Fraud Detection of Credit Card Payment System by Genetic Algorithm

K.RamaKalyani, D.UmaDevi
Department of Computer Science, Sri Mittapalli College of Engineering, Guntur, AP, INDIA.

Abstract— With an increase usage of credit cards for online purchases as well as regular purchases, causes a credit card fraud. In the mode of electronic payment system, fraud transactions are rising on the regular basis. The Modern techniques based on the Data Mining, Genetic Programming etc. has used in detecting fraudulent transactions. The technique of finding optimal solution for the problem and implicitly generate the results using genetic algorithm. The aim is to develop a method of generating test data and to detect fraudulent transaction with this algorithm. This algorithm is an optimization technique and evolutionary search based on the principles of genetic and natural selection, heuristic used to solve high complexity computational problems. This paper presents to find the detection of credit card fraud mechanism and examines the result based on the principles of this algorithm. The benefit of detecting fraud is to clear for both credit card companies and their clients. The fraudulent transactions are not prevented from being cleared; the company must accept the financial cost of that transaction. This reduces the cost associated with higher interest rates, and its charges.

Index Terms— Credit card, Electronic Payments system, Fraud detection, Genetic algorithm

1 INTRODUCTION

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 using genetic algorithm. 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.

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.

2 RELATED WORK

It is to develop a credit card fraud detection system using genetic algorithm. During the credit card transaction, the fraud is detected and the number of false alert is being minimized by using genetic algorithm. Instead of maximizing the numbers of correctly classified transactions we defined an objective function where the misclassification costs are variable and thus, correct classification of some transactions are more important than correctly classifying the others.

The high amount of losses due to fraud and the awareness of the relation between loss and the available limit

have to be reduced.

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 blocking the card, sending SMS or calling the card holder [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.

The various parameters are involved in the data set.
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
Data set T={t1,t2,t3,….,tn}, U is one data object, If p parts of data set named S in data set is far away from object U, S ∈ T, U ∈ T, then U is Common object.

The proposed system overcomes the above mentioned issue in an efficient way. Using genetic algorithm the fraud is detected and the false alert is minimized and it produces an optimized result.

The fraud is detected based on the customer's behavior. A new classification problem which has a variable misclassification cost is introduced.

Here the genetic algorithms is made where a set of interval valued parameters are optimized.

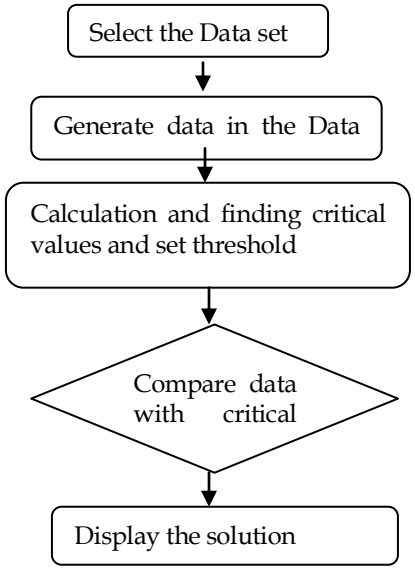
3 FIGURES AND TABLES

TABLE I. ATTRIBUTES OF TRAINING SAMPLE DATA SET

Attribute number	Attribute
1	Customer Id
2	Authentication type
3	Current balance
4	Average bank balance
5	Times of Overdraft
6	Credit card age
7	deducted amount
8	location of CC used
9	Time of the CC used with respect to location
10	Average daily Over draft
11	Amount of transaction
12	Credit card type
13	The Time of using credit card
14	Card holder income
15	Card holder age
16	Card holder position
17	Card holder profession
18	Card holder martial status
19	Average daily spending
20	Card frequency

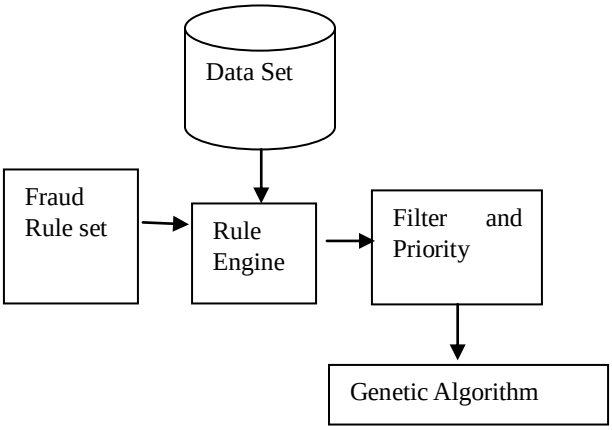
The current values of these parameters have been determined, and critical values are compared with the data set parameters, also maximizes the number of true alerts given that the number of alerts does not exceed a certain level.

