IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 1, January 2016

Credit Card Fraud Detection

Ishu Trivedi[1], Monika[2], Mrigya Mridushi[3]

Student, Dept. of Computer Science and Engineering, Sikkim Manipal Institute of Technology, Rangpo, India[1, 2, 3]

Abstract: The usage of credit cards for online and regular purchases is exponentially increasing and so is the fraud related with it. A large number of fraud transactions are made every day. Various modern techniques like Data Mining, Genetic Programming, etc. are used in detecting fraudulent transactions. This paper uses genetic algorithm which comprises of techniques for finding optimal solution for the problem and implicitly generating the result of the fraudulent transaction. The main aim is to detect the fraudulent transaction and to develop a method of generating test data. This algorithm is a heuristic approach used to solve high complexity computational problems. It is an optimization technique and evolutionary search based on the genetic and natural selection. The implementation of an efficient fraud detection system is imperative for all credit card issuing companies and their clients to minimize their losses.

Keywords: Credit card, Electronic commerce, Fraud detection, Genetic algorithms.

I. INTRODUCTION

A credit card is a thin handy plastic card that contains identification information such as a signature or picture, and authorizes the person named on it to charge purchases or services to his account - charges for which he will be billed periodically. Today, the information on the card is read by automated teller machines (ATMs), store readers, bank and is also used in online internet banking system. They have a unique card number which is of utmost importance. Its security relies on the physical security of the plastic card as well as the privacy of the credit card number.

There is a rapid growth in the number of credit card transactions which has led to a substantial rise in fraudulent activities. Credit card fraud is a wide-ranging term for theft and fraud committed using a credit card as a fraudulent source of funds in a given transaction. Generally, the statistical methods and many data mining algorithms are used to solve this fraud detection problem. Most of the credit card fraud detection systems are based on artificial intelligence, Meta learning and pattern matching.

The Genetic algorithms are evolutionary algorithms which aim to obtain the better solutions in eliminating the fraud. A high importance is given to develop efficient and secure electronic payment system to detect whether a transaction is fraudulent or not.

In this paper, we will focus on credit card fraud and its detection measures. A credit card fraud occurs when one individual uses other individuals' card for their personal use without the knowledge of its owner. When such kind of cases takes place by fraudsters, it is used until its entire available limit is depleted.

Thus, we need a solution which minimizes the total available limit on the credit card which is more prominent to frauds. And, a Genetic algorithm generates better solutions as time progresses. The complete emphasis is given on developing efficient and secure electronic payment system for detecting the fraudulent.

II. VARIOUS TECHNIQUES USED IN CREDIT CARD FRAUD

The advent of credit card has not just provided us with the comfort and convenience but has also attracted malicious characters as it is the easiest way to earn a huge amount of money over a very short span of time. Also it takes a while to realize such kind of fraud has occurred to the user.

A few common techniques that fraudster uses are:

- Copying a credit card and somehow getting hold of the secret pin of the user.
- Vendors charging more money from the user's credit card compared to what they have agreed to and without the latter being aware of the charged money.

So, not just the customers but, the bank issuing credit cards suffer from the losses and hence, it is their interest to reduce the illegitimate use of credit cards leading to development of various credit card fraud detection techniques.
Fraud detection is then carried out after observing a number of transactions and then identifying and classifying them into the genuine transaction and the fraudulent transaction.

III. PROBLEMS WITH CREDIT CARD FRAUD DETECTION

There are lots of issues that make this procedure tough to implement and one of the biggest problems associated with fraud detection is the lack of both the literature providing experimental results and of real world data for academic researchers to perform experiments on. The reason behind this is the sensitive financial data associated with the fraud that has to be kept confidential for the purpose of customer's privacy.

Now, here we enumerate different properties a fraud detection system should have in order to generate proper results:

- The system should be able to handle skewed distributions, since only a very small percentage of all credit card transactions is fraudulent.

