Università di Genova

OPERATION RESEARCH FOR STRATEGIC DECISIONS: MODELS, METHODS - 98230

Project:

Poste Italiane Facility Location Problem

Amaan Ala Ud Din

Matr. 5156681

Abstract:

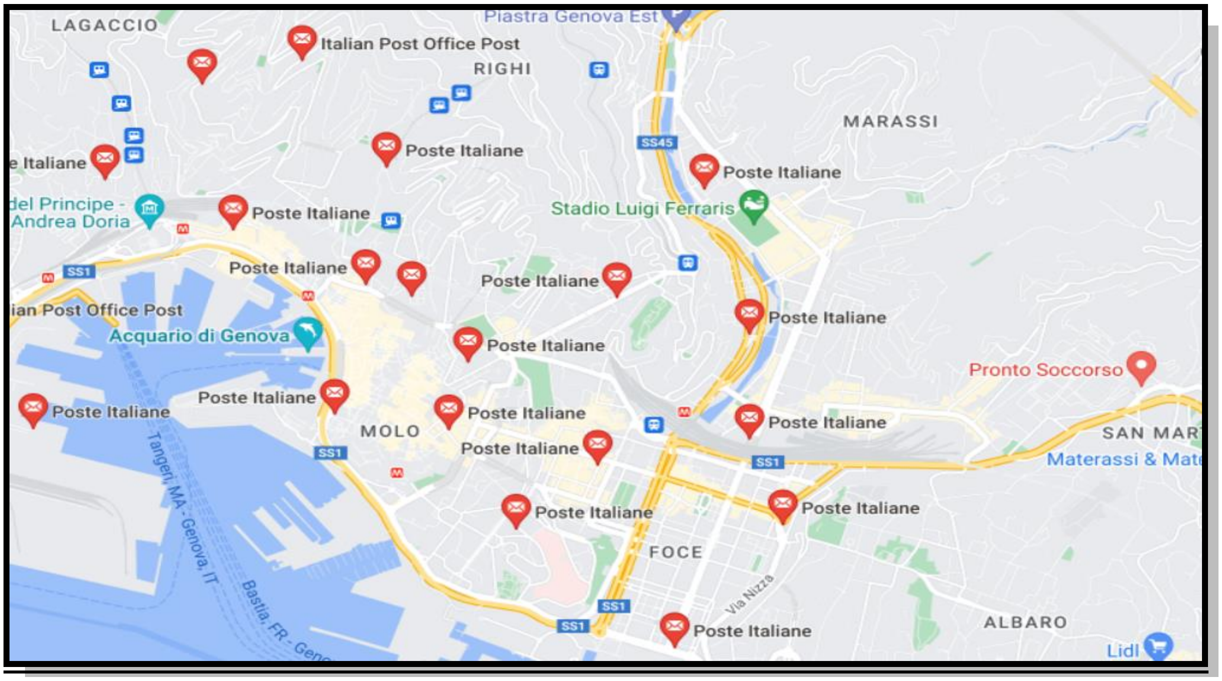
The facility location-allocation problem is an important topic that arises in many practical settings. FLP is the problem of finding optimal location for facilities within the given constraints. FLP has its widespread application in all areas. The problem of Facility location arises when it is to be decided where to locate a printer, server, warehouse or where to open a new store of a business chain. These applications arise in various areas such as transportation, distribution, production, supply chain decisions, telecommunication, and finance. This widespread application of Facility Location Problem has invited many researchers to try their hand to solve the problem. The problem of Facility Location is a well-known NP-hard problem and various heuristics have been proposed over the time to solve the problem. There exist multiple variants of the facility location problem like p-center problem or p-median problem. As a result, the studies on the facility location problems are steadily increasing in the literature. The warehouse site selection is a compelling decision for postal offices, and there are several studies in the literature that work on it, with different parameters and approaches. The main objective of the accurate warehouse site selection is to minimize the costs without compromising customer satisfaction. In this study, the concentration is on selecting the already present sites as a warehouse for storage of goods that are bound for inter-city distribution. The models are applied to Poste italiane branches located in Genoa. The results show that the location selection is an important factor and how much computational analysis goes into it.

