21st International Congress on Modelling and Simulation, Gold Coast, Australia, 29 Nov to 4 Dec 2015
www.mssanz.org.au/modsim2015

Effective Method for Locating Facilities

M. Dzator[a] and J.Dzator[b]

[a] ALSU, Central Queensland University, Mackay; [b] The University of Newcastle, NSW
Email: m.dzator@cqu.edu.au

Abstract: There has been an increasing interest in the problem of effective facility location over the past five decades. The location of these important facilities arises in the service and manufacturing industries. The fundamental questions that arise concern the number and location of facilities such as: schools; hospitals; ambulances; warehouses; factories; department stores; police stations; waste material dumps; fire stations; needed to achieve a prescribed level of service and output. The main concern of many location problems is to place facilities to optimize some spatially dependent objectives such as: minimize average travel time or distance between demand points and servers; minimize a cost function of travel or response time. These optimization problems are complicated with the need to meet a number of specified constraints that relate to safety, demand, available resources, level of service and time. Indeed, the optimization problems that arise in practice are computationally difficult (NP hard) to solve by exact methods.

An important problem is the ***p*-median problem** which is to find the location of p-facilities so as to minimize the average weighted distance or time between demand points and service centers. Many heuristic algorithms have been proposed for this problem due to the difficulty in obtaining solutions by the exact methods. We discussed below a reduction concept applied to p-median problem as follows.

Consider a weighted p-median problem with a distance matrix given as $D = (d_{ij})$. Note that each row (column) of D is associated with a demand (facility) location. We say that **column *k* dominates column *l*** if $d_{ik} \le d_{il}$ for all $i \ne k$. We use the term **strongly dominates** in the case of strict inequalities. Observe that locating a facility at a dominated location l would provide no advantage to locating a facility at k except possibly in serving the demands of customers in location l. Further, strongly dominated columns would only be used for 'self-serve'. Consequently, dominated column can be dropped to generate a feasible solution and the location can later be considered as a possible 'self-service' facility.

We extend the concept of dominance somewhat further as follows. We say **columns *k* and *l* dominate column *j*** if $d_{ij} \ge \min\{d_{ik}, d_{il}\}$ for all $i \ne j$. In this case there is no advantage in using location j (except for serving customers in location j) when locations k and l are used. So again we can drop the dominated column j if columns k and l are used. The term strongly is used as before.

We further extend this concept of dominance as follows. We say that **column *k* partially dominates column *l*** if $d_{ik} \le d_{il}$ for at least half or more of the entries for which $i \ne k$. Similarly, we say **columns *k* and *l* partially dominate column *j*** if $d_{ij} \ge \min\{d_{ik}, d_{il}\}$ for at least half or more of the entries for which $i \ne j$. Partially dominated columns correspond to nodes which may be assigned 'self-serve' facilities in the original and the reduced matrix.

In this paper, we developed a new greedy algorithm based on a concept known as dominance to obtain solutions for the p-median problem. This concept reduces the number of columns of a distance matrix by considering potential facilities that are near and those that are far from the population or demand. We illustrate our ideas and the algorithm with an example. We further applied the new algorithm to effectively locate additional ambulance stations in the Central and South East metropolitan areas of Perth to complement the existing ones. We also compare the performance of our new Greedy Reduction Algorithm (*GRA*) with the existing greedy algorithm of the p-median problem.

Keywords: *Facilities, reduction, greedy algorithm, dominance*

1. INTRODUCTION

Facility location problems are traditionally investigated with the assumption that all customers are to be provided services. In particular, for the location of emergency facilities all customers should be served within a specified time. However, some customers can be located much further away from a service facility than others. These customers (sometimes classified as outliers) can have a strong influence over the value of the final solution.

Problem size reduction is a widely used strategy in addressing large and computationally difficult optimization problems. However, the literature on the application of a reduction technique on the *p*-median problem is limited. For the few studies (Avella and Sforza; 1999, Rosing *et al.*; 1979, Church; 2003) on the reduction methods for the *p*-median problem, the original formulation of the *p*-median is modified before solving it as an optimization problem.

