International Journal of Innovative Research and Advanced Studies (IJIRAS)
Volume 7 Issue 10, October 2020

ISSN: 2394-4404

Performance Evaluation Of Selection Methods Of Genetic Algorithm In Outlier Detection

Efunboade, Ayodeji Oyeyinka

Department of Computer Science
Osun State College of Technology, Esa-oke, Osun State, Nigeria

Lawal, Nurudeen Tunde A.

Department of Computer Science
Osun State College of Technology, Esa-oke, Osun State, Nigeria

Adewale, Joseph Adekunle

Department of Computer Science
Osun State College of Technology, Esa-oke, Osun State, Nigeria

Sarumi, Olusegun Ajibola

Department of Computer Science
Osun State College of Technology, Esa-oke, Osun State, Nigeria

Abstract: ***This paper evaluates the performance of some feature selection methods of Genetic Algorithm in outlier detection on fingerprint images. Roulette wheel, Rank and Tournament selection approaches were considered and Histogram Equalization were used for image enhancement of selected features. K-nearest neighbour algorithm was employed for classification to detect outliers. The implementation of the experiment was carried out using Mat lab. Performance of the selection methods were evaluated based on metrics of accuracy and computation time.***

Keywords: Roulette wheel, Rank, Tournament, Genetic algorithm, outlier detection

I. INTRODUCTION

Genetic Algorithm (GA) is a search and optimization technique that belongs to the evolutionary algorithms class and its development was inspired through the process of natural genetic evolution. [1] Genetic Algorithms are optimization search algorithms that maximize or minimizes their given functions. GA is heuristic in procedures, so they are not guaranteed to find the optimal solutions to complex problems. However, literature review has shown that GA are able to find very good solutions for a wide range of problems such as outlier detection. Genetic algorithm was used for feature selection and classification in outlier detection on data sets. The result obtained claimed that the algorithm was effective for outlier detection in a wide research domain. [3]

Researchers have done lot of work in selection phase of GA but significant work is yet to be done with regards to outlier detection in fingerprint images. In this paper, the focus was on selection phase of GA considering roulette wheel, rank and tournament methods with the goal of comparing and determine the technique that performed best. In selection stage, individuals were chosen in the population that created offspring for the next generation and how many offspring each would create. The purpose of selection was to emphasize fittest individuals in the population in hopes that their offspring would in turn have even higher fitness.

II. GENETIC ALGORITHM FEATURE SELECTION USING ROULETTE WHEEL

In Roulette Wheel selection technique, all the chromosomes in the population are placed on the roulette wheel according to their fitness value. Each individual is assigned a segment of roulette wheel whose size is proportional to the value of the fitness of the individual [3]. This is a selection style where the selection probability is proportional to absolute fitness. In proportional roulette wheel, individuals are selected with a probability that is directly proportional to their fitness values i.e. an individual's selection

corresponds to a portion of a roulette wheel. The probabilities of selecting a parent in spinning a roulette wheel with the size of the segment for each parent being proportional to its fitness. Partially mapped crossover was selected to limits the possibility of invalid chromosomes.

The Genetic Algorithm for roulette wheel feature selection is as follows:

i Population size K ; Objective function F ( )
ii Crossover probability P_{co}
iii Mutation probability P_{mu}
iv Fitness threshold T
v Begin
vi do
vii Determine the fitness of each chromosome F(i)
i = 1,2,…….,K
viii Select the chromosome using roulette wheel
ix do
x Select two chromosomes with highest score
xi If (Rand[0,1] < P_{co}) then
xii Crossover the pair of chromosomes
xiv else
xv Change each chromosome with P_{mu}
xvi Remove the parent chromosomes
xvii until N offsprings have been created
xviii until any chromosome' s score fit F( ) exceeds T
xix return highest fitness chromosomes (best features)
xx end

III. GENETIC ALGORITHM FEATURE SELECTION USING RANK

Rank selection: In rank selection technique, the population and every chromosome receives fitness from the ranking. The worst has fitness 1 and the best has fitness N. It results in slow convergence but prevents too quick convergence. It also keeps up selection pressure when the fitness variance is low. It preserves diversity and hence leads to a successful search. In Linear Rank selection, individuals are assigned subjective fitness based on the rank within the population. The individuals in the population are sorted from best to worst according to their fitness values. Each individual in the population is assigned a numerical rank based on fitness, and selection is based on this ranking rather than differences in fitness [4]

The Genetic Algorithm for rank feature selection is as follows:

i Population size K ; Objective function F ( )
ii Crossover probability P_{co}
iii Mutation probability P_{mu}
iv Fitness threshold T
v Begin
vi do
vii Determine the fitness of each chromosome F(i)
i = 1,2,……,K
viii Select the chromosome using rank
ix do
x Select two chromosomes with highest score
xi If (Rand[0,1] < P_{co}) then
xii Crossover the pair of chromosomes
xiv else
xv Change each chromosome with P_{mu}
xvi Remove the parent chromosomes
xvii until N offsprings have been created
xviii until any chromosome' s score fit F( ) exceeds T
xix return highest fitness chromosomes (best features)
xx end

IV. GENETIC ALGORITHM FEATURE SELECTION USING TOURNAMENT

Tournament selection: GA uses this selection mechanism to select individuals from the population to insert into a mating pool. Individuals from the mating pool are used to generate new offspring, with the resulting offspring forming the basis of the next generation. A tournament selection mechanism in GA is simply a process that favours the selection of better individuals in the population for the mating pool. The selection pressure is the degree to which the better individuals are favoured: the higher the selection pressure, the better individuals are favoured [5]