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 1, January 2016

- There should be a proper means to handle the noise. Noise is the errors that is present in the data, for example, incorrect dates. This noise in actual data limits the accuracy of generalization that can be achieved, irrespective of how extensive the training set is.
- Another problem related to this field is overlapping data. Many transactions may resemble fraudulent transactions when actually they are genuine transactions. The opposite also happens, when a fraudulent transactions appears to be genuine.
- The systems should be able to adapt themselves to new kinds of fraud. Since after a while, successful fraud techniques decreases in efficiency due to the fact that they become well known because an efficient fraudster always find a new and inventive ways of performing his job.
- There is a need for good metrics to evaluate the classifier system. For example, the overall accuracy is not suited for evaluation on a skewed distribution, since even with a very high accuracy; almost all fraudulent transactions can be misclassified.
- The system should take care of the amount of money that is being lost due to fraud and the amount of money that will be required to detect that fraud. For example, no profit is made by stopping a fraudulent transaction that is way lesser than the amount of money that will be required to detect it.

These points direct us to the most important necessity of the fraud detection system, which is, a decision layer. The decision layer decides what action to take when fraudulent behavior is observed taking into account factors like, the frequency and amount of the transaction.

IV. CREDIT CARD FRAUD DETECTION METHODS

A proper and thorough literature survey concludes that there are various methods that can be used to detect credit card fraud detection. Some of these approaches are:

- Artificial Neural Network
- Bayesian Network
- Neural Network
- Hidden Markov Method
- Genetic Algorithm

In our research paper, as stated earlier, we will be emphasizing on the Genetic algorithm and how it is used in credit card fraud detection systems.

V. GENETIC ALGORITHM

Genetic Algorithm is an optimization technique that attempts to replicate natural evolution processes. The genetic pool of a specific population for a given problem potentially contains the solution, or a better solution. This is the basic idea behind the genetic algorithm. On the basis of genetic and evolutionary principles, the genetic algorithm repeatedly modifies a population of artificial structures through the application of initialization, selection, crossover, and mutation operators. This is done in order to obtain an evolved solution.

Artificial genetic algorithm aims at improving the solution to a problem. This improvement is carried out by keeping the best combination of input variables. It optimizes the problem definition and also generates an objective function that is the way of determining which individual produces the best outcome.

At first, from the sample space having many populations, the initial population is randomly selected and the fitness value is calculated and sorted. The tournament method is used in selection process and single point probability is calculated in the crossover. In mutation, the new offspring mutates using uniform probability measure. Always the best solution are selected and passed to the further generation, each time a new population is generated.

The operators of Genetic Algorithm are:

- Selection – It is the survival of the fittest and the preference is always given to better outcomes.
- Mutation – It is based on trying random combinations and evaluating the result (success or failure) of the outcome.
- Crossover- It is done by combining portions of good outcomes in the hope of creating an even better outcome.

A. 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.

B. System Design

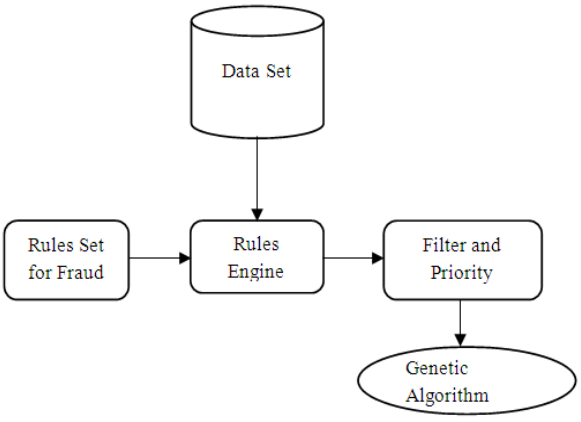

Fig. 1 Overall System Design

The above architectural design describes the work structure of the system:

- The data warehouse contains the customer data. This customer data is subjected to the rules engine and again, the rules engine comprises of the rules set.
- The filter and priority module sets the priority for the data and hence, plays a very important role in the system. Then the filtered data is sent to the Genetic Algorithm module which performs its functions and generates the output.

