Abstract—Credit card is one of the electronic payment mode and the fraud is committing of use of credit card in a fraudulent way either using credit or debit card. The purpose can be solved by purchasing the accessories without paying and giving unauthorized way of payment from account. In this paper, we are proposing the algorithm with the combined approach of hidden markov model and genetic algorithm of the data mining techniques and get the better result with respective of the individual approaches. Furthermore, there performance in term of precision, recall, F-measure has also increased comparative to the the state-of-the-art papers included in this paper.

Keywords-component; formatting; style; styling; insert (key words)

I. INTRODUCTION (HEADING 1)

The Credit Card fraud is defined as a card holder uses other credit card in its own, but the owner of the card and card issuer are not aware of the fact that the card has been using. It is like unauthorized account activity by a person in which that account was not intended for use.

In this study we are concerning the financial frauds and will particularly focus on detecting fraudulent credit card transactions. The measure is needed due to inherent structure of credit card (CC) transactions. This is about optimizing the parametric fraud detection solution. The amount of losses due to fraud and the awareness of the relation between loss and the available limit on the CC have forced us to develop a good performance solution. This solution is tested on the bases of data set. The results obtained on the sample data bases and selections of the best solution parameters.

The Traditional detection method mainly depends on database system and the education of customers, which usually are delayed, inaccurate and not in time. After that methods based on discriminate analysis and regression analysis are widely used which can detect fraud by credit rate for cardholders and credit card transaction. For a large amount of data it is not efficient.[5]

In recent years, the prevailing data mining concerns people with credit card fraud detection model based on data mining. Since our problem is approached as a classification problem, classical data mining algorithms are not directly applicable. So an alternative approach is made by using general purpose heuristic approaches like genetic algorithms. This paper is to propose a credit card fraud detection system with the fusion approach of hidden markov model and genetic algorithm.

Hidden Markov Model is probably the simplest and easiest models which can be used to model sequential data, i.e. data samples which are dependent from each other. An HMM is a double embedded random process with two different levels, one is hidden and other is open to all. An HMM based system is initially studied spending profile of the card holder and followed by checking an incoming transaction against spending behavior of the card holder, if it is not accepted by our proposed HMM with sufficient probability, then it would be a fraudulent transaction.

Genetic algorithms are evolutionary algorithms which aim at obtaining better solutions as time progresses. When a card is copied or stolen or lost and captured by fraudsters it is usually used until its available limit is depleted. Thus, rather than the number of correctly classified transactions, a solution which minimizes the total available limit on cards subject to fraud is more prominent. It aims in
minimizing the false alerts using genetic algorithm where a set of interval valued parameters are optimized. Generally Fraud is unauthorized activity taking place in electronic payments systems, these activities should be banned by laws and they are treated as illegal. Fraud can appear in various different domains like financial systems, telecommunications, public & private services. It is concerned with the financial frauds and focus on detecting fraudulent credit card transaction. Fraud detection problem is classification problem, in which some of statistical methods many data mining algorithms have proposed to solve it. Among decision trees are more popular. Fraud detection has been usually in domain of E-commerce, data mining. The Genetic algorithms are evolutionary algorithms in which the aim is to obtain the better solutions as it is technically to eliminate the fraud, a high importance has given to develop efficient and secure electronic payment system to detect whether a transaction is fraudulent or not.

Despite the fact that fraud only impacts a fraction of one percent of all purchases made with plastic, according to data from the Federal Reserve, it represents one of the biggest concerns among consumers. This can largely be attributed to the catastrophic impact of the worst-case scenarios that run through people’s minds as well as the notion that regardless of how low the incidence of fraud may be, no one wants to be the exception to the rule and find their hard-earned money siphoned away by criminals.

What consumers generally do not know is that they are shielded from liability for unauthorized transactions made with their credit cards via the combination of federal law issuer/card network policy. As a result, financial institutions and merchants assume responsibility for most of the money lost as a result of fraud. For example, card issuers bore a 63% share of fraudulent losses in 2012 and merchants assumed the other 37% of liability, according to the Nilson Report, August 2013.