Introduction:

Decisions about the distribution system are a strategic issue for every company. The facilities we plan today help an organization achieve supply chain excellence. Facility location problems deal with the question where to locate a single object or a set of objects. There are basically two groups of facility location problems: discrete location problems and continuous location problems. In many real-world applications, the facilities to be located can be warehouses, factories, depots, service centers distribution centers, antennas, or retailers. Problem that arises in a decision-making context of planning a distribution network or a production process (localization).

Problem Statement: Determine where among the available facility locations the warehouses could be positioned to minimize the total cost for the company.

There are 23 branches under consideration for the selection of warehouses for Poste Italiane. The warehouse selection is going to be of importance as all the goods bound for out of city delivery need to be stored at one or two locations from where they can be easily transported. For this purpose, a list of data regarding poste italiane offices in the genoa was collected.



<u>Some of the Facility Locations</u>

For making the model as accurate as possible, some real-world hypotheses were considered for calculation of costs. The data for the warehouses i.e., fixed (Activation Costs) costs was estimated according to the market trend and the minimum area required for the warehouse was set to be 100 sq. meters. In addition to this, the distance matrix (Distance Matrix) was constructed using the distance values (extracted from google maps) from one post office to another. These distance values were eventually scaled (considering the fuel costs) to estimate the monthly transportation costs so the analysis could be coherent when combined with fixed costs (Variable Costs) i.e., per month rental cost.

Monthly Fixed Costs (Activation Costs)				Area Required/ m^2
				100
Address	Monthly Rental Cost/€	Area/m^2	€/m^2	Monthly Rental Cost for 100 m^2 Depot
Via Vincenzo Blelè	1500	250	6.00	600.00
Via Donghi	1500	284	5.28	528.17
Via S. Fruttuoso	1680	230	7.30	730.43
Via Silvio Lagustena	400	66	6.06	606.06
Via Pisa	600	50	12.00	1200.00
Corso Sardegna	1250	30	41.67	4166.67
Via Fereggiano	500	60	8.33	833.33
Via Canevari	500	75	6.67	666.67
Via Francesco Pozzo	300	25	12.00	1200.00
Via Marassi	500	45	11.11	1111.11
Via Angelo Orsini	2500	400	6.25	625.00
Corso Europa	3000	250	12.00	1200.00
Via Aurelio Robino	500	60	8.33	833.33
Piazzale Marassi	2900	200	14.50	1450.00
Via Assarotti	1250	220	5.68	568.18
Via Dante	1250	120	10.42	1041.67
Via Granello	4300	260	16.54	1653.85
Corso Guglielmo Marconi	9000	700	12.86	1285.71
Via Redipuglia	1800	180	10.00	1000.00
Via Ilva	5400	720	7.50	750.00
Piazza Rovere	2800	290	9.66	965.52
Via Gaetano Colombo	1100	200	5.50	550.00
Piazza Cavour	1200	130	9.23	923.08

Conversion of Distance Values into Monetary Expenditure		
Fuel Consumption of a Mid-size Commercial Van made for delivery purposes	0.083	Liters per Km
Cost of Gasoline per Liter	1.83	Euro per Liter
Cost of Van's Fuel in a day assuming it travels multiple times between two nodes	2.27835	Euro per Km
Number of working days	22	Days per Month

Facility Location -LP Model

The problem of selecting the location was tackled by considering customers as nodes of a graph, and at the same time as possible places for opening a warehouse. The weights of the arcs connecting the nodes were obtained based on average distance on google maps. To convert the distance value into monetary expenditure, some hypotheses based on real data were considered.

We are going to use Linear Programming Model i.e., P-median model to find the number of facilities that should be used to minimize the total costs for the company. The theoretical approach used to solve is as follows.

