use of the viterbi algorithm. Let

$$\delta_t(i) = \max P(q_1 q_2 \cdots q_t = s_i, o_1, o_2, \cdots o_t \mid \lambda)$$

(3.4)

be the maximal probability of state sequences of the length *t* that end in state *i* and produce the *t* first observations for the given model. The variable $\delta_t(i)$ stores the probability of observing $o_1, o_2, \cdots o_t$ using the most probable path .The calculation is similar to the forward algorithm, except that the transition probabilities are maximized at each step, instead of summed.

The viterbi algorithm is as follows;

1. Initialization $\delta_1(i) = \pi_i b_i(o_1), 1 \le i \le N, \phi_1(i) = 0$

2. Recursion: $\delta_t(j) = \max[\delta_{t-1}(i) a_{ij}] b_j(o_t), 2 \le t \le T,, 1 \le j \le N$

$$\phi_t(j) = \arg\max[\delta_{t-1}(i) a_{ij}], 2 \le t \le T, 1 \le j \le N$$

3.Completion: $q_T^* = \arg\max[\delta_T(i)]$

4.Most probable state sequence backtracking: $q_t^* = \phi_{t+1}(q_{t+1}^*), t = T-1, T-2, \cdots, 1$

Learning

If we define $\lambda = (A, B, \pi)$ to signify set of HMM, then the algorithm developed by Baum and Welch for signal processing application (see Rabiner 1989) are applied to estimate the model parameters $\lambda = (A, B, \pi)$ that best explains the observation. Implementation of the forward-backward algorithm (Baum-Welch algorithm) works iteratively to improve the likelihood of $p(O \mid \lambda)$. This iterative process is the training of the model. The algorithm is calculated as follows;

1. Initialisation:
   Input initial values of λ and calculate $p(O \mid \lambda)$ using the forward algorithm.

2 . Estimate new values of λ iterate until convergence:

calculate $\gamma_t(i, j) = p(q_t = s_t, q_{t+1} = s_j \mid O, \lambda)$ for each t, i, j using the current λ

$$\gamma_t(i, j) = \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}$$

(3.5)

(a) Calculate new λ parameter estimates using $\gamma_t(i, j)$.

(b) Calculate $p(O \mid \lambda)$ with new λ values.

3. Go to step 4 if two consecutive calculations of $p(O \mid \lambda)$ are equal. Else repeat iterations.

4. Output λ

The parameters of the HMMs are estimated by using equation (3.5). Rabiner (1989) extensively describes the Baum-Welch procedure for parameter estimation, as well as the various implementation issues, are described in Rabiner (1989).

3.2 HMM for ATM Card Fraud Detection

In this study, we propose an ATM card fraud detection system based on Hidden Markov Model, taking the cardholder's spending habit as our point of departure. Basically, we take three different spending profiles of the card holder into consideration [depending upon price range, named High (H), Medium (M) and Low (L)]. In this set of symbols, we define Y = {L, M, H} and N =3. The price range of proposed symbols has taken as low (0, N20,000], medium (N30,000, N90,000] and high (N100,000, up to ATM 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 $\lambda = (A, B, \pi)$ 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 (Baum-Welch algorithm) Welch (2003).

Overall, the procedure of the HMM-based approach can be summarized as follows:

A. Training Phase

Step 1: Train the HMM parameters assuming a probability distribution for the counts for each (hidden) spending profile.

This is important phase of the fraud detection system. In this phase the HMM training starts which follow the following steps:

(i) Initialization of HMM parameters
(ii) Forward procedure
(iii) Backward procedure

For training the HMM, we convert the cardholder's transaction amount into observation symbols and form sequences out of them. At the end of the training phase, we get an HMM corresponding to the cardholder.

B. Detection Phase

At this phase, the proposed model based on HMM will verify fraudulent transactions. It includes two modules as follows:

3.2.1 Clustering

Clustering algorithm is a learning algorithm for grouping a given set of data based on the similarity in their attribute (often called feature) values. The group formed by Mean Clustering algorithm is called cluster. The grouping is formed based on the square of distance and centroid of their data values**.**

Step1: Compute the centroid of the cluster
Step2. Compute the distance between the object to the centroid
Step 3.Grouping is done on the basis of minimum distance between each point.