The following statistics will give you a better sense of the credit and debit card fraud landscapes as well as how both have changed over the years.

- Credit card and debit card fraud resulted in losses amounting to $11.27 billion during 2012. Card issuers and merchants incurred 63% and 37% of those losses, respectively, with the following transactional breakdown: [1]
  - Card issuer losses occur mainly at the point of sale from counterfeit cards while merchant losses occur mainly on card-not-present (CNP) transactions on the Web, at a call center or through mail order. [1]
- During 2012 credit card and debit card gross fraud losses accounted for roughly 5.22¢ per $100 in total volume, up from 5.07¢ per $100 in 2011. [1]
- In 2012, US accounted for 47.3% of the worldwide payment card fraud losses but generated only 23.5% of total volume. [1]
- Financial institutions incurred $955 million in losses due to debit card fraud in 2010—a 21% increase from the $788 million in losses incurred during 2008. [3]
- In 2011, 59% of the more than 37 billion debit card transactions that were made were verified by signature, 85% of all fraudulent debit card transactions involved signature “verification,” and $1.15 billion of the total $1.35 billion in debit card...
fraud losses (85%) stemmed from signature debit card transactions. [2]

- Identity theft is a form of fraud that often results in unauthorized credit card and debit card transactions. There were 13% more cases of identity theft in 2011 than in 2010, yet the costs incurred by consumers actually represented a 44% decrease over the course of the previous eight years.[4]
- 67% more Americans were impacted by financial data breaches in 2012 than in 2010[4]

II. FRAUD DETECTION METHODS
The detection of fraud is a complex computational task and still there is no system that surely predicts any transaction as fraudulent. They just predict the likelihood of the transaction to be a fraudulent.[6]

The properties of a good fraud detection system are:
1) It should identify the frauds accurately
2) It should detecting the frauds quickly
3) It should not classify a genuine transaction as fraud

A. Neural network
Fraud detection methods based on neural network are the most popular ones. An artificial neural network consists of an interconnected group of artificial neurons. The principle of neural network is motivated by the functions of the brain especially pattern recognition and associative memory [8] The neural network recognizes similar patterns, predicts future values or events based upon the associative memory of the patterns it was learned. It is widely applied in classification and clustering. The advantages of neural networks over other techniques are that these models are able to learn from the past and thus, improve results as time passes. They can also extract rules and predict future activity based on the current situation. There are two phases in neural network training and recognition. Learning in a neural network is called training. There are two types of NN training methods supervised and unsupervised. In supervised training, samples of both fraudulent and non fraudulent records are used to create models. In contrast, unsupervised training simply seeks those transactions , which are most dissimilar from the norm. On other hand, the unsupervised techniques do not need the previous knowledge of fraudulent and non fraudulent transactions in database. NNs can produce best result for only large transaction dataset. And they need a long training dataset.

Neural network can use following approaches:[6]
I)Back propagation neural network
II)Self Organizing Map

On the other side, there are still many disadvantages for the neural networks, such as:
1. Difficulty to confirm the structure.
2. Excessive training
3. Efficiency of training and so on

B. Bayesian network
The Bayesian belief network was first introduced by Cooper and Herskovits (1992) Bayesian belief networks are statistical techniques in data mining. Bayesian networks are very effective for modelling situations where some information is already known and incoming data is unsure or partially unavailable

Bayesian Network needs training of data to operate and require high processing speed. BN is more accurate and much faster than neural network [8], but BBNs are slower when applied to new instances

This techniques yield better results but having large cycle time to detect fraud. However, the time constraint is one main disadvantage of this technique, especially compared with neural network

C. Decision tree
Decision trees are statistical data mining technique that express independent attributes and a dependent attributes logically AND in a tree shaped structure. Classification rules, extracted from decision trees, are IF-THEN expressions and all the tests have to succeed if each rule is to be generated [9]. Decision tree usually separates the complex problem into many simple ones and resolves the sub problems through repeatedly using [9][10]. Decision trees are predictive decision support tools that create mapping from observations to possible consequences.

