Implementing Genetic Algorithm to solve Facility Location Problem

Monika Mangla[1], Rakhi Akhare[2],Smita Ambarkar[3]

[1] Asst. Prof. Computer Department, LTCE, Navi Mumbai, Maharashtra, INDIA
[2] Asst. Prof. Computer Department, LTCE, Navi Mumbai, Maharashtra, INDIA
[3] Asst. Prof. Computer Department, LTCE, Navi Mumbai, Maharashtra, INDIA

---***---

Abstract - ***Facility location problem is the problem of finding optimal location for facilities in a plane consisting of demand points so that each demand point has at least one facility at a distance no more than some permissible distance. Facility location problem has its wide spread application in all areas. The problem of Facility location arises when it is to be decided where to locate a printer, server, ware house or where to open a new store of a business chain. Facility location problem have several application in telecommunication, transportation, distribution etc. 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. Here in this paper, we have attempted to solve the well known p-median problem. We use genetic algorithm to solve the capacitated version of p-median problem.***

Key Words: *Euclidean space, Capacitated Facility, Optimization*

1. INTRODUCTION

Facility location problem has widespread application in various areas like telecommunication, networking or business. There is hardly any area left untouched by application of Facility Location Problem. The location problem arises when it is to be decided where to locate a server in a communication network or where to open a new store. Depending upon the application of location problem, there exists multiple variants of location problem. The most common variants are p-center and p-median problem. p-center problem deals with location of p emergency services like location of fire station or ambulance center. P-median problem is another variant in which the objective is to minimize the overall communication cost that incurs to connect to the nearest facility for all demand points. This problem is well known in the literature as Fermat Weber Problem. P-median problem consists of locating p facilities in a given Euclidean space that provides service to n demand points in such a way that the total sum of distances between each demand point and its nearest facility is minimized. Capacity restriction on the facility generates further classification of the location problem. Capacitated Facility location problem is the location of facilities with limited capacity and thus in turn can provide service to limited demand points. Whereas facilities having no capacity constraints has no limit on the number of demand points that can be served by facility and thus every demand point will always be seeking service from its nearest facility. The same is practically not possible and thus it necessitates considering capacity limits of the resources.

The location problem is a well known NP-hard problem and thus various researchers are getting attracted to carrying out research in the area. To solve NP-hard problem various techniques have been proposed by various researchers over the decade (Kariv and Hakimi, 1979). Approximation algorithms, computational geometry and heuristics are few important tools that have been utilized to solve the location problem. Here in this paper, we use Genetic algorithms to solve the facility location problem.

2. THE P-MEDIAN PROBLEM

As stated earlier the p-median problem is finding location of p facilities in a demand plane consisting of n demand points where the objective is to minimize the sum of distances from each demand point to its nearest facility. The p-median problem can be best described by the following equation.

$$\min \sum_{j=1}^{p} \sum_{i=1}^{n} d_{ij} x_{ij}$$

Where
p represents number of facilities
n represents number of demand points
d_{ij} represents the distance between demand node i and facility j

x_{ij} = 1 if demand node i is connected to facility j 0 otherwise

Thus, objective of the p-median location problem is to find location of p facilities such that sum of distances from each facility to its nearest facility is minimized. Here number of demand points exceeds number of facilities by a significant ratio (n>>p). There are multiple variants of the location problem. Most common of this classification exist continuous location problem and discrete location problem. Continuous location problem is the problem where demand plane is continuous and facility can be located anywhere in the plane. On the other hand, in discrete location problem, there exist a number of possible locations for facilities where it can be located and the objective is to choose p locations out of possible range of locations. An example of discrete location model is to find location of server is in a communication network. In this case the communication network is represented by a graph of vertices and edges where each vertex represents a node in the network and an edge signifies the connection among participating vertices. Now server can be located at any vertex only and not on the edges thus it falls in the category of discrete location model.

During the last decade, various researchers have developed a keen interest to solve multiple variants of the location problem. The most commonly used variants are continuous or discrete location problem. Then another classification on the basis of facilities are capacitated or incapacitated location problem. The presence of barriers in the demand plane also generates another classification of location problem that necessitates a special approach to solve the problem.

Various approaches have been used to solve the problem by different researchers. The common approaches to solve the location problem are Approximation algorithm, computation geometry and various heuristics based algorithms. Genetic algorithms have also proved to be an effective tool to solve the problem over the time [1] [2]. There exist infinite applications of location problem in the real world. One such application has been discussed in next section.

3. A REAL WORLD APPLICATION

We consider example of a school that provides bus service to its students. Now the school needs to decide the bus stops for all the students availing bus service. There exists few options for the bus stops thus it comes under the category of discrete location problem. The objective of the school is to decide p stops out of all available bus stops so that the total distance each students travels to its nearest bus stop is minimized. Now as the capacity of the bus is limited it can cater to only limited number of students thus falling in the category of capacitated location problem. Let us assume that total number of students is 1000 and the capacity of each school bus is 50. Now in order to obtain the distance between students' home and corresponding facility (bus stop in this case), addresses of all students in question should be precisely located in a digitized map of the city.

