A Fuzzy Set Covering-Clustering Algorithm for facility location problem

R. Sahraeian [1], M. S. Kazemi [2]

[1] Department of Industrial Engineering, Shahed University, Tehran, Iran

[2] Department of Industrial Engineering, Shahid Bahonar University of Kerman, Kerman, Iran

(sahraeian@shahed.ac.ir, sadeqkazemi@gmail.com)

Abstract – **mathematical models and solution algorithms which address the problem of locating facilities and allocating customers varies widely in terms of basic assumptions, mathematical complexity and computational performance. In this paper, we are concerned with a problem of locating the number of facilities among a finite number of sites such that all existing sites (customers) are covered by at least one facility. The problem was modeled and solved in three stages. In the first stage, an improved fuzzy set covering solution was proposed to determine the minimum number of facilities. In the second stage, the well known k-means clustering algorithm was applied for demand classification into groups. In the third stage, the assignment model was used to locate facilities in each cluster. Using extensive simulation studies, we also show that the proposed approach performs considerably well in all considered conditions in comparison to classic covering methods.**

Keywords - **Assignment Model, Facility Location, Fuzzy Set theory, k-means, Location Set Covering Problem**

I. INTRODUCTION

The first attempt to find an appropriate place for locating facilities was applied by Hakimi [1]. Since then, facility location problems have become an interesting subject for researchers and have been applied in variety of scientific fields [2]. Facility location problems can be categorized due to their objectives. Optimal placement of facilities in order to minimize transportation costs, locating facilities in the face of emergencies and finding the location of service facilities to maximize the number of customers are the examples. Covering problems are a particular form of facility location problems which aim to determine the minimum number of facilities required to cover the given number of customers. The Location Set Covering Problem (LSCP) is a main model within covering problems which has several important applications, including supply chain management, where a set of customers has to be covered by a minimum-cost set of supplier, locating facilities or services that are related to the public sector, such as emergency services school systems and postal facilities ,etc [3, 4].

One should note that facility location models are subjected to imprecise parameters. Therefore, researchers have considered fuzzy logic to address the problem appropriately. By application of fuzzy set theory in covering problems the flexibility and robustness of covering techniques are improved and we are able to handle the uncertainty and imprecision in parameters such as distance, capacity, cost and demand [5, 6].

As the LSCP is NP-hard to solve optimally, many heuristic methods have been developed and applied to provide the solution [7]. Several approximation algorithms have been developed for covering problems and many of its variants. Daskin (1983) have developed a heuristic to obtain better results for maximum expected covering location model which is closely related to LSCP. Haddadi [8] proposed a simple Lagrangian heuristic for the LSCP. However, it turned out to be efficient only for the low-density set covering problems.

Balasubramanian et al [9] have developed a set covering heuristic to determine the economical number of manufacturing cells and cell arrangements. Hwang (2004) studied a special case of the stochastic LSCP for both ameliorating and deteriorating item to determine minimum number of storage facility among a discrete set of location sites, so that the probability of each customer being covered is not less than a critical value.

Bell et al [10] developed a two step approach for locating aircraft alert sites for homeland defense. In the first step the minimum number of sites was identified using the LSCP and then the result is improved by finding the minimum aggregate network distance or p-median solution from the alternate optimal solutions to the LSCP.

In this paper, a three stages solution for LSCP is developed by introducing a novel fuzzy set covering problem, clustering algorithm and assignment model. At the first stage, we try to identify the minimum number of service sites by solving Fuzzy Location Set Covering Problem (FLSCP). Next, k-means algorithm is used to specify demand clusters and initial location for facilities and finally the appropriate locations for facilities are found by searching among candidate sites using assignment model.

This paper is organized as follows. In Section II, research methodology will be explained in three phases including application of fuzzy set theory in LSCP, demand clustering and assignment model as well as assumptions and notations in the model. The results of implementing the proposed methodology are summarized and discussed in Section III.

II. METHODOLOGY

Utilizing the FLSCP, the minimum number of facilities, k, needed for a complete coverage of demands is determined. In other words, FLSCP solution is the number of demand clusters. Demand points are distributed across k clusters using the k-means algorithm. Since the facilities location and demands individually are a set of discrete locations, the assignment model is used for local search and specifying the appropriate location for facilities in each cluster. The overall research process will be described in details in the following sections.