There are number of popular classifiers construct decision trees to generate class models. Decision tree methods C5.0,C&RT and CHAID.

The work demonstrates the advantages of applying the data mining techniques including decision trees and SVMs to the credit card fraud detection problem for the purpose of reducing the bank’s risk. The results show that the proposed classifiers of C&RT and other decision tree approaches outperform SVM
approaches in solving the problem under investigation. [12]

D. Dempster–Shafer theory and Bayesian learning

Dempster–Shafer theory and Bayesian learning combines evidences from current as well as past behavior. The fraud detection system (FDS) consists of four components, namely, rule-based filter, Dempster–Shafer adder, transaction history database and Bayesian learner. In the rule-based component, we determine the suspicion level of each incoming transaction based on the extent of its deviation from good pattern. Dempster–Shafer’s theory is used to combine multiple such evidences and an initial belief is computed. The transaction is classified as normal, abnormal or suspicious depending on this initial belief. Once a transaction is found to be suspicious, belief is further strengthened or weakened according to its similarity with fraudulent or genuine transaction history using Bayesian learning. Extensive simulation with stochastic models shows that fusion of different evidences has a very high positive impact on the performance of a credit card fraud detection system as compared to other methods.[11]

E. Artificial immune systems

Artificial immune systems (AIS) represent an important strategy inspired by biological systems and developed by Neal et al in 1998 [13]. The main developments within AIS have focused on three main immunological theories: clonal selection, immune networks and negative selection. The immune system can distinguish between self and non-self. In the concept of credit card fraud detection, self (S) represents all patterns in a finite space that is legitimate and non-self (Ŝ) represents all patterns that are not in self [14][15]. The AIS consists of artificial lymphocytes (ALCs) that able to classify any pattern as self or non-self by detecting only non-self patterns. AIS detection engines implements AIS based algorithms which can classify input data as normal or fraudulent [16].

F. K-Nearest Neighbor algorithm

The concept of nearest neighbor analysis has been used in several anomaly detection techniques. One of the best classifier algorithms that have been used in the credit card fraud detection is k-nearest neighbor algorithm that is a supervised learning algorithm where the result of new instance query is classified based on majority of K-Nearest Neighbor category. It was first introduced by Aha, Kibler, and Albert (1991) [17]. The performance of KNN algorithm is influenced by three main factors [Mohammed J. Islam]:

- The distance metric used to locate the nearest neighbors.
- The distance rule used to derive a classification from k-nearest neighbor.
- The number of neighbors used to classify the new sample.

Among the various credit card fraud detection methods of supervised statistical pattern recognition, the K Nearest Neighbor rule achieves consistently high performance, without a priori assumptions about the distributions from which the training examples are drawn. K-Nearest neighbor based credit card fraud detection techniques require a distance or similar the measure defined between two data instances.

The performance of KNN algorithm can be improved by using a genetic algorithm for optimizing the distance metric. This technique required legitimate as well as fraudulent samples of data for training. It is fast technique along with high false alarm [18].

G. Support vector machine

The basic idea of SVM classification algorithm is to construct a hyper plane as the decision plane which making the distance between the positive and negative mode maximum [19]. The strength of SVMs comes from two important properties they possess - kernel representation and margin optimization. Kernels, such as radial basis function (RBF) kernel, can be used to learn complex regions. A kernel function represents the dot product of projections of two data points in a high dimensional feature space. In SVMs, the classification function is a hyper-plane separating the different classes of data. The basic technique finds the smallest hypersphere in the kernel space that contains all training instances, and then determines on which side of hypersphere a test instance lies. If a test instance lies outside the hypersphere, it is confirmed to be suspicion SVM can have better prediction performance than BPN(Back propagation network) in predicting the future data. But in large data BPN has a good performance..
II. FUZZY LOGIC BASED SYSTEM

1) Fuzzy Neural Network
The aim of FNNs is to process the massive volume of uncertain information, which is widespread applied in our life [22]. Syeda et al (2002) [20] propose fuzzy neural networks on parallel machines to speed up rule production for customer-specific credit card fraud detection. His work can be related to Data mining and Knowledge Discovery in databases (KD). In this method syeda et al used GNN (Granular Neural Network) method that uses fuzzy neural network based on knowledge discovery (FNNKD), for how fast we can train the network and how fast a number of customers can be processed for detection in parallel. There are various fields in transaction table that include, the transaction amounts, time between transactions, statement date, transaction code, posting date, day, transaction description, and etc.

