A Novel Hidden Markov Model for Credit Card Fraud Detection

A.Prakash
MSC(CT),Mhil
Research Scholor
Manonmaniam Sundaranar University,
Tirunelveli

C.Chandrasekar, PhD.
Reader
Department of Computer Science
Periyar Universty
Salem

ABSTRACT

Nowadays the customers prefer 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 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. 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.

Index Terms—Hidden Markov Model, Semi Hidden Markov Model, Factorized Information Criterion, Maximum Entropy Principle, Average Information Entropy.

1. INTRODUCTION

As usage of credit cards become increasingly common in all field of the day by day life, credit card fraud has become greatly more uncontrolled. In an automatic and efficient way to get better security of the financial transaction systems, construction of an accurate and competent credit card fraud detection system is one of the key assignments for the financial institutions. A semi-Markov HMM (more suitably called a hidden Hidden-Markov model or SHMM) is comparable to HMM excluding that each state in SHMM can emit a sequence of observations. Since of this difference, the extent probability density of a state in SHMM can be a subjective distribution. Based on this SHMM, an algorithm of anomaly detection is provided in this research, which computes the distance between the processes monitored by credit card detection system and the perfect normal processes. We use the average information entropy (AIE) of fixed-length observed sequence as the anomaly detection metric, based on maximum entropy principle (MEP). Toward advance accuracy, the segmental K-means algorithm is applied as training algorithm for the SHMM. The system that has been developed can be employed in banks in their main server to make safe their customer's information. In this system the each and every transaction is loaded in the central database and can be ensured to find out whether transaction is ordinary or irregularity. If it senses as anomaly then customer may be inquired to enter a password provided by the bank. As an alternative of hidden markov model we can implement semi hidden markov model, which enhance the accuracy of detection process.

Another method for fraud detection is that having a key idea is to factorize marginal log-likelihood using a variation distribution over hidden variables. An asymptotic estimate, a factorized information criterion (FIC), is obtained by applying the Laplace method to each of the factorized factors. So as to evaluate FIC, we propose factorized asymptotic Bayesian inference (FAB), which exploits an asymptotically-consistent lower bound of FIC. FIC along with FAB have several enviable properties: 1) asymptotic reliability with the insignificant log-likelihood, 2) routine factor selection on the basis of an essential shrinkage mechanism, and 3) parameter identifiability in mixture modeling. We want to maximize the marginal log likelihood of the experimental data. The marginal distribution of the subset a collection of random variables is the probability distribution of the variables contained in the subset in probability theory and statistics. The term marginal variable is used to refer to those variables in the subset of variables being preserved. 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 gained by marginalizing over the distribution of the variables being discarded those are said to have been marginalized out.

The main contribution of the work is as follows:

1. Collecting the card holder's information and maintain that in a central database
2. Based on the spending behavioral 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

2. RELATED WORK

HMM-based purposes are common in different areas such as speech recognition, bioinformatics, and genomics. In recent years, Joshi and Phoba [1] have explored the capabilities of

HMM in abnormality detection. Cho and Park [2] proposed that improves the modeling time and performance which is an HMM-based intrusion detection system by considering only the privilege transition flows based on the domain knowledge of attacks. Ourston et al. [3] have suggested the application of HMM in identifying multistage network attacks. Hoang et al. [4] present a new method to process series of system calls for anomaly detection using HMM. Lane [5] has used HMM to model human behavior. Once human behavior is correctly modeled, any detected divergence is a cause for concern. Fujimaki and Morinaga [14] include recently a new Bayesian approximation inference method for fusion models. Through the factorized information criterion (FIC) and factorized asymptotic Bayesian inference (FAB). Each condition of the Markov chain gives rise to an release function of visible events (Rabiner and Juang, 1986)[15]. Hidden-Markov models are a version of HMMs capable of explicitly modeling the timing of state transitions is dealt in (Guedon, 2003) [16]. In cooperation HMMs and SHMMs have been shown to be competent of capturing time-varying signal characteristics by statistically sculpting the underlying active of the signal (Rabiner, 1989) [15]. Importantly, HMMs and SHMMs are interpretable because (1) the emission function of each hidden state is expressed explicitly over the space of observable events and (2) the changes between hidden states are as well clearly modeled. By reviewing all the literature details there is a detection of accuracy is lacking.