After the HMM parameters are learned, we form an initial sequence of the existing spending behavior of the card holder. Let $o_1, o_2, \cdots o_Q$ be the sequences of transaction done by the card holder, of length Q. This recorded sequence is formed from the cardholder's transactions up to time t. We put this sequence in HMM model to compute the probability of acceptance.
Let the probability be b_1, which can be calculated as follows.

$$b_1 = P(o_1, o_2, \cdots o_Q)$$

Let O_{Q+1} be the new generated sequence at time t + 1, when a transaction is going to process. To form another sequence of length Q, we drop O_1 and append O_{Q+1} in that sequence, generating $O_2, O_3, \cdots, O_Q, O_{Q+1}$ as the new sequence. We input this new sequence to the HMM and calculate the probability of acceptance by the HMM.
Let the probability of new Q sequences be b_2.

$$b_2 = P(O_2, O_3, \cdots O_{Q+1})$$

Hence, we find the differences in both the old and new sequences to identify whether the transaction is genuine or not. That is,

$$\Delta b = b_1 - b_2$$

If $Diff\ b > 0$, it means that the new sequence is accepted by the HMM with low probability, and therefore, this transaction will be considered a fraudulent transaction if and only if percentage change in probability is greater than a predefined threshold value.

$$\frac{\Delta b}{\Delta b_1} > 0 \quad Threshold \;\; value$$

Otherwise, O_{Q+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 so as to capture the changing spending behavior of a cardholder.

Additionally, the underlying distributions of the states which generate the observed time series (price range) are a priori unknown. Three distributions are of specific interest when we talk about modeling ATM card holder using the transaction pattern. They are as follows: The first HMM is based on the Poisson distribution, which is typically used to model counts. The second HMM uses the generalized Poisson distribution (Joe and Zhu (2005) that includes a further variance parameter to allow for a larger or smaller variation than the one assumed for a standard Poisson distribution and the Guassian based HMM. The Gaussian (or Normal) distribution is the most common (and easily analysed) continuous distribution. It is also a reasonable model in many situations.

In a HMM[Pois] one considers a sequence of discrete observation $\{Y_k\}_{k\geq 0}$ which are assumed to be generated from a sequence of unobservable finite state Markov chains $\{X_k\}_{k\geq 0}$ with a finite state space s = 1, 2, …m, and the random variable Y_t conditioned on X_t has a Poisson distribution for every t; when X_t is in state $i \;\; (i \in S_x; \;\; t \in M)$, then the conditional distribution of Y_t is a Poisson random variable with parameter λ_i; for any $y \in M$, the state dependent probabilities are given by

$$p_{i\,j} = p(Y_t = y \mid X_t = i) = \frac{e^{-\lambda}\lambda_i^y}{y!}$$

The generalized Poisson distribution has the density

$$p(x) = \lambda_1(\lambda_1 + \lambda_2\,.\,x)^{x-1}\frac{\exp(-\lambda_1 - \lambda_2\,.\,x)}{x!}$$

for $x = 0, 1, 2, \cdots b$

with $E(X) = \frac{\lambda_1}{1-\lambda_2}$ and variance $\mathrm{var}(X) = \frac{\lambda_1}{(1-\lambda_2)^3}$

The output probability distribution $b_i(o)$ of the observational data of state i can be discrete or continuous depending on the observations. In continuous distribution HMM for the continuous observational data, the output probability distribution is modeled by a mixture of multivariate Gaussian distributions as follows:

The Gaussian distribution with μ^i and covariance matrix $\sum^i$

$$b_i(o) = p(Y_t = y \mid X_t = i) = N(y, \mu^i, \sum{}^i)$$

4. Application of HMM in credit card fraud detection

4.1 Data:

We apply the above-described methodology to model the ATM card fraud detection on last 100 transactions of a card holder and also calculate percentage of each transaction (low, medium and high) based on total number of transactions. Table 1 contains the transaction that is done by the customer. The amount that is spent by the customer based on which the transaction can be considered as genuine or fraudulent. The most recent transaction is placed at the first position and correspondingly first transaction is placed at the last position in the table and so-on.

