

A New Selection Operator - CSM in Genetic Algorithms for Solving the TSP

Wael Raef Alkhayri

Dept. of Computer Science
Fahed Al duwairi High School
Aljabereyah, Kuwait

Suhail Sami Owais

Dept. of Computer Science
Applied Science Private University
Amman, Jordan

Mohammad Shkoukani

Dept. of Computer Science
Applied Science Private University
Amman, Jordan

Abstract—Genetic Algorithms (GAs) is a type of local search that mimics biological evolution by taking a population of string, which encodes possible solutions and combines them based on fitness values to produce individuals that are fitter than others. One of the most important operators in Genetic Algorithm is the selection operator. A new selection operator has been proposed in this paper, which is called Clustering Selection Method (CSM). The proposed method was implemented and tested on the traveling salesman problem. The proposed CSM was tested and compared with other selection methods, such as random selection, roulette wheel selection and tournament selection methods. The results showed that the CSM has the best results since it reached the optimal path with only 8840 iterations and with minimum distance which was 79.7234 when the system has been applied for solving Traveling Salesman Problem (TSP) of 100 cities.

Keywords—Genetic Algorithm; Traveling Salesman Problem; Genetic Algorithm Operators; Clustering; Selection Operator

I. INTRODUCTION

Genetic Algorithm (GA) is one of the Evolutionary Algorithms (EAs), which is an optimization technique based on natural evolution [2, 4, 6]. GA has evolved into a powerful method for solving hard combinatorial optimization problems that uses a stochastic search technique. It is considered as one of the most important technologies in the search for the perfect choice of a set of solutions available for a particular design. GA proposed by John Holland and it relies on Darwin's principle of eclecticism [1, 6].

Traveling Salesman Problem (TSP) is one of the most intensively studied problems in computational mathematics, and it is a non-deterministic polynomial time problem [2]. TSP requires the most efficient path between the set of cities (i.e., least total distance). There is no general method of such solution known for TSP. TSP focused on an optimize path solution, where the salesman can take through each of the n cities and visits each city exactly only once and finally returning back to the home city [2, 7].

In GA process the selection method is the process of choosing parent(s) from the population for recombination, and mainly the selection step is preceded by the fitness value assignment which is based on the objective value. Where the performance of GA using differing selection methods is usually evaluated in terms of convergence rate and the number of generations to reach the optimal solution of a problem [5].

There are many types of selection methods used in GAs, in this paper the authors will propose a new selection operator.

This research paper proposes a new selection method in GA. The proposed method has been tested and compared with other selection methods which indicates that the new method increases the GA performance by reducing the number of generations.

The paper includes the following sections: introduction is the first part, GA discussed in the second section, TSP has been introduced in the third section. The fourth section explains the research problem statement, then the selection methods were described in the fifth section. The proposed method was discussed in section six, followed by the experimental test in section seven. Finally section eight clarified the conclusions.

II. GENETIC ALGORITHM

Genetic algorithm generally composed of two processes. The first process is the selection to select parents for reproduction to the next generation [4, 6]. The second process is the reproduction by manipulating the selected parents by crossover and mutation operators with respect to the fitness value for the new generation children [3].

Fig. 1 demonstrates the processes of Genetic algorithm [1], where GA consists of the following steps: [6, 7]

- 1) **[Start]** Generate random population of n chromosomes and encoding them.
- 2) **[Fitness]** Evaluate the fitness $f(x)$ of each chromosome x in the population.
- 3) **[New population]** Create a new population by repeating the following steps until the new population is complete:
 - a) **[Selection]** Select parents' chromosomes from a population according to their fitness.
 - b) **[Crossover]** is performed with a probability that crossover the selected parents to produce a new offspring
 - c) **[Mutation]** is performed with a probability that mutates new offspring.
 - d) **[Accepting]** place the new offspring in a new population.
 - e) **[Replace]** use newly generated population for a further run of algorithm

f) [Test] if the termination conditions are satisfied, then stop the algorithm, and return the best solution in current population

