



Cairo University,
Faculty Of Engineering,
Dep. Of Computer Engineering,
2nd year.



Faculty of Engineering
Cairo University

Credit Card Fraud Detection using Genetic Algorithm

By: Mathletes team

MathlΣtΣs

Team Members:

- 1-Areej Raafat
- 2-Asmaa Sayed
- 3-Aya Sameer
- 4-Aya Adel
- 5-Bassant Mohamad
- 6-Shaza Mohamad

Table Of Contents:

1. Cover page	1
2. Abstract	4
3. Introduction	4
4. Various Techniques used for the credit card fraud	5
4.1. Lost or stolen card	5
4.2. Rigging card fraud	5
4.3. Free wi-fi trap	5
4.4. Fake call or Email	5
5. Problems in Detection	6
6. Detection methods	6
7. Evolutionary Algorithms (EAs)	6
7.1. Population -Based	6
7.2. Fitness-Oriented	6
7.3. Variation-Driven	6
8. Genetic Algorithm	7
9. Fraud detection using Genetic Algorithm	10
10. Results	12
11. Conclusion	13
12. Future work	13
13. References	13

List Of Figures:

1. Basic algorithm for genetic algorithm	8
2. Representation of number 7 in binary	8
3. Simple Section of the used data set	11
4. Interface of the algorithm we used	13

Abstract:

With the fast development of economy, credit cards are more popular in commercial transactions. The problem of the credit card fraud appears accordingly. Credit card fraud emerges when someone uses other persons' card without the knowledge of the card holder. Machine learning and artificial intelligence approaches have been suggested to defeat these problems. There are many methods used to detect fraud and one of them is genetic algorithm. The main goal of the genetic algorithm is to get the better solution so as to remove the fraud and to develop secure electronic payment system to detect if a transaction is fraudulent or not[2].

Introduction:

The credit card is a plastic card supplied to users by banks as a system of payment and contains a unique identification. It allows the user to buy goods and services based on their promise to pay for these goods and services.

Fraud by definition is the illegal usage of any system or goods. Fraud can appear in a variety of different domains including finance, telecommunications, health care and public services. Increased use of the internet for online shopping has resulted in a considerable proliferation of credit card transactions throughout the world. Thus a rapid growth in the number of credit card transactions has led to a substantial rise in fraudulent activities. Credit card fraud is an extensive term for the theft and fraud committed using credit card as a fraudulent source of funds in the given transactions. Credit card fraud happens when one person uses other persons' card for their personal use without the knowledge of the card owner.

When a card is captured, or stolen or lost, it is used by the fraudsters until the entire available limit of the credit card is depleted. Fraud detection problem is a classification problem, in which some of statistical methods many data mining algorithms have proposed to solve it but since fraud detection is approached as a classification problem, classical data mining algorithms are not directly applicable. So an alternative approach is made by using general purpose approaches like genetic algorithms. Genetic algorithms are evolutionary algorithms which aim at obtaining better solutions to eliminate the fraud problem as time progresses. The main goal is to develop efficient and secure electronic payment system to detect whether a transaction is fraudulent or not. The traditional detection method mainly depends on database system and the education of customers, which usually are delayed, inaccurate and not in-time

To thoroughly test credit card frauds one would require a large set of realistic transactions, both legitimate and fraudulent. Therefore real data should not be used, so instead a way

to generate synthetic data must be found.

In this paper, we are going to talk about credit card fraud and the measures to detect the fraud.

Various Techniques Used For The Credit Card Frauds[6]:

1. Lost Or Stolen Card:

Many times, person lose their wallet and lose the credit card and maybe it gets stolen.

So, your credit card will be taken from your ownership, either through theft or because you might have lost it.

The criminals who get their hands on your credit card will use it for shopping or for various payments. It is difficult to do this through machines, as in most of the credit cards they will require a pin number. However, it is easy enough to use a found or stolen card to make online purchases

1.1.1 2. Rigging Card Fraud:

Credit Card Skimming is done with a device usually placed with a regular credit card processing terminal.

The places like Gas stations and ATMs have been a long time favorite for scammers to place skimming devices.

The scammer uses that device to steal your credit card information when you swipe it or hand it to a cashier to pay for goods or services. Once your credit card information has been stolen, the Scammers can use it to create fake credit cards And also, Scammers can simply withdraw this card at payment terminals for goods or services purchased.

Places such as gas stations and ATMs have long been a favorite of scammers for skimming machines.

1.1.2 3. Free Wi-Fi Trap

Credit card thieves exploit that free Wi-Fi is preferred by many people by setting up a free Wi-Fi hotspot that does not require a password. Once connected, hackers or phishers can actually access any information you send over the network.

