A Revised Weighted Fuzzy C-Means and Center of Gravity Algorithm for Probabilistic Demand and Customer Positions

Engin Bayturk^{1(✉)}, Sakir Esnaf², and Tarik Kucukdeniz²

¹ American College of the Middle East,
Dasman 220, 15453 Kuwait City, Kuwait
Engin.Bayturk@acm.edu.kw

² Istanbul Cerrahpasa University, Avcilar, Istanbul, Turkey
{sesnaf, tkdeniz}@istanbul.edu.tr

Abstract. This study proposes four probabilistic fuzzy c-means algorithms which include a probabilistic fuzzy c-means algorithm (Probabilistic FCM), a probabilistic revised weighted fuzzy c-means algorithm (Probabilistic RWFCM) and hybrid algorithms that combine these algorithms with the center of gravity methods for the un-capacitated planar multi-facility location problem when customer positions and customer demands are probabilistic with predetermined service level. The performance of proposed algorithms was tested with 13 data sets and compared with each other. Experimental results indicate that Probabilistic RWFCM-COG algorithm performs better than other compared algorithms in terms of cost minimization.

Keywords: Probabilistic demand and position · Probabilistic fuzzy c-means · Center of gravity · Multi-facility location problem

1 Introduction

One of the strategic decisions of a company is facility location selection. The correct facility location selection will lead to less deviation in the cost estimates taken into account in the investment decisions made over the years. In addition to the costs, order delivery from suppliers to facilities and from facilities to customers at the desired time will be directly affected by these decisions. It is one of the most frequently studied subjects on the problem of facility location due to the important role outlined above.

The facility location problem was first proposed by [3]. This study has gained much interest in the academic field, and then Kuenne and Soland [11], proposed the branch and bound algorithm method for this particular problem. Murtagh and Niwattisyawong [17] have developed a model on the problem of determination of facility location. Megiddo and Supowit [14] conducted a study that proved that the facility location problems were NP-Hard problems. Murray and Church's [16] simulated annealing, Ohlemuller's [18] tabu search, Hansen et al.'s [9] p-median studies can be listed as some of these studies. Hansen et al. [9] discuss the deterministic facility location

problems. In addition to these methods, solutions were suggested with a fuzzy approach to facility location problems. Yang et al. [25] used the chance constraint programming method for the capacitated facility location where the demand of customers are fuzzy. Zhou and Liu [27] discussed three different facility location models with fuzzy demand. Wen and Iwamura [21] developed the fuzzy cost model with the Hurwicz criterion. Peidroa et al. [19] developed fuzzy supply chain models. Gao [7] developed the uncertain shortest path problem. A facility location model with random fuzzy demand proposed by Wen and Kang [22], while Wang and Watada [20] suggested a model in a fuzzy random environment. Gao [8] developed a solution for the uncertain single facility location problem.

Wen et al. [23] developed a mathematical model that determines the best possible plant location with the mixed model in which the simplex method and genetic algorithm are used together. Customer demands are determined by expert judgment. Wen et al. [24] proposed a mathematical model for determining the capacity of a facility in a problem where customer demands are not certain. The proposed model has a mixed model with genetic algorithm, the simplex algorithm and Monte Carlo simulation. Uncertain customer demands are determined by expert judgment. Markovic et al. [13] discussed the problem of choosing a stochastic facility location where customer demands are independent. Diabat et al. [4] developed a mathematical model in which demand is not known, and lead time was determined jointly by plant location and stock decisions.

In literature, some researchers also proposed a solution for the facility location problem where the customers' locations and demands are probabilistic. Zhou and Liu [26] solved the facility location problem, where customer demands have random parameters. Altinel et al. [1] developed solution for the capacitated facility location problem where positions and demands of the customer are probabilistic. Mousavi et al. [15] suggested a solution for the same problem using chance-constrained programming and genetic algorithms.

Facility locations' problems are NP-hard problems [14]. Continuous location-allocation problem or Multisource Weber problem focuses on to determine locations of c facilities to serve n customers with minimum total cost. In this particular problem, customer positions are fixed. Revised Weighted Fuzzy C-Means (RWFCM) is first discussed by [5]. This algorithm was shown as an improved version of Fuzzy C-Means (FCM), which was first developed by [2].

Esnaf and Kucukdeniz [6] have improved the results by combining FCM and the Center of Gravity (COG) methods. In this method; first, the FCM and the set of demand points are divided into c clusters, after that the COG is calculated for each. These centers are selected as new cluster centers. Kucukdeniz et al. [12] applied convex programming and FCM methods for the capacitated facility problem. In their method; cluster centers are determined using the FCM in the first step within the facility capacity limit, then the centers of the new clusters are determined by convex programming, which is treated as the problem of a single facility location selection. The RWFCM method discussed by Esnaf and Kucukdeniz [5] used a different metric in which the weight values are considered constant. It is calculated by multiplying the Euclidean distance between the two points on the metric plane in the objective function

by the demand quantity of the demand point. This method used the assumption that the demands are fixed in the problem.