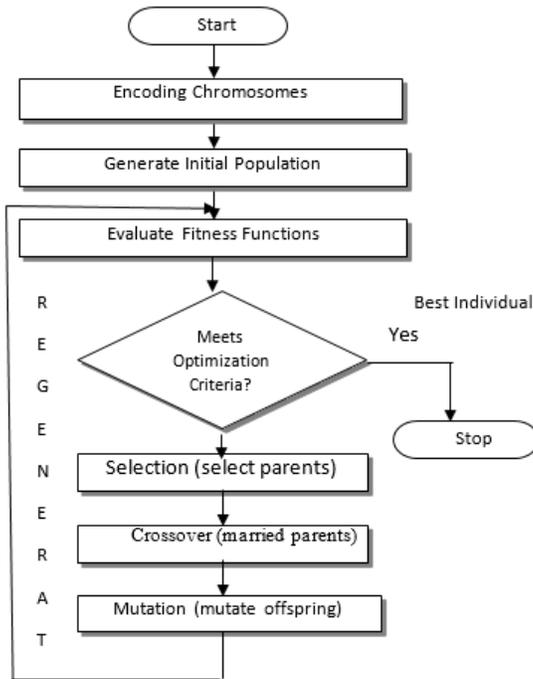


Fig. 1. Genetic Algorithm Processes [1]

III. TRAVELING SALESMAN PROBLEM

The Traveling Salesman Problem consists of a number of cities, where each pair of cities has an equivalent distance. The goal for the salesman is to visit all the cities. So, the total distance travelled will be minimized [7].

TSP is conceptually simple even though it is difficult to get an optimal solution which it is the shortest path between n cities. The major difficulty of this problem is the large number of possible paths; which is estimated by $\frac{(n-1)!}{2}$ for n cities. The problem will be considered when the number of cities increased, the numbers of variations of the valid paths are also increased. Several researchers try to find the optimal path for the TSP using GA [2, 7].

IV. PROBLEM STATEMENT

Parent selection for next generation is the most important operator in GA because the solution of a problem depends on that operator for reproduction; its importance is to choose the successive individuals in the population that will create offspring for the next generation.

There are many selection methods such as roulette, rank, etc. Some of these methods select the parents randomly while others select the parents by sorting them, then choose the highest fitness value parents the thing which is not guaranteed to get the optimal solution with a minimum number of generations. Therefore, the number of generations that evolve depends on whether an acceptable solution is reached as optimal or as the number of iterations is exceeded. Thus, for the best performance expected is how to get the optimal

solution using the minimum number of generations to ensure more efficiency and high speed.

V. SELECTION METHODS

The selection mechanism determines which individuals will be chosen for mating (reproduction) to reproduce new generation [3]. Where the main purpose of selection method is to select parents that can find the better parents in the population to reach the optimal solution.

The most important decision that should be taken into consideration in the selection method in GA is how to decide the most appropriate selection method that should increase the speed of evolution for reaching the optimal path with minimum distance in TSP [8].

There are many selection methods such as Roulette wheel selection; Tournament selection; Elitism Selection; Rank Selection; Stochastic universal sampling; Local selection; Truncation selection and others [8].

VI. PROPOSED CLUSTER SELECTION METHOD

The proposed method Clustering Selection Method (CSM) based on selecting parents from clusters with varying number of individual per each cluster, the cluster will be created by collecting all the closest individuals with respect to their fitness values together in the same cluster for all individuals in the population as shown in Fig. 2. Then parents will be chosen from the cluster that has more elements with respect to the highest or lowest fitness values depending on fitness function.