For the problem of size n the original *p*-median model has n^2 variables. In a study by Rosing *et al.* (1979) with the aim of lowering the computational time required to obtain a solution to the *p*-median problem, they considered the variables of the *p*-median model. Rosing *et al.* (1979) modified the original formulations by reducing the number of variables from n^2 to $n\ (n\text{-}p+1)$ where p is the number of facilities. Therefore the variable reduction for a problem of size n with p facilities is $n^2 - n\ (n\text{-}p+1) = n\ (p\text{-}1)$. This variable reduction is based on eliminating variables that are associated with facilities not assigned to any demand (customer). If one facility out of p facilities is assigned to a customer the remaining p-1 facilities are not assigned to that demand (customer). Therefore, for n customers, n (p-1) variables are eliminated. For example, a problem with 50 nodes and 5 facilities yields a reduction in the number of 50(5-1) = 200. The new formulation was used to solve problems of size 49, 75 and 100 and the time for solving them ranged from 0.92 minutes to 41.73 minutes.

Church (2003) introduced another concept in which demands (customers) which are served by the same facility are represented by only one variable. For example if a facility serves two demands (customers) then only one variable is used in the model formulation instead of two. This new concept depends on the proximity of demands from the selected facility. If, for instance, two demands (customers) have the same distance or proximity to a facility, then if the facility serves one of the customers it must serve the other one also. Church (2003) incorporates this procedure into the Rosing *et al.* formulations to reduce the number of variables further. We note from the study that the new formulation is able to reduce the number of variables by up to 60% compared to the formulation of Rosing *et al.* (1979). Therefore the number of variables in the Church (2003) formulation is significantly smaller than that of Rosing *et al.* (1979) formulation leading to a significant reduction in the computational time for solving the *p*-median problem. This new formulation was applied (using an integer programming package) to problems with 79, 150 and 372 nodes requiring the location of 5, 5 and 10 facilities, respectively. The computational times for solving the problem ranged from 0.01 seconds to 14.8 seconds.

Caccetta and Dzator (2001) introduced a reduction technique to solve the *p*-median problem which considers facilities which are far away from demand as dominated compared to the ones that are close. The idea of using the dominated facilities in a 'self-serve' manner was discussed by outlining the advantages. An example was used to confirm these advantages. This method was further enhanced by Caccetta and Dzator (2005) locate emergency facilities. Dzator and Dzator (2013) developed a new reduction method to locate ambulance stations.

Perth Metropolitan Area

Perth, the metropolitan capital of Western Australia covers approximately 5000 square kilometers, extends 70 kilometers along the coast and had an estimated 2.30 million residents in 2010. The Perth metropolitan area is divided into five major statistical divisions, namely Central Metropolitan, East Metropolitan, North Metropolitan, South East Metropolitan and South West Metropolitan. We consider the Central Metropolitan area and the South East Metropolitan area for this study. In the Central Metropolitan area which includes the city center we currently have three ambulance stations to serve the community (25 suburbs). In the South East Metropolitan area we currently have seven ambulance stations to serve the 26 suburbs.

Data

All location-allocation heuristics need information about the distance or travel cost between the demand locations and the service or candidate locations. For this study the distance matrix from node to node for the study is determined by distance data developed by a company by the name travelmate. This distance is the road network among the various suburbs in the Perth metropolitan area. Hence the distance values are the shortest road travel distances (equivalent to distance on a road map) between the origin and the destination.