Now in order to maintain operating convenience and economic reasons, the school decides to limit the number of bus stops to 25 only among a set of 50 candidate locations for bus stops. We cast this problem as a capacitated p-median problem, as follows:

The set of 50 facilities (potential bus stops locations) is the set V (|V| = 50) of all facilities candidate to median (actual exam locations)

Let V_p = 25 be the set of the 25 selected locations for bus stops. Each of the 50 potential exam locations can satisfy only a limited number of candidate students. The goal is to select a set Vp Ɛ V that minimizes the total sum of distances between each candidate student's home and its nearest bus stop.

4. GENETIC ALGORITHM FOR CAPACITATED P-MEDIAN PROBLEM

This section describes usage of Genetic Algorithm to solve the capacitated p-median location problem. our proposed GA for the capacitated p-median problem(Cap-p-Med-GA) is as follows:

A. Individual Representation

Each individual (chromosome) has exactly p genes, where p is the number of medians, and the allele of each gene represents the index (a unique id number) of a facility selected as median. For instance, consider a problem with 15 facilities (potential medians) represented by the indexes 1, 2,..., 15. Suppose one wants to select 5 medians. In our GA, the individual [2, 7, 5, 15, 10] represents a candidate solution for the problem where facilities 2, 5, 7, 10 and 15 have been selected as medians. In Cap-p-Med-GA the genome is interpreted as a set of facility indexes, in the mathematical sense of set - i.e. there are no duplicated indexes and there is no ordering among the indexes.

B. Fitness Evaluation

Fitness function of a chromosome or individual represents its closeness to the optimal solution. Higher is the value of fitness, it is closer to the optimal solution. In essence, the fitness of an individual is given by the value of the objective function for the solution represented by the individual - as measured by equation given above. However, there is a caution in the computation of the fitness of an individual. Note that Capacitated-p-Median-GA is used only to optimize the choice of the 25 medians, out of the 50 facilities. However, the computation of equation requires that each of the 1000 candidate students be assigned to exactly one of the selected medians (i.e. the facility bus stop from where the student will board the school bus). This assignment is done by a procedure that is used by Cap-p-Med-GA as a black box.

Here we just mention the basic idea of this procedure. Once the 25 medians are selected, the procedure tries to assign each candidate student to the median (bus stop) that is the nearest one to its home. The problem is that, since each median has a fixed capacity, some candidate students will have to be assigned to the second (or third, fourth ...) nearest median to their homes. Suppose there is an assignment conflict -e.g. there is just one position in one median, and that median is the nearest one for two candidate students.

In this case the student-assignment procedure prefers to assign to that median the student that would be most intolerant if it was assigned to its second nearest median. A student is "intolerant" to the extent of the difference between two distances, namely the distance between her home and her nearest median and the distance between her home and her second nearest median. Once the student-assignment procedure is complete, the fitness of an individual can be computed by given equation.

C. Selection

The first step consists of selecting individuals for reproduction. This random selection is done with a probability depending on the relative fitness of the individuals so that best ones are often chosen for reproduction than poor ones. Typically we can distinguish two types of selection scheme, proportionate selection and ordinal-based selection. Proportionate-based selection picks out individuals based upon their fitness values relative to the fitness of the other individuals in the Population. Ordinal-based selection schemes select individuals not upon their raw fitness, but upon their rank within the population.

D. Crossover

Crossover is the process of taking two parent solutions and generating an offspring. After the selection (reproduction) process, the population is maintained by replacing the weaker individual by stronger ones. Here stronger individuals are in terms of better fitness function value. Reproduction makes clones of good strings but does not create new ones. Crossover operator is applied to the mating pool with the hope that it creates a better offspring.
Crossover is a recombination operator that proceeds in three steps:

i). The reproduction operator selects at random a pair of two individual strings for the mating.
ii). A cross site is selected at random along the string length.
iii). Finally, the position values are swapped between the two strings following the cross site.

As a preprocessing step for the possible application of crossover, Cap-p-Med-GA computes two exchange vectors, one for each parent, as follows. For each gene of parent 1, Cap-p-Med-GA checks whether the allele (facility index) of that gene is also present (in any position) at the genome of parent 2. If not, that facility index is copied to the exchange vector of parent 1.

This means that facility index may be transferred to parent 2 as a result of crossover, since this transfer would not create any duplicate facility indexes in parent 2's genotype. The same procedure is performed for each facility index in the genotype of parent 2. For instance, let the two parents be the facility-index vectors [1, 2, 3, 4, 5] and [2, 5, 9, 10, 12]. Their respective exchange vectors are: vp1 = [1, 3, 4] and vp2 = [2,9, 10].