3. CREDIT CARD FRAUD DETECTION USING HMM

Based on Hidden Markov Model The credit card fraud detection system is not require fraud signatures and still it is able to sense frauds just by accepting in mind a cardholder's spending routine [9]. The exacting of acquired items in single transactions is broadly unidentified to any Credit card Fraud Detection System management moreover at the bank that issues credit cards to the cardholders or at the merchant site where goods is going to be purchased [11]. As trade processing of credit card fraud detection system runs on a credit card issuing bank site or commercial site. Each incoming transaction is surrendered to the fraud detection system for verification purpose [10]. The fraud detection system accept the card details such as credit card number, card type, end date and the quantity of items purchase to authenticate, whether the transaction is genuine or not [12].

The implementation techniques of Hidden Markov Model, it makes clusters of training set and identifies the spending profile of cardholder in order to sense fraud transaction through credit cards [9]. In that procedure the number of items purchased by customers, categories of items that are bought in a particular transaction focuses on the quantity of item purchased and use for supplementary processing [13] 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 conflict in the transaction based on the expenditure behavioral profile of the cardholder, transport address, and billings address and so on [8]. Based on the spending behavioral 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 [8].

The Security information module will get the information features for the security purpose and its store's in database [8]. The Security information component form arises to accept the security information if the card lost. There will be a number of security questions in the security form like account number, date of birth, mother name, other personal question and their answer, etc. where the user has to react it properly to shift to the transaction section [7] 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 resolve that are accounted evenly by the originality giving people and entities a trusted means to user, secure, search, process, and substitute personal and/or confidential information [9]. The information are collected in the type of network data in the database only when if an accurate individual recognition authorized code is used with the communication [6] the cardholder can follow-up with further steps with the credit card. The vendor does not need to see or transmit an accurate since the transaction is pre-authorized, individual recognition code [10].

An HMM can be characterized by the following [15]:

1. The number of states in the model is N. The set of states is $S = \{S_1, S_2, \dots S_N\}$, where $\mathrm{S_i}$, $i = 1, 2, \dots, N$ is an individual state. The state at time instant t is referred by $\mathrm{q_t}$.
2. The number of distinct observation symbols per state is M. The set of symbols is $V = \{V_1, V_2, \dots V_3\}$, where V_i, $i = 1;\ 2;\ \dots;\ M$ is an individual symbol.
3. The state transition probability matrix $A = [a_{ij}]$ where $a\_ij = P(q_t + 1 = S_j \mid q_t = S_i);\ 1 \le i \le N;\ 1 \le j \le N;\ t = 1, 2, \dots : N$ where $a_{ij} > 0$ for all i, j. Also, $\sum_{j=1}^{N} a_{ij} = 1{,}1 \le i \le N$.
4. The observation symbol probability matrix $B = [b_j(k)]$, where $b_j(k) = P(V_k \mid S_j){,}1 \mid \le j \le N, 1 \le k \le M$ and $\sum_{k=1}^{M} b_j\,(k) = 1{,}1 \le j \le N$
5. The initial state probability vector $\pi = [(\pi_i)]$, where $\pi_i = P(q_1 = S_i){,}1 \le j \le N$, such that $\sum_{k=1}^{M} \pi_i = 1$
6. The observation sequence $\mathrm{O} = \mathrm{O_1, O_2, O_3 \dots O_R}$, where each observation $\mathrm{O_t}$ is one of the symbols from V, and R is the number of observations in the sequence.

It is obvious that a complete specification of an HMM requires the estimation of two model parameters, N and M, and three probability distributions $\mathrm{A, B}$, and π. We use the notation $\leftthreetimes = (\mathrm{A, B}, \pi)$ to indicate the complete set of parameters of the model, where $\mathrm{A, B}$ 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 $\mathrm{Q} = \mathrm{q_1, q_2} \dots\dots \mathrm{q_R}$ Where $\mathrm{q_1}$ is the initial state.

