MASTER DEGREE PROGRAM IN DATA SCIENCE AND ADVANCED ANALYTICS | NOVA IMS

Computational Intelligence for Optimization Project Report

Vehicle Routing Problem

Github Repository: https://github.com/Routzahn30/CIFO.2022.git

May, 2022

Diogo Hipólito, m20210633
Mohamed Ali Felfe,l m20211322
Sergio Alfonso Corral Sobrevilla, m20211011
Jessica Routzahn, m20210987

NOVA IMS Information Management School

1. OBJECTIVE

For this project our group attempted a well-known optimization problem known as the Vehicle Routing Problem (VRP). This problem seeks the optimal set of routes for a fleet of vehicles, in this case trucks, to travel in order to deliver goods to customers. The intended goal of this project is just that, to achieve the optimal routes for multiple vehicles visiting a set of locations. Our intended definition by “optimal routes” for VRP is to minimize the length of the longest single logic route among all vehicles. This is the intended goal and definition as we want to complete all deliveries as quickly as possible.

Before jumping into the project, the similarities between the Traveling Salesperson Problem (TSP) and the Vehicle Routing Problem (VRP) must be noted. TSP considers a single vehicle visiting multiple customer locations before returning to the starting point with the goal of minimizing the total travel time or vehicle distance of the driver. VRP on the other hand considers multiple vehicles and generates multiple routes to pass through all customer locations. In order to achieve this goal our group assigned encoding rules, a fitness function, and developed different approaches to reach an optimal solution using various combinations of genetic operators and genetic algorithm parameters.

2. METHODOLOGY

Due to the nature of the problem and how it is commonly analyzed it was decided to opt for a representation composed of an array of X lists (depending on the number of vehicles) and each of these containing integers as the destinations each truck needs to go within its route.

We used the functions already created during the Computational Intelligence for Optimization Project Report practical classes and added some functions note taught to us. The pipeline that was carried out in this work was the one that is commonly used in genetic algorithms, which is composed as follows: initialization; followed by a fitness calculation/first population including selection, crossover, and mutation; and lastly new population (best fitness).

3. CONCEPTUALIZATION

For this project our group did not implement elitism. Our implementations are abstract and work with both maximization and minimization, although for the purpose of this project only minimization fitness scores will be discussed as this is a minimization problem. The following selection will be the conceptualization of the initial approach for the genetic algorithm.

3.1. *Encoding*

The encoding of chromosomes is often the first step in solving a genetic algorithm problem and varies with each problem. Just as in TSP, each gene is a number, but in VRP implementation each gene is an array of numbers and represents the route taken by each truck. The sizes of the arrays do not have to be equal. Below is an example of the chromosome representation of two individuals.

Individual 1 representation: [[6,10,5,7],[9,2,3,8],[11,12,1,4]]
Individual 2 representation: [[6,10]],[5,7,9,2,3,8,11],[12,1,4]]

	Truck 1				Truck 2				Truck 3			
Individual 1	6	10	5	7	9	2	3	8	11	12	1	4

	Truck 1		Truck 2							Truck 3		
Individual 2	6	10	5	7	9	2	3	8	11	12	1	4

3.2. *Fitness function*

The fitness function for this project represents the sum of the distances made by each individual. The starting point and ending point of each truck is always the DEPOT. Different fitness functions were not tried for this project. If our group had additional time, they would try to penalize trucks with only one or very few points in its route.

Truck 1				Truck 2				Truck 3			
6	10	5	7	9	2	3	8	11	12	1	4

The function will iterate through each truck's route and calculate the distance from the current point to the previous. For the first point, the DEPOT is considered point 0, and after iterating the array, the distance from the final point to the DEPOT is added.

	Truck 1					Truck 2					Truck 3			
0	6	10	5	7	0	9	2	3	8	0	11	12	1	4	0

Please Note: The DEPOT is configurable and isn't necessarily 0 in the distance matrix. However, this is not something our group configured for this project.

3.3. Reparations

We also included reparations in our genetic algorithm. The reparation method took place after the crossover operator. Reparations were included so there were no repeated values. It replaces repeated values with unseen values.

3.4. Data

Data used for this project is a Distance Matrix of all Districts of Bangladesh[1]. The *Kaggle* dataset is an indicative distance chart for 64 districts within the country of Bangladesh. All the distances have been measured in kilometers(km).

