CHAPTER 4

EFFICIENT METHODS FOR IMPROVING CREDIT CARD FRAUD DETECTION

Nowadays the customers preferring, the most accepted payment mode via credit card for the convenient way of online shopping, paying bills in easiest way. At the same time the fraud transaction risks using credit card is a main problem which should be avoided. There are many data mining techniques available to avoid these risks effectively. In existing research they have implemented the concept of Hidden Markov model. In that research, they modelled the sequence of operations in credit card transaction processing using a Hidden Markov Model (HMM) and shown how it can be used for the detection of frauds. An HMM is initially trained with the normal spending profile of the cardholder. If an incoming credit card transaction is not accepted by the trained HMM with sufficiently high probability and with threshold value, it is considered to be fraudulent.

To provide better accuracy and to avoid computational complexity in fraud detection in proposed work semi hidden Markov model (SHMM) algorithm of anomaly detection is presented which computes the distance between the processes monitored by credit card detection system and the perfect normal processes. With this we are implementing another method for fraud detection is that having a key idea is to factorize marginal log-likelihood using a variation distribution over latent variables. An asymptotic approximation, a factorized information criterion (FIC) obtained by applying the Laplace method to each of the factorized components. Our experimental
results demonstrate that we can significantly reduce loss due to fraud through distributed data mining of fraud models.

4.1 INTRODUCTION

As usage of credit cards become more and more common in every field of the daily life, credit card fraud has become much more rampant. To improve security of the financial transaction systems in an automatic and effective way, building an accurate and efficient credit card fraud detection system is one of the key tasks for the financial institutions. A semi hidden Markov HMM (more properly called a SHMM) is similar to HMM except that each state in SHMM can emit a sequence of observations. Because of this difference, the duration probability density of a state in SHMM can be an arbitrary distribution. Based on this SHMM, an algorithm of anomaly detection is presented in this research, which computes the distance between the processes monitored by credit card detection system and the perfect normal processes.

In this algorithm, we use the average information entropy (AIE) of fixed-length observed sequence as the anomaly detection metric based on maximum entropy principle (MEP). To improve accuracy, the segmental K-means algorithm is applied as training algorithm for the SHMM. The system that has been developed can be used in banks in their main server to secure their customers. In this system the each and every transaction is loaded in the main data base and can be checked to determine whether transaction is normal or anomaly. If it detects as anomaly then user may be asked to enter a password provided by the bank. Instead of hidden markov model can implement semi markov model, which increase the accuracy of detection process.
The multiple observation probability is expressed as a combination of individual observation probabilities without losing generality in the Multiple HMM (Hidden Markov Model) technique. The independence-dependence property of the observations is characterized by the combinatorial weights, and hence it gives us more freedom in making different assumptions and also in deriving corresponding training equations. The maximum valuable will be selected from those results. From those results the deviation of user selection will be detected as a fraud.

Another method for fraud detection is that having a key idea is to factorize marginal log-likelihood using a variation distribution over latent variables. An asymptotic approximation, a factorized information criterion (FIC), is obtained by applying the Laplace method to each of the factorized components. In order to evaluate FIC, we propose factorized asymptotic Bayesian inference (FAB), which maximizes an asymptotically-consistent lower bound of FIC. FIC and FAB have several desirable properties:

1) Asymptotic consistency with the marginal log-likelihood,

2) Automatic component selection on the basis of an intrinsic shrinkage mechanism, and

3) Parameter identifiably in mixture modelling. We want to maximize the marginal log likelihood of the observed data.

In probability theory and statistics, the marginal distribution the subset of a collection of random variables is the probability distribution of the variables contained in the subset. The term marginal variable is used to refer to those variables in the subset of variables being retained. These terms are dubbed "marginal" because
they used to be found by summing values in a table along rows or columns, and writing the sum in the margins of the table. The distribution of the marginal variables (the marginal distribution) is obtained by marginalizing over the distribution of the variables being discarded those are said to have been marginalized out. The main contribution of their work is as follows:

1. Collecting the card holder’s information and maintain that in a central database

2. Based on the spending behavioural profile of card holder the probabilities of initial set have been chosen and construct a sequence observation

3. The Semi Hidden Markov is used to train the stored dataset and using this the testing data will be verified to find the anomaly

4. With that we are combining the features of the FIC method to improve the detection accuracy

### 4.2 CREDIT CARD FRAUD DETECTION USING HMM

The credit card fraud detection system is based on Hidden Markov Model, which does not require fraud signatures and still it is capable to detect frauds just by bearing in mind a cardholder’s spending habit [10]. The particulars of purchased items in single transactions are generally unknown to any Credit card Fraud Detection System running either at the bank that issues credit cards to the cardholders or at the merchant site where goods is going to be purchased [12]. As business processing of credit card fraud detection system runs on a credit card issuing bank site or merchant site. Each arriving transaction is submitted to the fraud detection system for verification purpose [11]. The fraud detection system accept the card details such as
credit card number, CVV number, card type, expiry date and the amount of items purchase to validate, whether the transaction is genuine or not [13].

