2020 5th International Conference on Mechatronics, Control and Electronic Engineering (MCEE 2020)

P-median problem solution and research

Yuankun Chen[1,a], Yigao Li[2,b]

[1]Huazhong University of Science and Technology

[2]No.1 Middle School Affiliated to Central China Normal University

Email: [a]chenyuankun1108@163.com, [b]liyigao516@163.com

Keywords: P-median problem; CFLP problem; Lingo; genetic algorithm

Abstract: The location of facilities is a common decision-making problem in the field of operations research and management science. The correctness of the location decision depends mainly on whether the economic benefits, utility, personal or social satisfaction and social value can be brought about after the location decision. P-median problem is to study how the Pfacility location such that the product of the distance between the demand and the needs of the service facilities and the smallest part of NP-Hard problem. In addition to this article to solve the problem with the lingo P-median and genetic algorithm, the further study of the multi-facility location problem with capacity constraints, and the performance indicators lingo outcome of the genetic algorithm derived performance comparison, also Compare the performance of the two solutions in solving the problem in dense and sparse graphs. Finally, this paper discusses the use scenarios of lingo and heuristic algorithms in solving operational research problems.

1. Background

The location of facilities is a problem of decision-making that is prevalent in the field of operations research and management science. The problem of facility location is to study how to choose the number of facilities and the optimal location to provide users with corresponding services. The correctness of the location decision depends mainly on whether the economic benefits, utility, personal or social satisfaction and social value can be brought about after the location decision.

The location decision is to study the location of human organizations at different levels, from individuals and families to companies, government agencies and even international institutions.„

Location decision making is a strategic decision. Long-term capital utilization and economic benefits need to be considered. For private facilities, location decisions directly affect the company's competitive strength in the market; for utilities, location decisions affect the efficiency of providing public services and the fairness of public access to public services.

The location decision also covers the extended meaning of the economy. Including pollution, traffic congestion and economic development potential.

Since most of the siting problems are NP-hard problems, it is difficult to solve the optimal solution of the location model, especially the large problem.„

The location problem has a corresponding application background, and the structure of the model (objective functions, variables and constraints) is determined by the corresponding application background. There is nouniversal model that can solve all siting problems.

2. Key to location problem

The location problem involves a comprehensive consideration of many factors, and we need to weigh the various advantages and disadvantages, and on this basis, we can find a satisfactory solution. The usual location model provides multiple, contradictory goals when providing decision-making advice to decision makers. In general, a concise model is better than a complicated model, and the essence of the real problem is simply reflected with the least amount of variables.

Through modeling, weigh the conflicts of interest between the objectives while satisfying the real constraints as much as possible. The problem of facility location is to consider the following factors:

1) How many facilities are needed;
2) Where to build each facility;
3) The cost scale of each facility;
4) How the location point meets the demand for traffic distribution.

The coverage problem and the central problem are based on the facilities that the facility can enjoy the facilities within the radius of coverage. To increase the efficiency of the system, the demand point must be increased within the coverage radius. Due to the existence of the weight relationship, the comprehensive utilization of the system is often wasted. The media nangle is standing problem of the system of interest to characterize the relationship between the distance and cost, which is the overall cost of the system targetminimum.

P-median problem is to study how the Pfacility location such that the product of the distance between the demand and the needs of the service facilities and the smallest part of NP-Hard problem, the solution may be the case with the variable increased from variable geometry growth.

The mathematical model of the problem is given below:

$$
\begin{aligned}
\min \quad & \sum_{i=1}^{n}\sum_{j=1}^{n} w_i d_{ij} x_{ij} \\
\text{s.t.} \quad & \sum_{j=1}^{n} x_{ij} = 1 && \forall i, \\
& x_{ij} \leqslant y_j && \forall i, j, \\
& \sum_{j=1}^{n} y_j = p, \\
& x_{ij} = 0 \text{ or } 1 && \forall i, j, \\
& y_j = 0 \text{ or } 1 && \forall j,
\end{aligned}
$$

- Indicates facility capacity (service demand), in the data set corresponding to demand column, simple P-median problem Equal to 1, that is, regardless of the impact of system capacity on the problem.

- Indicates the distance between the facility and the demand point, and corresponds to the transportation cost column in the dataset. In the following process, we will approximate the distance between the points as the operation or transportation cost between the two points.

- Binary variables, When =1, it means that it is confirmed that j point is confirmed as a construction facility, otherwise j point is not a facility (service point).

- Binary variable, indicating the connection between demand point i and facility j, only in =1 makes sense.

P-Building facilities (total number of service points).