The probability that O is generated from this state sequence is given by $\mathrm{P(O \mid Q{,}\leftthreetimes)} = \prod_{\mathrm{t=1}}^{\mathrm{R}} \mathrm{P(\,O_t \mid, q_t \leftthreetimes)}$ where statistical independence of observations is assumed. Above equation can be described as

$$\mathrm{P(O \mid Q{,}\leftthreetimes)} = \mathrm{b_{q_1}(O_1).\,b_{q_2}(O_2)} \dots\dots \mathrm{b_{q_R}(O_R)}$$

The probability of the state sequence Q is given as $\mathrm{P(Q \mid \leftthreetimes)} = \pi_{\mathrm{q1}} \mathrm{a_{q1q2} a_{q2q3}} \dots\dots \mathrm{a_{qR-1R}}$

Thus, the probability of creation of the observation sequence O by the HMM specified by ⋋can be defined as an equation as follows:

$$\mathrm{P}(\mathrm{Q} \mid ⋋) = \sum_{\text{all Q}} \mathrm{P}(\mathrm{O} \mid \mathrm{Q},⋋)\mathrm{P}(\mathrm{Q} \mid ⋋)$$

Deriving the value of $\mathrm{P}(\mathrm{Q} \mid ⋋)$ using the direct definition of above equation is divisionally exhaustive. Subsequent to the HMM parameters are discovered, we take the symbols from a cardholder's training data and form an initial state series of symbols. Let $\mathrm{O}_1, \mathrm{O}_2, \mathrm{O}_3 \dots \mathrm{O}_\mathrm{R}$ be one such sequence of length R. This observation sequence results is formed from the cardholder's transactions based on mainly time t. They produce this input sequence to the HMM and compute the probability of acceptance in training stage calculated by HMM. Let the probability α_1 can be formulated as follows:

$$\alpha_1 = \mathrm{P}(\mathrm{O}_1, \mathrm{O}_2, \mathrm{O}_3 \dots \mathrm{O}_\mathrm{R} \mid \lambda)$$

This probability computation was performed for all $\alpha_{\mathrm{R}+1}$ and maintained these results. If $\Delta\alpha > 0$ the new sequence is verified and accepted by the HMM with low probability, and it could be a fraud by the way of if the percentage change in the probability is above a threshold, that is,

$$\frac{\Delta\alpha}{\alpha_1} \geq \text{Threshold}$$

4. SEMI HIDDEN MARKOV MODEL BASED CREDIT CARD FRAUD DETECTION MECHANISM

4.1. Semi Hidden Markov Model

Using state transition probability, state duration probability and observation probability in this research, the major states of components were designed using a semi hidden -Markov chain (SHMM). A modified forward–backward algorithm for SHMMs was used to develop the parameters of SHMMs using discriminant function investigation the probability will be determined. For prediction of remaining existence, a state interval model-based prediction computation procedure was offered. The results show that the SHMMs can endow with valuable timing information in the single operator case, but HMMs have a tendency to be more robust to increased team complexity.

4.1.1 Detection algorithm

Commencing maximum entropy principle (MEP), we identify that when a computer system is successively in normal state, the review data it produces contains less information than that it creates when operation in fraud state. That is to say, the information entropy of fraud state is larger than that of standard state; consequently the information entropy can proceed as the metric in credit card fraud detection. Other than when the span of visible symbol succession increases, the information entropy of observable symbol sequence will turn out to be larger and larger. It simply makes sense to contrast the value of information entropy among the same-length sequences. We work out the average information entropy (AIE) of visible symbol sequences, and use it as the metric to distinguish between normal behavior and anomaly behavior in order to employ entropy metric variable-length symbol sequences.

4.1.2 Training algorithm

For the reason that the normal state of a computer system possibly will change over the time, consequently preparation of hidden semi-Markov model is also a crucial part in anomaly detection algorithms. For the hidden semi-Markov model $k = (N, M, V, A, B, \pi)$ we created in earlier section, both the allocation of state transfer probabilities A and the original distribution of normal state and fraud state p are preset values, so only the allocation of visible symbol for regular behavior $B_0 = \{b_0(k)\}$, $1 \leq k \leq M$ need to be brought up to date. The training can be implemented by system administrator on normal data sequences.