The implementation techniques of Hidden Markov Model in order to detect fraud transaction through credit cards, it create clusters of training set and identify the spending profile of cardholder [10]. In that process the number of items purchased by customers, types of items that are bought in a particular transaction concentrates on the amount of item purchased and use for further processing [14] that are not known to the Fraud Detection system absolutely. It stores higher amount of different data transactions in form of clusters depending on transaction amount which will be either in low, medium or high value ranges. It tries to find out any variance in the transaction based on the spending behavioural profile of the cardholder, shipping address, and billing address and so on [9]. Based on the spending behavioural profile of card holder the probabilities of initial set have been chosen and construct a sequence for further processing. If the fraud detection system makes sure that the transaction to be of fraudulent, it raises an alarm, and the issuing bank declines the transaction [9].

For the security purpose, the Security information module will get the information features and its store’s in database [9]. If the card lost then the Security information module form arises to accept the security information. The security form has a number of security questions like account number, date of birth, mother name, other personal question and their answer, etc. where the user has to answer it correctly to move to the transaction section [8] in which all those information must be known by the card holder only and can proceed only by the card holder. It has informational privacy and informational self determination that are addressed evenly by the
innovation affording people and entities a trusted means to user, secure, search, process, and exchange personal and/or confidential information [10].

The system and tools for pre-authorizing business provided that a connections tool to a retailer and a credit card owner [15]. By communicating to a credit card number, card type with expiry date and storing it into database, a distinctive piece of information that characterizes a particular transaction to be made by an authoritative user of the credit card at a later time [9] the cardholder will be initiating a credit card transaction process. The details are received in the type of network data in the database only when if an accurate individual recognition code is used with the communication [7] the cardholder can precede with further steps with the credit card. Since the transaction is pre-authorized, the vendor does not need to see or transmit an accurate individual recognition code [11].

4.3 HMM BACKGROUND

An HMM is a double embedded stochastic process with two hierarchy levels. It can be used to model much more 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 [16]. In recent years, Joshi and Phoba [2] have investigated the capabilities of HMM in anomaly detection. They classify TCP network traffic as an attack or normal using HMM. Cho and Park [3] suggest considering only the privilege transition flows based on the domain knowledge of attacks of an HMM-based intrusion detection system that improves the modelling time and performance.
Ourston et al. [4] have proposed the application of HMM in detecting multistage network attacks. Hoang et al. [5] present a new method to process sequences of system calls for anomaly detection using HMM. Based on both HMMs and enumerating methods for anomaly detection this is providing a multilayer model of program behaviours. Lane [6] has used HMM to model human behaviour based on this if human behaviour is correctly modelled once, any detected deviation is a cause for concern since an attacker is not expected to have a behaviour similar to the genuine user. Hence, an alarm is raised in case of any deviation. An HMM can be characterized by the following [16]:

1. The number of states in the model is \( N \). The set of states is \( S = \{ S_1, S_2, \ldots, S_N \} \), where \( S_i \), \( i = 1, 2, \ldots, N \) is an individual state. The state at time instant \( t \) is denoted by \( q_t \).

2. The number of distinct observation symbols per state is \( M \). The set of symbols is \( V = \{ V_1, V_2, \ldots, V_M \} \), where \( V_i \), \( i = 1, 2, \ldots, M \) is an individual symbol.

3. The state transition probability matrix \( A = [a_{ij}] \) where \( a_{ij} = P(q_{t+1} = S_j \mid q_t = S_i) ; 1 \leq i \leq N; 1 \leq j \leq N; t = 1, 2, \ldots; \) where \( a_{ij} > 0 \) for all \( i,j \).