We weighted the distance by taking into consideration only the population of the origin suburb since we can also weigh by considering the nature of the road. The fact is that the better the road network the easier it is for a vehicle to move from one suburb to another. The weighted distance is thus the product of the weight assigned to the origin suburb and the distance between the origin suburb and the destination suburb. This weight is proportional to the population of each suburb. We note that the larger the population the larger the weight. This weighted distance is calculated for each of the twenty-six suburbs in South East Metropolitan region and the twenty-five suburbs in Central Metropolitan region. This calculation is done by noting the distance from a suburb to all other suburb and each value is multiplied by the weight of the origin suburb. This is repeated for every suburb and values are recorded as a 26×26 and 25×25 matrix representing South East Metropolitan and Central Metropolitan area respectively. We note from these calculations that the weighted distance of twenty-six suburbs of South East Metropolitan region ranges from about 4 to 1009 kilometers. In the case of Central Metropolitan region the minimum value is 1 kilometer while the maximum is 171 kilometer resulting in a range of 1 to 171 kilometers.

We present the *p*-median problem in the next section followed the greedy algorithm in Section 3. The algorithm is presented in Section 4. The application of the new algorithm is presented in Section 5 and conclusion in Section 6.

2. THE *P*-MEDIAN PROBLEM

The objective of the *p*-median problem is to find the locations of *p* facilities to minimize the demand weighted total distance (total cost) between each demand node and the nearest facility. For the *p*-median problem the cost of serving demands at node *i* is the product of the demand at node *i* and the distance between demand node *i* and the nearest facility to node *i.*

I = the set of demand nodes indexed by i

J = the set of candidate facility locations, indexed by j

p = the number of servers to be deployed or facilities to be located

a_i = the population at the demand node i

d_{ij} = distance between demand node $i \in I$ and candidate sites $j \in J$

$$Y_{ij} = \begin{cases} 1, & \text{if demands at node } i \in I \text{ are assigned to a facility at candidate site } j \in J \\ 0, & \text{otherwise} \end{cases}$$

$$X_j = \begin{cases} 1, & \text{if we locate at candidate site } j \in J \\ 0, & \text{otherwise} \end{cases}$$

Minimize $\sum_i \sum_j a_i d_{ij} Y_{ij}$ (1)

subject to

$$\sum_{j\in J} Y_{ij} = 1\ ,\quad \forall i \in I \tag{2}$$

$$\sum_{j\in J} X_j = p \tag{3}$$

$$Y_{ij} \le X_j\ ,\quad \forall i \in I, j \in J \tag{4}$$

$$Y_{ij} \in \{0,1\},\quad \forall i \in I, j \in J \tag{5}$$

$$X_j \in \{0,1\},\quad \forall j \in J \tag{6}$$

The objective (1) is to minimize the total distance from customers or clients to their nearest facility. Constraint (2) shows that the demand of each customer or client must be met. From constraint (3), the number of facilities to be located is *p*. Constraint (4) shows that customers must be supplied from open facility. Constraints (5) and (6) present the problem as a binary integer programming. The above formulation assumes that the potential facility sites are nodes on the network. Hakimi (1964) showed that allowing facilities to be located on the arcs of the network instead of the nodes would not reduce total travel cost

3. GREEDY ALGORITHM OF THE *P*-MEDIAN PROBLEM

The *p*-median problem, as stated in earlier, is a difficult problem to solve exactly so there are many heuristics which have been developed to solve it. Hence we describe only the greedy heuristic which is Myopic Algorithm. This Algorithm is an example of primary algorithm, applied to solve the *p*-median problem. The algorithm uses all the values of the weighted distance matrix without deleting any extreme values or dominated values before locating facilities. In considering column totals we may not select a location because of a very large cost entry. Since in practice one would never allocate a customer (demand) to very distant facility, it would make sense to eliminate large entries before taking column sums. Our new algorithm will implement this idea.

3 . 1 Myopic Algorithm (MA) for the *p*-Median Problem

The myopic heuristic is a greedy type which works in the following way. Firstly, a facility is located in such a way as to minimize the total cost for all customers. Facilities are then added one by one until *p* is reached. For this heuristic, the location that gives the minimum cost is selected. The main problem with this approach is that once a facility is selected it stays in all subsequent solutions. Consequently, the final solution attained may be far from optimal. This heuristic is specifically known as Greedy-Add since facilities are added one-by-one to attain the required number of facilities. The reverse approach is known as Greedy-Drop which starts with facilities located at all potential facility sites and then eliminate (drop) the facility that has the least impact on the objective function. This method eliminates the facilities one-by-one until the required number of facilities *p* remains.

