

Modified Genetic Algorithm to Traveling Salesman Problem for Large Input Datasets

Vladimir Vladimirov, Fatima Sapundzhi, Radoslava Krалеva, Velin Krалev
Faculty of Mathematics and Natural Sciences
South-West University "Neofit Rilski", Blagoevgrad, Bulgaria
v.g.vladimirov@abv.bg, sapundzhi@swu.bg, rady_krалеva@swu.bg,
velin.krалev@swu.bg

Keywords: Hamiltonian cycle, Genetic algorithm, backtracking, graph

The use of graphs is widely applied in modeling and solving problems in the field of computer science and bioinformatics. Therefore, it is essential to develop and improve algorithms reducing their computational complexity and increasing the precision of the solutions generated by them as well as the size of the input data. In this study two well-known algorithms for solving the problem for finding a minimum Hamiltonian cycle in weighted, undirected and complete graph (also known as Travelling Salesman Problem - TSP) are analyzed. The first algorithm is based on the backtracking method and it always finds the optimal solution, while with the second one, the genetic algorithm (GA), finding the optimal solution is not always guaranteed. The aims of the study are to determine: (1) which of the algorithms can be used so that the resulting solution is optimal or near-optimal and the execution time be reasonable depending on the size of the input data; (2) the influence of GA parameter values on the quality of the resulting solutions for large size of the input data. The parameters determine the number of solutions in each population and the number of all generations. The analysis of the results revealed that: (1) the algorithm that finds all possible solutions can be used for graphs with a small number of vertices (not more than 20), whereas GA can be used for graphs with a large number of vertices; (2) in graphs with a small number of vertices: $n < 20$ (and $n*(n-1)/2$ edges) GA always finds the optimal solution as long as enough solution space is set. However, the number of all Hamiltonian cycles in a complete graph with n vertices ($(n-1)!/2$) is bigger than the solution space; (3) all input datasets showed that with the number increase of vertices in the graph it is necessary to increase the number of the current solutions in the population. In this way GA reaches a certain rate of convergence faster, i.e., a generation after which the space of solutions contains only optimal solutions or near optimal ones.

Acknowledgments: This work is partially supported by the project of the Bulgarian National Science Fund, entitled: Bioinformatics research: protein folding, docking and prediction of biological activity, NSF I02/16, 12.12.14.