3. Introduction to multi-facility location problem with capacity limitation

The CFLP (capacitated facilities location problem) model is a multi-facility location problem with capacity limitations. The problem is described as follows: A company has n sales regions, and the demand for each sales region is known. Company decided to be established several distribution centers, confirmed by the inspection that there are m candidate locations, each candidate has the capacity limit, and has a fixed cost (e.g., construction costs or cost lease), the problem is how to ground the candidate m Choose a location to build a distribution center to minimize logistics costs.

The problem model is given below:

$$\min Z = \sum_{i=1}^{m}\sum_{j=1}^{n} C_{ij}X_{ij} + \sum_{i=1}^{m} F_i Y_i$$

s.t.

$$\sum_{i=1}^{m} X_{ij} = D_j \quad j=1,2,3,\ldots,n$$

$$\sum_{j=1}^{n} X_{ij} \le W_i Y_i \quad i=1,2,3,\ldots,m$$

$$\sum_{i=1}^{m} Y_i \le k$$

$$Y_i \in \{0,1\}$$

$$X_{ij} \ge 0$$

I-distribution center candidate, i=1, 2, 3..., m

J-sales area: j=1, 2, 3..., n

K-the number of proposed distribution centers;

Dj-the demand for sales j;

Fi-distribution center candidate site i fixed cost;

Wi-distribution center capacity;

Cij - unit transportation cost from the distribution center candidate i to the sales place j;

Xij-unit transportation volume from the distribution center candidate i to the sales location j (decision variable)

Yi-Distribution Center candidate i is selected when it is selected, otherwise it is 0 (0-1 decision variable)

4. Experiment

At the beginning of the report, I hope that I can understand the simple p-median. However, with the deepening of the research on the problem and considering the practical application in real life, I decided to continue to explore the CFLP problem after solving the basic p-median problem. CPMP problem, I hope to solve similar problems with similar heuristic algorithms. Explore the similarities and differences between the two issues.

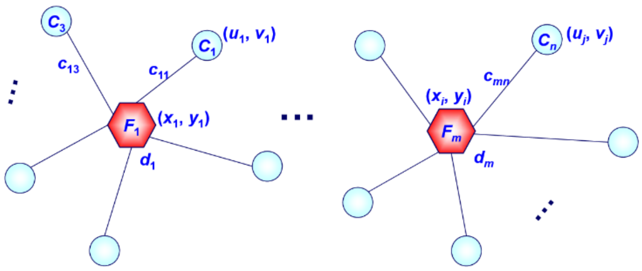
4.1 Simple P-median problem

OR Library->P-median Problem->Instanseon perfect codes, etc. (Other data such as checkpoints, FFP problems, etc. can also be calculated, there are errors).

First process the data, standardize the data, if the two points can not be connected, set the distance between them to infinity (99999), here use simple Matlab code to import the data into the excel table, easy to lingo data Read and output. If you need to read other data, you only need to modify the cell part.

4.2 Genetic algorithm design

Data set is given directly facility with transportation cost between users, where transport costs can be viewed as a collection of distance with other transport costs, so this did not need to use the coordinates and then again to solve the Euclidean distance, can be used directly. At the same time, because the data on the OR Library is too sparse, each facility serves 8 clients, and only 16 facilities can be selected. Unless the client of each facility service cannot be duplicated, it is a legal solution. It is very complicated when generating the initial population. in late is not easy to eliminate bad solution, and therefore **to each facility is considered one of the nodes in the diagram, the service between the two facilities if** the same customer, then add an edge between two nodes, generate a undirected FIG, 16 and then find the nodes in the graph, there is no edge between each node can be reduced datainaccuracies Solutions to some extent the problems caused by such optimization.



4.3 Chromosome encoding and representation

For encoding facility using 0-1 where 0 in dicatesnoopen, 1 open; point for the user is arandomly generated number between 1-J.Taking P=5, n=10 as an example, the actual data set is much larger than this, so it will be slower when generating the initial population. At the same time, according to the dataset on the Or Library, it can be seen that the data is sparse, so the data is first adjusted to generate a matrix. If one point cannot serve another point, the distance between them is set to infinity.

1	2	3	4	5	1	2	3	4	5	6	7	8	9	10
1	0	1	0	1	4	5	5	3	4	3	1	3	2	5

Because this article divides the chromosome into two parts, we use different crossover methods for the different two segments. For a single point crossing of the facility segment, use double tangent crossing for the user segment .Improve operational efficiency

Similarly, the mutation uses two methods. The use of inversion variants for facilities and the use of substitution variants for user points is based on different data types. At the same time, in order to prevent falling into local optimum, the mutation rate of the gene mutation is set to 100% when calculating the OR Library data.