Also, \( \sum_{j=1}^{N} a_{ij} = 1, 1 \leq i \leq N. \)

4. The observation symbol probability matrix \( B = [b_j(k)] \), where \( b_j(k) = P(V_k \mid q_t = S_j); 1 \leq j \leq N, 1 \leq k \leq M \) and \( \sum_{k=1}^{M} b_j(k) = 1, 1 \leq j \leq N. \)

5. The initial state probability vector \( \pi = [(\pi_i)] \), where \( \pi_i = P(q_1 = S_i); 1 \leq j \leq N, \) such that \( \sum_{k=1}^{M} \pi_i = 1 \)

6. The observation sequence \( O = O_1, O_2, O_3 \ldots O_R \), where each observation \( O_t \) is one of the symbols from \( V \), and \( R \) is the number of observations in the
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. An observation sequence \( O \), as mentioned above, can be generated by many possible state sequences. Consider one such particular sequence \( Q = q_1, q_2, \ldots, q_R \) where \( q_1 \) is the initial state.

The probability that \( O \) is generated from this state sequence is given by

\[
P(O \mid Q, \lambda) = \prod_{t=1}^{R} P(O_t \mid q_t, \lambda)
\]

where statistical independence of observations is assumed. Above equation can be expanded as

\[
P(O \mid Q, \lambda) = b_{q_1}(O_1) b_{q_2}(O_2) \cdots b_{q_R}(O_R)
\]

The probability of the state sequence \( Q \) is given as

\[
P(Q \mid \lambda) = \pi a_{q_1} a_{q_2} \cdots a_{q_R-1}
\]

Thus, the probability of generation of the observation sequence \( O \) by the HMM specified by \( \lambda \) can be written as follows:

\[
P(Q \mid \lambda) = \sum_{\text{all } Q} P(O \mid Q, \lambda) P(Q \mid \lambda)
\]

Deriving the value of \( P(Q \mid \lambda) \) using the direct definition of above equation is computationally intensive. After the HMM parameters are learned, we take the symbols from a cardholder’s training data and form an initial sequence of symbols. Let \( O_1, O_2, O_3 \ldots O_R \) be one such sequence of length R. This result of recorded sequence is formed from the cardholder’s transactions up to time t. They given this
input sequence to the HMM and compute the probability of acceptance by the HMM. Let the probability be \( \alpha \), which can be written as follows:

\[
\alpha = P(O_1, O_2, O_3 \ldots O_R | \lambda)
\]

This probability calculation was done for all \( \alpha_{R+1} \) and stored these results. If \( \Delta \alpha > 0 \), it means that the new sequence is accepted by the HMM with low probability, and it could be a fraud. The newly added transaction is determined to be fraudulent if the percentage change in the probability is above a threshold, that is,

\[
\frac{\Delta \alpha}{\alpha} \geq \text{Threshold}
\]

### 4.4 CREDIT CARD SYSTEM USING SEMI HIDDEN MARKOV MODEL

In this research, the health states of components were modelled using state transition probability, state duration probability and observation probability using a semi hidden Markov chain (SHMM). A modified forward–backward algorithm for SHMMs was used to estimate the parameters of SHMMs using discriminate function analysis the weight will be determined. For prognosis of remaining life, a state duration model-based prediction calculation procedure was provided. The results show that the SHMMs can provide valuable timing information in the single operator case, whereas HMMs tend to be more robust to increased team complexity.

#### 4.4.1 Detection algorithm

From maximum entropy principle (MEP), we know that when a computer system is running in normal state, the audit data it generates contains less information than that it generates when running in fraud state. Namely, the information entropy of fraud state is larger than that of normal state, so the information entropy can act as the
metric in credit card fraud detection. But when the length of visible symbol sequence increases, the information entropy of visible symbol sequence will become larger and larger. It only makes sense to compare the value of information entropy among the same-length sequences. In order to use entropy metric variable-length symbol sequences, we compute the average information entropy (AIE) of visible symbol sequences, and use it as the metric to distinguish between normal behaviour and anomaly behaviour.

4.4.2 Training algorithm

Because the normal state of a computer system may change over time, so training of hidden semi-Markov model is also an important part in anomaly detection algorithms. For the hidden semi-Markov model \( k = (N, M, V, A, B, \pi) \) we constructed in previous section, both the distribution of state transfer probabilities \( A \) and the initial distribution of normal state and fraud state \( p \) are fixed values, so only the distribution of visible symbol for normal behaviour \( B_0 = \{b_0(k)\}, 1 \leq k \leq M \) need to be updated. The training can be implemented by system administrator on normal data sequences.
Figure 4.1: System Flow Diagram
4.5 ALGORITHM STEPS

TRAINING PHASE: Cluster creation

STEP 1: Identify the profile of cardholder from their purchasing

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

STEP 3: Construct the training sequence for training model

DETECTION PHASE: Fraud detection

STEP 1: Generate the observation symbol $B_{R+1}$

STEP 2: Form new sequence by adding $B_{R+1}$ in existing sequence

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

STEP 4: If both are same it will be a normal customer else there will be fraud signal will be provided.