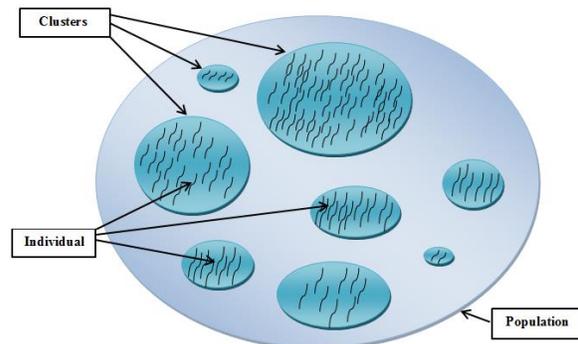


Fig. 2. Clustering individuals in population

The CSM starts by sorting the individuals in the population in ascending or descending order with respect to their fitness values. After that evaluate the ALFA value (α) by using the equation (1), using the standard deviation (STD) because it is the measure of spread out numbers are or how much the individuals in a population (pop) are scattered around the mean. The value of α will be used to determine the clusters, (that means to evaluate the amount of dispersion between individuals in the population). Where, a low standard deviation means that the data are tightly clustered; a high standard deviation means that they are widely scattered, and by using the square root for STD will give more chance to determine the closest individuals to create the clusters.

$$ALFA = \text{SQRT}(\text{STD}(\text{pop})) \quad (1)$$

After that, to divide the population for clusters with respect to the amount of dispersion between the individuals in the population. The population is divided based on α value that will creates many clusters, where each cluster contains different number of individuals.

Then, for each cluster, evaluate the Cluster_Value (CV) by calculating the STD and the differences (D) using equation (2). By computing the STD for all individuals in a cluster_sub_population (CS_pop) and by computing D between the fitnesses values for all individuals in the CS_pop. To avoid the division by zero (when the fitness's are equals in a CS_pop or there is no differences) by adding 1 to D.

$$CV = \frac{\sqrt{(STD(CS_pop)) * Count(CS_pop)}}{(D * CS_pop) + 1} \quad (2)$$

Then by comparing the CV values for all clusters, the "highest CV" means that this cluster contains the closest individuals with a high number of individuals which is our goal. Then we will consider this cluster with high CV to select the parents from it for the next generations. For maximization problem consider selecting the highest fitness values from this cluster as parents, and for minimization problem consider selecting the lowest fitness value as parents. As shown in Fig. 3.

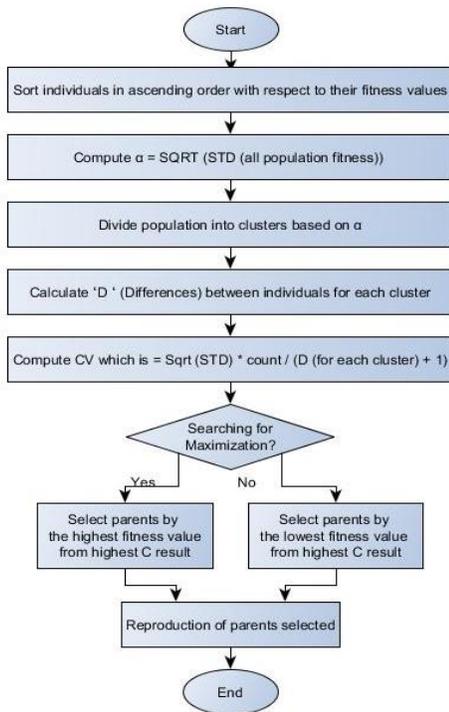


Fig. 3. The Proposed Cluster Selection Method (CSM)

VII. EXPERIMENTAL TEST

The proposed CSM was applied on GA and tested by using MATLAB for solving the TSP on a different number of cities

10, 20, 30, 40, 50, 60, 70, 80, 90 and 100 cities. The system was executed for 10,000 generations for each case or till finding the optimal path, with different selection methods listed in table 1.