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 1, January 2016

C. Process Flow of Genetic Algorithm

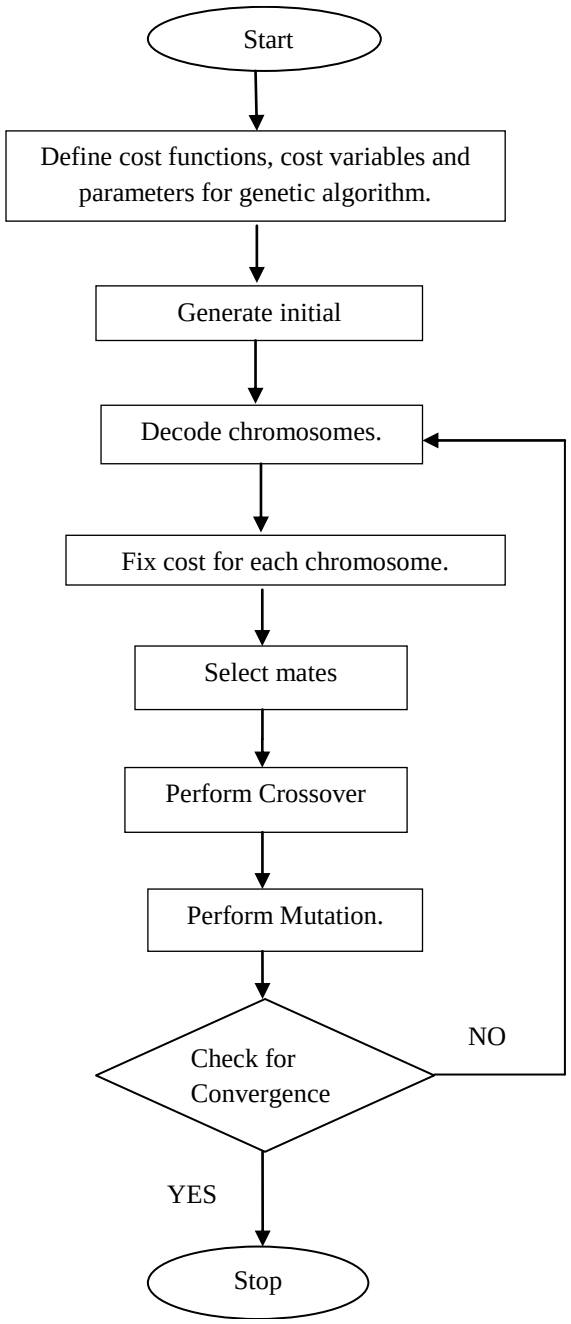

Fig. 2 Process Flow Diagram

D. Data Set Used for Fraud Detection System

The various parameters that are involved in the sample data set for fraud detection are:

CCfreq = number of times card is being used
CCage = age of credit card
CCloc = location at which the credit card is used by fraudsters
CCoverdraft = the rate of overdraft time
CCbank_balance = the balance available at bank of credit card
CCdailyspending = the average daily spending amout
U = it is one data object
Data Set T= {

If p parts of data set named S is far away from object U, S belongs to T, and U becomes a common object. The proposed system overcomes the above mentioned issue in an efficient way using genetic algorithm for fraud detection and also minimizes the false alert for generating an optimized result.

1). Number of times credit card used so far (CCfreq) obtained from dataset:
CCfreq = Total number card used (CU) / CCage
If CCfreq is less than 0.2, it means this property is not applicable for fraud andcritical value = CCfreq
Otherwise, it checks for condition of fraud:
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

2). Number of locations credit card used so far (CCloc) obtained from dataset:
If CCloc 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:
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 CCloc/CUT
If false, no fraud occurrence and critical value =0.01

3). Number of times CCoverdraft w.r.t credit card usage occurred so far can be found as:
Overdraft w.r.t CU = OD/CU
If Overdraft w.r.t CU is less than 0.02, it means this property is not applicable for fraud and critical value = Overdraft w.r.t CU
Otherwise, it check for condition of fraud is,
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 * Overdraft w.r.t CU
If false, no fraud occurrence and critical value = Overdraft w.r.t CU