– Let's define p-median of the graph every sub-set P of the set of nodes V that minimize:

$$d(i, P) = min_{j \in P}\, d_{ij}$$

– Decision variables: To activate the facilities we define binary variables:

$$y_i = \begin{cases} 1 & if\ i \in p\ median > 0 \\ 0 & otherwise \end{cases}$$

– To associate each customer with the nearest facility we define binary variables:

$$x_{ij} = \begin{cases} 1\ if\ node\ i\ is\ served\ by\ node\ j \\ 0 \qquad otherwise \end{cases}$$

– Objective function: to minimize costs

$$\min \sum_{i=1}^{n} f_j\, y_j + \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij}\ x_{ij}$$

– Constraints: one customer can be served by only one facility:

$$\sum_{i=1}^{n} x_{ij} = 1$$

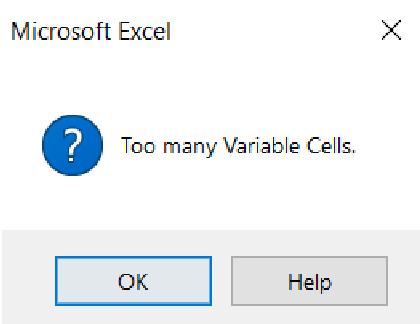
– Customer I can be served by j only if j is a facility:

$$\sum_{i=1}^{n} x_{ij} \leq My_j$$

– Total p facilities required:

$$\sum_{j=1}^{n} y_j = p$$

The problem was addressed first using the Excel tool. To my surprise, Excel [(LP Model)] couldn't compute more than 200 decision variables. But the model was constructed anyways.



However, With the use of Python language with the Pulp library, the problem was solved. The data of the problem have been passed to the program in two ways. As regards the fixed costs, the insertion was done manually inside the code. For the variable costs data was read form csv command using Matrix file which can be managed by Python. Once all the necessary data has been passed, the Python program evaluates the costs for all the possible combinations and outputs the optimal situation with the given number of P-value (Number of facilities to be active). At this point the problem is defined according to the methods defined by the procedure described above, a variable x which represents the activation of the customer facility arcs, and the variable y which represents the activation of a facility. Consequently, the constraints listed before with the objective function are defined, and the optimizer solves the problem.

Results:

P-value =3

```
Status: Optimal

[True, False, False, False, False, False, False, True, False, False, False, False, False, False, True, False, False, False, False, False, False, False, False]

Cost Evaluation MILP: 4220.73812
```

P-value =2

```
Status: Optimal
[True, False, False, False, False, False, False, False, False, False,
False, False, False, False, True, False, False, False, False, False,
False, False, False]
Cost Evaluation MILP: 4150.54015
```

P-value =1

```
Status: Optimal
[False, False, True, False, False, False, False, False, False, False,
False, False, False, False, False, False, False, False, False, False,
False, False, False]
Cost Evaluation MILP: 4334.32403
```

Facility Location - Heuristic Approach

We are going to now apply Heuristic Method on the problem described earlier, that although it shouldn't offer the absolute optimal solution, it is still able to find a fair enough solution with excellent resolution time. The basic idea is to be able, through multiple iterations, to determine the facilities that reduces overall costs as much as possible. A possible facility is added to each iteration, and it is verified if this can contribute to decrease the expenses. Where this is not more convenient, the procedure has reached its best solution.

- Initialization: no facility is active. If F is the set of the chosen facilities: F = ϕ

It is calculated the activation cost of the first facility. Let be V the set of nodes:

$$L:\ L_j = f_j + \sum_{i \in V} c_{ij}$$

Where fj is the activation cost of facility j and cij is the transport cost from node j to node i

- Selection: add the facility corresponding to the j * node so that:

$$L_{j^*} = \min_{j \in N} L_j$$

- Iteration: for each possible facility to be added, the possible savings corresponding to the opening of a new facility in node j are calculated:

$$\Delta L_j = \sum_{i \in V} \max(\ min_{k \in F}\ c_{ik} - c_{ij}, 0) - f_i$$

It is calculated $\Delta Lj*$ = max ΔLj

If $\Delta Lj*$ > 0 it means that there is convenience in opening facility j * and therefore j * is added to F.

The tool chosen to solve the problem was the Excel worksheet. Since this is a heuristic method and the problem does not consist of a very large number of variables, we believed that this tool could be excellent with regards to clarity. The excel file (OR-Project) contains sheet ‘Initialization’ and ‘Heuristic’ which contains the calculation for this problem. It follows the procedure described above, with the initialization and possible addition of a facility to the solution of the problem.