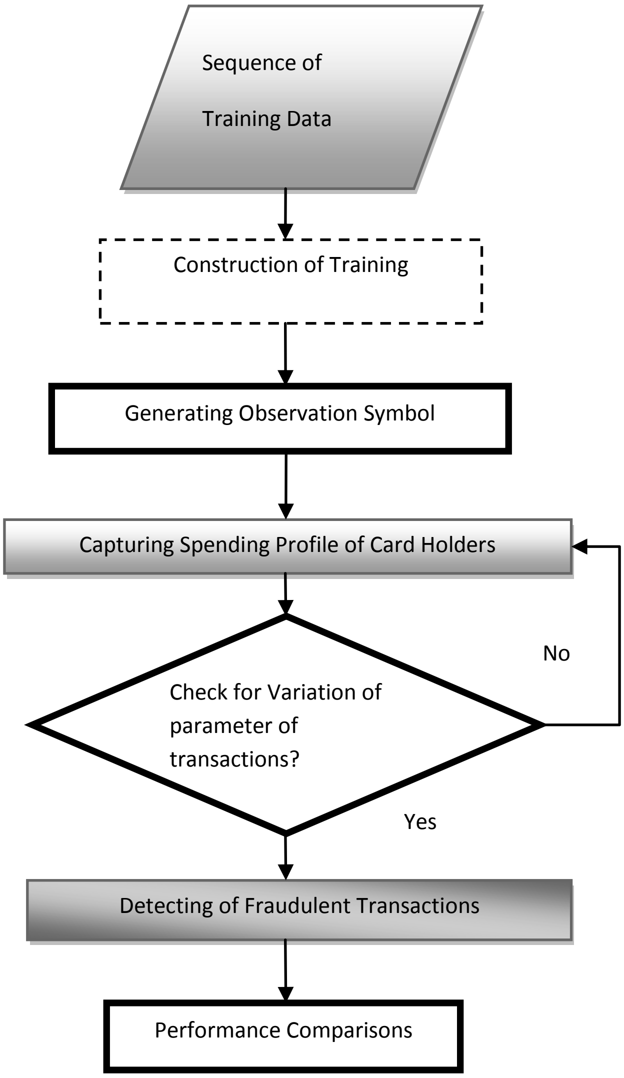
Sequence of
Training Data
Construction of Training
Generating Observation Symbol
Capturing Spending Profile of Card Holders
Check for Variation of parameter of transactions?
No
Yes
Detecting of Fraudulent Transactions
Performance Comparisons

Fig 1: System Flow Diagram

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.

4.2 Factorized Information Criterion

We have proposed estimation of marginal log likelihood (FIC) and a presumption method (FAB) for Bayesian model selection for mixture models, as a substitute to VB inference. We have given their justifications (asymptotic consistency, convergence, etc) and analyzed FAB mechanisms in terms of overfitting mitigation (shrinkage) and identifiability. Experimental results have shown that FAB does better than state-of the- art VB methods for a realistic number of data in terms of together model selection performance and computational efficiency.

Our main idea is to factorize marginal log-likelihood using a variational distribution over latent variables. An asymptotic approximation, a factorized information criterion (FIC), is got hold of by relating the Laplace method to each of the factorized modules. 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 by means of the marginal log-likelihood, 2) automatic component collection on the origin of an intrinsic shrinkage mechanism, and 3) parameter identifiability in mixture modeling. Experimental results show that FAB outperforms state-of-the-art VB methods.

4.3 Maximum Loglikelyhood

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 minimax 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.

4.4. Distributed Datamining

Credit card transactions continue to grow in number, taking an ever-larger share of the US payment system and leading to a higher rate of stolen account numbers and subsequent losses by banks. Improved fraud detection thus has become essential to maintain the viability of the US payment system. Banks have used early fraud warning systems for some years. Large scale data-mining techniques can improve the state of the art in commercial practice. Scalable techniques to analyze massive amounts of transaction data that efficiently compute fraud detectors in a timely manner is an important problem, especially for e-commerce. Besides scalability and efficiency, the fraud-detection task exhibits technical problems that include skewed distributions of training data and nonuniform cost per error, both of which have not been widely studied in the knowledge-discovery and data mining community. In this article, we survey and evaluate a number of techniques that address these three main issues concurrently. Our proposed methods of combining multiple learned fraud detectors under a "cost model" are general and demonstrably useful; our empirical results demonstrate that we can significantly reduce loss due to fraud through distributed data mining of fraud models.

5 PERFORMANCE EVALUATION

5.1 Precision accuracy

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 1.