TABLE I. SELECTION METHOD AND DESCRIPTION

No.	Method	Description
1	Random selection (Ga)	Random selection
2	Roulette Wheel	Roulette wheel selection
3	Tournament	Tournament selection
4	CSM	Clustering Selection Method using ALFA

TABLE 2. Demonstrates the final results from execution the system implementation with different selection methods and with the proposed CSM, where:

- *Dis* : the average distances between n cities.
- *O.D* : the order of distances in an ascending order from all paths, if there are any equal values they will have the same order.
- *Itr* : the average number of iterations to reach an optimal path.
- *O.I* : the order of iteration, if there are any equal values they will have the same order.
- *O.T* : the multiplication of O.I by O.D.

TABLE 3. summarized the final analysis for all cities from 10 – 100 for the TSP. It is obvious that the proposed CSM was the best method based on both values *Sum Dis* and *Sum O.T* because they have the minimum values than the other selection methods. Whereas the *Sum Itr* was the best for the Roulette Wheel selection method.

Other test for the system were done for solving the TSP for 100 cities using three different selection methods in addition to the proposed one.

Fig. 4 shows the system result for the random selection method (Ga), where the optimal path distance was 80.5908 with 9159 iterations.

The distance for the optimal path was 79.9358 and iterations needed was 8864 for the Roulette Wheel selection method as shown in Fig. 5.

Tournament selection method reached the optimal path with 9458 iterations and with distance equals to 82.0521 as presented in Fig. 6.

The proposed method CSM has the best results since it reached the optimal path with only 8840 iterations and with minimum distance which was 79.7234 as shown in Fig. 7.

TABLE II. SYSTEM RESULTS FOR SELECTION METHODS AND THE PROPOSED ONE

Method	10-Cities					20-Cities					30-Cities					40-Cities					50-Cities				
	Dis	O.D	Itr	O.I	O.T	Dis	O.D	Itr	O.I	O.T	Dis	O.D	Itr	O.I	O.T	Dis	O.D	Itr	O.I	O.T	Dis	O.D	Itr	O.I	O.T
Ga	22.1	1	80.0	2	10	32.8	1	977.9	3	3	41.1	1	2157.4	9	9	48.1	1	3478.4	8	8	54.2	2	4456.5	3	6
Roulette Wheel	22.1	1	98.8	4	8	32.8	1	1063.3	6	6	41.1	1	2042.9	4	4	48.1	1	3389.8	1	1	54.1	1	4436.8	2	2
Tournament	22.1	1	78.5	1	4	32.8	1	1083	8	8	41.1	1	1952.7	1	1	48.2	2	3452.9	6	12	54.2	2	4536.5	7	14
CSM	22.1	1	82.0	3	3	32.8	1	1081.1	7	7	41.2	2	1999.1	2	4	48.1	1	3470.8	7	7	54.1	1	4511.7	5	5

Method	60-Cities					70-Cities					80-Cities					90-Cities					100-Cities				
	Dis	O.D	Itr	O.I	O.T	Dis	O.D	Itr	O.I	O.T	Dis	O.D	Itr	O.I	O.T	Dis	O.D	Itr	O.I	O.T	Dis	O.D	Itr	O.I	O.T
Ga	59.8	2	5559.9	8	16	65.1	2	6526.7	4	8	69.9	1	7622	2	2	74.4	4	8534.7	2	8	79.4	2	9160.0	5	10
Roulette Wheel	59.8	2	5564.7	9	18	65.1	2	6559.7	7	14	70	2	7657.6	4	8	74.3	3	8587.5	5	15	79.5	3	9137.3	3	9
Tournament	59.7	1	5330.9	1	1	65.1	2	6509.5	2	4	70	2	7648.2	3	6	74.3	3	8596.3	6	18	79.4	2	9136.2	2	4
CSM	59.7	1	5474.9	6	6	65	1	6517.6	3	3	70	2	7713.5	7	14	74.3	3	8558.7	4	12	79.3	1	9160.0	5	5