The literature review encouraged us to develop a solution with a successful fuzzy clustering approach tried before to this problem.

In this paper, the probabilistic version of the RWFCM and a hybrid version of the RWFCM and Center of Gravity methods are proposed, when customer demands and positions are assumed to be distributed normally. Comparison of the proposed algorithms are made in terms of total transportation cost. The probabilistic version of the RWFCM and its hybrids for the multi-facility location problems have not been studied in the literature yet according to the best of our knowledge.

The purpose of this paper to propose the probabilistic version of the RWFCM algorithm combined with the center of gravity method and how to reach facility locations that provide minimum total transportation costs when customer demands and positions are probabilistic.

The rest of the paper is organized as follows; in the second section, problem definition is provided. Third section, contains explanation of the proposed algorithms. The results of the proposed algorithms using thirteen data sets are discussed in the fourth section. The last section concludes with discussions about paper and future studies.

It should be noted that this hybrid not yet used in this way.

It should be noted that this new hybrid approach presented for the first time and not yet used before in this manner.

2 Problem Definition

Multi-facility location problems (MFLP) focus on determining locations of c facilities to serve n demand points with the minimum total cost when customer locations are fixed. But if the locations and demands of the customers are variable, the MFLP can be reformulated as follows:

$$\min_{c_1, c_2, \dots, c_k} \sum_{j=1}^k \sum_{a_i \in V_j} w_i d(\bar{a}_i, c_j) \quad (1)$$

Where, $\bar{a}_i$ is the average location of customer i in a plane, $i = 1, 2, \dots, n$, w_i is the probabilistic demand of customer i , $w_i > 0$, $i = 1, 2, \dots, n$, c_j is the location of facility j , V_j is the customer cluster that assigned to j^{th} facility, and $d(\bar{a}_i, c_j)$ is the probabilistic distance (Euclidean) between the customer i and facility j .

3 Proposed Algorithms

In this section, Probabilistic FCM, Probabilistic RWFCM and hybridized versions with the Center of Gravity method are discussed. Since demand and customer positions are not fixed, the probabilistic approach is employed to find better facility locations to reduce total costs with predetermined service level.

3.1 Probabilistic RWFCM

Esnaflar and Kucukdeniz [5] developed RWFCM. In RWFCM, weights which are symbolized as w_i in equation applied as customer demand when customer demand is fixed. Customer positions which are coordinates of demand points X, Y are fixed data points in the set a_i , and Euclidean distance is used. Unlike original RWFCM, demand and coordinates in the Probabilistic RWFCM are not fixed values. Considering demand values fit the Normal distribution and the weights for each customer are determined according to the predetermined service level with the Eqs. (5) and (6). Distances between demand points and facilities are calculated as in Eq. (7) and all remaining calculations are the same as in the RWFCM.

Objective function is given as follows:

$$J_p(U, c) = \sum_{i=1}^n \sum_{j=1}^k w_i (u_{ij})^p d(\bar{a}_i, c_j) \quad (2)$$

The optimization problem is to minimize (2) under the constraint:

$$\sum_{j=1}^k u_{ij} = 1, \quad \forall_i \quad (3)$$

Membership values, u_{ij} , are calculated with the following equation:

$$u_{ij} = \frac{1}{\sum_{j=1}^k \left(\frac{\bar{a}_i - c_j}{\bar{a}_i - c_i} \right)^{\frac{2}{p-1}}} \quad (4)$$

Probabilistic RWFCM algorithm is given as follows in (5)–(8):

Step 1: At the beginning of the algorithm, the number of cluster, k , cluster centers of subsets, $\{c_1, c_2, \dots, c_k\}$, coefficient of fuzziness, p , and stopping criterion, $\varepsilon > 0$, are determined.

Step 2: Probabilistic demand of the customers (w_i) are calculated.

$$P\left(z \leq \frac{w_i - \bar{w}_i}{\sigma_{w_i}}\right) = T \quad (5)$$

$$w_i = \bar{w}_i + z\sigma_{w_i} \quad (6)$$

Step 3: Probabilistic Euclidean distance of the customers $d(\bar{a}_i, c_j)$ are calculated as [10]:

$$d(\bar{a}_i, c_j) = \sqrt{\frac{(\bar{a}_{1i} - c_{1j})^2}{s_{1i}} + \frac{(\bar{a}_{2i} - c_{2j})^2}{s_{2i}}} \quad (7)$$

Step 4: Calculate membership values u_{ij} using Eq. (4).