If you sign in to your bank online or check your credit card balance, the fraudster can get your username and password. A fraudster can steal all your credit cards and personal information if you place an order from your mobile phone.

1.1.3 4. Fake Call or Email

Fraudsters and credit card thieves make a fake call or send an email to get your credit card information. They can make a call to give you the best benefit on your credit card or to request credit card information to update in their system or warn the cardholder to update their personal information. Responding to these emails with

personal information, even clicking the link and entering personal information can give the fraudster enough information to commit the theft.

Problems in Detection:

The credit card fraud detection is very difficult to implement in practice because of many reasons. One of them, there is not enough information available that provides experimental results on the real-world data so that researchers can do experiments on it. The reason of this is that the financial data which is sensitive is related with the fraud that has to be kept confidential for the purpose of customers' privacy [3]

To achieve good results, Fraud detection system should have the following properties:

1. It should be able to handle skewed distribution, because a very small percentage of all credit card is fraudulent.[4]

2. There should be a way to handle noise to refer to the error present in the data.[3]

3. Another problem related to the credit card fraud detection is overlapping data like False alarm or mistakes. [3]. The system should be able to end and minimize these problems.[5]

4. the system should have good metrics to estimate the classifier system. the overall accuracy is not appropriate for estimating the skewed distribution.

Detection Methods

These methods are used to find out if the given transaction is genuine or

fraudulent when the transaction is fraudulent, the detection system must recognize it and produce an alert for fraudulent transactions, there are a lot of methods that can be used to detect credit card fraud.

And here are some of them[1]:

- Artificial neural network
- Bayesian network
- Decision tree
- Neural network
- Hidden Markov method
- Outlier detection
- Genetic algorithm

Evolutionary Algorithms (EAs):

1. Population-Based:

Evolutionary algorithms are to optimize a process in which current solutions are bad to generate new better solutions. The set of current solutions from which new solutions are to be generated is called the population.

2. Fitness-Oriented: If there are some several solutions, how to say that one solution is better than another? There is a fitness value associated with each individual solution calculated from a fitness function. Such fitness value reflects how good the solution is.

3. Variation-Driven: If there is no acceptable solution in the current population according to the fitness function calculated from each individual, we should make something

to generate new better solutions. As a result, individual solutions will undergo a number of variations to generate new solutions.

We will move to GA and apply these terms.

1.2 Genetic Algorithm (GA):

The genetic algorithm is a random-based classical evolutionary algorithm. By random here we mean that in order to find a solution using the GA, random changes applied to the current solutions to generate new ones. Note that GA may be called Simple GA (SGA) due to its simplicity compared to other EAs.

GA is based on Darwin's theory of evolution. It is a slow gradual process that works by making changes to the making slight and slow changes. Also, GA makes slight changes to its solutions slowly until getting the best solution.

Here is the description of how the GA works:

GA works on a population consisting of some solutions where the population size (popsize) is the number of solutions. Each solution is called individual. Each individual solution has a chromosome. The chromosome is represented as a set of parameters (features) that defines the individual. Each chromosome has a set of genes. Each gene is represented by somehow such as being represented

as a string of 0s and 1s as in the next diagram.

Also, each individual has a fitness value. To select the best individuals, a fitness function is used. The result of the fitness function is the fitness value representing the quality of the solution. The higher the fitness value the higher the quality the solution. Selection of the best individuals based on their quality is applied to generate what is called a mating pool where the higher quality individual has higher probability of being selected in the mating pool.

The individuals in the mating pool are called parents. Every two parents selected from the mating pool will generate two offspring (children). By just mating high-quality individuals, it is expected to get a better quality offspring than its parents. This will kill the bad individuals from generating more bad individuals. By keeping selecting and mating high-quality individuals, there will be higher chances to just keep good properties of the individuals and leave out bad ones. Finally, this will end up with the desired optimal or acceptable solution.