Table 1: List of transaction amount of different state

No. of Transaction	Amount	No. of Transaction	Amount	No. of Transaction	Amount	No. of Transaction	Amount
1	6996	26	76891	51	59323	76	587
2	8126	27	78232	52	13392	77	10983
3	12075	28	498209	53	9907	78	5840
4	14478	29	299826	54	10311	79	12120
5	15460	30	442832	55	14748	80	4248
6	80864	31	112918	56	19972	81	13828
7	64953	32	19348	57	19780	82	4636
8	46779	33	10825	58	18316	83	10879
9	50736	34	12467	59	4748	84	2537
10	57514	35	15472	60	5942	85	393704
11	114475	36	2624	61	11004	86	118614
12	196502	37	5769	62	214973	87	68268
13	18975	38	7589	63	173755	88	68309
14	88842	39	4598	64	176584	89	56644

15	17995	40	9885	65	43282	90	36955
16	70722	41	5728	66	36000	91	9752
17	496551	42	61603	67	32856	92	17133
18	11709	43	44904	68	56183	93	8167
19	10924	44	40192	69	84947	94	1210
20	15174	45	88675	70	53091	95	13717
21	16395	46	46527	71	86288	96	11672
22	10303	47	50121	72	80385	97	7824
23	62787	48	83036	73	35119	98	10466
24	81386	49	74197	74	73916	99	16615
25	44722	50	38480	75	13635	100	4609

As indicated in Table 2, to find the observation symbols matching with the cardholder's transactions dynamically, we run a clustering algorithm Montague, (2010) on the values of the cardholder's transaction with cl, cm, and ch as the respective centroids. It may be noted that the naira amounts (0, N20,000] have been clustered together as cl resulting in a centroid of 10797.94. The percentage (p) of total number of transactions in this cluster is thus 52 percent. Similarly, naira amounts (N30,000, N90,000] have been grouped in the cluster cm with centroid 61214.69, whereas amounts (N100,000, up to ATM card limit) have been grouped together in cluster ch with centroid 269911.9. cm and ch, thus, contain 36 percent and 12 percent of the total number of transactions.

Table 2: Output of k-means clustering

Cluster mean/centroid name	CI	Cm	Ch
Observation symbol	$o_1 = L$	$o_2 = M$	$o_3 = H$
Mean value	10797.94	61214.69	269911.9
Percentage of total Transactions	52%	36%	12%