4.6 Factorized Information Criterion (FIC)

We have proposed approximation of marginal log likelihood (FIC) and an inference method (FAB) for Bayesian model selection for mixture models, as an alternative to VB inference. We have given their justifications (asymptotic consistency, convergence, etc) and analyzed FAB mechanisms in terms of over-fitting mitigation (shrinkage) and identify. Experimental results have shown that FAB outperforms state-of-the-art VB methods for a practical number of data in terms of both model selection performance and computational efficiency.
Our key idea is to factorize marginal log-likelihood using a variation distribution over latent variables. An asymptotic approximation, a factorized information criterion (FIC), is obtained by applying the Laplace method to each of the factorized components. In order to evaluate FIC, we propose factorized asymptotic Bayesian inference (FAB), which maximizes an asymptotically-consistent lower bound of FIC. FIC and FAB have several desirable properties: 1) asymptotic consistency with the marginal log-likelihood, 2) automatic component selection on the basis of an intrinsic shrinkage mechanism and 3) parameter identify in mixture modelling. Experimental results show that FAB outperforms state-of-the-art VB methods.

4.7 Maximum Loglikelihood

Since the introduction of the NML universal model in the context of MDL, there has been significant interest in the evaluation of NML stochastic complexity for different practically relevant model classes, both exactly and asymptotically. For discrete models, exact evaluation is often computationally a sum over all possible data-sets. For continuous cases, the normalizing coefficient is an integral which can be solved in only a few cases. Under certain conditions on the model class, different versions of stochastic complexity (which include two part, mixture, and NML forms) have the same asymptotic form, the so called Fisher information approximation. However, for small data-sets and for model classes that do not satisfy the necessary conditions, the asymptotic form is not accurate.

For many interesting model classes, such as Bayesian networks, the mini-max regret optimal normalized maximum likelihood (NML) universal model is
computationally very demanding. We suggest a computationally feasible alternative to NML for Bayesian networks, the factorized NML universal model, where the normalization is done locally for each variable. This can be seen as an approximate sum-product algorithm. We show that this new universal model performs extremely well in model selection, compared to the existing state-of-the-art, even for small sample sizes.

4.8. Experimental results

The performance of the proposed approach is evaluated by using the credit card transaction data set.

4.8.1. Credit Card Transaction Dataset

The credit card data set is a transaction data set which includes a sequence of transaction details which are made by the users. Data set consists of the transaction attributes and the details of persons who are doing corresponding transaction. These data sets consists of transaction attributes, for example user details who are doing transaction, date of transaction, type of transaction and the date and place of transaction. By analysing these attributes one can learn the fraudulent behaviour.

MATLAB (MATrix LABoratory) is used for the computation of the numerical analysis and is considered as a fourth generation programming language. It is a feasible Matrix Laboratory package which functions as an interactive programming environment. Hence, for the present research MATLAB has been taken into consideration and all the four techniques have been implemented using MATLAB.

The evaluation of the proposed approach in Fraud detection in credit card transaction system was performed based on the following factors.
• Precision
• Recall
• F-Measure

4.8.2 Precision Comparison

We analyze and compare the performance offered by HMM, SHMM and SHMM with FIC. Here if the no of data sizes increased the precision accuracy also increased linearly while transaction. The precision accuracy of the proposed SHMM with FIC is high. Based on the comparison and the results from the experiment shows the proposed approach works better than the other existing systems with higher rate. The values are represented in the Table 4.2.

<table>
<thead>
<tr>
<th>S. No</th>
<th>No. of datasets</th>
<th>HMM</th>
<th>SHMM</th>
<th>SHMM with FIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100</td>
<td>0.75</td>
<td>0.85</td>
<td>0.92</td>
</tr>
<tr>
<td>2</td>
<td>200</td>
<td>0.69</td>
<td>0.75</td>
<td>0.89</td>
</tr>
<tr>
<td>3</td>
<td>300</td>
<td>0.58</td>
<td>0.64</td>
<td>0.76</td>
</tr>
<tr>
<td>4</td>
<td>400</td>
<td>0.45</td>
<td>0.56</td>
<td>0.69</td>
</tr>
<tr>
<td>5</td>
<td>500</td>
<td>0.39</td>
<td>0.49</td>
<td>0.62</td>
</tr>
<tr>
<td>6</td>
<td>600</td>
<td>0.33</td>
<td>0.47</td>
<td>0.58</td>
</tr>
</tbody>
</table>
The above graph indicates that the precision value is improved in the proposed methodology of SHMM-FIC than the existing researches. In this graph numbers of data’s are predicted in the x axis and the precision value is predicted in the y axis. The precision value is used to denote the correctness where it is higher in the proposed methodology.

