Finding an Optimal Restaurant Location Using a Genetic Algorithm

Coursera IBM Data Science Capstone Project : Business Problem Statement

Ryan Herchig

March 30, 2020

The business problem I have chosen to address is the ideal placement of a new restaurant within a city chosen by the user. Once the city is chosen, the city center will be considered the origin of the grid and the search for an optimal location will be confined to a radius also chosen by the user. Additional input from the user will be the type of restaurant they are interested in opening (e.g. Chinese, American, German, etc.). Once the ideal location has been calculated, it will be displayed on a map of the city along with the relevant input data for the algorithm.

Input data for the optimization algorithm will be found using the FourSquare API, and will consist of locations of grocery stores and markets as well as the locations of other restaurants of the same type as that chosen by the user. The goal of the algorithm will be to simultaneously 1) minimize the distance between all markets and grocery stores and 2) maximize the distance to the existing restaurants of the same type. This type of optimization problem, known as multi-objective or Pareto optimization, is difficult in practice, mostly owing o the fact that it is not a well-defined optimization problem. Stated differently, while trying to optimize with respect to one parameter, the solution usually becomes less optimized with respect to the other parameter. This typically results in a multitude of "good" solutions which represent trade-offs between the parameters with no truly optimal solution existing. The method used to solve this problem will be explained in detail in the final project report.

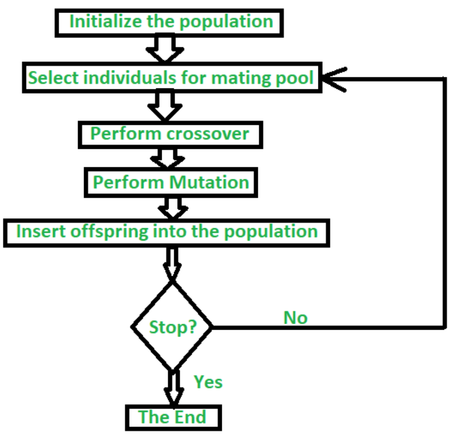

Figure 1: Schematic diagram depicting the phases of a genetic algorithm. [1]

To solve the defined business problem, a genetic algorithm will be used to search the space of possible solutions for one which provides an acceptable trade-off between the two optimization goals. The different phases which make up the genetic algorithm are shown in Fig 1. The members of the population represent possible solutions (candidate restaurant locations). The individuals from the population which represent the current most optimal solutions (best solutions to the cost function) are chosen to produce the next generation. The worst individuals from the population will be replaced by the offspring of the best individuals. The best individuals will be paired up and a linear mixture of each member of the pairs selected will be used to generate the new offspring. This tends to create new individuals which also represent "good" solutions. To ensure that the solution space continues to be explored, helping to avoid the algorithm settling into a local optimum, random mutations are applied to the newly created individuals. If the change in the average fitness (value of the cost function averaged over all members of the population) falls below some threshold for a predetermined number of steps, the algorithm is said to be converged.

$$f(lat, lng, G, R) = \sum_{n=1}^{N_{pop}} \left\{ \sum_{i=1}^{G_{max}} L_2(v_n, G_i) - \sum_{j=1}^{R_{max}} L_2(v_n, R_j) \right\} \quad (1)$$

The parameters represented by the best individual are considered to be the optimal solution. In this problem, the cost function will be given by equation 1 where lat and lng represent the latitude and longitude coordinates, v_n is the vector from the origin to the candidate location, G is the array containing the lat, lng of the grocery stores, and R is the array storing the lat, lng coordinates of the restaurant locations. The Euclidean distance L_2 is calculated between all the points in v_n and the R and G arrays respectively. Then, the distances are summed between v_n and G, and v_n and R. The difference between the sums is taken so that the minimum between v_n and G is calculated, while the maximum (negated minimum) between v_n and R is calculated. Therefore, minimizing equation 1 should find the location which simultaneously minimizes the distance to all the grocery stores while maximizing the distance to the competitors restaurants.

References

[1] Avik Dutta. GeeksforGeeks : Encoding methods in genetic algorithm. LINK.