The outline of the Myopic algorithm as indicated by Daskin (1995) is presented below as follows:-

Step 1: Initialize $k = 0$ and $X_k = \{\ \}$, the empty set.

Step 2: Increase *k*, the counter on the number of facilities located.

Step 3: Compute $Z_j^k = \sum_i h_i d(i, j \cup X_{k-1})$ for each node *j*, which is not in the set X_{k-1}, where h_i is the demand at node *i*.

Step 4: Find the node $j^*(k)$ that minimizes Z_j^k. Add node $j^*(k)$ to the set X_{k-1} to obtain the set X_k.

Step 5: If $k = P$ stop. Go to step 2 if $k < P$.

4. DOMINANCE ALGORITHM FOR LOCATING EFFICIENT FACILITIES

The Greedy Reduction Algorithm (*GRA*) for the *P*-Median Problem is presented as follows

Step 1: Input the distance matrix of size $n \times n$ and let it be denoted by *D*.

Step 2: Identify the columns that are strongly dominated, dominated and partially dominated and eliminate them from the matrix *D*.

Step 3: Call the reduced matrix D' and eliminate partially dominated columns. Continue to step 4 if we are left with one column. Otherwise, for a tie or more than two columns choose the column with the largest total and continue to step 4. (For the two cases, we consider the last remaining column as the solution for the 1-median problem).

Step 4: Calculate the objective values for columns in turn, to correspond to 'self-serve' nodes that are assumed to be a potential second facility. Choose the column yielding to the minimum value as corresponding to the second facility.

Step 5: Use the columns which were dropped as dominated columns in addition to the original dominated columns and allocate the most expensive customer corresponding to these dominated columns in the previous number of facilities already located. That is, the facility is located corresponding to a column which will result in the largest reduction in the objective value, based on the previous number of facilities when the customers are allocated to their current closest facilities.

Step 6: Repeat Step 5 until the required number of facility is attained.

We illustrate this algorithm with an example below.

4.1 Illustration of the Algorithm - Illustrative Example

Step 1: The *p*-Median Problem with distance matrix *D*

$$\begin{bmatrix} 0 & 45 & 86 & 35 & 23 & 63 & 96 & 43 & 48 & 10 \\ 45 & 0 & 49 & 62 & 70 & 87 & 69 & 90 & 89 & 77 \\ 86 & 49 & 0 & 68 & 54 & 33 & 40 & 85 & 99 & 70 \\ 35 & 62 & 68 & 0 & 29 & 81 & 21 & 10 & 97 & 40 \\ 23 & 70 & 54 & 29 & 0 & 74 & 54 & 53 & 30 & 78 \\ 63 & 87 & 33 & 81 & 74 & 0 & 41 & 36 & 32 & 95 \\ 96 & 69 & 40 & 21 & 54 & 41 & 0 & 16 & 37 & 42 \\ 43 & 90 & 85 & 10 & 53 & 36 & 16 & 0 & 78 & 51 \\ 48 & 89 & 99 & 97 & 30 & 32 & 37 & 78 & 0 & 98 \\ 10 & 77 & 70 & 40 & 78 & 95 & 42 & 51 & 98 & 0 \end{bmatrix}$$

Step 2: From the matrix column 1 partially dominates columns 2, 3, 4, 5, 6, 8, 9 and 10. Column 7 dominates columns 1, 2, 3, 5, 6, 8, 9, and 10.

Step 3: We consider column 1 and 7 as the elements of the reduced matrix since they dominated the highest number of columns. Therefore, the reduced matrix is given as:

$$\begin{bmatrix} 0 & 96 \\ 45 & 69 \\ 86 & 40 \\ 35 & 21 \\ 23 & 54 \\ 63 & 41 \\ 96 & 0 \\ 43 & 16 \\ 48 & 37 \\ 10 & 42 \end{bmatrix}$$

Column 7 partially dominates column 1 so we eliminate column 1 and consider column 7 as the solution for the 1-median problem.

