



Cairo University

MathlΣtΣs

CREDIT CARD FRAUD DETECTION USING GENETIC ALGORITHM

TEAM: MATHLETES

Members:

1. Areej Raafat Zahir
2. Asmaa Sayed Saad
3. Aya Sameer Ahmed
4. Aya Adel Hassan
5. Bassant Mohamad Ali
6. Shaza Mohamad Abdulrahman

AGENDA:

- Introduction
- Techniques used for credit card frauds
- Problems in detecting
- Detection methods
- Evolutionary algorithms (Genetic Algorithm)
- CC Fraud Detection using genetic algorithm
- Results
- Conclusion

INTRODUCTION

- What is **fraud**?
 - Criminal activity
 - Compromising someone's rights for personal gain
- Credit Card is a small card issued to users as a system of payment
- Credit card security relies on the physical security of the card as well as the privacy of the credit card number



INTRODUCTION

- Online shopping is one of the largest and fast going trends.
- Increased usage of the internet for such shopping has resulted in a considerable increase in credit card transactions throughout the world.
- Credit Card fraud is a term for theft and fraud committed using a credit card as a fraudulent source of funds

TECHNIQUES USED FOR CREDIT CARD FRAUDS

1. Lost Or Stolen Card
2. Rigging Card Fraud
3. Free Wi-Fi Trap
4. Fake Call or Email



PROBLEMS IN DETECTING

- 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

PROBLEMS IN DETECTING

To achieve good results:

- It should be able to handle skewed distribution
- There should be a way to handle noise to refer to the error present in the data
- The system should be able to end or minimize data overlapping problems
- The system should have good metrics to estimate the classifier system

DETECTION METHODS

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

EVOLUTIONARY ALGORITHMS (GENETIC ALGORITHM)

Training a Machine Learning Model

Data

x_1	x_2	x_3	x_4	x_5	x_6	y
4	-2	7	5	11	1	44.1

$$y = w_1 x_1 + w_2 x_2 + w_3 x_3 + w_4 x_4 + w_5 x_5 + w_6 x_6$$

- Goal is to find the set of parameters ($w_1:w_6$) that maps the following input to its output.

$$y' = 4w_1 - 2w_2 + 7w_3 + 5w_4 + 11w_5 + w_6$$

EVOLUTIONARY ALGORITHMS (GENETIC ALGORITHM)