4). Based on Credit Card Book Balance (BB):
Standard Book balance can be found as,
BB = current BB / AverageBB
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,
If true, there may be chance for fraud using this property and its critical value is currBB * BB.
If false, no fraud occurrence and critical value = BB (book balance)

E. Use of 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,

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 1, January 2016

Age of cc by month = CCage/30
Total money being spent from the available limit (1 lakh = 100000)
Bal = 100000 – averageBB
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 checks for condition of fraud is,
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

1). Evaluation of the fitness of a chromosome:

```
int fitness ( Chromosome input )
{
        int fitness = 0 ;
        int [ ] ideal = new int [A : E ] ;
//Array of ideal inputs A to E .
        int [ ] actual = input . toArray ( ) ;
// Retrieved at a from chromosome .
        for ( int = A ; E ; i++)
        {
                fitness = absolute ( actual −ideal ) ;
        }
        return fitness ;
}
```

2). Pseudo code for Crossover:

```
Chromosomecrossover (Chromosome parentX, Chromosome parentY)
{
        int crossPoint = 8 ;
        string x First Half = parentX .substring (0 , cross Point ) ;
        string x Second Half = parentX .sub
        string (cross Point, parentX. length ) ;
        string y First Half = parentY .substring (0 , cross Point ) ;
        string y Second Half = parentY .substring (cross Point , parentY. length ) ;
        Chromosome crossed X = x FirstHalf + y Second Half;
        Chromosome crossed Y = y FirstHalf + x Second Half;
}
```

3). Pseudo code for Mutation:

```
Chromosome mutate (Chromosome)
{
        int randomValue = new Random(Chromosome . length );
        if ( Chromosome . valueAt ( randomValue ) == 0 )
        {
        Chromosome .valueAt ( randomValue ) = 1;
        }
        else
        {
        Chromosome .valueAt (randomValue ) = 0 ;
        }
        return Chromosome ;
}
```

VI. RESULTS

This detection process constitutes of four steps. These steps are mentioned below:

- Input all the transactions record and standardize the data. Finally get the sample which includes the confidential information about the card holder in the data set with their consent.
- In this step the CCusage frequency count, CC location, CC overdraft, Current bank balance and average daily spending is computed.
- Generating critical values after finding out the limited number of generations for critical fraud detected, monitorable fraud detected, ordinary fraud detected, etc. using Genetic Algorithm.
- Generate fraud transactions detected in the final step. It is done by applying detection mining on critical values obtained in the process of fraud detection.

VII. CONCLUSION

This method proves accurate in finding out the fraudulent transactions and minimizing the number of false alert. Genetic Algorithm is appropriate in such kind of application areas. The use of this algorithm in credit card fraud detection system results in detecting or predicting the fraud probably in a very short span of time after the transactions has been made. This will eventually prevent the banks and customers from great losses and also will reduce risks.

VIII. FUTURE ENHANCEMENTS

The findings obtained here is not in a generalized form that can be directly used in the global fraud detection problem, here, we have considered a sample data set. As future work, some effective algorithm could be developed for the classification problem with variable misclassification costs.

REFERENCES

[1] Nitu Kumari, S. Kannan and A. Muthukumaravel, “Credit Card Fraud Detection Using Genetic-A Survey” published by Middle-East Journal of Scientific Research , IDOSI Publications, 2014
[2] Satvik Vats, Surya Kant Dubey, Naveen Kumar Pandey, “A Tool for Effective Detection of Fraud in Credit Card System”, published in International Journal of Communication Network Security ISSN: 2231 – 1882, Volume-2, Issue-1, 2013.
[3] Rinky D. Patel and Dheeraj Kumar Singh, “Credit Card Fraud Detection & Prevention of Fraud Using Genetic Algorithm”, published by International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-6, January 2013.
[4] M. Hamdi Ozcelik, Ekrem Duman, Mine Isik, Tugba Cevik, “Improving a credit card fraud detection system using genetic algorithm”, published by International conference on Networking and information technology, 2010.
[5] Wen-Fang YU, Na Wang,“ Research on Credit Card Fraud Detection Model Based on Distance Sum”, published by IEEE International Joint Conference on Artificial Intelligence, 2009.