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

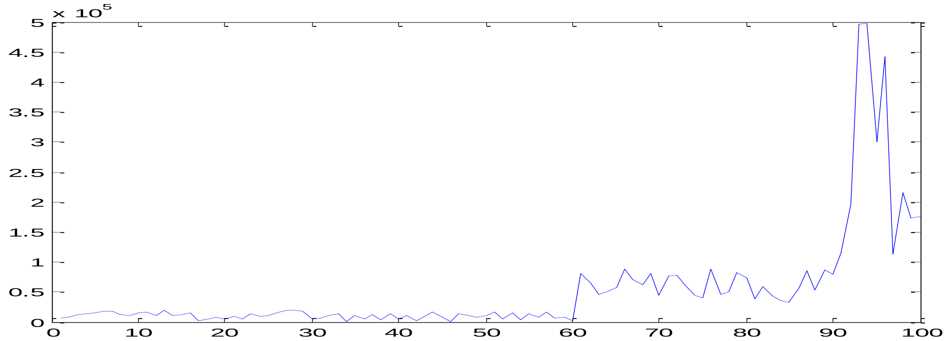

Figure 2: Spending profile of all transactions

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

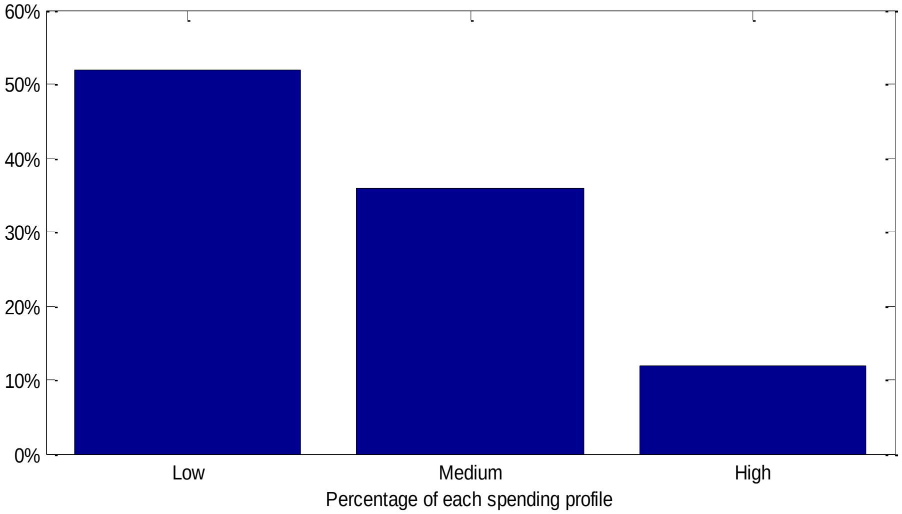

Figure 3: Spending profile of all transactions

It has been noticed that low spending profile has maximum percentage of 52, followed by medium profile 36% and then 12% of high spending profile as per details of transactions in Table 1. Thus, we conclude that the user comes under the cluster 1 or he/she is in low spending profile.

4.2 Evaluation statistic-distribution comparison on technique based on Poisson, Generalised Poisson and Guassian distribution

By calculating the spending pattern of customer Fraud detection of incoming transaction by transition probabilistic calculation, HMM discovers whether the transaction is genuine or fraudulent.

Here, three distribution were used, namely Poisson, Generalised Poisson and Guassian distribution. The performance of the HMM based on the Poisson distribution (HMM[Pois], the generalized Poisson distribution (HMM[GenPois]) and the Gaussian distribution (HMM[Gauss]) for which optimal detection of patterns of anomalies is computed. If it justifies a predefined threshold value then the transaction is decided to be legitimate else declared as fraudulent. In other words, if it is not accepted by our proposed HMM with sufficiently high probability, then it would be a fraudulent transaction.

The performance of all three distributions was assessed in terms of sensitivity (i.e. the fraud detection rate) and false positive (i.e. the misclassification rate- MCR). All models are assessed via measures: “sensitivity” and “false positive”. While sensitivity measures the number of correctly classified positive samples (e.g. fraud) as a proportion of all positive samples in the data, false positive calculates the number of negative samples.

Table 3: Evaluation statistics distribution comparison

Measure	HMM[GenPois]	HMM[Guass]	HMM[Pois]
Sensitivity (%)	70.2	85.6	10
False Positive (MCR) (%)	29.8	14.4	90