Fig. 1 The simple method of Genetic Algorithm



Genetic algorithm is the procedure is repeated until a pre specified number of generations has passed, and the best solution found. It is parametric procedure and it needs to be problem undertaken to get a better performance. The list of these parameters and the settings are needed to generate fraud transaction.Such parameters are needed to compute the critical values, to calculate the CC usage frequency count, CC usage location, CC overdraft, current bank balance, average daily spending etc. as shown in Fig.1

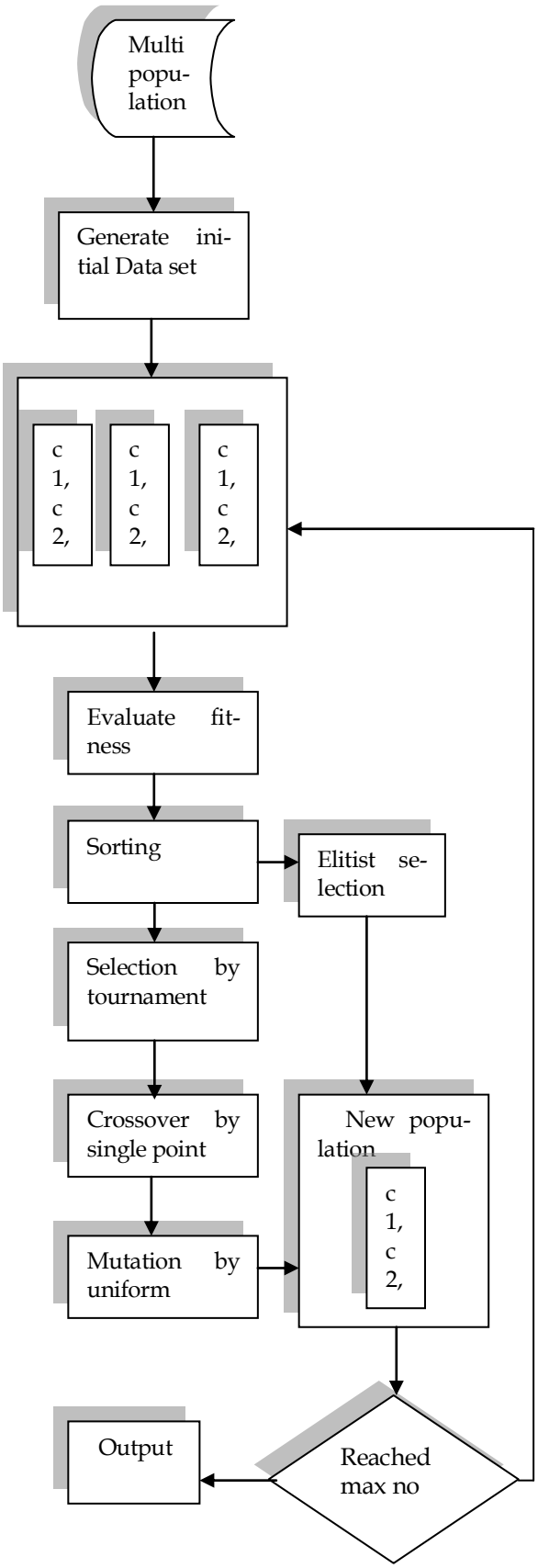
Fig. 2 System Design



As shown in Fig 2, Genetic algorithms are evolutionary algorithms which aim at obtaining better solutions as time progresses. It also been used in data mining mainly for variable selection and are mostly coupled with other data mining algorithms.

In this study, we are solving the classification problem by using only a genetic algorithm solution.

Fig. 3 The flow of Genetic Algorithm



4 EXPERIMENT PROCESS

The Experiment process has four 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, store in the data set.
Step2. 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, Monitorable Fraud Detected, Ordinary Fraud Detected etc. using Genetic algorithm
Step4. Generate fraud transactions using this algorithm.
This is to analyze the feasibility of credit card fraud detection based on technique, applies detection mining based on critical values into credit card fraud detection and proposes this detection procedures and its process.