A. Location set covering problem formulation

The classic LSCP was developed by Toregas et al [4]. Since then; it has been widely used to model the variety of problems in different areas such as supply chain, assembly line balancing and facility routing. In this research, the classical LSCP is applied in order to determine the minimum number of facilities. The following notation is adopted from [11]:

I: set of demand points.

J: set of candidate sites or services.

i: the index of demand points.

j: the index of candidate sites or services.

d_{ij}: a distance between a demand points i and a candidate site j.

s_{ij}: a maximum allowable critical distance, in which a facility at site j can cover demand point i.

a_{ij}: a binary variable which indicate the feasibility of covering demand point i by facility at site j.

$$a_{ij} = \begin{cases} 1 & d_{ij} \le s_{ij} \\ 0 & d_{ij} > s_{ij} \end{cases} \tag{1}$$

In LSCP, the distance between any demand and service point is calculated according to Euclidean distance. LSCP formulation is as follow:

$$\min \sum_{j=1}^{n} x_j \tag{2}$$

$$\text{s.t.} \sum_{j=1}^{n} a_{ij} x_j \ge 1 \qquad i \in I \tag{3}$$

$$x_j \in \{0,1\} \qquad j \in J \tag{4}$$

The objective function (2) minimizes the number of facilities. Equation constraint (3) introduces constraints to the problem that ensure all demand points are supplied by at least one facility and (4) is the integrality constraint.

B. Fuzzy set theory application in Set covering problem modeling

Since the introduction of fuzzy set theory by Zadeh [12], it has been widely use to model different kinds of problems with vague or imprecision data. This theory tries to model uncertain and vague boundaries for subset of a universe. A fuzzy set is a set on a universe X which has an element in the form of

$$A = \{ (x, \mu_A(x)) \mid x \in X \}. \tag{5}$$

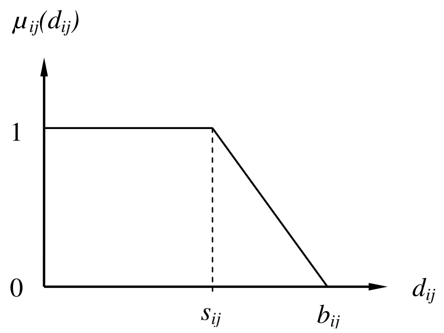

Fig. 1. The membership function of the fuzzy coefficient.

For each element x belongs to a fuzzy set A, a membership function $\mu_A(x)$ will define which takes value between 0 and 1.

In the conventional LSCP, for candidate site j, demand point i should be within its critical distance, s_{ij}, to be covered. Even smallest violation from critical distance completely leaves out the possibility of covering the demand point i. To overcome this limitation, the degree of coverage is considered as a fuzzy value. Demand points within critical distance are considered to be covered completely. Demands within critical distance and the extended backup distance, $s_{ij}+b_{ij}$, are covered with a specific degree, $\mu(d_{ij})$. It is clear that the demands beyond this distance, $s_{ij}+b_{ij}$, can not be covered. Each a_{ij} represents the covering degree of the demand point i by the candidate site j. These subsets are associated with membership function $\mu(d_{ij})$. As the distance between demand point i and service point j increases, the degree of covering decreases.

$$\tilde{a}_{ij} = \{ (d_{ij}, \mu_{a_{ij}}(d_{ij})) \mid d_{ij} \ge 0) \} \tag{6}$$

The following right hand membership function demonstrated in Fig .1 is assumed for coefficients:

$$\mu_{a_{ij}}(d_{ij}) = \begin{cases} 1 & d_{ij} \le s_{ij} \\ \dfrac{s_{ij}+b_{ij}-d_{ij}}{b_{ij}} & s_{ij} < d_{ij} \le s_{ij}+b_{ij} \\ 0 & d_{ij} > s_{ij}+b_{ij} \end{cases} \tag{7}$$

FLSCP formulation is given as follow:

$$\min \sum_{j=1}^{n} x_j \tag{8}$$

$$\text{s.t.} \sum_{j=1}^{n} \mu_{a_{ij}}(d_{ij}) x_j \ge 1 \qquad i \in I \tag{9}$$

$$x_j \in \{0,1\} \qquad j \in J \tag{10}$$

The objective function (8) minimizes the number of facilities needed to cover demands. Constraints with fuzzy coefficient (9) guarantees covering every demand points by at least one facility and (10) are integrality constraint.

C. Demand points clustering

The objective of cluster analysis is the classification of objects according to similarities among them and organizing of data into groups. The main capability of clustering is to detect the underlying structure in data, not only for classification and pattern recognition, but for model reduction and optimization [13].