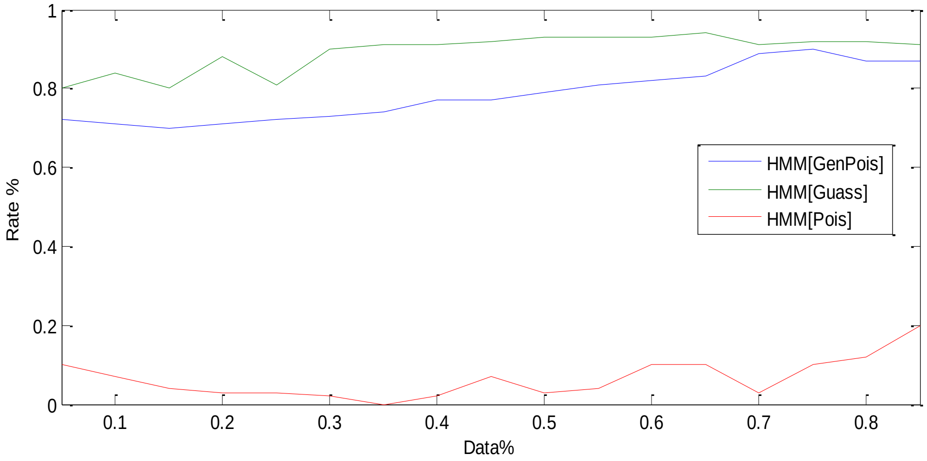

Figure 4: Spending profile of all transactions

Figure 4 reveals how HMM[Gauss] outperforms both the HMM[GenPois] and the HMM[Pois] for the dataset. HMM[Gauss] and HMM[GenPois] attain 85.6% and 70.2% sensitivity consistently after seeing 40% of the database, whereas HMM[Pois] displays the lowest sensitivity rate of 10%. This result shows the ability of HMM[Gauss] search to employ link analysis to consistently detect fraudulent activity and focus on it as the dataset increases.

The MCR shows that less than 15% for HMM[Gauss] along with HMM[GenPois] with about 30% and the HMM[Pois] which shows the highest misclassification rate with about 90%; which means it misclassifies non-fraud samples frequently.

5 Conclusions

This study puts forward a HMM using transaction patterns for ATM card fraud detection. It modelled the sequence of transactions with a HMM based on the Poisson distribution

(HMM[Pois]), the generalized Poisson distribution (HMM[GenPois]), and the Gaussian distribution (HMM[Gauss]) for which optimal detection of patterns of anomalies is computed using the forward-backward algorithm. The suggested estimation procedure based upon the three distributions for the HMM model is used to construct a sequence of operations in ATM card transaction processing, and detect fraud by studying spending profile of the cardholder. If the transaction satisfies a predefined threshold value, then the transaction is confirmed legitimate; else, the transaction is declared fraudulent.

In our implementation, we took three observation symbols which are spending ranges of a cardholder that are low, medium, and high. To find the observation symbols corresponding to the cardholder's transactions dynamically, we run a clustering algorithm on the values of each cardholder's transaction with cl, cm, and ch as the respective centroids. An HMM is trained with forward-backward algorithm (Baum-Welch algorithm) for the cardholder. The functions offered by MATLAB facilitated us to develop the techniques based on the Poisson distribution (HMM[Pois]), the generalized Poisson distribution (HMM[GenPois]), and the Gaussian distribution (HMM[Gauss]) for fraudulent ATM card use. The experimental result of the data analyses confirms that the proposed method is viable. The evaluation statistics are calculated to compare the fit of distributions. Of the HMM-based techniques, HMM[Gauss] proved to be the most suitable choice in detecting ATM card fraudulent transactions as demonstrated by the sensitivity value, having 85.6% sensitivity consistently after seeing 40% of the database and MCR value having the best MCR with less than 15%.

5. References

Aleskerov, E., Freisleben B., and Rao B. (1997), “CARDWATCH”: A Neural Network Based Database Mining System for Credit Card Fraud Detection”, Proc. IEEE/IAFE: Computational Intelligence for Financial Eng., 220-226.

Ashphak, P. K, Vinod, S. M. ,Shehzad H. S., Akash B. K. (2013). Credit Card Fraud Detection System through Observation Probability Using Hidden Markov Model. International Journal of Thesis Projects and Dissertations 1(1): 7-16.

Bhar R., & Hamori S. (2004) Hidden Markov Models: Application to Financial Economics. Dordrecht, kluwer Academic Publishers.

Bhusari, V. and Patil, S. (2011). Study of Hidden Markov Model in Credit Card Fraudulent Detection. International Journal of Computer Applications, 20(5):33-36.

Brause R., Langsdorf, T., and Hepp M., (1999) “Neural Data Mining for Credit Card Fraud Detection”, Proc. IEEE Int’l Conf. Tools with Artificial Intelligence, 103-106.

Bolton R., and Hand D., (2001). "Unsupervised Profiling Methods for Fraud Detection", Credit Scoring and Credit Control VII.

Chen R., Chiu M., Huang Y., and Chen L., "Detecting Credit Card Fraud by Using Questionaire- Responded Transaction Model Based on Support Vector Machines", Proceedings of IDEAL, 800-806.

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

Cho S. B. and Park, H. J. (2003). "Efficient Anomaly Detection by Modeling Privilege Flows Using Hidden Markov Model," Computer and Security, 22(1), 45-55.