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 shown in Fig 3.

Pseudo code of genetic algorithm

Initialize the population
Evaluate initial population
Repeat
Perform competitive selection
Apply genetic operators to generate new solutions
Evaluate solutions in the population
Until some convergence criteria is satisfied.

Selection process

Selection is used for choosing the best individuals, that is, for selecting higher fitness values. The selection operation takes the current population and produces a 'mating pool' which contains the individuals which are going to reproduce. There are several selection methods, like biased selection, random selection, roulette wheel selection, tournament selection. In this work the following selection mechanisms are used.

Tournament Selection

Tournament selection has been used in this as it selects optimal individuals from diverse groups. It selects t individuals from the current population uniformly at random, forms a tournament and the best individual of a group wins the

tournament and is put into the mating pool for recombination. This process is repeated the number of times necessary to achieve the desired size of intermediate population. The tournament size controls the selection strength. The larger the tournament size, the stronger is the selection process.

Elitist Selection

In order to make sure that the best individuals of the solution are passed to further generations, and should not be lost in random selection, this selection operator is used. So we used this algorithm, based on the higher fitness value and are passed to the next generation of population.

Reproduction

To generate a second generation population of solutions from those selected through genetic operators: crossover (also called recombination), and/or mutation.

These processes ultimately result in the next generation population different from the initial generation. Generally the average fitness will have increased by this procedure for the population, since only the best organisms from the first generation are selected for breeding, along with a small proportion of less fit solutions, for reasons already mentioned above.

Although Crossover and Mutation are known as the main genetic operators, it is possible to use other operators such as regrouping, colonization-extinction, or migration in genetic algorithms.

Termination

This generational process is repeated until a termination condition has been reached. Common terminating conditions are:

- A solution is found that satisfies minimum criteria
- Fixed number of generations reached
- Allocated budget (computation time/money) reached
- The highest ranking solution's fitness is reaching or has reached a plateau such that successive iterations no longer produce better results
- Manual inspection
- Combinations of the above

This Experiment process may contains

- User GUI
- Critical Value Identification
- Fraud Detection using Genetic Algorithm

i. User GUI:

In this module, User Interface module is developed using Applet Viewer. This module is developed to user to identify the credit card fraud using genetic algorithm technique. So the user interface must be capable of providing the user to upload the dataset and make manipulations and finally must show the user whether fraud has been detected or not. Only final output will be in applet screen. All the generation details (crossover and mutation) will be in the console screen of eclipse.

ii. Critical Value identification:

Based on CC usage Frequency

CCfreq = Total number card used (CU) / CC age

If CCfreq is less than 0.2, it means this property is not applicable for fraud and critical value =CCfreq
Otherwise, it check for condition of fraud (i.e) =
Fraud condition = number of time Card used Today (CUT) >(5 * CCfreq)
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

Based on CC usage Location

Number of locations CC used so far (loc) obtained from dataset (loc)
If loc is less than 5, 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) > (5 * 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

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 with respect to CU
If false, no fraud occurrence and critical value = Od with respect to CU

Based on CC Book Balance

Standard Book balance can be found as,
Bb = current BB / Avg. BB
If bb is less or equals than 0.25, it means this property is not applicable for fraud and critical value = BB
Otherwise, it check for condition of fraud (i.e) =
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

iii. Fraud Detection using Genetic Algorithm

In this module the system must detect whether any fraud has been occurred in the transaction or not. It must also display the user about the result. It is calculated based on following:
Age of CC in months can be calculated using CCage (from dataset) by,

Age of cc by month = CCage/30
Total money being spent from the available limit (1 lakh _ 100000)
Bal = 100000 – avg BB
So, total money spent can be found as, Tot = Age of cc by month * Bal
Total money spent on each month can be calculated as,
Ds=tot* Age of cc by month
it check for condition of fraud (i.e) =
Fraud condition = (10 * DS) is amount spent today (AmtT in dataset)
If true, there may chance for fraud using this property and its critical value is AmtT/(10*DS)
If false, no fraud occurrence and critical value 0.01
It is to prevent financial institutions from great losses before and reduce risks associated with electronic payment system. And this is to prove accurate in predicting fraudulent transaction. If this algorithm is applied into bank credit card fraud detection system, the probability of fraud transactions can be predicted after credit card transactions by the financial institution.

