

An Ensemble Approach for Credit Card Fraud Detection

A.Prakash
MSC(CT),M.Phil
Research Scholar
Manonmaniam Sundaranar University,
Tirunelveli

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

ABSTRACT

The most important moral issue in the credit card trade is fraud involvement. The main aspires are, primarily, to recognize the different types of credit card fraud, and, secondly, to evaluate unconventional techniques that have been used in fraud detection. The sub-aim is to present, compare and examine recently published discovering in credit card fraud detection. Credit card fraud detection has developed a number of techniques via bunch of investigate interest and, with special importance on, data mining and distributed data mining have been recommended. In our existing research we proceeded with the semi hidden markov model (SHMM) where we got efficient result in credit card fraud detection. That is also having a larger class of practical problems that can be properly modeled in the setting of SHMM. Also major constraint is found, conversely, in mutually HMM and SHMM, i.e., it is generally imagined that there survives at least one observation connected with every state that the hidden Markov chain takes on. To improve the efficiency of SHMM in our proposed research we are combining the multiple observation of SHMM called Multiple Semi Hidden Markov Model (MSHMM) through this we can improve the detection accuracy better than the SHMM. Our suggested methods of combining multiple learned fraud detectors under a “cost model” are common and obviously useful; our experimental results make obvious that we can significantly reduce loss due to fraud through distributed data mining of fraud models.

Keywords: Hidden Markov Model, Semi Hidden Markov Model, Multiple Semi Hidden Markov Model, Distributed Data Mining, Credit card.

1. INTRODUCTION

The hidden Markov model (HMM) technique has circled out to be one of the greater part successful techniques in the field of evaluation and detection (e.g., fraud detection, speech recognition detection)). In the conventional HMM move toward the state duration is moreover of a unit interval or unreservedly unspecified to be geometrically dispersed to construct the fundamental method Markovian. A Semi hidden Markov model (SHMM) is an expansion of HMM designed to allow all-purpose allocations for the state intervals. A number of instigators use expressions such as “HMM with variable duration” and “HMM with explicit duration” to denote what we call in this paper an SHMM. To the most extraordinary of our knowledge Ferguson is the first that observed the SHMM. In the ordinary discrete-time HMM and SHMM, an apparent output is produced at each discrete time,

still at the same time as the hidden Markov state leftovers unchanged.

In various applications, though, observations may not of essential be made frequently sufficient for one reason or a supplementary. In such cases, assessment of the state sequence and/or model factors have to be completed depended on insufficient observations. The multiple observation probability is conveyed as a combination of individual observation probabilities without losing majority in the Multiple HMM (Hidden Markov Model) method. The independence-dependence belongings of the observations are characterized by the combinatorial weights, and hence it gives us more autonomy in making different assumptions and also in deriving corresponding training equations. Only one observable state is attached with the hidden state is an additional assumption usually made in the conventional HMM and HSMM. In some other requests, multiple observations may be available associated with the hidden state sequence. Moreover, these multiple observation sequences may not be synchronous to each other.

In existing research paper, the fitness states of constituents were formed using state evolution probability, state duration probability and observation probability using a hidden semi-Markov chain (HSMM). A modified forward-backward algorithm for HSMMs was used to estimate the parameters of HSMMs using discriminant 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 HSMMs can provide valuable timing information in the single operator case, whereas HMMs tend to be more robust to increased team complexity. 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. So as to use entropy metric variable-length symbol sequences, we calculate the average information entropy (AIE) of visible symbol sequences, and use it as the metric to distinguish between normal behavior and anomaly behavior.