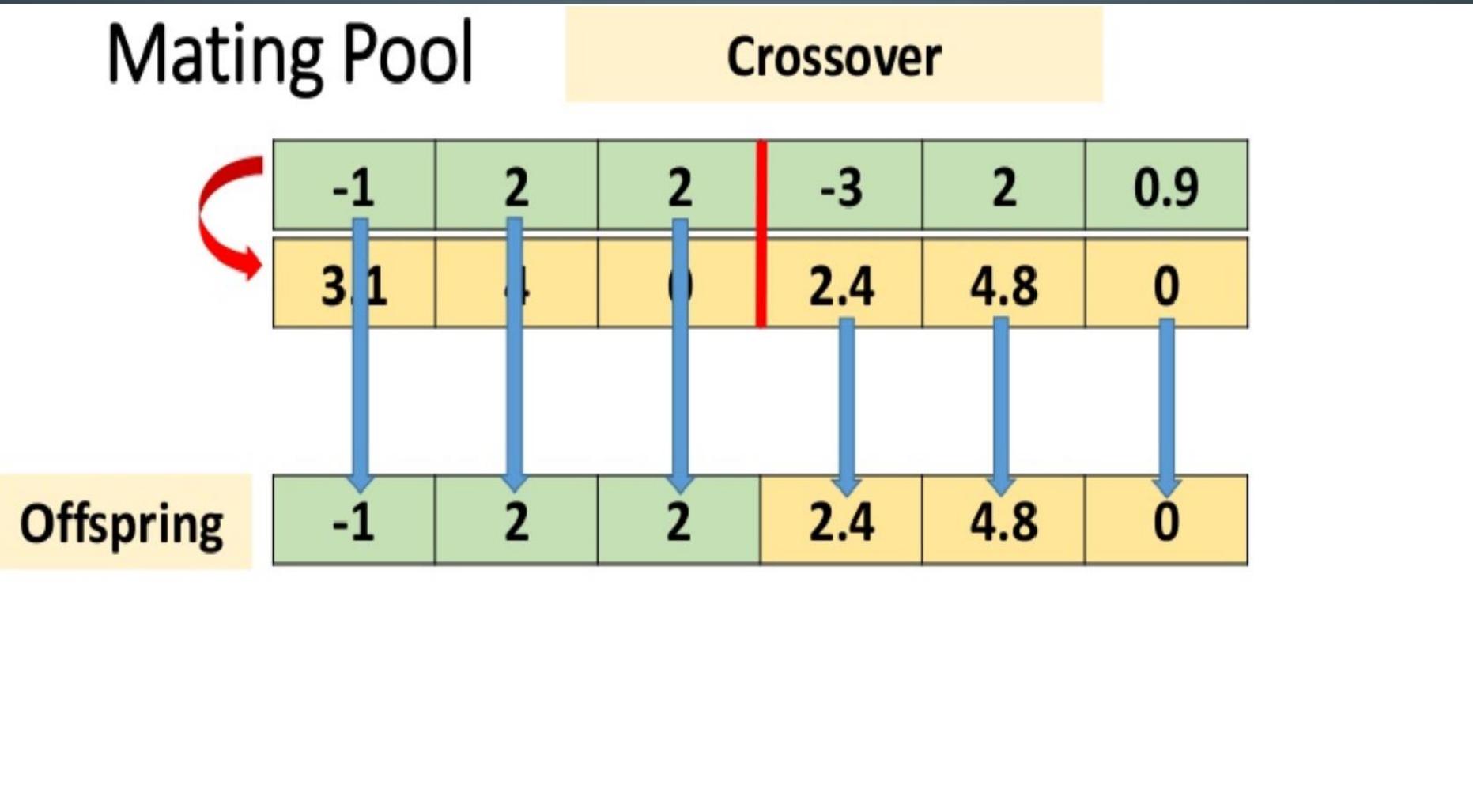
Mating Pool

y'	$F(C)$
110.3	0.015
100.1	0.018
13.9	0.033
127.9	0.012
69.2	0.04
3	0.024

2.4	0.7	8	-2	5	1.1
-0.4	2.7	5	-1	7	0.1
-1	2	2	-3	2	0.9
4	7	12	6.1	1.4	-4
3.1	4	0	2.4	4.8	0
-2	3	-7	6	3	3

Add best 3 individuals to the mating pool for producing the next generation of solutions.

EVOLUTIONARY ALGORITHMS (GENETIC ALGORITHM)

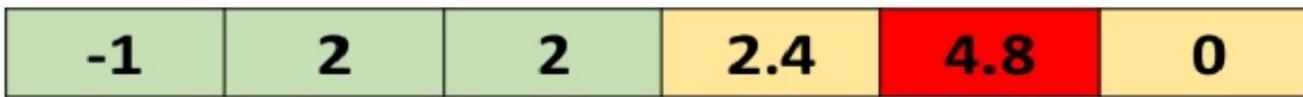


EVOLUTIONARY ALGORITHMS (GENETIC ALGORITHM)

Mating Pool

Mutation

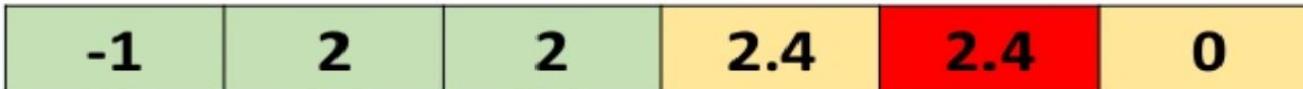
Offspring



$$= 4.8/2$$

$$= 2.4$$

Mutant



CC FRAUD DETECTION USING GENETIC ALGORITHM

- GA and its usage in **detecting** one of the most common problems, which is credit card fraud, and **minimizing** the number of the false alerts.
- The main **concept** of this analysis is obtained through the **behavior** of the credit card holder and the transactions happening.

CC FRAUD DETECTION USING GENETIC ALGORITHM

- **Generic genetic algorithm:**

```
BasicGeneticAlgorithm ( )  
{  
    Initialize population  
    Evaluate population;  
    While (termination criteria not met) {  
        Select solutions for next generation;  
        Perform crossover and mutation;  
        Evaluate population;  
    } }
```

CC FRAUD DETECTION USING GENETIC ALGORITHM

- Preparing for the first step in the algorithm:

First we need to calculate a few variables to help generate the initial **data set** we will use in the code.