The credit card fraud detection system there is a need of very large amount of previous data related to card holder made during credit card use in purchase, we must have to design some system that may control credit card fraud before any real transaction is made.

5 RESULTS

The data set may contains the following parameters
CardID, Auth, Cur.BB, CU, Avg.BB, OD, CCAge, CUT, Loc, LocT, ODT, AmtT
11111,111,20000,13,60000,4,125,0,3,0,0,0
11112,112,25000,40,55000,20,264,6,4,2,0,9000
11113,113,15000,21,45000,3,111,2,10,2,1,15000
11114,114,100000,90,60000,29,350,1,11,14,0,8500
11115,115,15000,85,61000,17,211,3,3,7,0,12000
11116,116,72000,51,60000,19,321,5,9,0,1,12000
11117,117,20000,43,40000,12,261,0,6,1,0,0
11118,118,23000,31,35000,9,259,4,7,4,0,19000
11119,119,12000,29,45000,7,183,1,10,2,0,16000
11120,120,35000,189,70000,30,269,5,4,10,1,11000
11121,121,77000,31,60000,7,311,2,8,2,0,11000
11122,122,50000,31,65000,9,208,0,2,11,0,0
11123,123,29000,51,55000,16,291,1,6,12,0,14000
11124,124,81000,62,70000,18,196,2,6,3,0,9000
11125,125,13000,83,55000,12,138,4,3,1,1,19000
11126,126,70000,32,50000,9,173,0,2,12,0,0
11127,127,54000,51,75000,9,275,6,9,0,1,7000
11128,128,72000,46,40000,12,271,1,7,2,0,19000
11129,129,14000,103,30000,22,318,1,11,4,1,22000
11130,130,20000,111,61000,29,201,6,5,11,0,14000

FRAUD DETECTED

Based on CC usage Frequency

In CC ID: 11115 - Usage Freq. Fraud is found with value - 1.2085308
In CC ID: 11120 - Usage Freq. Fraud is found with value - 3.513011
In CC ID: 11124 - Usage Freq. Fraud is found with value - 0.63265306
In CC ID: 11125 - Usage Freq. Fraud is found with value - 2.405797
In CC ID: 11130 - Usage Freq. Fraud is found with value - 3.313433
Based on CC usage Location

In CC ID: 11115 - Usage Location Fraud is found with value - 0.42857143
In CC ID: 11120 - Usage Location Fraud is found with value - 0.4
In CC ID: 11122 - Usage Location Fraud is found with value - 0.18181819
In CC ID: 11126 - Usage Location Fraud is found with value - 0.16666667
In CC ID: 11130 - Usage Location Fraud is found with value - 0.45454547
Based on CC OverDraft

In CC ID: 11113 - CC OverDraft Fraud is found with value - 0.14285715
In CC ID: 11120 - CC OverDraft Fraud is found with value - 0.15873016
In CC ID: 11125 - CC OverDraft Fraud is found with value - 0.14457831
In CC ID: 11127 - CC OverDraft Fraud is found with value - 0.1764706
Based on CC Book Balance

In CC ID: 11115 - CC Book Balance Fraud is found with value - 0.4918033
In CC ID: 11125 - CC Book Balance Fraud is found with value - 0.47272727
Based on CC Average Daily Spending

In CC ID: 11120 - CC Daily Spending Fraud is found with value - 1.1
In CC ID: 11125 - CC Daily Spending Fraud is found with value - 1.2666667
In CC ID: 11130 - CC Daily Spending Fraud is found with value - 1.0769231

FRAUD TRANSACTIONS

Fraud Detected used Genetic Algorithm:

Critical Fraud Detected:

Credit Card with ID 11120.0 is detected as fraud with 4.0 occurrences and its critical value is 5.171741

Credit Card with ID 11125.0 is detected as fraud with 4.0 occurrences and its critical value is 4.289769

Credit Card with ID 11130.0 is detected as fraud with 3.0 occurrences and its critical value is 4.8449016

Monitorable Fraud Detected:

Credit Card with ID 11115.0 is detected as fraud with 3.0 occurrences and its critical value is 2.1289055

Ordinary Fraud Detected:

ACKNOWLEDGMENT

I render my sincere thanks to Assoc Prof. D. Uma Devi M.Tech, (PhD) Head of the Department of Computer Science in Sri Mittapalli College of Engineering, for his encouragement.