Although there is a problem is that first the observation data possibly will be missing for a number of intervals. Following that there are multiple observation streams that are not necessarily synchronous to each other and possibly will have

different “production distributions” for the same state. So in projected research we are using multiple observation sequences which are associated with the semi hidden state sequence and these observations may not be synchronized to each other. We divide a large data set of labeled transactions (either fraudulent or legitimate) into smaller subsets by applying distributed data mining techniques to generate classifiers in parallel, and come together the resultant base models by metalearning from the classifiers’ performance to produce a metaclassifier. In addition extensibility, combining multiple models computed over all available data constructs metaclassifiers that can counterbalance the loss of predictive presentation that usually occurs when mining from data subsets or sampling. Furthermore, when we use the learned classifiers (for example, during transaction authorization), the base classifiers can carry out in parallel, with the metaclassifier then combining their results. So, our approach is highly efficient in generating these models and also relatively efficient in applying them. The contribution of our work is as follows:

1. In training phase obtain the card holder profile and calculate the probability.
2. Using the MSHMM create the observation model with multiple observations.
3. There is millions of credit card transactions processed each day. Mining such efficient techniques that scale using the distributed datamining
4. In testing phase the detection of fraud is obtained If both probability value from multiple observation are same it will be a normal customer else there will be fraud signal will be provided

The remainder of this paper is as follows: The related work is discussed in section 2. The Semi hidden markov model is explained in section 3. In section 4 the proposed work is explained. The section 5 is deal with experimental results and discussion. In section 6 the conclusion of the paper is described.

2. RELATED WORK

Jared O’Connell [1] the main features of the mhsmm package are expanded as follows: Observations are permitted to be multi-variate. Missing values are tolerated. Observations must be recorded at equidistant times. The package is proposed to allow the specification of custom emission distributions. It is possible to have multiple sequences of data. [2] They believe that in most real applications we need models that combine the features of all of these models and introduce a new class of models, namely the hierarchical multichannel hidden semi-Markov models. [3] The multi-view approach is based on the principle of maximizing the consensus among multiple independent hypotheses; they developed this principle into a semi supervised hidden Markov Perceptron algorithm. [4] Models that combine Markovian states with implicit geometric state occupancy distributions and semi-Markovian states with explicit state occupancy distributions are investigated.

This type of model retains the flexibility of hidden semi-Marko chains for the modeling of short or medium size homogeneous zones along sequences but also enables the modeling of long zones with Markovian states. The forward–

backward algorithm, which in particular enables to implement efficiently the E-step of the EM algorithm, and the Viterbi algorithm for the restoration of the most likely state sequence are derived.

Lee-Min Lee [6] The HSMM is equivalent to a form of expanded state HMM (ESHMM). Expansion of a state to several connected states with the same output probability distribution allows for more flexible duration model. Yi Xie and Shun-Zheng Yu [7] an extended hidden semi-Markov model is proposed to describe the browsing behaviors of web surfers. In order to reduce the computational amount introduced by the model’s large state space, a novel forward algorithm is derived for the online implementation of the model based on the M-algorithm.

In [13] a single repository data base where data is stored in central site, then applying data mining algorithms on these data base, patterns are extracted, which is clearly implausible and untenable for many realistic problems and databases. To deal with these complex systems has revealed opportunities to improve distributed data mining systems in a number of ways. Additionally, in certain situations, data may be essentially distributed and cannot be merged into a single database for a variety of reasons including security, fault tolerance, legal constraints, competitive reasons, etc. In such cases, it may not be possible to examine all of the data at a central processing site to compute a single global model. Here, we develop techniques that scale up to large and possibly physically distributed databases.

3. CREDIT CARD FRAUD DETECTION USING SEMI HIDDEN MARKOV MODEL

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 hidden semi-Markov model (HSMM) 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. Because of 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 $k = (N, M, V, A, B, \pi)$ they 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 behavior $B_0 = \{b_0(k)\}$, $1 \leq k \leq M$ need to be updated. The training can be implemented by system administrator on normal data sequences.