To simplify the algorithm, the fitness is to use the 1/object function (1/cost).At the same time for evaluating fitness, and the chromosomehereinare obtained from the above step Cij, while each chromosome will have to calculate the fitness.

Similarly, when choosing this article, we also choose to use two different methods to select the population. For the first half of the facility, the elite selection method is used, and the second half of the facility uses the roulette selection method.

The program terminates after a certain number of runs. The program will temporarily set the number of iterations to 50000.

The P-median data on the OR Library is very sparse, and there are very few points that each point can serve. Therefore, the genetic algorithm is used to initialize the population slowly, but once the population is initialized successfully, the number of iterations will be very fast. Taking 50,000 iterations as an example, it takes 28.58 seconds, and using lingo software solves 1 minute. 02 seconds.

5. Conclusion

There are a variety of location issues, including P-median, as well as CFLP issues, CPMP issues, barrier location issues, multi-layer location issues, and more.

The existence of lingo and heuristic algorithms is not repeated. They are suitable for different objects, such as lingo packaged software, their algorithms are fixed, so their accuracy is very high and inefficiency is difficult to highlight in the face of small-scale data. But when the amount of data expands, lingo's disadvantages appear. He may fall into the unknown deadlock, but he can't get an answer without a solution, which is caused by the embedded embedded calculation method. In the case of large data volume, the advantages of the heuristic algorithm are highlighted, and the calculation speed is fast. Although the calculated solution is not the optimal solution, the solution is always stronger than the solution. The heuristic algorithm can always be this paper. We provide a solution, which is why this article personally looks at the design of restarting the algorithm instead of the lingo application.

Thinking after the process and after the completion. Decided to start the use of genetic algorithms to solve the problem, but the more backward the more think there should be more suitable algorithm, this paper argue sant colony algorithm is more suitable for the calculation of such problems, the next step will be to use ant colony algorithm such problems again .

References

[1] Densham,P.J. and G.Rushton.(1992). “Amore Efficient Heuristic for Solving Large p-Median Problems.” Papers in Regional Science 71, 307–329.

[2] Dibble, C. and P.J. Densham. (1993). “Generating Interesting Alternatives in GIS and SDSS Using Genetic Algorithms.” In GIS/LIS 1993.

[3] Dowsland, K.A. (1996). “Genetic Algorithms – A Tool for OR?” Journal of Operational Research Society 47, 550–561.

[4] Erkut, E., T. Myroon, and K. Strangway. (2000). “Trans Alta Redesigns Its Service Delivery Network.” Interfaces 30(2), 54–69.

[5] Fitzsimmons, J.A. and A.L.Austin. (1983). “A Warehouse Location Model Helps Texas Comptroller Select Out-of-State Audit Offices.” Interfaces 13(5), 40–46.

[6] Galvão, R.D.(1980). “A Dual-Bounded Algorithm for the p-Median Problem.” Operations Research 28(5), 1112–1121.

[7] Galvão, R.D. and L.A. Raggi. (1989). “A Method for Solving to Optimality Uncapacitated Location Problems.” Annals of Operations Research 18, 225–244.

[8] Galvão, R.D. and C. ReVelle. (1996). “A Lagrangean Heuristic for the Maximal Covering Location Problem.” European Journal of Operations Research 88, 114–123.

[9] Hosage, C.M. and M.F. Goodchild. (1986). “Discrete Space Location–Allocation Solutions from Genetic Algorithms.” Annals of Operations Research 6, 35–46.

[10] Koerkel, M. (1989). “On the Exact Solution of Large-Scale Simple Plant Location Problems.” European Journal of Operations Research 39, 157–173.

[11] Moreno-Perez, J.A., J.M. Moreno-Vega, and N. Mladenovic. (1994). “Tabu Search and Simulated Annealing in p-Median Problem.” In Proceedings of the Canadian Operational Research Society Conference, Montreal.

[12] Murray, A.T. and R.L. Church. (1996). “Applying Simulated Annealing to Location-Planning Models.” Journal of Heuristics 2, 31–53.

[13] Narula, S.C., U.I. Ogbu, and H.M. Samuelsson. (1997). “An Algorithm for the p-Median Problem.” Operations Research 25, 709–712.

[14] Pizzolato, N.D. (1994). “A Heuristic for Large-Size p-Median Location Problems with Application to School Location.” Annals of Operations Research, 50 473–485.

[15] Reeves, C.R. (1993). “Genetic Algorithms.” In C.R. Reeves (ed.), Modern Heuristic Techniques for Combinatorial Problems. Chapter 4, pp. 151–196.