2) Fuzzy Darwinian System
Fuzzy Darwinian Detection [22] is Evolutionary-Fuzzy system which uses genetic programming for evolving fuzzy logic rules. It classifies the transactions into suspicious and non-suspicious. It comprises of Genetic Programming (GP) search algorithm and a fuzzy expert system. This approach has very high accuracy and produces a low false alarm. But it is not applicable in online transactions. Also it is highly expensive and processing speed is low.[20]

I. HIDDEN MARKOV MODEL
A Hidden Markov Model is a finite set of states; each state is linked with a probability distribution. Transitions among these states are governed by a set of probabilities called transition probabilities. In a particular state a possible outcome or observation can be generated which is associated symbol of observation of probability distribution. It is only the outcome, not the state that is visible to an external observer and therefore states are “hidden” to the outside; hence the name Hidden Markov Model.
Hidden Markov Model will be helpful to find out the fraudulent transaction by using spending profiles of user. It works on the user spending profiles which can be divided into major three types such as:
1) Lower profile;
2) Middle profile;
3) Higher profile.
For every credit card, the spending profile is different, so it can figure out an inconsistency of user profile and try to find fraudulent transaction. It keeps record of spending profile of the card holder by both way, either offline or online. Thus analysis of purchased commodities of cardholder will be a useful tool in fraud detection system and it is assuring way to check fraudulent transaction, although fraud detection system does not keep records of number of purchased goods and categories. Every user represented by specific patterns of set which containing information about last 10 transaction using credit card [23].The set of information contains spending profile of card holder, money spent in every transaction, the last purchase time, category of purchase etc. The potential threat for fraud detection will be a deviation from set of patterns.

Fig(1). Architecture of HMM

III. GENETIC ALGORITHM
Genetic Algorithm (GA) is an optimization technique that attempts to replicate natural evolution processes in which the individuals with the considered best characteristics to adapt to the environment are more likely to reproduce and survive. These advantageous individuals mate between them, producing descendants similarly characterized, so favourable characteristics are preserved and unfavourable ones destroyed, leading to a progressive evolution of the species.
In other words The basic idea of genetic algorithms is that given a problem, the genetic pool of a specific population potentially contains the solution, or a better solution. Based on genetic and evolutionary principles, the genetic algorithm repeatedly modifies a population of artificial structures through the application of initialization, selection, crossover, and
mutation operators in order to obtain an evolved solution.

In Genetic algorithm The initial population is selected randomly from the sample space which has many populations. The fitness value is calculated in each population and is sorted out. In selection process is selected through tournament method. The Crossover is calculated using single point probability. Mutation mutates the new offspring using uniform probability measure. In elitism selection the best solution are passed to the further generation. The new population is generated and undergoes the same process it maximum number of generation is reached as The basic GA operators are crossover, selection and mutation.

i. Selection— or survival of the fittest. The key to selection is to give preference to better outcomes.

ii. Mutation— or randomly trying combinations and evaluating the success (or failure) of the outcome.

iii. Crossover— or combining portions of good outcomes in the hope of creating an even better outcome.

A. A comprehensive review of various fraud detection methods on the basis of speed accuracy and cost

As the results show, the fraud detection systems based on Fuzzy Darwinian, has a very high accuracy with 100% true positive but with very low processing speed. In another view, HMM has a fast processing speed with low accuracy. And also BN is very high in speed processing with good accuracy in comparing to other techniques. At the same time, the processing speed in decision tree is very fast enough to enable detection of credit card fraud. AIS also has a good result in between other techniques, because is a fast technique with good accuracy. For comparing other classifiers such as KNN, SVM and DT: DT has a very fast processing speed in comparing with other classifiers, KNN also with large value for K, has a good result.