4. VARIATIONS TABLE

In order to increase the performance of the algorithm, our group applied various different variations and generic operators summarized in the following table.

Approaches/Genetic Operators	Implemented Variations
Selection Approach	1. Tournament 2. Ranking 3. Fitness Proportional Selection (FSP)
Crossover Operator	1. Single Crossover 2. Multiple Crossover 3. Uniform Crossover
Mutation Operator	1. Swap Mutation 2. Inversion Mutation 3. Scramble Mutation

Table 1. Variations

4.1. Selection Approach

Selection algorithms choose a selected individual(s) from the population. Our group implemented three selection approaches based on course lectures [2][3] and applied them to this VRP. Those approaches are Fitness Proportional Selection (FPS, or Roulette Wheel) PS, Ranking, and Tournament.

4.2. Crossover Operator

To start the modification, process our group generated and applied various crossover genetic operators. Crossover exchanges some characteristics of a set of individuals, in order to generate offspring individuals, that are a recombination of their parents. The crossover operators our group chose based on the lecture notes include Single, Multiple, and Uniform Crossover.

4.3. Mutation Operator

The second group of genetic operators our group generated and applied were three mutations that appeared in the course materials including Swap Mutation, Inversion Mutation, and Scramble Mutation. Mutations create a new individual randomly modifying a small portion of an existing individual.

5. METHODS COMPARISON

In order to find the optimal genetic algorithm to solve this VRP our group approached this task in phases. Each phase attempted to improve upon information gained in the previous phase. Before jumping into each phase it is worth noting which configurations worked best together, how different operators affected the convergence of the genetic algorithm, and how our group determined the best possible configuration.

Our group determined that the configurations that worked best together were Tournament Selection, Single Crossover, and Swap Mutation; Tournament Selection, Single Crossover, and Inversion Mutation; and lastly Tournament Selection, Multiple Crossover, and Swap Crossover. These combinations constantly showed good results across multiple tests. In order to find an efficient way to test combinations of genetic operators and parameters our group developed a three-step process into what will be called phases. Each phase or stage had the goal of removing possibilities and narrowing the search field for the optimal genetic algorithm.

5.1. Phase I

For this first initial step and phase I of the project our group attempted to set a baseline of fitness for all combinations of selection, crossover, and mutation. This resulted in 27 different combinations. During this phase no parameters for the genetic algorithm were changed. This we would adjust later. Parameters used for this initial phase included a population size 300, genes 100, crossover probability = 0.9, mutation probability = 0.3, and number of trucks set to 3. In this phase our group determined that FSP selection produced very poor results. In order to fully rule this selection method out the group performed one-off tests with FSP changing some of the parameters but resulting fitness was consistently poor. In the next phase these baseline results would be used to further rule out possibilities.

5.2. Phase II

As stated previously our group continued with testing combinations of genetic operators only with the Tournament Selection and Rank Selection. In phase II of the work for this report our group applied different parameters to the 18 combinations (9 for Tournament, 9 for Rank) previously identified. The parameters that were adjusted and changed in this phase included the number of trucks and genes. Each of the 18 combinations was tested with 3 trucks with 100, 200, then 300 genes; 5 trucks with 100 genes, 200 genes, and 300 genes; 7 trucks with 100 genes, 200 genes, and 300 genes; and lastly 10 trucks with 100 genes, 200 genes, and 300 genes. We quickly learned that increasing the number trucks past 5 dramatically reduced the fitness score, regardless of other parameter shifts. Our group also quickly realized that Rank Selection was computationally heavy and always resulted in a poor fitness score regardless of trucks and genes. Rank Selection was also dropped as a result of testing from this phase. It is through this phase our group learns which combinations work best together.

5.3. Phase III

In this last phase of testing our group applied additional parameter adjustments only for the three combinations mentioned above for truck sizes 3 and 5 as these produced the best overall results. The top ten results can be found in the table below.