TABLE III. TOTAL ANALYSIS FOR ALL CITIES FROM 10 - 100

Method	Sum Dis	Sum Itr	Sum O.T
Ga	546.9387	48553.474	80
Roulette Wheel	546.9381	48538.407	85
Tournament	546.9383	48324.69	72
CSM	546.638	48569.338	66
Minimum	546.638	48324.69	66

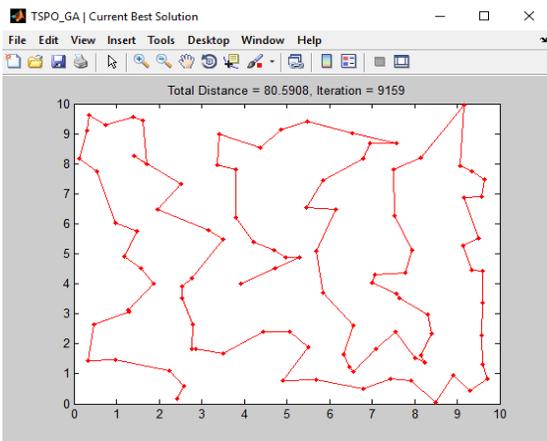


Fig. 4. Random selection method (Ga) results

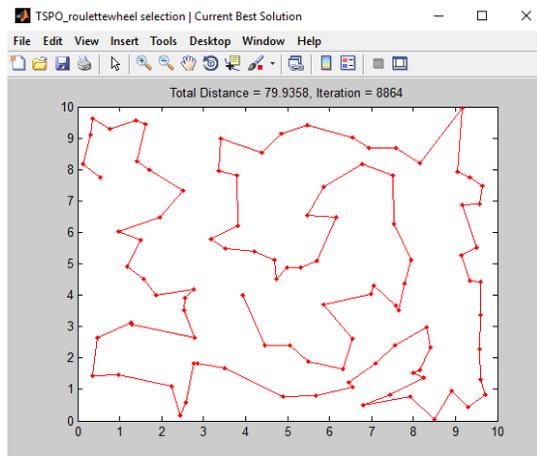


Fig. 5. Roulette Wheel selection method results

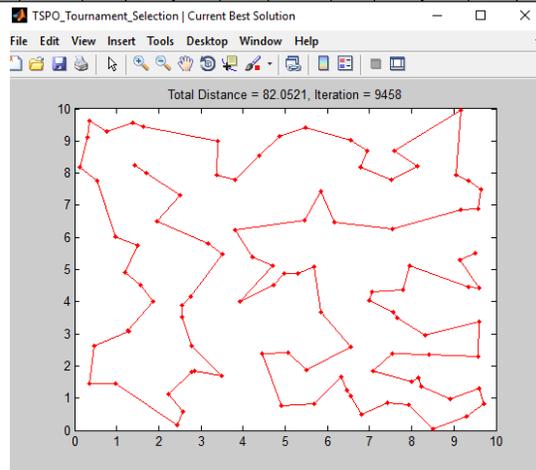


Fig. 6. Tournament selection method results

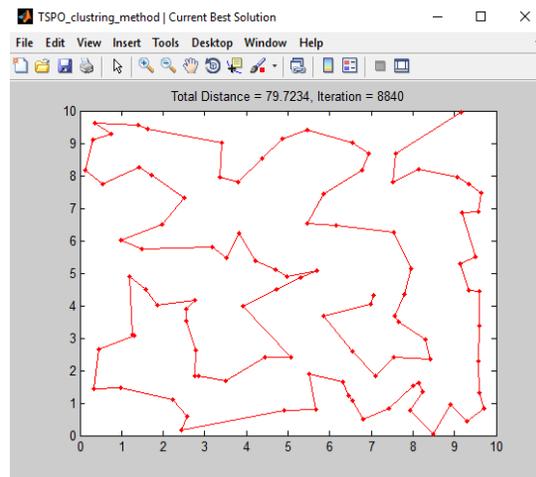


Fig. 7. The proposed CSM selection method results

Table 4 summarizes the results from the figures 4, 5, 6 and 7 which demonstrate the distances and the number of iterations for each method.

