Application of genetic algorithm in solving the travelling Salesman problem

Fatka Kulenović and Azra Hošić
University of Bihać, Faculty of Technical Engineering, Bosnia and Herzegovina

E-mail: kullenovic.fatka@gmail.com

Abstract. The Travelling Salesman Problem is categorized as NP-complete problems called combinatorial optimization problems. For the growing number of cities it is unsolvable with the use of exact methods in a reasonable time. Genetic algorithms are evolutionary techniques used for optimization purposes according to survival of the fittest idea. These methods do not ensure optimal solutions, however they give good approximation usually in time. Studies have shown that the proposed genetic algorithm can find a shorter route in real time, compared with the existing manipulator model of path selection. The genetic algorithm depends on the selection criteria, crosses, and mutation operators described in detail in this paper. Possible settings of the genetic algorithm are listed and described, as well as the influence of mutation and crossing operators on the efficiency of the genetic algorithm. The optimization results are presented graphically in the MATLAB software package for different cases, after which a comparison of the efficiency of the genetic algorithm with respect to the given parameters is performed.

1. Introduction
The travelling salesman problem is one of the best known problems of combinatorial optimization with the goal of minimization. It represents a logistical problem of everyday life whose solution is found in graph theory. The mathematical model of travelling salesman problem comes down to searching for the Hamiltonian cycle of least weight in a weight graph. The weight of a given edge can represent the amount of cost, distance, time or any other measure that characterizes that edge. There is a well-known set of cities to visit and their mutual distances. The optimal travel route must be found so that each city is visited only once, and then returned to the starting city that represents the starting point and with the shortest possible distance. The more cities to visit, the more complicated this problem becomes. Since exact methods do not provide an optimal solution to a given problem, heuristic methods are used that can find solutions very close to the optimum in a reasonable time. The aim of this paper is to define the methodology and apply it to solve the travelling salesman problem. The complexity of the problem itself is reflected in the set of partial tasks that need to be solved in order to obtain the most rational solution. In other words, the aim of the paper is to determine the shortest distances between network nodes, determine locations and determine optimal routes. Solving this problem is a major challenge, partly because of its extreme difficulty and partly because of its connection to interesting practical issues. Confirmation of the popularity and applicability of The Travelling Salesman Problem is found in many papers [1], [2] and [3].
2. Mathematical formulation of the travelling Salesman problem

There are numerous formulations of TSP in the literature whose basic form is most often formulated as follows [4]:

1. The travelling salesman has a given set of cities, each of which he must visit exactly once and return to the starting point. The question is in what order he should tour the cities so that the total length of the journey is minimal.

2. Find the Hamiltonian cycle of minimum weight in the weight graph.

To better understand the theory of the travelling salesman problem, it is necessary to get acquainted with the basic definitions of graphs [4].

Graph definitions are given:

1. A graph is an ordered triple $G = (V, E, \varphi)$, comprising:
   - $V = V(G)$ is a nonempty set whose elements are called vertices,
   - $E = E(G)$ is a set disjoint with $V$ whose elements are called edges and,
   - $\varphi$ is a function that associates a pair $u, v \in V$ with each edge $e$ of $E$.

2. Graph is in short marked as $G = (V, E)$ or just $G$.

3. Expressed in the terminology of graph theory, let $V$ be a set of $n + 1$ vertices (locations) $V = \{0, 1, 2, \ldots, n\}$, the problem is to find a route (cycle) that begins and ends at vertex 0, and passes through all vertices from 1 to $n$ exactly once, so that the total distance is minimal.

4. The Hamiltonian path (cycle) in graph $G$ is the path (cycle) that contains all the vertices of the graph.

5. In other words, the travelling salesman problem is finding the Hamiltonian cycle of minimum length.

6. A graph containing a Hamiltonian cycle is called a Hamiltonian graph.

7. In complete graphs with $n$ vertices the number of Hamiltonian cycles is $(n - 1)!$, while in undirected graphs $\frac{(n - 1)!}{2}$, since the two routes are identical, only inverse, and grows exponentially with the number of vertices.

If the variables $x_{ij}$ are considered, which are defined as follows [5]:

$$x_{ij} = \begin{cases} 1, & \text{if branch } (i,j) \text{ belongs to the optimal route of the travelling salesman} \\ 0, & \text{otherwise} \end{cases}$$

The travelling salesman problem can be formulated as follows [5]:

Minimize

$$\sum_{i=1}^{m} \sum_{j=1}^{n} d_{ij}$$

under restrictions
\[
\sum_{j=1}^{m} x_{ij} = 1 \quad i = 1, 2, ..., m
\]
\[
\sum_{i=1}^{m} x_{ij} = 1 \quad j = 1, 2, ..., m
\]
\[
\sum_{i \in S, j \in S} x_{ij} \leq |S| - 1 \quad S \subset \{1, 2, ..., m\}
\]

\[x_{ij} = 0 \text{ or } 1 \quad \forall i, j\]

where \(S\) is any subset of cities \(1, 2, ..., m\). The first two restrictions take care of the movement of travelling salesman and indicate that only one branch can come out of each node visited by travelling salesman, i.e. that only one branch can come to each node in the route of the travelling salesman. The third constraint prevents the creation of cycles that do not represent the complete route of the travelling salesman [5]. Methods of solving the travelling salesman problem can be divided into four main groups: exact, approximation, heuristic and metaheuristic.