B. Fusion approach of hidden markov model and genetic algorithm:

In the process of HMM each incoming transaction is submitted to the FDS for verification. FDS receives the card details and the value of purchase to verify whether the transaction is genuine or not. If the FDS confirms the transaction to be malicious, it raises an alarm and the issuing bank declines the transaction. The concerned cardholder may then be contacted and alerted about the possibility that the card is misused. HMM never check the original user as it maintains a log. The log which is maintained will also be a proof for the bank for the transaction made. HMM reduces the tedious work of an employee in bank since it maintains a log. HMM produces high false alarm as well as high false positive. That is overcome by using fusion of this with genetic algorithm which identify fraud accurately and prevent them to classify a genuine transaction as fraud provided that only the relevant fields from the database were extracted into a simple text file by applying appropriate SQL queries which reduce the accessing time and help to identify the fraud easily.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Speed of detection</th>
<th>Accuracy</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMM</td>
<td>Fast</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>FDS</td>
<td>Very low</td>
<td>Very high</td>
<td>High</td>
</tr>
<tr>
<td>AIS</td>
<td>Very fast</td>
<td>Good</td>
<td>Inexpensive</td>
</tr>
<tr>
<td>FNN</td>
<td>Very Fast</td>
<td>Good</td>
<td>Expensive</td>
</tr>
<tr>
<td>NN</td>
<td>Fast</td>
<td>Medium</td>
<td>Expensive</td>
</tr>
<tr>
<td>DT</td>
<td>Fast</td>
<td>Medium</td>
<td>Expensive</td>
</tr>
<tr>
<td>BN</td>
<td>Very Fast</td>
<td>High</td>
<td>Expensive</td>
</tr>
<tr>
<td>KNN</td>
<td>Good</td>
<td>Medium</td>
<td>Expensive</td>
</tr>
<tr>
<td>SVM</td>
<td>Low</td>
<td>Medium</td>
<td>Expensive</td>
</tr>
<tr>
<td>SOM</td>
<td>Fast</td>
<td>Medium</td>
<td>Expensive</td>
</tr>
<tr>
<td>EP</td>
<td>Low</td>
<td>Low</td>
<td>Expensive</td>
</tr>
<tr>
<td>GA</td>
<td>Good</td>
<td>Medium</td>
<td>Inexpensive</td>
</tr>
</tbody>
</table>

Table 1: comparison of different methods [6]

Genetic algorithms, inspired from natural evolution were first introduced by Holland (1975). Genetic algorithm (GA) is a search technique used in computing to find exact or approximate solutions to optimization and search problems.

GA is used in data mining mainly for variable selection and is mostly coupled with other DM algorithms [24]. And their combination with other techniques has a very good performance. They have
been used in a number of applications in engineering and social science.

The Hidden Markov Model is a finite set of states, each of which is associated with a probability distribution. Transitions among the states are governed by a set of probabilities called transition probabilities. In a particular state an outcome or observation can be generated, according to the associated probability distribution. It is only the outcome, not the state visible to an external observer and therefore states are “hidden” to the outside; hence the name Hidden Markov Model [25].

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 [28]. The number of items purchased, types of items that are bought in a particular transaction are not known to the Fraud Detection system, but it only concentrates on the amount of item purchased and use for further processing [31]. It stores data of different amount of transactions in form of clusters depending on transaction amount which will be either in low, medium or high value ranges.

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 [29]. 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 [30].

The fraud has to be deducted in real time and the number of false alert has to be minimized.

There are different devices helpful to do about that transaction. The possible actions are sending security code to registered mobile no. either by sending SMS or calling the card holder, blocking the card[2].

In financial institutions, use the fraud detection which is based on customer behavior variables. The Sample data set has been considered for the generating the fraud transactions and detection of fraud in the electronic payment systems.

**Some attributes in the training data set**

customer id, authentication type, current balance, average bank balance, time of overdraft, credit card age, deducted amount, location of credit card used, time of credit card used with respect to location, average daily overdraft, amount of transaction, credit card type, time of using credit card, card holder income, card holder age, card holder position, card holder profession, card holder marital status, average daily spending, card frequency.