Table 1: Precision Accuracy

S. No	No. of datasets	HMM	SHMM	SHMM with FIC
1	100	0.75	0.85	0.92
2	200	0.69	0.75	0.89
3	300	0.58	0.64	0.76
4	400	0.45	0.56	0.69
5	500	0.39	0.49	0.62
6	600	0.33	0.47	0.58

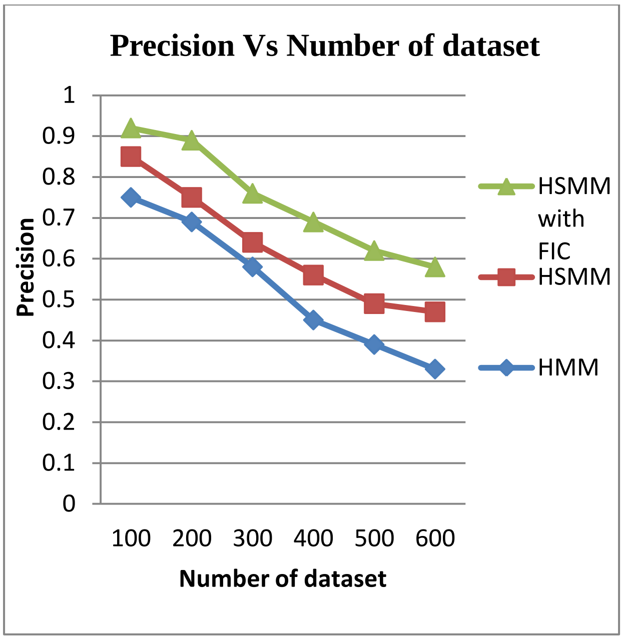

Fig 2: Precision Vs Number Of Dataset

5.2 Recall rate:

We analyze 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 2.

Table 2: Recall Rate

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

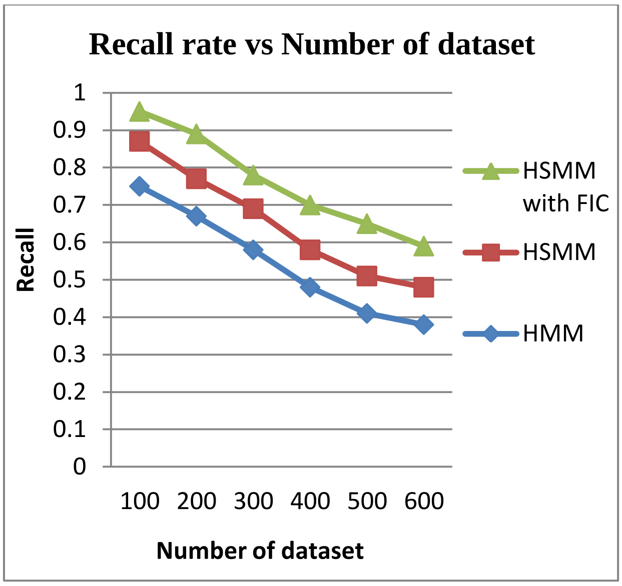

Fig 3: Recallrate Vs Number Of Dataset

5.3 F Measure

We analyze and compare the Fmeasure offered by HMM, SHMM and SHMM with FIC. Here if the no of datasets increased the recall rate also increased linearly. The Fmeasure 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 3.

Table 3: F Measure

S. No	No of datasets	HMM	SHMM	SHMM with FIC
1	100	0.67	0.75	0.92
2	200	0.56	0.69	0.85
3	300	0.48	0.62	0.76
4	400	0.40	0.53	0.66
5	500	0.34	0.45	0.59
6	600	0.23	0.39	0.56