There are many clustering techniques for classification of data into different groups. These techniques are different according to their input data and the results which they yield. Regardless of these differences, clustering algorithm takes set of samples and similarity criterion as an input data and return groups of data as output.

In this study, we used k-means algorithm which is one of the most commonly used clustering techniques [14]. It starts with random initial cluster centers and in several runs; points are assigned to clusters in such a way that the sum squared distance for each cluster is minimized. Input data for k-means algorithm can be multi dimensional arrays.

In facility location problems, it is common to consider two dimension points. K-means choose initial clusters randomly among demands and classify demands into k cluster such that minimize the sum squared distance in each one. As k-means is significantly sensitive to the initial randomly selected cluster centers, the candidate sites which obtained by solving FLSCP are considered as the initial centers in the algorithm. K-means finds the location to place each facility and define demand clusters.

It should be noted that in the process of assigning demands to clusters, the cluster centers array may change and not located in candidate site any more. To confine cluster centers to candidate site the following mathematical model is used.

D. Assignment model

The assignment model is a special case of transportation problem. Its aim is to find the optimized matching between a number of agents and a number of tasks. In this research, after clustering of the demand points, if the cluster centroid coordinate lies on the candidate sites, solution has been found, otherwise the nearest candidate site to cluster center is chosen to place facility considering its critical distance. So we decided to use an assignment model to compare a cluster center with all qualified candidate sites. The variable x_{kj} represents the assignment of cluster center k to candidate site j.

K: set of cluster centers.

J: set of candidate sites or services.

k: the index of cluster centers.

j: the index of candidate sites or services.

d_{kj}: a distance between a cluster center k and a candidate site j.

The linear assignment model is expressed as follow:

$$Minimize \sum_{k=1}^{n}\sum_{j=1}^{n} d_{ij}x_{ij} \tag{11}$$

$$\text{s.t.} \sum_{j=1}^{n} x_{kj}=1 \qquad k\in K \tag{12}$$

$$\sum_{k=1}^{n} x_{kj}=1 \qquad j\in J \tag{13}$$

$$x_{kj}=0,1 \quad k\in K;\, j\in J \tag{14}$$

The objective function (11) minimizes the total distance of assignment. Constraints (12) guarantee that every cluster center is assigned to exactly one candidate site. Constraints (13) require that every candidate site is assigned exactly to one cluster center and constraints (14) are integrality constraint.

The following assumption makes it possible to use a linear assignment model for candidate matching in our proposed model:

1) Any qualified candidate site can be assigned to any cluster center.

2) It is required to assign exactly one candidate site to each cluster center in such a way that the total distance of the assignment is minimized.

III. RESULTS

In order to evaluate the performance of proposed methodology, extensive simulation studies are conducted. A list of demand points and candidate sites was generated normally distributed over a 100×100 network. Both demands and candidate sites were considered to follow normal distribution with mean 50 and variance 15. Critical distance for every candidate site was considered to be 20 with the backup distance 5.

Considering fuzzy coefficients instead of crisp values increase the flexibility of model. Simulation result for a set of 10 demand points and 30 candidate sites shows that by increasing backup distance, in some cases, the number of facilities needed to cover all the demands decreases (see Table I).

TABLE I
Number of facilities needed for covering the demands

		Degree of membership=α									
		1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
Number of facilities	1	16	16	16	16	16	16	16	16	16	16
	2	364	366	367	368	370	373	373	379	386	394
	3	501	501	500	503	503	501	502	501	499	496
	4	107	106	106	103	101	100	99	94	90	85
	5	12	11	11	10	10	10	10	10	9	9

TABLE II
Average distance for different number of candidate sites and demand points for the classic LSCP and the proposed algorithm