Step 4: We calculate the objective values for columns which are supposed to correspond to self-serve facility nodes and we list them in the table below

Self-Serve Facility	2	3	4	5	6	8	9	10
Objective Function Value	296	338	294	282	330	330	298	288

The table above shows that the minimum corresponds to locating the second facility at node 5. Therefore the best location for the next facility after locating the first at node 7 is at node 5, giving a facility set of {5, 7}.

Step 5: The dominated columns are1, 2, 3, 4, 6, 8, 9 and10. Therefore, we allocate the next facility to the most expensive customer corresponding to the dominated columns in the facilities 5 and 7 to give three facilities. This results in the facilities {1,5,7} with an objective function value of 203.

Step 6: We repeat the process for 4 and 5 facilities which results in the following optimal facility sets, with their respective objective values as follows. Four Facilities: {1,5,6,7}, 155; Five Facilities: {1,2,5,6,7}, 110.

We compare the new algorithm with the optimal value and the results are shown in Table 5.5

Table 1: Comparison of *GRA* and Optimal Value

Number of Facilities	*GRA*	Optimal Value
2	{5,7}-282	{1,7}-233
3	{1,5,7}-203	{1,6,8}-169
4	{1,5,6,7}-155	{1,2,6,8}-124
5	{1,2,5,6,7}-110	{1,2,3,6,8}-91

Table 1 shows the result for the comparison of *GRA* and the optimal value. The results show that the optimal value performs better in all cases. While the optimal performs better, the strong point about this algorithm is that after the location of the first facility the subsequent ones are optimally added to the previous facility. This algorithm will therefore be useful in locating new facilities optimally when existing facilities are to be retained.

5. APPLICATION OF THE GREEDY ALGORITHM TO AMBULANCE STATION LOCATION IN PERTH

Currently they are three ambulance stations in the Central metropolitan area and seven in South East metropolitan area. In future, there will be a need for more stations because of a steady population increase. It is clearly more practicable to simply optimally locate new facilities and leave the old ones in place rather than relocate the old ones. We use C++ to obtain the results for the location of additional stations. SITATION software (Daskin; 1995) was used for the optimal value and the results for Myopic Algorithm. The results are presented below.

Table 2: Comparison of Myopic Algorithm with Greedy Reduction Algorithm (*GRA*)

Number of Additional Stations	Myopic Algorithm (*MA*)	Greedy Reduction Algorithm (*GRA*)	$\frac{MA-GRA}{GRA}\times 100$
Central Metropolitan Area			
4	281	304	-7.5
5	238	209	13.8
6	196	174	12.6
7	169	147	14.9
South East Metropolitan Area			
8	485	487	-0.4
9	423	418	1.2
10	365	361	1.1