Results:

Selected Nodes	Total Cumulative Cost
Node 3	4334.33
Node 15	*4170.70*

Final Cost	*4170.70*

Conclusion:

The preliminary analysis of the LP model shows that the appropriate choice would be to activate 2 locations for warehouses i.e., **Node 1** and **Node 15**. Which results in the minimum costs to be paid by the company. However, the heuristic approach has a different result i.e., which is **Node 3** and **Node 15** with a slightly higher cost. This result was expected, as the heuristic method is based on the intuitive and fast procedure which does not frame the overall situation in the case of multiple facilities. Probably the cause of this analogy is to be derived from the geographical configuration of the nodes as well as from the weight of its arcs. Surely with a more relaxed graph the results would have been different between the two methods. This however confirms, in this case in terms of expenses, the validity of the chosen heuristic method because the difference between the total costs is almost negligible. On the contrary, the costs of managing the selected warehouses in the possible geographical locations i.e., utilities bills, labor force, taxes etc. should have been considered. For these reasons, a more detailed study of the variables and external factors involved would have allowed a more careful strategic choice.

Facility Location - Metaheuristic Approach

In the metaheuristic solution to the problem, we are going to use genetic algorithms to find the minimum travelling distance between all the post offices. This scenario is like the most conventional problem in operations research commonly known as TSP or Travelling Salesman Problem. As it is obvious that a delivery van would need to travel between all the locations, hence we are going to find the shortest route possible route that minimizes the travelling costs for the company, Under the constraint that it visits each office only once. It should be noted that, the results of this algorithm will only incorporate the distances between the locations.

GA approach:

Let’s start with a few definitions, rephrased in the context of our problem:

- Gene: a city (represented as (Longitude, Latitude) coordinates)
- Individual (aka “chromosome”): a single route satisfying the conditions
- Population: a collection of possible routes (i.e., collection of individuals)
- Parents: two routes that are combined to create a new route
- Mating pool: a collection of parents that are used to create our next population (thus creating the next generation of routes)
- Fitness: a function that tells us how good each route is (in our case, how short the distance is)
- Mutation: a way to introduce variation in our population by randomly swapping two cities in a route
- Elitism: a way to carry the best individuals into the next generation

Our GA will proceed in the following steps:

1. Create the population
2. Determine fitness
3. Select the mating pool (Fitness Proportionate Selection)
4. Crossover (3 Operators have been used)
5. Mutate
6. Repeat

We are using the coordinates (coordinates sheet) from google maps for each location (which are in decimal degrees) and using python command to convert them into radians. Subsequently, to model the exact distance, we are using the radial distance formula.

$$Distance = \cos^{-1}(\cos(Radians(90-Lat1)) \times \cos(Radians(90-Lat2)) + \sin(Radians(90-Lat1) \times \sin(Radians(90-Lat2)) \times \cos(Radians(Long1-Long2))) \times R$$

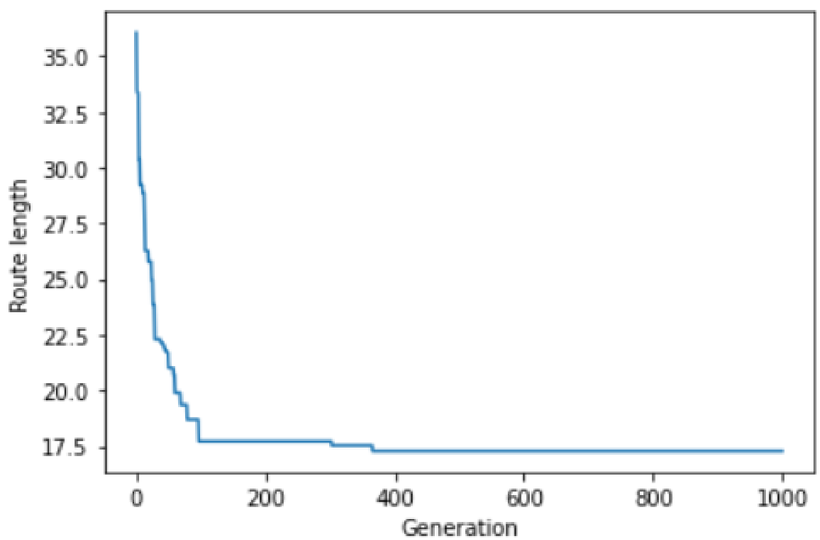
The input arguments for the algorithm includes the manual insertion of the coordinates of the locations in the same order as written in (coordinates sheet). Furthermore, Population size, Elite size, Mutation Rate and Generations can be altered to get different results, higher the number of generations more accurate the optimal solution but with a tradeoff of more computation time.