Average distance		Number of candidate sites									
		10		15		20		25		30	
		LSCP	proposed	LSCP	proposed	LSCP	proposed	LSCP	proposed	LSCP	proposed
Number of demand points	5	12.1064	11.5545	13.1777	11.4714	12.7951	11.1868	13.3186	10.9320	13.8932	11.5567
	6	12.0876	11.1559	12.9430	11.3716	13.6513	11.7501	12.4291	10.8195	13.3367	11.6053
	7	12.4605	11.3130	13.2029	11.3072	12.8001	11.4195	12.2010	11.1810	13.6450	11.1362
	8	12.1689	11.1783	12.3746	11.4429	12.5994	10.8726	12.6750	10.6016	13.2576	11.5105
	9	12.1326	11.3461	12.0181	11.1523	12.6135	11.1811	12.9719	11.3256	13.1305	11.5433
	10	12.1864	11.2373	12.3478	11.3765	12.2766	10.7542	12.1519	11.1372	13.2233	11.4709
	11	12.0009	11.0547	12.5368	10.9772	11.9750	10.9540	12.8199	11.4782	12.7451	10.9084
	12	12.2224	11.3982	11.7794	10.7707	12.0258	10.6807	12.8095	11.2805	12.8614	11.3573
	13	11.4305	10.7053	12.6402	11.1023	12.2213	11.0089	12.4231	10.7072	12.9306	11.1734
	14	11.4239	10.5885	12.0985	11.0487	11.5263	10.7733	12.6126	11.2198	12.9365	11.2701
	15	10.8147	10.3817	11.7062	10.7237	11.7234	10.2991	12.8343	11.2797	11.9373	10.4939

To show the effectiveness of the proposed model in selecting appropriate candidate site, after each assignment, the average distance between selected candidate sites and demand points were calculated. This measure was calculated once for LSCP solutions and another time after applying the proposed algorithm. The average distance between demand points and facilities for 1000 test problem under previous assumptions were calculated (see Table II).

The algorithm has been evaluated by different combination of demand points and candidate sites. The results show the decrease in the average distance between the facilities and demand points within its coverage after applying the algorithm.

V. CONCLUSION

This paper has proposed a three stage solution for the LSCP. It has shown how to employ fuzzy approach to find a satisfying solution for the LSCP. Simulation results verify our assumption that considering fuzzy critical distances may lead to lower locating cost. Also we apply k-means clustering algorithm to cluster demand points according to their distance into several groups. This makes it possible to use an assignment model to match service points and demands. Application of the proposed algorithm has led to decrease in the distance between demand points and candidate sites.

The model proposed in this paper considers coefficients as fuzzy values; however for future study considering different fuzzy variables or membership functions will extend the model to formulate different facility location problems. Incorporating different clustering algorithms in demand classification will also provide additional challenges in future research.

REFERENCES

[1] S. L. Hakimi, "Optimum locations of switching centers and the absolute centers and medians of a graph," *OperationsResearch,* vol. 12, pp. 450-9, 1964.

[2] M. Daskin, *Network and Discrete Location: Models, Algorithms, and Applications*. NewYork: John Wiley and Sons, 1995.

[3] Z. Drezner and H. W. Hamacher, *Facility Location: applications and theory*. Berlin: Springer, 2004.

[4] C. Toregas, R. Swain, C. S. ReVelle, and L. Bergman, "The location of emergency service facilities," *Operations Research,* vol. 19, pp. 63–73., 1971.

[5] M. J. Hwang, C. I. Chiang, and Y. H. Liu, "An Alternative Formulation for Certain Fuzzy Set-Covering Problems," *Mathematical and Computer Modelling,* vol. 42, pp. 363-365, 2005.

[6] Y. J. Lai and C. L. Hwang *Fuzzy Mathematical Programming: Methods and Applications* Springer-Verlag New York, Inc. Secaucus, NJ, USA, 1995

[7] M. Garey and D. Johnson, *Computers and Intractability:A Guide to the Theory of NP-Completeness*. New York: W.H.Free man and Co, 1979.

[8] S. Haddadi, "Simple Lagrangian heuristic for the set covering problem," *European Journal of Operational Research,* vol. 97, pp. 200-204, 1997.

[9] K. N. Balasubramanian and R. Panneerselvam, "Covering technique based algorithm for machine grouping to form manufacturing cells," *International Journal of Production Research,* vol. 31, pp. 1479-1504, 1993.

[10] J. E. Bell, S. E. Griffis, W. C. Cunningham, and J. A. Eberlan, "Location optimization of strategic alert sites for homeland defense," *Omega,* vol. 39, pp. 151–158, 2011.
[11] C. S. ReVelle and J. C. Williams, "Reserve design and facility siting," in *Facility Location: Applications and Theory*, Z. Drezner and H. H. W., Eds. New York: Springer-Verlag, 2002, pp. 307-328.
[12] L. A. Zadeh, "Fuzzy sets," *Information and Control,* vol. 8, pp. 338-353, 1965.
[13] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data Clustering: A Review," 1999.
[14] J. Mcqueen, "Some methods for classification and analysis of multivariate observations," in *In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1967, pp. 281-297.