The various parameters are also involved in the data set[5].

CCfreq= number of times card used
CCloc = location at which CCs in the hands of fraudsters
CCoverdraft = the rate of overdraft time
CCbank balance = the balance available at bank of CC
CCdailyspending =the average daily spending amount

It tries to find out any variance in the transaction based on the spending behavioural profile of the cardholder, shipping address, and billing address, CCfreq, CCloc, CCoverdraft, CCbank balance, CCdailyspending and so on [27]. The probabilities of initial set have chosen based on the spending behavioural profile of card holder 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 [12].

C. **Technique & Algorithm Used:**

In this section, we first describe the basic operating principles of the fusion approach of hidden markov model and genetic algorithms and then explain the steps of the suggested hidden genetic model.

To record the credit card transaction dispensation process in conditions of a Hidden Markov Model (HMM), it creates through original deciding the inspection symbols in our representation.

For implementation of this credit card fraud detection method, only the relevant fields from the database were extracted into a simple text file by applying appropriate SQL queries. In this detection method the transaction amounts for any customer is the key input data. This pre-processing of data has been helped in reducing the data size and efficient processing, thus speeding up the training and making the dataset more concise.

We quantize the purchase values x into M price ranges V1, V2. . . VM, form the study symbols by the side of the issuing bank. The genuine price variety for each symbol is configurable based on the
expenditure routine of personal cardholders. HMM determine these prices range dynamically by using clustering algorithms (like K clustering algorithm) on the price values of every card holder transactions. It uses cluster Vk for clustering algorithm as k = 1, 2 . . . . M, which can be represented both observations on price value symbols as well as on price value range. In this prediction process it considers mainly three price value ranges such as 1) low (l) 2) Medium (m) and 3) High (h). So set of this model prediction symbols is V { l, m, h}, so V ¼ f as l (low), m (medium), h (high) which makes M ¼ 3.

E.g. If card holder perform a transaction as $ 550 and card holders profile groups as l (low) = (0, $ 100], m (medium) = ($ 200 , $ 500], and h (high) = ($ 500, up to credit card limit], then transaction which card holder want to do will come in high profile group. So the corresponding profile group or symbol is M and V (2) will be used.

In various period of time, purchase of various types with the different amount would make by credit card holder and Compute the critical values, Calculate the CC usage frequency count, CC usage location, CC overdraft, current bank balance, average daily spending. It uses the deviation in a purchasing amount of new generated transaction sequence ( and adding one new transaction in that sequence) which is one of the possibilities related to the probability calculation.

In initial stage, model does not have data of last 10 transactions, in that case, model will ask to the cardholder to feed basic information during transaction about the cardholder such as mother name, place of birth, mailing address, email id etc. Due to feeding of information, HMM model acquired relative data of transaction for further verification of transaction on spending profile of cardholder after setting of the data set. The initial population is selected randomly from the sample space which has many populations. The fitness value is calculated in each population and is sorted out.

Artificial genetic algorithm aims to improve the solution to a problem by keeping the best combination of input variables. It starts with the definition of the problem to optimize, generating an objective function to evaluate the possible candidate solutions , i.e., the objective function is the way of determining which individual produces the best outcome[26].

D. Equations

The Experiment process has these steps:-

Step1. Input group of data credit card transactions, every transaction record with n attributes, and standardize the data, get the sample finally, which includes the confidential information about the card holder along with other different parameter related to the card holder and the credit card, store in the data set.

Step2. Only the relevant fields from the database were extracted into a simple text file by applying appropriate SQL queries.

Step3. Compute the probabilities of transition from a particular state (1, 2, or 3) to different spending habits h, m, or l (for example, b1-h, b1-m, etc.).

Step3. Compute the critical values, Calculate the CC usage frequency count, CC usage location, CC overdraft, current bank balance, average daily spending.

Step3. Generate critical values found after limited number of generations. Critical Fraud Detected, Monitor able Fraud Detected, Ordinary Fraud Detected etc. using Genetic Algorithm.