3. Optimization route of the travelling Salesman problem using genetic algorithm

The solution of the travelling salesman problem using a genetic algorithm was done in the MATLAB software package. Using a graphical user interface, two basic tasks of creating an application are integrated: setting up visual components and programming the operation of the application, which allows a quick transition between visual design on the screen and code development in the MATLAB editor. An application has been created that calculates the shortest route and selects a route to tour the cities of Bosnia and Herzegovina, so that each city he visits only once and returns to the source point. For a detailed description of the problem, UML (Unified Modelling Language) modelling was used, the main goal of which is to present a detailed procedure, as well as the functioning of genetic algorithms and how to solve problems by applying them [6]. With the development of an application to solve this problem, it was determined that one of the most important components for the successful operation of this algorithm is the selection of the correct genetic operators and parameters that will determine the behaviour of these operators. If set correctly, the algorithm gives excellent results, but if the parameters are set incorrectly the algorithm will tend to a local optimum, which is closer or farther from the true optimum, depending on the parameters. The initial view of the application is given in Figure 1.

![Initial view after starting the application.](image-url)
In order to create a network of cities to visit, the user must select the type of map to create. The program offers the ability to select two types of maps that are implemented using a drop-down menu: an arbitrary map and a map created by randomly selecting positions. The performance of the algorithm depends on the parameters. Regardless of the genetic algorithm, it has the following parameters: population size, number of generations (number of iterations, repetitions) and probability of mutation and probability of crossing. One of the earlier studies on parameter values was performed by De Jong, and according to his results, the best population size is between 50 and 100 individuals, the probability of crossing is about 0.6, and he determined the probability of mutation as 0.001 per bit in Figure 2.

![Figure 2. Display of genetic algorithm parameters.](image)

The probability function measures the goodness of a statistical model for given parameter values. In mathematics, computing and especially in graph theory, the distance matrix is a square matrix (two-dimensional array) that contains the distances between pairs, or elements of a set. If there are \(N\) elements, the size of the matrix will be \(N \times N\). The data can be displayed in graphical form as a heat map.

As it can be seen in Figure 3, a distance matrix was created based on a randomly formed map of random cities. There are several types of heat maps. It is possible to define a heat map as a map of five colors: red, green, blue, yellow and black. The user has the ability to change the properties of the map, where using different functions will change the appearance of the map, create edges on the map, interpolate colors with each other and the like. Testing of the genetic algorithm was performed on the example of a randomly created map of 30 cities. The initial parameters are programmatically defined in Figure 4.
Figure 3. Example of heat maps based on two created maps.

Figure 4. Creating randomly selected 30 cities for path optimization using a genetic algorithm.
Figure 5. Changing the path of the travelling salesman through iterations for a randomly created map of 30 cities.

By running the program, the initial locations are drawn with mutual paths that will change through iterations all the way to a certain number of given generations, trying to reach the ideal path as shown in Figure 5. Figure 6 shows the completed process of optimizing the route of a commercial traveller on the example of a randomly created map of 30 cities whose parameters are population size: 60 and number of generations: 300. The minimum distance between the given locations is 543.1525. The obtained results do not represent the optimal path and visually do not represent the best solution, therefore the parameters will be changed.

Figure 6. Completed process of optimizing the route of the travelling salesman on the example of a randomly created map of 30 cities.
Figure 7. Changing path optimization parameters on the example of a randomly created map.

By adjusting the parameters of population size to 200 and the number of generations to 500, the final solution is obtained as in Figure 7 and in comparison with the previously set parameters, a significant reduction in the distance between the points is visible.

4. Conclusion
This paper presents a practical implementation of an application for optimizing the path of a travelling salesman using a genetic algorithm in the software package MATLAB. The travelling salesman problem and the parameters used in the realization are described, and the cases of use in the work of the application whose solutions are graphically presented. Genetic algorithms represent an easy approach to complex large space search problems. Using models of learning and evolution, they enable modelling of the desired properties of the system, and automatic finding of solutions. In the described optimized cases, it is shown that abstract models of the genetic algorithm can be applied to real problems from different areas. The implementation of GA involves the use of various parameters: population size, number of generations (number of iterations, repetitions) and the probability of mutation and the probability of crossing. Genetic algorithms have various advantages and disadvantages that make them more or less applicable to various problems. The disadvantage of the genetic algorithm is that there is no single combination of parameters that is best for all problems, or at least for different instances of the same problem, but in each case we have to adjust them experimentally [7]. The achieved results, although they show satisfactory solutions, do not always ensure finding the optimal solution. As the commercial traveler problem is an NP-complete problem, and the genetic algorithm has proven to be suitable for solving this problem, it is obvious that the genetic algorithm is suitable for solving related problems from the class of NP-complete problems, provided that the problem can be represented according to the genetic algorithm.

References
[1] Matai R, Singh S and Mittal M L 2010 Traveling salesman problem: An overview of applications, formulations, and solution approaches IntechOpen
[2] Iqbal Ali, A., & Kennington, J. L. (1986). The asymmetric M-travelling salesmen problem: A Duality based branch-and-bound algorithm. Discrete Applied Mathematics, 13(2-3), 259–276. https://doi.org/10.1016/0166-218x(86)90087-9
[3] Applegate, D. L. (2007). The traveling salesman problem: A computational story. Princeton Univ. Press.
[4] Golemac, A., & Vojković, T. (1970, January 1). Kako pomoći trgovačkom Putniku Velga
Bosančić. Semantic Scholar. Retrieved October 13, 2021.

[5] Teodorović Dušan. (2007). Transportne mreže. Univerzitet u Beogradu, Saobračajni fakultet.

[6] Azra Hošić, (2021). Optimizacija putanje trgovačkog putnika primjenom genetskog algoritma u programskom paketu Matlab, diplomski rad, Univerzitet u Bihaću, Tehnički fakultet. (in Bosnian)

[7] Fatka Kulenović, (2014). Metode matematičkog programiranja za rješavanje lokacijskih i alokacijskih problema, Doktorska disertacija, Univerzitet u Sarajevu, PMF. (in Bosnian)