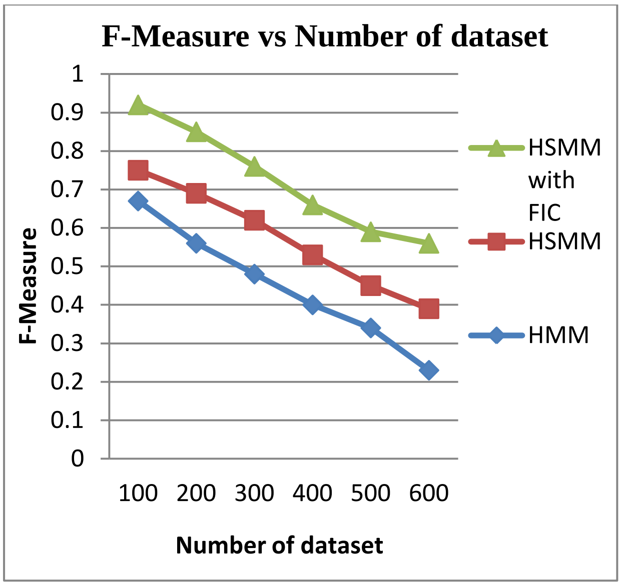

Fig 4: F Measure Vs Number of Dataset

6. CONCLUSION & FUTURE WORK

In this research, Semi Hidden Markov Model is introduced keen on credit card fraud detection systems. We presented an algorithm of fraud detection based on SHMM, which calculates the distance between the processes monitored by intrusion detection system and the perfect normal processes. In this algorithm, based on maximum entropy principle (MEP), we establish the concept of average information entropy (AIE), which is used as detection metric via analyzing variable-length observed symbol sequences. To get better accuracy, we propose a new approximation implication algorithm and refer to it as a factorized asymptotic Bayesian inference (FAB) with the SHMM.

REFERENCES

[1] S.S. Joshi and V.V. Phoha, “Investigating Hidden Markov Models Capabilities in Anomaly Detection,” Proc. 43rd ACM Ann. Southeast Regional Conf., vol. 1, pp. 98-103, 2005.

[2] S.B. Cho and H.J. Park, “Efficient Anomaly Detection by Modeling Privilege Flows Using Hidden Markov Model,” Computer and Security, vol. 22, no. 1, pp. 45-55, 2003.

[3] D. Ourston, S. Matzner, W. Stump, and B. Hopkins, “Applications of Hidden Markov Models to Detecting Multi-Stage Network Attacks,” Proc. 36th Ann. Hawaii Int’l Conf. System Sciences, vol. 9, pp. 334-344, 2003.

[4] X.D. Hoang, J. Hu, and P. Bertok, “A Multi-Layer Model for Anomaly Intrusion Detection Using Program Sequences of System Calls,” Proc. 11th IEEE Int’l Conf. Networks, pp. 531-536, 2003.

[5] 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.

[6] Fan, W., Prodromidis, A. L., and Stolfo, S. J., 1999. Distributed Data Mining in Credit Card Fraud Detection, IEEE Intelligent Systems, vol. 14, no. 6 (1999), pp. 67-74.

[7] Brause, R., Langsdorf, T., and Hepp, M., 1999. Neural Data Mining for Credit Card Fraud Detection, Proceedings of IEEE International Conference Tools with Artificial Intelligence (1999), pp. 103-106.

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

[9] Phua, C., Lee, V., Smith, K., and Gayler, R., 2007. A Comprehensive Survey of Data Mining-Based Fraud Detection Research (2007), March.

[10] Rabiner, L.R. 1989. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, Proceedings of IEEE, vol. 77, no. 2 (1989), pp.257-286.

[11] Ourston, D., Matzner, S., Stump, W., and Hopkins, B., 2003. Applications of Hidden Markov Models to Detecting Multi- Stage Network Attacks, Proceedings of 36th Annual Hawaii International Conference System Sciences, vol. 9 (2003), pp. 334-344.

[12] Cho, S.B., and Park, H.J., 2003. Efficient Anomaly Detection by Modeling Privilege Flows Using Hidden Markov Model, Computer and Security, vol. 22, no. 1 (2003), pp. 45-55.

[13] Kim, M.J., and Kim, T.S., 2002. A Neural Classifier with Fraud Density Map for Effective Credit Card Fraud Detection, Proceedings of International Conference on Intelligent Data Eng. and Automated Learning, (2002), pp. 378-383.

[14] Fujimaki, Ryohei and Morinaga, Satoshi. Factorized asymptotic bayesian inference for mixture models. In AISTATS, 2012.

[15] Rabiner, L. R. A tutorial on hidden markov models and selected applications in speech recognition. Proceedings of IEEE, 77:257{86, 1989.

[16] Guedon, Y. (2003). "Estimating Hidden Semi-Markov Chains From Discrete Sequences " Journal of Computational & Graphical Statistics **12**(3): 604-639.

[17] M. Beal. Variational Algorithms for Approximate Bayesian Inference. PhD thesis, Gatsby Computational Neuroscience Unit, University College London, 2003.

[18] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from imcomplete data via the EM algorithm. Journal of the Royal Statistical Society, B39(1):1–38, 1977.

[19] H. Jeffreys. An invariant form for the prior probability in estimation problems. Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences, 186:453–461, 1946.