TABLE IV. SYSTEM RESULTS FOR 100 CITIES

Method	Distance	Iterations
Ga	80.5908	9159
Roulette Wheel	79.9358	8864

Tournament	82.0521	9458
CSM	79.7234	8840

VIII. CONCLUSION

GA has several operators for solving the optimization problems such as selection, crossover, and mutation operator. Where selection operator considered as one of the most important operators in GA process because it selects parents for recombination and effects on the other operators in the regeneration process.

A new selection method was proposed for selecting parents called Clustering Selection Method (CSM). It divides the population into clusters by collecting all the closest individuals' together with respect to their fitness values in the same cluster, and then selects parents (individuals) from clusters with varying number of individuals per each cluster. Then CSM will choose parents with highest or lowest fitness values based on the chosen cluster depending on Cluster_Value (CV).

TSP has been considered for a different number of cities to test the proposed method, the cities were 10, 20, 30, 40, 50, 70, 80, 90 and 100 cities. The system tested by running it 10,000 generations in each case till finding the best tour path, and the results showed that the proposed method was the best one among the other different types which are Random Selection, Roulette wheel selection, and Tournament selection method.

The result of the proposed CSM was as following: *Sum Dis* is 546.638 and *Sum O.T* is 66; which are the minimum values.

When the system has been tested again for 100 cities also the CSM has showed the best results since it reached the optimal path with minimum distance which was 79.7234 and with only 8840 iterations.

For the future work authors can modify the proposed method by generating new equations for ALFA in order to get better results.

ACKNOWLEDGMENT

The authors are grateful to the Applied Science Private University, Amman, Jordan, for the financial support granted to this research project.

REFERENCES

- [1] S. Owais, V. Snasel, P. Kromer, and A. Abraham, "Survey: Using Genetic Algorithm Approach in Intrusion Detection Systems Techniques", 7th Computer Information Systems and Industrial Management Applications, IEEE, 2008, pp. 300-307, June 2008.
- [2] F. H. Khan, N. Khan, S. Inayatulla, And S. T. Nizami, "Solving ISP Problem by Using Genetic Algorithm", International Journal of Basic & Applied Sciences IJBAS-IJENS vol. 09, no. 10, 2009, pp. 55-60.
- [3] N. M. Razali, and J. Geraghty, "Genetic Algorithm Performance with Different Selection Strategies in Solving TSP", Proceedings of the World Congress on Engineering, 2011, London, U.K., vol. II, July 2011.
- [4] S. Karakatič, and V. Podgorelec, "A survey of genetic algorithms for solving multi depot vehicle routing problem", Applied Soft Computing, vol. 27, pp. 519-532, 2015.
- [5] A. Sharma, and A. Mehta, "Review Paper of Various Selection Methods in Genetic Algorithm", International Journal of Advanced Research in Computer Science and Software Engineering, vol. 3, no 7, July 2013.
- [6] S. Owais, and J. Atoum, "New Method for Mutation Operator in Genetic Algorithm", Journal of Association for the Advancement of Modelling & Simulation Techniques in Enterprises (A.M.S.E), D Computer Science and Statistics", vol. 14 , no. 2, pp. 17-26, 2009.
- [7] S. Khattar, and Dr.P. Gosawmi, "An Efficient Solution of Travelling Salesman Problem Using Genetic Algorithm", International Journal of Advanced Research in Computer Science and Software Engineering, vol. 4, no. 5, 2014, pp. 656-660.
- [8] C. Chudasama, S. M. Shah, and M. Panchal, "Comparison of Parents Selection Methods of Genetic Algorithm for TSP", Proceedings published by International Journal of Computer Applications, International Conference on Computer Communication and Networks CSI-COMNET-2011, pp. 85-87, 2011.