This sequence is recorded from cardholder’s transaction till time t. We put this sequence in model to compute the probability of acceptance. Let us assume be this probability is α1, which can be calculated as α1 = P (O1, O2, O3, ...OR | λ), Let OR+1 be new generated sequence in next generation at time t+1, when a transaction is going to process. The total number of sequences is R+1. To consider R sequences only, we will drop O1 sequence and we will have R sequences from O2 to OR+1. Let the probability of new R sequences be α2 α2 = P (O2, O3, O4, ....OR+1 | λ),

Hence, we will find

\[ \Delta \alpha = \alpha_1 - \alpha_2 \]

If \( \Delta \alpha > 0 \), it means that it consider new sequence i.e. OR+1 with low probability and therefore, this transaction will be considered as fraud transaction if and only if percentage change in probability is greater than a predefined threshold value.

\[ \Delta \alpha / \alpha_1 \geq \text{threshold value,} \]

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

Step 4. We took the weighted average of the parameter values of the previous generation (parent) solutions and obtained the next generation (child) solution.

Step 5. Sort them and fitness function is computed on the comparison basis with its threshold value of a parameter in a dataset.

Step 6. Selection will be done from last generation with the highest fitness function value.

Step 7. We decided to run the generations until no improvements are observed and obtain the optimized result with combination of the current and the previous generation and update the dataset.

**Pseudo code of fusion approach of hidden markov model and genetic algorithm**

1. Initialize the population.
2. Extract the data from database by applying SQL query.
3. Repeat
   (i) Perform selection.
   (ii) Determine the price range dynamically using k-clustering algorithm and classify into three categories i.e. low(l), medium(m), High(h).
   (iii) Compute the critical value from the dataset.
4. Apply genetic operator to generate new solution.
5. Optimized solution is obtained and updates the dataset.

**K-Means Clustering Technique**

K-Clustering algorithm used to determine the clusters. K-means is an unsupervised learning algorithm for grouping a given set of data based on the similarity in their attribute values. Fig shows three clusters.

Transactions below 1.5 forms low spending group, transactions in between (1.5,2.5) form medium spending group, and transactions above 2.5 form high spending group. These groups are observation symbols in our implementation. Fig indicates that clustering probability of each observation symbol. In this Fig clustering probability of high spending is highest among three. It can be said that spending profile of given cardholder is high spending.

**One Time Password (OTP)**

A one-time password (OTP) is a password that is valid for only one login session or transaction. OTPs avoid a number of shortcomings that are associated with traditional passwords.
Critical Value identification:

a) Based on CC usage Frequency
- F_max = mode(Number of times card used in a day)
- CCfreq = F_max / Total number of times card used.
- If CCfreq is less than 0.2, it means this property is not applicable for fraud and critical value = CCfreq. Otherwise, it checks for condition of fraud (i.e.) = Fraud condition = number of time Card used Today (CUT) > 2 * F_max
- If true, there may chance for fraud using this property and its critical value is CUT * CCfreq.
- If false, no fraud occurrence and critical value = CCfreq.

b) Based on CC usage Location
- Total number of locations CC used so far (loc) obtained from dataset (T_loc) number of locations CC used in a day (loc)
- CCloc = mode(loc)/T_loc
- If CCloc is less than 0.01, it means this property is not applicable for fraud and critical value = 0.01. Otherwise, it checks for condition of fraud (i.e.) = Fraud condition = number of locations Card used Today (CUT) > loc If true, there may chance for fraud using this property and its critical value is loc/CUT. If false, no fraud occurrence and critical value = 0.01.

c) Based on CC OverDraft
- Number of times CC overdraft with respect to CU occurred so far. Consider the (OD) can be found as, OD with respect to CU = OD/CU. If OD with respect to CU is less than 0.02, it means this property is not applicable for fraud and critical value = OD with respect to CU. Otherwise, it checks for condition of fraud (i.e.) = Fraud condition = check whether overdraft condition occurred today from (ODT dataset) If true, there may chance for fraud using this property and its critical value is ODT * OD. If false, no fraud occurrence and critical value = OD with respect to CU.