6 CONCLUSION

In this paper, we present to find the detection of credit card fraud mechanism and examine the result based on the principles of this algorithm. In this paper we saw genetic algorithm that are being used to execute credit card fraud how credit card fraud impact on financial institution as well as merchant and customer, fraud detection technique by genetic algorithm. The Genetic algorithms are evolutionary algorithms in which the aim is to obtain the better and optimal solutions. In this study fraud detected and fraud transactions are generated with the given sample data set. If this algorithm is applied into bank credit card fraud detection system, the probability of fraud transactions can be predicted soon after credit card transactions by the banks. And a series of anti-fraud strategies can be adopted to prevent banks from great losses before and reduce risks.

7 REFERENCES

[1] S.Benson Edwin Raj, A. Annie Portia, "Analysis on Credit Card Fraud Detection Methods", IEEE International Conference on Computer, Communication and Electrical Technology, IEEE March 2011
[2] M.Hamdi Ozcelik, Mine Isik, "Improving a credit card fraud detection system using Genetic algorithm", IEEE International Conference on Networking and Information Technology, IEEE 2010.
[3] Genetic algorithms for credit card fraud detection by Daniel Garner, IEEE Transactions May 2011.
[4] Research on credit card fraud detection model based on distance sum IEEE 2009 International Joint Conference on Artificial Intelligence.
[5] Credit card fraud detection using neural network, Raghavedra Patidar, Lokesh Sharma, ISSN: 2231-2307, Volume, Issue-NCAI211,JUNE2011.
[6] Panigraili, S., Kundu, A., Sural, S. & Majumdar, A. (2009). Credit Card Fraud Detection: A Fusion Approach Using Dempster-Shafer Theory and Bayesian Learning. Information Fusion , 354-363.
[7] Dr Markus Roggenbach. CS364 Software testing slides. Swansea University, 2011.
[8]. D.WHITLEY,"Genetic Algorithm And Neural Network."2003.
[9] Wang Xi. Some Ideas about Credit Card Fraud Prediction China Trial. Apr. 2008, pp. 74-75.
[10] Amlan Kundu, Suvasini Panigrahi, Shamik Sural and Arun K. Majumdar, "BLAST-SSAHA Hybridization for Credit Card Fraud Detection," IEEE Transactions On Dependable And Secure Computing, vol. 6, Issue no. 4, pp.309-315, October-December 2009.
[11] Amlan Kundu, Suvasini Panigrahi, Shamik Sural and Arun K. Majumdar, "Credit card fraud detection A fusion approach using Dempster–Shafer theory and Bayesian learning," Special Issue on Information Fusion in Computer Security, Vol. 10, Issue no 4, pp.354- 363, October 2009.
[12] Liu Ren, Zhang Liping, Zhan Yinqiang. A Study on Construction of Analysis Based CRM System. Computer Applications and Software. Vol.21, Apr. 2004, pp. 46-47.
[13] A. Chiu, C. Tsai, "A Web Services-Based Collaborative Scheme for Credit Card Fraud Detection," Proceedings of the IEEE International Conference on e-Technology, e-Commerce and e-Service, pp.177-181, 2004.
[14] M. Mehdi, S. Zair, A. Anou and M. Bensebti," A Bayesian Networks in Intrusion Detection Systems," International Journal of Computational Intelligence Research, Issue No. 1, pp.0973-1873 Vol. 3, 2007.
[15] Ezawa.K. & Norton.S,"Constructing Bayesian Networks to Predict Uncollectible Telecommunications Accounts," IEEE Expert, October; 45-51, 1996.
[16] Blickle, T., & Thiele, L. (1995). A Comparison of Selection Schemes used in Genetic Algorithms (Vol. 2). Zurich: Swiss Federal Institute of Technology.
[17] Jitendra Dara,Laxman Gundemoni, "Credit Card Security and E-Payment." 2006.
[18] Wang Xi. Some Ideas about Credit Card Fraud Prediction China Trial. Apr. 2008, pp. 74-75.
[19] M. Hamdi Ozcelik, Ekrem Duman, Mine Isik, Tugba Cevik, Improving a credit card fraud detection system using genetic algorithm, International conference on Networking and information technology 2010.
[20] Wen-Fang YU, Na Wang, Research on Credit Card Fraud Detection Model Based on Distance Sum, IEEE International Joint Conference on Artificial Intelligence 2009.