Selection	Crossover	Mutation	Fitness	Trucks	gens	co_p	mu_p	Population
Tournament	Single point	Inversion	4417	3	1000	0.8	0.3	1000
Tournament	Single point	Inversion	4670	3	150	0.8	0.3	1000
Tournament	Single point	Inversion	5319	3	1000	0.9	0.3	1000
Tournament	Single point	Inversion	5336	5	1000	0.9	0.3	500
Tournament	Single point	Inversion	5447	3	300	0.9	0.4	300
Tournament	Multiple point	Inversion	5489	3	300	0.9	0.4	300
Tournament	Multiple point	Swap	5650	3	500	0.9	0.3	500
Tournament	Single point	Swap	5861	3	500	0.9	0.3	500
Tournament	Multiple point	Swap	5886	5	800	0.9	0.3	1000
Tournament	Single point	Inversion	5910	3	500	0.9	0.3	500

Table2. Top 10 Fitness results

5.4. Final Genetic Algorithm

After running the model repeatedly with different combinations of parameters, the best model (Fig.1) was found in which, unlike the second-best model (Fig.2), it uses a very large number of genes, which in real situations would require a large amount of energy and time, which would cause a loss in performance. For this reason, it was decided to reduce this characteristic to the point where it is appreciated that the fitness reaches a minimum and stabilizes. The first graph is being included as a demonstration that it is possible to continue improving the process, however this improvement will entail a cost that depending on the situation and the improvement, could be (or not) profitable.

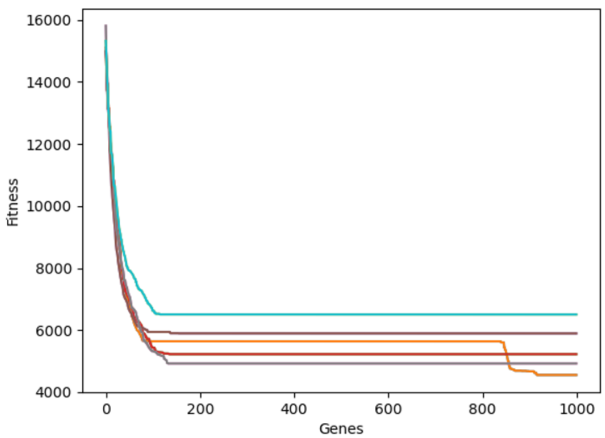

Fig.1 Best fitness

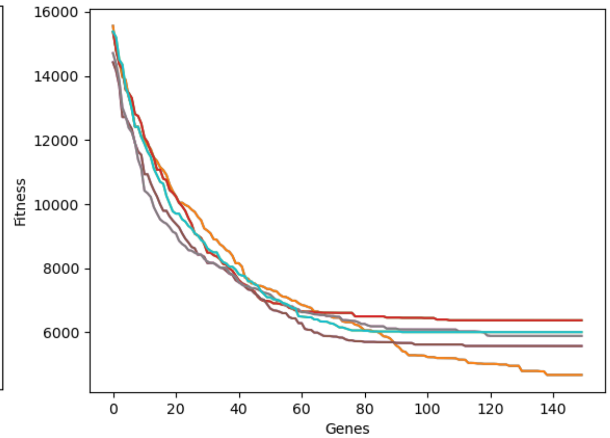

Fig.2 Second best fitness

6. CONCLUSION

This report depicts a good solution to the presented VRP using district location data from Bangladesh. Overall, our group received good results. A comment that our group had about the project is the variation of the results. While large variations in results are assumed due to the nature of the problem, different algorithms would incur in these differences. That said tournament selection is the parameter that provided the most substantial improvement in fitness and consistently produced good results. Despite the results received many improvements can be made to this project with additional time.

Given the chance to improve the work of this project several things would happen. One of the opportunity costs that our group experienced was the analysis of experiments, when performed manually, the number of combinations made vs. the possible ones was limited; although the experience when carrying out the tests led to understanding where relevant improvements could be obtained. Among the most important points to consider if it were decided to delve into this problem would be the logistic time, since it is one of the most important factors for any client and the operating and capital costs, since although more trucks could improve the time and fitness, does not mean that for a company that is economically feasible.

BIBLIOGRAPHY

[1] Distance Matrix of all Districts of Bangladesh. Accessed on 5/20/22 at:
https://www.kaggle.com/datasets/tasfiapasha/distance-matrix-of-all-districts-of-bangladesh
[2] Code used by Lectures
[3] Lecture Textbook