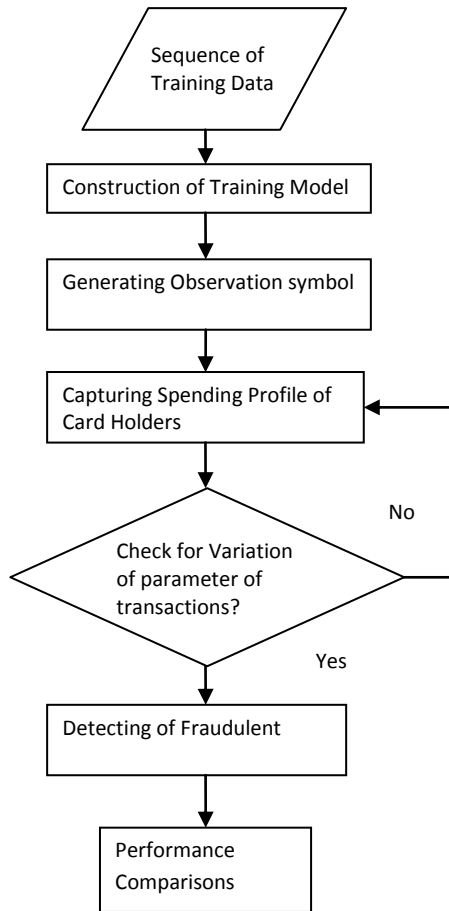


Fig 1: Flow diagram of SHMM

With this they 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. They 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 outperforms state-of-the-art VB methods for a practical number of data in terms of both model selection performance and computational efficiency. Their key idea is to factorize marginal log-likelihood using a variational 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, they proposed 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 identifiability in mixture modeling.

4. MULTIPLE SEMI HIDDEN MARKOV MODEL FOR CREDIT CARD FRAUD DETECTION

Consider a Markov chain with M states that are labeled as $\{1, 2, \dots, M\}$, in which the probability of transition from state m' to state m is denoted $a_{m,m'}$, where $m, m' = 1, 2, \dots, M$, and the initial state probability distribution is given by $\{\pi_m\}$.

The Markov state is called a “hidden” state, when the state is not directly observable. If some output sequence that is probabilistically associated with the underlying hidden Markov chain is observable, then this doubly stochastic process is referred to as a hidden Markov model or an HMM. Let s_t denote the state that the system takes at time t , where $t = 1, 2, \dots, T$. We denote the state sequence as $\{s_t\}$, but when we wish to be explicit about the interval, we adopt the notation s_a^b , meaning $\{s_t: a \leq t \leq b\}$. Similarly, let o_t denote the observable output at time t associated with state s_t , and let $b_m(o_t)$ be the probability of observing o_t , given $s_t = m$. We assume the “conditional independence” of outputs so that $b_m(o_a^b) = \prod_{t=a}^b b_m(o_t)$, where o_a^b represents the observation sequence from time a to b . If, for instance, $\{o_t\}$ is a function of $\{s_t\}$ observed through a channel with additive white noise, the above simple product form holds.

We present a special case of observation type (f) defined above. We assume two sequences $\{o_t\}$ and $\{q_t\}$ are available as the outputs of an HSMM state sequence. The conditional probability that o_t appears when the state is at m is given by $b_m(o_t)$ and the corresponding conditional probability for the second output is given by $c_m(q_t)$. If we introduce some random delay τ between the two output sequences, these two sequences are no longer synchronized. The symbol \emptyset_t represents the missed observation (i.e., null observation) of the output at time t . Because the observation may not necessarily be made at every time interval, we denote the set of the observation time instants $G = \{t_1, t_2, \dots, t_n\}$, where $1 \leq t_1, t_n \leq T$. Then we can denote the observation sequence. $o_a^b = \{o_t: a \leq t \leq b, \text{ and } t \in G$