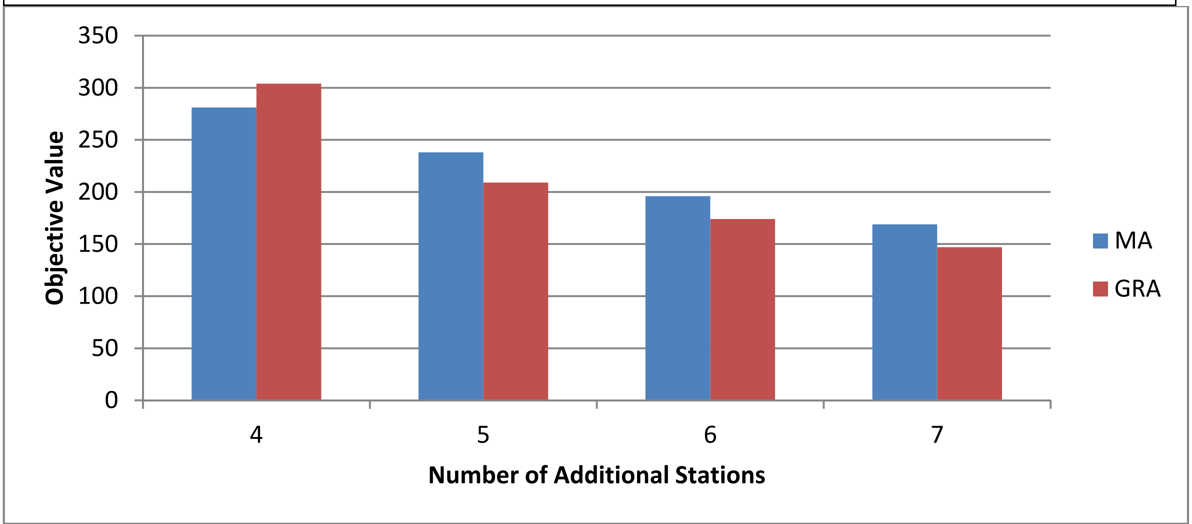

Figure 1: Additional Stations in Central Metropolitan Area using MA and GRA

The results in Table 2 show that *GRA* performs better than the existing *MA* in most cases. For Central Metropolitan area the percentage benefit for using GRA is from 12.6% to 14.9% while for South East Metropolitan area the benefit is -0.4% for locating 8 additional facilities, 1.2% for locating 9 facilities and

1.1% for locating 10 additional facilities. Figure 1 also shows the superiority of using GRA to locate additional ambulance stations in the Central Metropolitan area.

6. CONCLUSION

This paper focuses on the effective way of locating facilities by the use of the *p*-median. We discussed a new reduction rules for the *p*-median problem. In discussing the reduction rules, we present an algorithm based on the principle of dominance as outline for the *p*-median problem.

We applied the new algorithm to locate additional ambulance stations in Perth. The performance of the new greedy algorithm as compared to the existing greedy algorithm of the *p*-median showed that it is more effective.

REFERENCES

Avella, P. and A. Sforza (1999), Logical reduction tests for the *p*-median problem, *Annals of Operations Research,* 86, 105-115

Caccetta, L. and M. Dzator (2001), Models for the location of emergency facilities, In *Proceedings Modsim 2001.*

Caccetta, L. and M. Dzator (2005), Heuristics methods for locating emergency facilities, In *Proceedings Modsim 2001.*

Church, R.L. (2003), COBRA: A new formulation of the classic *p*-median location problem, *Annals of Operations Research,* 122, 103-120.

Church, R.L. and M.E. Meadows (1979), Location modeling utilizing maximum distance criteria, *Geographical Analysis,* 11, 358-373

Church, R.L. and R.A Gerrard (2003), The multi-level location set covering model**,** *Geographical Analysis,* 35, 249-256.

Colome, R., H.R. Lourenco, and D. Serra (2003), A new chance-constrained maximum capture location problem, *Annals of Operations Research*, 122, 121-139.

Daskin, M.S. (1995), *Network and Discrete Location: Models, Algorithms and Applications,* John Wiley and Sons, Inc., 498 pp.,New York.

Downs, B.T. and J.D. Camm (1996), An exact algorithm for the maximal covering location problem, *Naval Research Logistics,* 43, 435-461.

Dzator, M. (2008). *Facility Location in Cities: The Optimal Location of Emergency Units within Cities,* VDM Verlag.

Dzator, M., and J. Dzator (2013). An Effective Heuristic for the *P*-median problem with Application to Ambulance Location. *Opsearch* 50 (1), 60-74.

Hakimi, S.L. (1964), Optimisation locations of switching centres and the absolute centres and medians of a graph, *Operations Research,* 12, 450-459.

Rosing, K.E., C.S. ReVelle and H. Rosing-Vogelaar (1979), The *p*-median and its linear programming relaxation: An approach to large problems, *Journal of the Operational Research Society,* 30, 815-823.

Toregas, C. and C. ReVelle (1973), Binary logic solutions to a class of location problems, *Geographical Analysis,* 5, 145-155.

.