The Genetic Algorithm for rank feature selection is as follows:

i Population size K ; Objective function F ( )
ii Crossover probability P_{co}
iii Mutation probability P_{mu}
iv Fitness threshold T
v Begin
vi do
vii Determine the fitness of each chromosome F(i)
i = 1,2,……,K
viii Select the chromosome using Tournament
ix do
x Select two chromosomes with highest score
xi If (Rand[0,1] < P_{co}) then
xii Crossover the pair of chromosomes
xiv else
xv Change each chromosome with P_{mu}
xvi Remove the parent chromosomes
xvii until N offsprings have been created
xviii until any chromosome' s score fit F( ) exceeds T
xix return highest fitness chromosomes (best features)
xx end

V. EXPERIMENT DESIGN, RESULT AND DISCUSSION

This section focuses on the experimental design, results that we collected and their comparative analysis. The implementations in Mat Lab 2016 environment. The data collected for this work includes three hundred (300) fingerprint images acquired from selected students of Osun State College of Technology, Esa-Oke, Osun State, Nigeria. The images were in jpeg format with resolution of 100 X 100. The images were enhanced using Histogram equalization. Two hundred (200) fingerprint images were trained and one hundred (100) fingerprint images were used for testing. For our experiments, the selected features using roulette wheel,

rank and tournament were loaded into *K*-NN, after which k-fold cross validation method was employed as classifier for training and testing to determine outliers. The value of *k*, which was the nearest data point was selected using Euclidean distance function stated below.

Using $x_i \ldots x_k$ to represent the *k* instances from training samples that are nearest to x_q in Euclidean distance

$$d(x, y) = \sqrt{\sum_i (x_i - y_i)^2}$$

Where *x* and *y* are two instances with i^{th} input attributes x_i and y_i, .

The supervised machined learning experiment was carried out using several user-set threshold values to determine the best result for outliers. From the experiment, the best result was achieved for accuracy in outlier detection at threshold 0.56 as shown in Table 1. The graph of selection methods in terms of accuracy is also shown in Figure 1.

GA SELECTION TECHNIQUES	Threshold	ACCURACY (%)	COMPUTATION TIME (SEC)
Roulette Wheel	0.56	86.43	121
Rank	0.56	84.72	116
Tournament	0.56	79.20	132

Table 1: Result for GA selection methods for outlier detection on fingerprint images

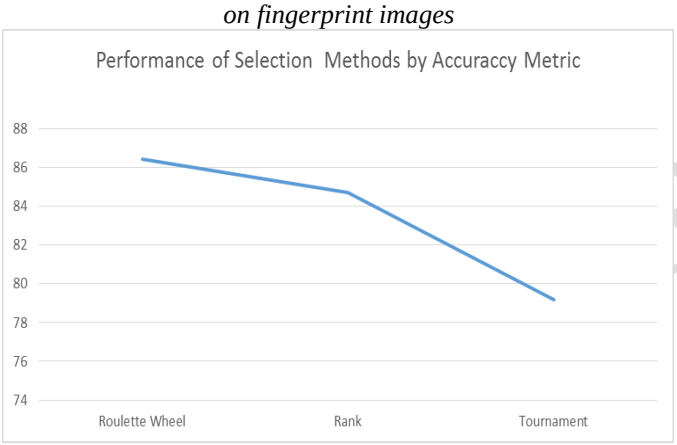

Figure 1: Graph of selection methods in terms of accuracy

In conclusion, according to the experiment we conducted, roulette wheel gave the best result in term of accuracy. It was then followed by rank and tournament selection techniques respectively. However, rank selection method gave the best result in term of computation time, followed by roulette wheel and tournament selection respectively. It was observed that the three selection methods performed very well for the purpose of outlier detection in fingerprint images with roulette wheel performing best in terms of accuracy.

VI. SUGGESTION FOR FURTHER STUDY

The data collected for this work was acquired locally from randomly selected students Osun State College of Technology, Esa-Oke, Nigeria. Further research can be geared towards the use of external database of fingerprint images available on the Internet. These fingerprint images can be downloaded, selection methods of Genetic Algorithm applied and the result obtained can be compared with the one obtained in this study.

ACKNOWLEDGEMENT

The authors wish to thank the staff members and students of Computer Science of Osun State College of Technology, Esa-Oke, Nigeria, for giving us the necessary support during the acquisition of the fingerprint images used in this study and during the testing of the system.

REFERENCES

[1] Holland J.H. (1975) "Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence." University of Michigan Press, Oxford.

[2] Maria Afzal and S. M. Arif Ashraf. (2016). "Genetic Algorithm for Outliers Detection" International Journal of Computer Science and Information Technology. 7(2)

[3] Noraini Mohd Razali, John Geraghty, (2011)" Genetic Algorithm Performance with Different Selection Strategies in Solving TSP", Proceedings of the World Congress on Engineering 2011 Vol II.

[4] Saneh Lata Yadav and Asha Sohal (2017) " Comparative Study of Different Selection Techniques in Genetic Algorithm" International Journal of Engineering, Science and Mathematics Vol. 6 Issue 3, July 2017, ISSN: 2320-0294

[5] Sivanandam, S.N. and .Deepa, S.N. (2008) "An Introduction to Genetic Algorithms" Springer-Verlag Berling Heidelberg

[6] Yadav and Sohal (2017). "Comparative Study of Different Selection Techniques in Genetic Algorithm" International Journal of Engineering, Science and Mathematics Vol. 6 Issue 3, ISSN: 2320-0294