But the offspring currently generated using the selected parents just have the characteristics of its parents and no more without changes. There is no new added to it and thus the same drawbacks in its parents will actually exist in the new offspring. To overcome such problem, some changes will be applied to each offspring to create new individuals. The set of all newly generated individuals will be the new population that replaces the previously used old

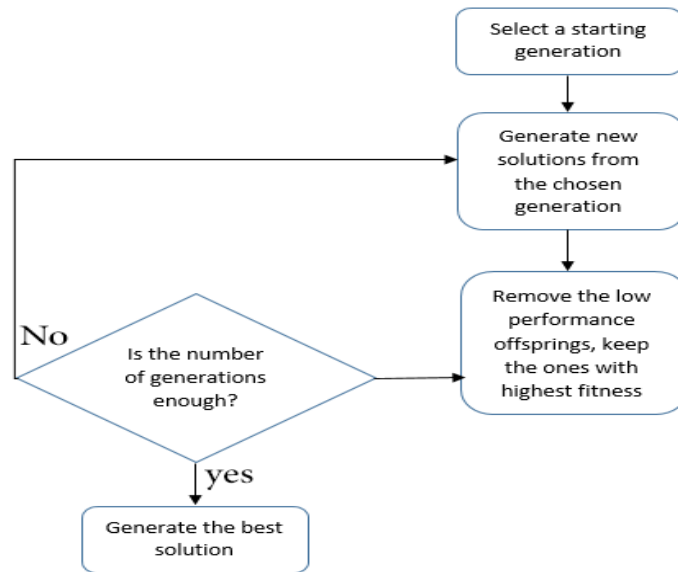


Figure 1: Basic algorithm for genetic algorithm

population. Each population created is called a generation. The process of replacing the old population by the new one is called replacement. The following diagram summarizes the steps of GA.

There are two questions to be answered to get the full idea about GA:

1. How the two offspring are generated from the two parents?
2. How each offspring gets slightly changed to be an individual?

We will answer these questions later.

Chromosome Representation and Evaluation

There are different representations available for the chromosome and the selection of the proper representation is problem specific. The good representation is what makes the

search space smaller and thus easier search.

The representations available for the chromosome including:

- **Binary:** Each chromosome is represented as a string of zeros and ones.
- **Permutation:** Useful for ordering problems such as travelling salesman problem.
- **Value:** The actual value is encoded as it is.

For example, if we are to encode the number 7 in binary, it might look as follows:

3	2	1	0
0	1	1	1

Figure 2: Representation of number 7 in binary

Each part of the above chromosome is called gene. Each gene has two properties. The first one is its value (allele) and the second one is the location (locus) within the chromosome which is the number above its value.

Each chromosome has two representations.

1. **Genotype:** The set of genes representing the chromosome.

2. **Phenotype:** The actual physical representation of the chromosome.

In the above example, binary of 0111 is the genotype and 7 is the phenotype representation.

After representing each chromosome the right way to serve to search the space, next is to calculate the fitness value of each individual. Assume that the fitness function used in our example is:

$F(x) = 2x+2$ Where x is the chromosome value

Then the fitness value of the previous chromosome is:

$$F(7) = 2(7) + 2 = 16$$

The process of calculating the fitness value of a chromosome is called evaluation.

Initialization

After getting how to represent each individual, next is to initialize the population by selecting the proper number of individuals within it.

Selection

Next is to select a number of individuals from the population in the mating pool. Based on the previously calculated fitness value, the best individuals based on a threshold are selected. After that step, we will end selecting a subset of the population in the mating pool.

Variation Operators

Based on the selected individuals in the mating pool, parents are selected for mating. The selection of each two parents may be by selecting parents sequentially (1-2, 3-4, and so on). Another way is random selection of the parents.

For every two parents selected, there are a number of variation operators to get applied such as:

1. Crossover (recombination).
2. Mutation.

Crossover

Crossover in GA generates new generation the same as natural mutation. By mutating the old generation parents, the new generation offspring comes by carrying genes from both parents. The amount of genes carried from each parent is random. Remember that GA is random-based EA. Sometimes the offspring takes half of its genes from one parent and the other half from the other parent and sometimes such percent changes. For every two parents, crossover takes place by selecting a random point in the

chromosome and exchanging genes before and after such point from its parents. The resulting chromosomes are offspring. Thus operator is called single-point crossover.

Note that crossover is important and without it, the offspring will be identical to its parent.

Mutation

Next variation operator is mutation. For each offspring, select some genes and change its value. Mutation varies based on the chromosome representation but it is up to you to decide how to apply mutation. If the encoding is binary (i.e. the value space of each gene have just two values 0 and 1), then flip the bit value of one or more genes.

But if the gene value comes from a space of more than two values such as 1,2,3,4, and 5, then the binary mutation will not be applicable and we should find another way. One way is by selecting a random value from such set of values.

Note that without mutation the offspring will have all of its properties from its parents. To add new features to such offspring, mutation took place. But because mutation occurs randomly, it is not recommended to increase the number of genes to be applied to mutation.

The individual after mutation is called mutant.