Those Main variables are:

- 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.

CC FRAUD DETECTION USING GENETIC ALGORITHM

The previous variables and some extra ones help generate the data set as follows:

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

RESULTS

- 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 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.

RESULTS

Quick analysis for step 3:

- 1-The initial population or generation is selected randomly from the data set
- 2- the fitness value is calculated in each population, and those fitness values are sorted out
- 3- The Crossover is calculated using single point probability.
- 4- Mutation mutates the new generation using uniform probability measure
- 5- Then the best solution is passed to the further generation
- 6- repeating the previous 5 steps on the new population, until maximum number of generation is reached.

RESULTS

Applet Viewer: cc.Creditcardfraud.class

Applet

CREDIT CARD FRAUD DETECTION SYSTEM

Browse DataSet E:\second year\linear algebra\Fraud-Detection-master\Fraud-Detect Browse Find Clear Exit

DATASET SELECTED

11113	113	15000	21	45000	3	111	2	10	2	1
11114	114	100000	90	60000	29	350	1	11	14	0
11115	115	15000	85	61000	17	211	3	3	7	0
11116	116	72000	51	60000	19	321	5	9	0	1
11117	117	20000	43	40000	12	261	0	6	1	0
11118	118	23000	31	35000	9	259	4	7	4	0
11119	119	12000	29	45000	7	183	1	10	2	0
11120	120	35000	189	70000	30	269	5	4	10	1
11121	121	77000	31	60000	7	311	2	8	2	0
11122	122	50000	31	65000	9	208	0	2	11	0
11123	123	29000	51	55000	16	291	1	6	12	0
11124	124	81000	62	70000	18	196	2	6	3	0
11125	125	13000	83	55000	12	138	4	3	1	1
11126	126	70000	32	50000	9	173	0	2	12	0
11127	127	54000	51	75000	9	275	6	9	0	1
11128	128	72000	46	40000	12	271	1	7	2	0
11129	129	14000	103	30000	22	318	1	11	4	1
11130	130	20000	111	61000	29	201	6	5	11	0

FRAUD TRANSACTIONS

Credit Card with ID 11125.0 is detected as fraud with 4.0 occurrence and its critical value is 4.289769
Credit Card with ID 11130.0 is detected as fraud with 3.0 occurrence and its critical value is 4.8449016
Monitorable Fraud Detected:
Credit Card with ID 11115.0 is detected as fraud with 3.0 occurrence and its critical value is 2.1289055
Ordinary Fraud Detected:

FRAUD DETECTED

In CC ID: 11126 - CC OverDraft Fraud is found with value - 0.16666667
In CC ID: 11130 - CC OverDraft 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

RESULTS

DATASET SELECTED

CardID	Auth	Cur.BB	CU	Avg.BB	OD	CCAge	CUT	Loc	LocT	Co	^
11111	111	20000	13	60000	4	125	0	3	0	0	
11112	112	25000	40	55000	20	264	6	4	2	0	
11113	113	15000	21	45000	3	111	2	10	2	1	
11114	114	100000	90	60000	29	350	1	11	14	0	
11115	115	15000	85	61000	17	211	3	3	7	0	
11116	116	72000	51	60000	19	321	5	9	0	1	
11117	117	20000	43	40000	12	261	0	6	1	0	
11118	118	23000	31	35000	9	259	4	7	4	0	
11119	119	12000	29	45000	7	183	1	10	2	0	
11120	120	35000	189	70000	30	269	5	4	10	1	
11121	121	77000	31	60000	7	311	2	8	2	0	
11122	122	50000	31	65000	9	208	0	2	11	0	
11123	123	29000	51	55000	16	291	1	6	12	0	
11124	124	81000	62	70000	18	196	2	6	3	0	
11125	125	13000	83	55000	12	138	4	3	1	1	
11126	126	70000	32	50000	9	173	0	2	12	0	
11127	127	54000	51	75000	9	275	6	9	0	1	▼

RESULTS

FRAUD DETECTED

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

RESULTS

FRAUD TRANSACTIONS

Monitorable Fraud Detected:

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

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

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

Ordinary Fraud Detected:

CONCLUSION

Genetic algorithm can be used to predict the credit card fraud in a very short time after doing the credit card transactions.



ACKNOWLEDGEMENT

Dr. Samah

For being our mentor and advisor, and for giving us the wonderful opportunity to do this research, so thank you so much.



Any Questions?