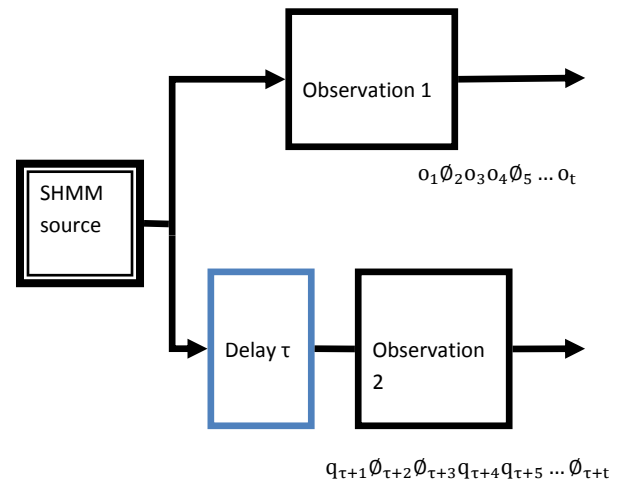


Fig 2: An example of multiple observation sequences

4.1 Evaluation of HSMM with Multiple Observation Sequences

We now discuss a case where multiple sequences of observations are available. These multiple observation sequences may have their observation intervals, starting points, etc. different from others. In Fig. 2 we show only two observation sequences $\{o_t\}$ and $\{q_t\}$. There is a delay τ between the two observation sequences, where τ takes on a value from $\{0, \pm 1, \pm 2, \dots\}$. Either of the two streams can be any type of the observation or missing patterns discussed.

4.1.1 Dynamic Generation of Observation Symbols

For each cardholder, we train and maintain an HMM. To find the observation symbols corresponding to individual cardholder's transactions dynamically, we run a clustering algorithm on his past transactions. On the whole, the transactions that are stored in the issuing bank's database include many attributes. For our work, we consider only the amount that the cardholder spent in his transactions. The clustering process is done by using the distributed datamining technique. In this section we describe previous research in meta-learning and in particular address the following specific research issues:

1. Using a variety of statistical, information-theoretic and the characterization of datasets is performed.
2. By applying a set of algorithms at the base level and combining these through a metalearner information is extracted.
3. To accelerate the rate of learning process, knowledge is extracted through a continuous learner.

Meta-learning (learning from learned knowledge) technique is dealing with the problem of computing a global classifier from large and inherently distributed databases. A number of independent classifiers "base classifiers" are computed in parallel. The base classifiers are then collected and combined to a "meta-classifier" by another learning process. Meta-classifiers can be defined recursively as collections of classifiers structured in multi-level trees. Such structures, however, can be unnecessarily complex, meaning that many classifiers may be redundant, wasting resources and reducing system throughput.

- The predictive accuracy of base classifiers is improved.
- Assuming that a system consists of several databases interconnected through an intranet or internet, the goal is to provide the means for each data site to utilize its own local data and, at the same time, benefit from the data that is available at other data sites without transferring or directly accessing that data.

4.1.2 Spending Profile of Cardholders

The spending profile of a cardholder suggests his normal spending behavior. Cardholders can be broadly categorized into three groups based on their spending habits, namely, high-spending (hs) group, medium-spending (ms) group, and low-spending (ls) group. Spending profiles of cardholders are determined at the end of the clustering step.