Fraud Detection Using Genetic Algorithm (GA):

Genetic Algorithms are evolutionary algorithms that aim at finding or obtaining the best solution to a problem, they were first introduced by Holland and have been applied to many problems in many various fields such as, computer science, sports, etc. in our case we're using this algorithm to find the best solution or method that we can use to detect fraud in credit cards transactions [7].

Every successful bank nowadays exerts the best effort they can to build a secure and efficient e-payment system that can easily identify whether the transactions happening are fraudulent or not, and often the main concept of this analysis is obtained through the behavior of the credit card holder and the transactions [8]. Genetic algorithm is not used to just detect the fraud, it's also used to minimize the number of false alerts.

But first of all let's take a look at a **generic genetic algorithm** understand well how the algorithm work before applying it on our problem [9]:

```
BasicGeneticAlgorithm ( )
{
    Initialize population;
    Evaluate population;
    While (termination criteria not met) {
        Select solutions for next generation

        Perform crossover and mutation;

        Evaluate population;
```

```
}
}
```

```
//end of algorithm.
```

So in brief the main concept of the genetic algorithms is to generate a new generation, from an existing one, that has better characteristics to adapt to the environment, the process keeps happening until we reach the best generation (**solution**). In other words the basic idea of genetic algorithms is that given a problem, the genetic pool of a specific generation actually contains the solution, or at least a better solution. And based on genetic and evolutionary principles, the genetic algorithm repeatedly modifies a generation through applying some operations which are, **initialization**, **selection**, **crossover**, and **mutation** in order to obtain an evolved and a better solution [9].

Now back to what happens when we try to detect fraud transactions in credit cards, our aim is to develop a credit card fraud detection system using genetic algorithm. This system as we mentioned previously is able to detect fraud in credit card transactions and minimize the number of false alerts as much as possible. And when actually detecting fraud in any transaction quick actions are taken, such as: blocking the card, sending SMS or calling the card holder.

Also as mentioned before we mainly detect fraud through the behavior of the transactions, so we need to calculate some behavior variables, and to calculate these variables we need

to have data about the transactions, each one of them, and then we generate what we call **data set** to use it in the algorithm or the system to complete the detection. Those variables are as follows:

- 1-Number of times card is used.
- 2-Location at which the credit card is in the hands of fraudsters.
- 3- The rate of overdraft time.
- 4-The balance available at bank of credit card.
- 5-The average daily spending amount.

Then the data set is generated and the algorithm starts to optimize the variables and keeps the best combinations of input variables (generations) through calculating what's called the objective function, which is the way to determine which individual produces the best outcome.

CardID,Auth,Cur.BB,CU,Avg.BB,OD,CCAge,C,T,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

Figure 3: Simple Section of the used data set

Results:

Generally, the process is completed in four steps which are:

Step 1: Get the data of the transactions, each transaction with n attributes, and finally generate the data set.

Step 2: Then we calculate the critical values or the previously mentioned five variables.

Step 3: Using the genetic algorithm we then recalculate those critical values after limited number of generations.

Step 4: Then the fraud transactions are generated and busted.

But what actually happened in **step 3**?
Let's clear that in brief:

The initial population or generation is selected randomly from the data set, the fitness value is calculated in each population, and those fitness values are sorted out. The Crossover is calculated using single point probability. Mutation mutates the new generation using uniform probability measure. Then the best solution is passed to the further generation. The new population is generated and undergoes the same process until maximum number of generation is reached.

To clarify some of the mentioned terms in the previous section, we can relist the basic operators of the GA [10]:

1- Selection: Here we give preference to better outcomes, whoever has the

higher fitness value survives till the next operation.

2- Mutation: which is randomly trying combinations and evaluating the success (or failure) of the outcome.

iii. Crossover: which is combining parts of good outcomes hopefully, to create an even better outcome.

Below is a figure that shows the results after loading data set in the algorithm and running it.