Dorronsoro J.R., Francisco G., Carmen S., and Carlos S.C., (1997). "Neural Fraud Detection in Credit Card Operation." IEEE Transaction on Neural Network, 8(4), 827-834.

Montague, D. A. (2010). Fraud Prevention Techniques for Credit Card Fraud.

Ehramikar S., (2000). "The Enhancement of Credit Card Fraud Detection Systems using Machine Learning Methodology", MASc Thesis, Department of Chemical Engineering, University of Toronto

Elliott R. J., Aggoun, L. and Moore, J. B. (1995). Hidden Markov Models: Estimation and Control. Springer - Verlag New York.

Foster D., and Stine R., (2004) . Variable Selection in Data Mining: Building a Predictive Model for Bankruptcy", Journal of American Statistical Association, 303-313.

Ghosh S. and Reilly, D. L. (1994). Credit Card Fraud Detection with a Neural-Network," Proceedings of the 27th Hawaii International Conference on System Science: Information Systems: Decision Support and Knowledge Based Systems, 3, 621-630.

Hoang, X.D., Hu, J. (2004). An Efficient Hidden Markov Model Training Scheme for Anomaly Intrusion Detection of Server Applications Based on System Calls. In: Proc. of 12th IEEE Conference on Networks, vol. 2, 470-474.

Jadhav, S. N and Bhandari, K. (2013). Anomaly Detection Using Hidden Markov Model. International Journal of Computational Engineering Research 3(7): 28-35

Joe H. and Zhu R. (2005). Generalized Poisson distribution: the property of mixture of Poisson and comparison with negative binomial distribution. Biometrical 47(2), 219–229.

Joshi S. S. and V.V. Phoha, V. V. (2005). "Investigating Hidden Markov Models Capabilities in Anomaly Detection," Proc. 43rd ACM Ann. Southeast Regional Conf., vol. 1, 98-103.

Kokkinaki, A. (1997). "On Atypical Database Transactions: Identification of Probable Frauds using Machine Learning for User Profiling." Knowledge and Data Engineering Exchange Workshop. IEEE, 107-113.

Kim, M, and Kim T., (2002). A Neural Classifier with Fraud Density Map for Effective Credit Card Fraud Detection", Proceedings of IDEAL, 378-383.

Lane, T. (1999). "Hidden Markov models for human/computer interface modeling,". [Online]. Available: citeseer.nj.nec.com/lane99hidden.html.

MacDonald I. L. and Zucchini W. (1997): Hidden Markov and other Models for Discrete-Valued Time Series. Vol., 70 of Monographs on Statistics and Applied Probability. London: Chapman & Hall.

Maes, S., Karl T., Bram V., Bernard M., (1993). Credit card fraud detection using Bayesian and neural networks", Interactive image-guided neurosurgery, 261-270.

Mohdavesh, Z. K., Jabir, D. P., Ali, H. E. A. (2014). Credit Card Fraud Detection System Using Hidden Markov Model and K-Clustering. International Journal of Advanced Research in Computer and Communication Engineering, 3(2), 5458-5461.

Stolfo, S. J., Fan, D.W., Lee W., (1997). Prodromidis A.L., and Chan P.K., "Credit Card FraudDetection Using Meta-Learning: Issues and Initial Results", Proc. AAAI Workshop AI Methods in Fraud and Risk Management, 83-90.

Singh P. and Singh M. (2015). Fraud Detection by Monitoring Customer Behavior and Activities. International Journal of Computer Applications, 111(11), 23-32.

Rabiner, L.R. and Juang, B.H., (1986). An Introduction to Hidden Markov Models, *IEEE ASSP Magazine, 3* (1), 4-16.

Rabiner, L. R. (1989). A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, in Proceedings of the IEEE, 77 (2), 257-286.

Welch, L. R. (2003). "Hidden Markov Model and Baum-Welch algorithm" IEEE Information Theory Society Newsletter Vol. 53, No.4

Wheeler R., and Aitken S., (2000). "Multiple algorithms for fraud detection", Knowledge-Based Systems, no. 13, 93-99.