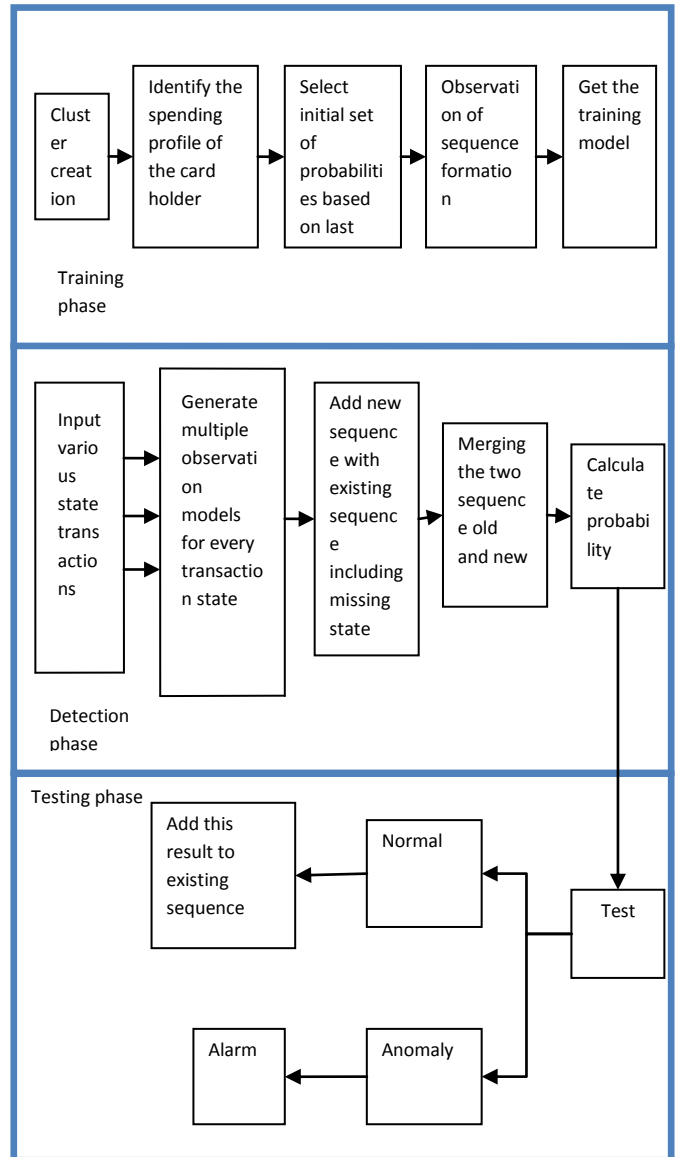


Fig 3: Block Diagram of proposed work

5. RESULTS AND DISCUSSION

5.1 Precision accuracy

This graph shows the precision rate of existing and proposed system based on two parameters of precision and the number of Dataset. From the graph we can see that, when the number of Dataset is advanced the precision also developed in proposed system but when the number of number of Dataset is improved the precision is reduced somewhat in existing system than the proposed system. From this graph we can say that the precision of proposed system is increased which will be the best one. The values are given in Table 1:

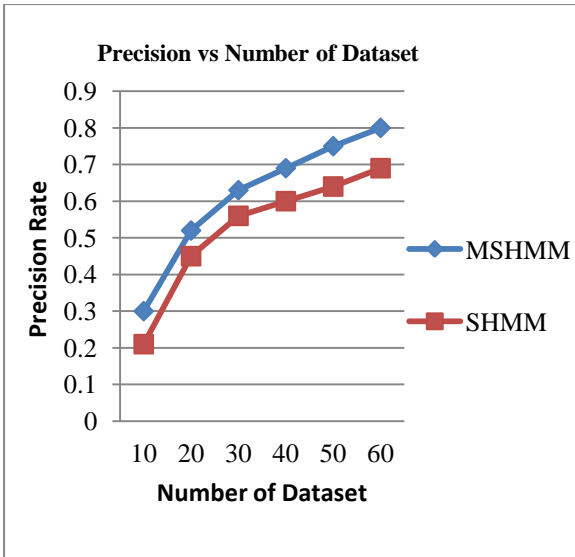


Fig 4: Precision vs. Number of Dataset

Table 1: Precision vs. Number of Dataset

SNO	Number of Dataset	MSHMM	SHMM
1	10	0.3	0.21
2	20	0.52	0.45
3	30	0.63	0.56
4	40	0.69	0.6
5	50	0.75	0.64

In this graph we have chosen two parameters called number of Dataset and precision which is help to analyze the existing system and proposed systems. The precision parameter will be the Y axis and the number of dataset parameter will be the X axis. The blue line represents the existing system and the red line represents the proposed system. From this graph we see the precision of the proposed system is higher than the existing system. Through this we can conclude that the proposed system has the effective precision rate.