### 4.8.3 Recall Comparison

We analyse and compare the performance offered by HMM and SHMM. Here if the no of datasets increased the recall rate also increased linearly. The recall rate of the proposed SHMM with FIC is high. Based on the comparison and the results from the experiment show the proposed approach works better than the other existing systems. The values are given below as a table form in Table 4.2.

---

![Size of Training data set Vs Precision](image-url)

**Figure 4.2: Number of Dataset vs. Precision**
Table 4.2: Number of Dataset vs. Recall Rate

<table>
<thead>
<tr>
<th>S. No</th>
<th>No of datasets</th>
<th>HMM</th>
<th>SHMM</th>
<th>SHMM with FIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100</td>
<td>0.75</td>
<td>0.87</td>
<td>0.95</td>
</tr>
<tr>
<td>2</td>
<td>200</td>
<td>0.67</td>
<td>0.77</td>
<td>0.89</td>
</tr>
<tr>
<td>3</td>
<td>300</td>
<td>0.58</td>
<td>0.69</td>
<td>0.78</td>
</tr>
<tr>
<td>4</td>
<td>400</td>
<td>0.48</td>
<td>0.58</td>
<td>0.7</td>
</tr>
<tr>
<td>5</td>
<td>500</td>
<td>0.41</td>
<td>0.51</td>
<td>0.65</td>
</tr>
<tr>
<td>6</td>
<td>600</td>
<td>0.38</td>
<td>0.48</td>
<td>0.59</td>
</tr>
</tbody>
</table>

Figure 4.3: Number of Dataset vs. Recall rate

The above graph indicates that the precision recall value is improved in the proposed methodology of SHMM-FIC than the existing researches. In this graph numbers of data’s are predicted in the x axis and the recall value is predicted in the y axis.
axis. The recall value is used to denote the correctness where it is higher in the proposed methodology.

4.8.4 F-Measure Comparison

We analyze and compare the F measure offered by HMM, SHMM and SHMM with FIC. Here if the no of datasets increased the recall rate also increased linearly. The F measure rate of the proposed SHMM with FIC is high. Based on the comparison and the results from the experiment show the proposed approach works better than the other existing systems. The values for the graph are given below as a table form in Table 4.3.

Table 4.9.3: Number of Dataset vs. F- Measure

<table>
<thead>
<tr>
<th>S. No</th>
<th>No of datasets</th>
<th>HMM</th>
<th>SHMM</th>
<th>SHMM with FIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100</td>
<td>0.67</td>
<td>0.75</td>
<td>0.92</td>
</tr>
<tr>
<td>2</td>
<td>200</td>
<td>0.56</td>
<td>0.69</td>
<td>0.85</td>
</tr>
<tr>
<td>3</td>
<td>300</td>
<td>0.48</td>
<td>0.62</td>
<td>0.76</td>
</tr>
<tr>
<td>4</td>
<td>400</td>
<td>0.40</td>
<td>0.53</td>
<td>0.66</td>
</tr>
<tr>
<td>5</td>
<td>500</td>
<td>0.34</td>
<td>0.45</td>
<td>0.59</td>
</tr>
<tr>
<td>6</td>
<td>600</td>
<td>0.23</td>
<td>0.39</td>
<td>0.56</td>
</tr>
</tbody>
</table>
The above graph indicates that the precision-Measure value is improved in the proposed methodology of SHMM-FIC than the existing researches. In this graph numbers of data’s are predicted in the x axis and the F-Measure value is predicted in the y axis. The F-Measure value is used to denote the correctness where it is higher in the proposed methodology.

4.9 SUMMARY

In this research, SHMM is introduced into credit card fraud detection systems. We present an algorithm of fraud detection based on SHMM, which computes the distance between the processes monitored by intrusion detection system and the perfect normal processes. In this algorithm, based on MEP, we introduce the concept AIE, which is used as detection metric via analyzing variable-length observed symbol sequences. To improve accuracy, propose a new approximation inference algorithm and refer to it as a factorized asymptotic Bayesian inference (FAB) with the SHMM.