Results:

Popsize=200, Elite size =10, Mutation Rate = 0.002, Generations = 100:

```
Optimal route is: [20, 7, 18, 9, 12, 11, 19, 5, 4, 2, 1, 3, 6, 8, 10, 17, 13, 14, 15, 21, 22, 16, 23]
```

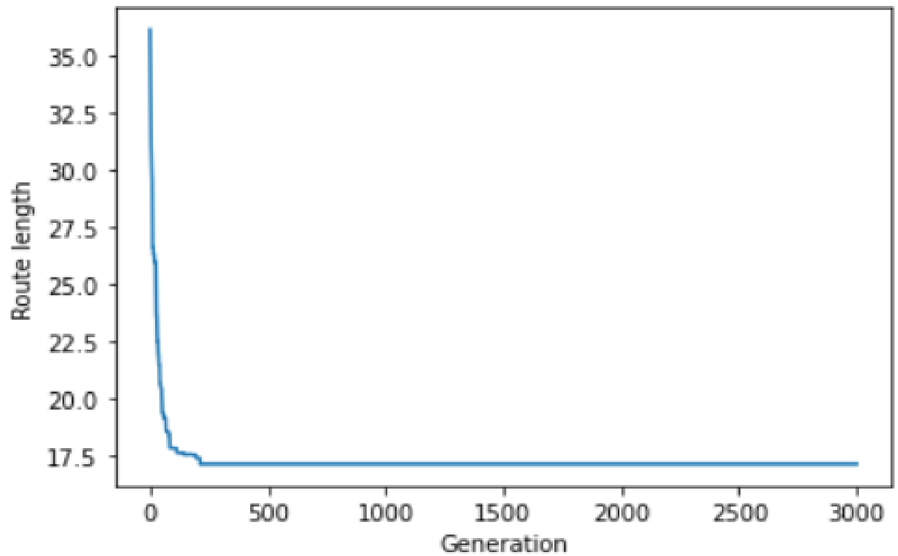
```
Final route length:      17.27
  Time =  417.96  seconds
```



Popsize=200, Elite size =10, Mutation Rate = 0.002, Generations = 3000:

```
Optimal route is: [7, 2, 14, 13, 8, 10, 21, 15, 23, 22, 20, 16, 6, 17, 18, 9, 11, 5, 12, 19, 3, 4, 1]
```

```
Final route length:      17.15
  Time =  1247.52  seconds
```



Conclusion:

Genetic Algorithm is an evolutionary process, it's going to result in varying and suboptimal solutions when the number of generations are not enough. To get a concrete idea of an optimal solution, a higher number of generations must be produced, which will minimize the distance (travelling costs). The sequence generated by the algorithm shows that in what order the locations must be visited or how the vehicle route must be planned. In the courier service company, distances hold more relevance than other costs. As their business model is built around logistics and transport terminologies. Hence, the first node can be selected as an option for the warehouse when we are assuming that the collection vehicle will need to visit all the offices and come back. This approach will most definitely, reduce the transportation costs. But again, a more detailed study of the external factors, geographical analysis will result in a better strategic choice.

Data Collection References:

https://commons.wikimedia.org/wiki/File:Quartieri_di_Genova.png

https://www.city-facts.com/genova

https://www.sust-it.net/miles-per-gallon-mpg-fuel-efficient-cars.php?type=van&fuel=petrol&manufacturer=fiat

https://www.globalpetrolprices.com/Italy/gasoline_prices/

Case in vendita - Immobiliare.it