d) Based on CC Bank Balance
- Standard Bank balance can be found as, Bb = current BB / Avg. BB. If bb is less or equals than 0.2, it means this property is not applicable for fraud and critical value = BB. Otherwise, it checks for condition of fraud (i.e.) = Fraud condition = check whether overdraft condition occurred today from (ODT dataset) If true, there may chance for fraud using this property and its critical value is currBB * BB.
- If false, no fraud occurrence and critical value = BB.
IV. RESULTS AND DISCUSSION

Fraud detection will be check on the basis of the combination of the latest 10 transaction as well the result obtained from the previous transaction. Previous transaction result is obtained by taking the average value of amount that occurred in each profile i.e. high(h), medium(m), low(l) and prompt this result to obtained the optimized result to the next generation of the transaction.

<table>
<thead>
<tr>
<th>Transaction</th>
<th>Amount</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>240</td>
<td>Medium</td>
</tr>
<tr>
<td>2.</td>
<td>535</td>
<td>High</td>
</tr>
<tr>
<td>3.</td>
<td>225</td>
<td>Medium</td>
</tr>
<tr>
<td>4.</td>
<td>205</td>
<td>Medium</td>
</tr>
<tr>
<td>5.</td>
<td>12</td>
<td>Low</td>
</tr>
<tr>
<td>6.</td>
<td>222</td>
<td>Medium</td>
</tr>
<tr>
<td>7.</td>
<td>15</td>
<td>Low</td>
</tr>
<tr>
<td>8.</td>
<td>540</td>
<td>High</td>
</tr>
<tr>
<td>9.</td>
<td>55</td>
<td>Low</td>
</tr>
<tr>
<td>10.</td>
<td>300</td>
<td>Medium</td>
</tr>
</tbody>
</table>

Computed average value of low=27, medium=238 and high=537

Apply the k-mean clustering technique to evaluate the result in the next generation of the transaction by considering computed average value of the previous generation as well new generated values.

Calculate the CC usage frequency count, CC usage location, CC overdraft, current bank balance, average daily spending.

A. Probability of False Alarm compared with Fraud Transaction Mean Distribution.

It is noted that when probability of genuine transaction is going down correspondingly probability of false transaction is going to increase and vice versa. It helps to find out the false alarm for the detection of fraud transaction. Hence, when the probability of false alarm will be more than threshold probability, then it will generate an alarm for fraudulent and also decline the transaction.
V. CONCLUSION

In this paper, it has been discussed that how the fusions approach of Hidden Markov Model and genetic algorithm will facilitate to stop fraudulent online transaction through credit card. The Fraud Detection System is also scalable for handling vast volumes of transactions processing. Its approach based credit card fraud detection system is not taking long time and having complex process to perform fraud check like the existing system and it gives better and fast result than existing system. The Hidden Markov Model makes the processing of detection very easy and tries to remove the complexity. At the initial state HMM checks the upcoming transaction is fraudulent or not and it allow to accept the next transaction or not based on the probability result. Artificial genetic algorithm aims to improve the solution to a problem by keeping the best combination of input variables. It starts with the definition of the problem to optimize, generating an objective function to evaluate the possible candidate solutions, i.e., the objective function is the way of determining which individual produces the best outcome. The different ranges of transaction amount like low group, medium group, and high group as the observation symbols were considered and card holder behaviour related to card is also observed and maintain in a database. The types of item have been considered to be states of the Hidden Markov Model. It is recommended that a technique for finding the spending behavioural habit of cardholders, also the application of this knowledge in deciding the value of observation symbols and initial estimation of the model parameters. The relative studies and our results sure that the correctness and effectiveness of the proposed system is improved to large extend.

REFERENCES

[1] (Source: Nilson Report, August 2013)
[3] (Source: ABA Deposit Account Fraud Survey, 2011)


[21] Clifton Phua, Vincent Lee, Kate Smith, and Ross Gayler “A comprehensive survey of data mining-based fraud detection research”. In Artificial Intelligence Review.(2005)