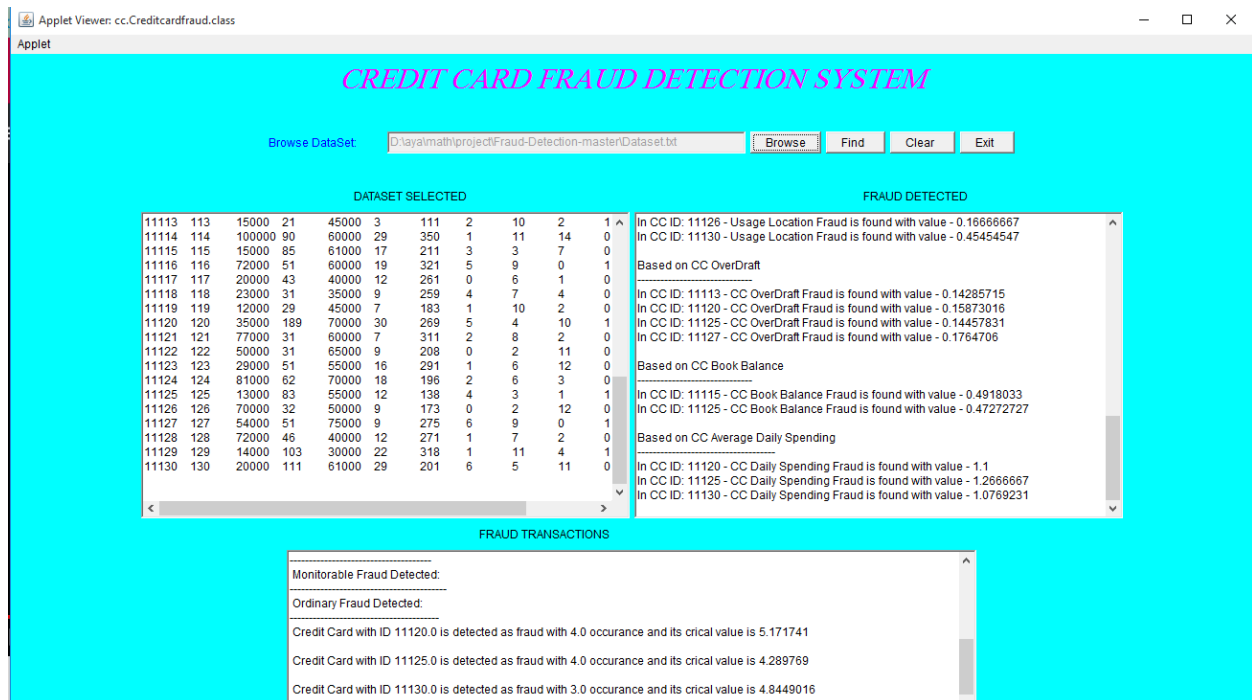


Figure 4: Interface of the algorithm we used

Conclusion:

In this paper, we have known what Genetic Algorithm is a method that can be used to detect credit card fraud and what is the main problem in the credit card fraud detection and also saw the result based on the Genetic Algorithm. Genetic Algorithm is appropriate in such kind of application areas. This algorithm can be used to predict the credit card fraud in a very short time after doing the credit card transactions. Finally, this help the banks and customers not to loss and reduce risks.

Future Work:

There are many ways one can continue this journey, we can use a more updated and modified version of the genetic algorithm to obtain better

results, or we can even modify it ourselves and this could be achieved, we won't say it's easy, but it's most definitely possible.

References:

- [1] Aravindh, M. S., & Bai, M. W. J. S. TECHNIQUES TO PREVENT CREDIT CARD FROM VARIOUS FRAUDLENT BY GENETIC ALGORITHM AND HIDDEN MARKOV MODEL.
- [2] Trivedi, I, Monika& Mridushi,M. (2016, Jan).Credit Card Fraud Detection, International Journal of Advanced Research in Computer and Communication Engineering (Vol. 5)
- [3] Al-Khatib, A. (2012). Electronic payment fraud detection techniques. World of Computer Science and Information

Technology Journal (WCSIT), 2(4), 137-141.

[4] CliftonPhua, DammindAlahakoon, and Vincent Lee; Minority Report in Fraud Detection Classification of

Skewed Data; Sigkdd Explorations, Vol. 6, Issue 1, p.p. 50-51.

[5] Andreas L. Prodromidis and Salvatore J. Stolfo; Agent-Based Distributed Learning Applied to Fraud Detection.

Department of Computer Science- Columbia University; 2000.

[6] BuzzingFacts: Top 5 Most Common Credit Card Fraud Techniques And How To Protect Yourself From This.

[7] Holland J., Adaptation in Natural and Artificial Systems. Ann Harbor, MI: University of Michigan Press. 1975.

[8] Sethi, N., & Gera, A. (2014). A revived survey of various credit card fraud detection techniques. International, Journal of Computer Science and Mobile Computing, 3(4), .780-791

[9] K.RamaKalyani, D.UmaDevi, „Fraud Detection of Credit Card Payment System by Genetic International Journal of ,”Algorithm Scientific & Engineering Research .Volume 3, Issue 7, July-2012

[10] Srinivas M., Patnaik L., Genetic algorithms – a survey. IEEE Computer Society, 1994.