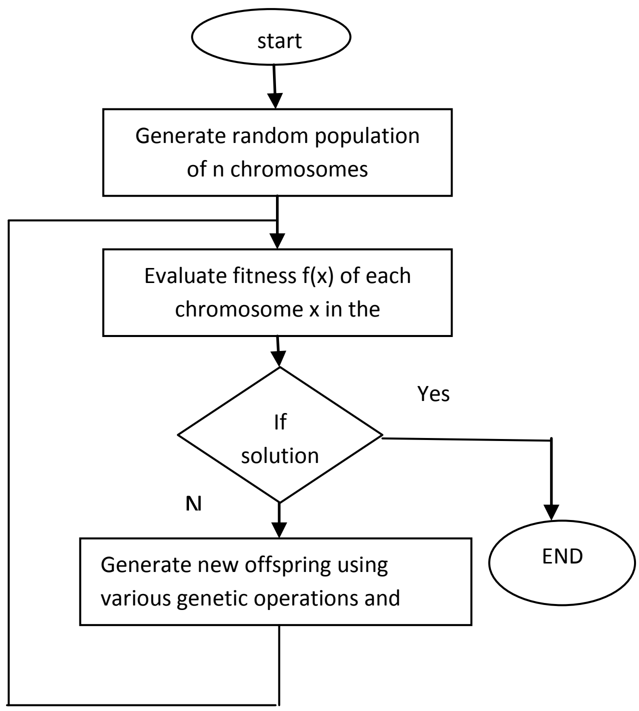

Fig 1. Flow Diagram for Genetic Algorithm

Once the facility indexes that can be exchanged have been identified, the crossover operator can be applied using evaluation followed by replacement that replaces some individuals from old population and adding new better individuals.
The basic Genetic Algorithm stops when the population

converges toward the optimal solution. It can be better described by the flow diagram shown in Fig. 1.

5. COMPUTATIONAL RESULTS

The discussed Genetic Algorithm has been implemented. The data for implementation was selection 26 locations out of 43 candidate locations. Therefore, there are 26C43 = 421,171,648,758 (roughly 421 billion) candidate solutions (Table 1). The results obtained have been shown in following table 1.

Table -1: Computational Results using GA

No. of eval. solutions	24,300
Run time	01:43:21
Avg distance	2.40Km
Total distance	47,313Km
Percent near facility	79%

6. CONCLUSIONS AND FUTURE WORK

Here in this paper we have discussed various location problems and the numerous techniques that exists to solve the location problems. We have shown that how Genetic algorithm can be used to solve a discrete p-median location problem. and have applied it to a real-world problem with a quite large search space, containing roughly 421 billion (4, 21 x 1011) candidate solutions. Our GA uses an individual representation and genetic operators developed specifically for the p-median problem. Capacitated p-median problem (CPMP) is an important deviation of facility location problem in which p capacitated medians are economically selected to serve a set of demand vertices so that the total assigned demand to each of the candidate medians must not exceed its capacity.

REFERENCES

[1] BOOKER, L. B., “Improving Search in Genetic Algorithms”, In: Genetic Algorithms and Simulated Annealing (Edited by L. Davis), 61-73, Morgan Kauffmann, Los Altos, CA, 1987.
[2] CORREA, Elon S., “Algoritmos Geneticos e Busca Tabu Aplicados ao Problema das P-Medianas”, (In Portuguese) M.Sc. Mestrado em Metodos Numericos em Engenharia. Universidade Federal do Parana (UFPR). Curitiba, Brazil 2000.
[3] DIBBLE, C., DENSHAM, P. J., “Generating Interesting Alternatives in GIS and SDSS Using Genetic Algorithms”, GIS/LIS symposium, University of Nebraska, Lincoln, 1993.
[4] ERKUT, Erhan; BOZKAYA, Burçin; ZHANG, Jianjun., “An Effective Genetic Algorithm for the p-median Problem”, (In press) 2001.
[5] P. C. Gilmore and R. E. Gomory, “A linear programming approach to the cutting stock problem,” Operations research, vol. 9, pp. 849-859, Nov-Dec. 1961.
[6] R. E. Kuenne and R. M. Soland, “Exact and approximate solutions to the multisource weber problem”, Mathematical Programming, vol. 3, pp. 193-209, Dec. 1972.
[7] T. Murray and R. L. Church, “Applying simulated annealing to location-planning models,” Journal of Heuristics, vol. 2, pp. 31-53, June 1996.
[8] J. Brimberg and N. Mladenovic, “Solving the continuous location-allocation problem with tabu search,” Studies in Locational Annals, vol. 8, pp. 23-32, Jan. 1996.
[9] X. Li and A. G. Yeh, “Integration of genetic algorithms andgis for optimal location search,” International Journal of Geographical Information Science, vol. 19, pp. 581-601, May.2005.
[10] D. E. Goldberg, “Genetic algorithms in search optimization and machine learning”, Menlo Park, CA: Addison-Wesley,1989.