5.2 Recall vs. Number of Dataset

This graph shows the recall rate of existing and proposed system based on two parameters of recall and number of Dataset. From the graph we can see that, when the number of number of Dataset is improved the recall rate also improved in proposed system but when the number of number of Dataset is improved the recall rate is reduced in existing system than the proposed system. From this graph we can say that the recall rate of proposed system is increased which will be the best one. The values of this recall rate are given below:

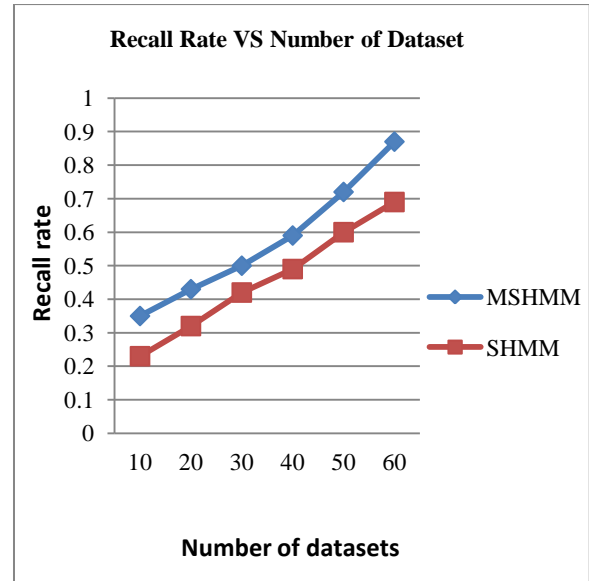


Fig 5: Recall vs. Number of Dataset

Table 2: Recall vs. Number of Dataset

SNO	Number of Dataset	MSHMM	SHMM
1	10	0.35	0.23
2	20	0.43	0.32
3	30	0.5	0.42
4	40	0.59	0.49
5	50	0.72	0.6
6	60	0.87	0.69

In this graph we have chosen two parameters called number of Dataset and recal which is help to analyze the existing system and proposed systems on the basis of recall. In X axis the Number of dataset parameter has been taken and in Y axis recal parameter has been taken. From this graph we see the recal rate of the proposed system is in peak than the existing system. Through this we can conclude that the proposed system has the effective recall.

5.3 Fmeasure vs. Number of Dataset

This graph shows the Fmeasure rate of existing and proposed system based on two parameters of Fmeasure and number of Dataset. From the graph we can see that, when the number of number of Dataset is improved the Fmeasure rate also improved in proposed system but when the number of number of Dataset is improved the Fmeasure rate is reduced in existing system than the proposed system. From this graph we can say that the Fmeasure rate of proposed system is increased which will be the best one. The values of this Fmeasure rate are given below:

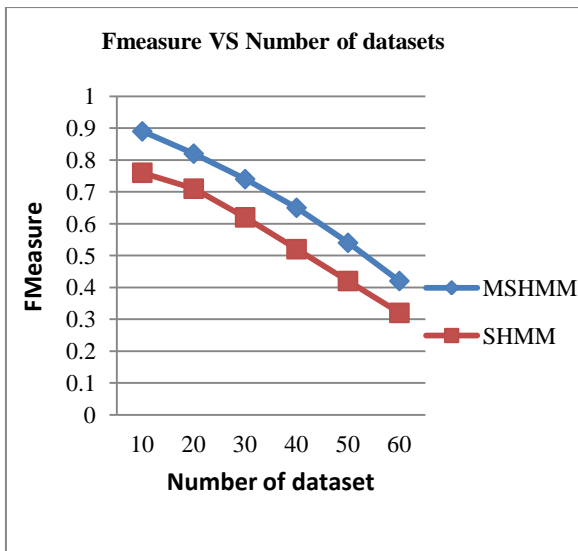


Fig 6: Fmeasure vs. Number of Dataset

Table 3: Fmeasure vs. Number of Dataset

SNO	Number of Dataset	MSHMM	SHMM
1	10	0.89	0.76
2	20	0.82	0.71
3	30	0.74	0.62
4	40	0.65	0.52
5	50	0.54	0.42
6	60	0.42	0.32

In this graph we have chosen two parameters called number of Dataset and recal which is help to analyze the existing system and proposed systems on the basis of Fmeasure. In X axis the Number of dataset parameter has been taken and in Y axis Fmeasure parameter has been taken. From this graph we see the Fmeasure of the proposed system is in peak than the existing system. Through this we can conclude that the proposed system has the effective Fmeasure.

6. CONCLUSION

The fundamental assumption in the existing HMM and SHMM models is that there is at least one observation produced per state visit and that observations are exactly the outputs of states. In some applications, these assumptions are too restrictive. We extended the ordinary HMM and SHMM to the model with missing data and multiple observation sequences. In our research Multiple Semi Hidden Markov Model is implemented to provide better fraud detection mechanism. Where the multiple observations will be collected using distributed datamining technique and the detection phase is successfully executed. From the calculation of the probability values the training value and the testing value will be compared. If two values are same then there is no anomaly detection other than that there can be a false alarm generation to denote the anomaly involvement.

REFERENCES

- [1] Jared O'Connell, Soren Hojsgaard "Hidden Semi Markov Models for Multiple Observation Sequences: The mhsmm Package for R", Journal of Statistical Software, March 2011, Volume 39, Issue 4.
- [2] Pradeep Natarajan and Ramakant Nevatia, "Hierarchical Multi-channel Hidden Semi Markov Models*", Institute for Robotics and Intelligent Systems, University of Southern California, Los Angeles, CA 90089-0273
- [3] Ulf Brefeld and Christoph Buscher and Tobias Scheffer ,Humboldt-Universitat zu Berlin,"Multi-View Hidden Markov Perceptrons ", Department of Computer Science, Unter den Linden 6, 10099 Berlin, Germany
- [4] Yann Guédon*, "Hidden hybrid Markov/semi-Markov chains", Computational Statistics & Data Analysis 49 (2005) 663 – 688
- [5] Shi Zhong, "Semi-supervised Sequence Classification with HMMs", American Association for Artificial Intelligence, Copyright 2004
- [6] Lee-Min Lee, "High-Order Hidden Markov Model and Application to Continuous Mandarin Digit Recognition*" Journal of Information Science and Engineering 27, 1919-1930 (2011) 1919
- [7] Yi Xie and Shun-Zheng Yu, "A Large-Scale Hidden Semi-Markov Model for Anomaly Detection on User Browsing Behaviors", IEEE/ACM Transactions on Networking, Vol. 17, No. 1, February 2009
- [8] K. Riedhammer, T. Bocklet, A. Ghoshal, D. Povey, "Revisiting Semi-Continuous Hidden Markov Models", IEEE, 2012
- [9] Xuedong Huang, Fil Allewa, Satoru Hayamizu, Hsiao-Wuen Hon, Mei-Yuh Hwang, Kai-Fu Lee, "Improved Hidden Markov Modeling for Speaker-Independent Continuous Speech Recognition", School of Computer Science, Carnegie Mellon University
- [10] Takashi Yamazaki, Naotake Niwase, Junichi Yamagishi, Takao Kobayashi, "Human Walking Motion Synthesis Based on Multiple Regression Hidden Semi-Markov Model", Tokyo Institute of Technology, Proceedings of the International Conference on Cyber worlds, 2005
- [11] Qinfeng Shi, Yasemin Altun, Alex Smola, S.V. N. Vishwanathan, "Semi-Markov Models for Sequence Segmentation", 2004
- [12] Wei-Ho Chung and Kung Yao "Modified Hidden Semi-Markov Model for Modelling the Flat Fading Channel" IEEE Transactions on Communications, Vol. 57, No. 6, June 2009
- [13] Sanjay Kumar Sen, dr Sujata Dash, Subrat P Pattanayak, "Agent Based Meta Learning In Distributed Data Mining System ", Vol. 2, Issue 3, May-Jun 2012, pp. 342-348
- [14] David Tweed, Robert Fisher, Jos´e Bins, Thor List, "Efficient Hidden Semi-Markov Model Inference for Structured Video Sequences", University of Edinburgh, Institute for Perception, Action & Behaviour, School of Informatics
- [15] Shun-Zheng Yu and Ning Li, "An Anomaly Detection System Based on a Hidden Semi-Markov Model*", Department of Electrical and Communication Engineering.