Step 5: Cluster centers are calculated as follows:

$$c_j = \frac{\sum_{i=1}^n w_i u_{ij}^p \bar{a}_i}{\sum_{i=1}^n w_i u_{ij}^p} \quad (8)$$

Step 6: If the difference of the calculation between consecutive cluster centers is bigger than ε , the algorithm continues starting from step 2, else, the algorithm is stopped.

Where P is the probability, $\bar{w}_i$ is the average demand and σ_{w_i} is the standard deviation of demand for the customer i , T is the threshold value for customer service level, $\bar{a}_i$ is the average coordinates of customer i , and s_i is the variance of the coordinates of customer i . The Probabilistic RWFCM becomes to the Probabilistic FCM in the case that all weights are equal. Probabilistic FCM is the first proposed and used as a benchmark algorithm in this study.

3.2 Probabilistic Center of Gravity

Improving location of the cluster centers $(\bar{X}, \bar{Y})$ by Probabilistic Center of Gravity method is as follows in (9)–(12):

$$\bar{X} = \frac{\sum_i w_i \bar{a}_{1i}}{\sum_i w_i} \quad (9)$$

$$\bar{Y} = \frac{\sum_i w_i \bar{a}_{2i}}{\sum_i w_i} \quad (10)$$

$$\bar{X} = \frac{\sum_i w_i \bar{a}_{1i} / d(\bar{a}_i, c_j)}{\sum_i w_i / d(\bar{a}_i, c_j)} \quad (11)$$

$$\bar{Y} = \frac{\sum_i w_i \bar{a}_{2i} / d(\bar{a}_i, c_j)}{\sum_i w_i / d(\bar{a}_i, c_j)} \quad (12)$$

3.3 Probabilistic RWFCM-COG

This method first matches plants and customers using Probabilistic RWFCM, after which optimum plant locations within each created cluster are determined by a special center of gravity approach.

Step 1: Probabilistic customer demands (w_i) are calculated.

Step 2: Probabilistic Euclidean distance between customers and facilities $d(\bar{a}_i, c_j)$ are calculated.

Step 3: Using Probabilistic RWFCM, customers are divided in to k number of clusters (c_k).

Step 3.1: Using Probabilistic RWFCM, k number of cluster center is determined.

Step 3.2: Customers are assigned to closest facilities. Sub clusters are created by assigned customers.

Step 4: For every sub clusters, cluster centers are calculated by Probabilistic Center of Gravity.

Step 5: Membership degrees (u_{ij}) are calculated.

Step 6: If the difference of the calculation between consecutive cluster centers is bigger than ε , algorithm continues starting from step 2, else, algorithm is stopped.

When algorithm stops, final versions of the cluster centers are determined and customers are divided into sub clusters accordingly. Thereby, n number of customers are divided in to k number of subsets with minimum total cost. If Probabilistic FCM applied instead of Probabilistic RWFCM, the method called as Probabilistic FCM with the Center of Gravity (Probabilistic FCM-COG).

4 Experimental Study

The performance of the Probabilistic RWFCM-COG algorithm against FCM-based benchmark algorithms is conducted in this section.

Probabilistic FCM, Probabilistic FCM-COG and Probabilistic RWFCM are methods employed for comparison. Equation (1) is used for calculating total transportation costs and threshold value, T , is taken as 0.9 in all methods.

Probabilistic RWFCM-COG is compared with other proposed methods using 13 different datasets and 21 different trials. Data were obtained from a company in the telecommunication sector in Turkey. Table 1 includes the details of the total transportation costs of all methods for each trial.

Table 1. Total transportation costs of proposed algorithms.

Dataset	Demand Points	Clusters	Probabilistic FCM	Probabilistic FCM-COG	Probabilistic RWFCM	Probabilistic RWFCM-COG
1	100	4	31242.66	31567.75	30383.11	30107.38
2	400	10	27421.6	27043.75	27140.48	27037.82
2	400	20	18108.51	17909.93	17716.87	17688.36
3	500	30	28118.66	27423.99	26657.18	26517
4	800	10	99675.04	98921.52	99166	98654.5
5	1200	100	30285.64	29128.2	28593.57	28298.05
5	1200	120	26637.7	25743.55	25917.26	25500.58
6	2000	120	169868	166142.2	165692.1	164029.2

(continued)

Table 1. (continued)

Dataset	Demand Points	Clusters	Probabilistic FCM	Probabilistic FCM-COG	Probabilistic RWFCM	Probabilistic RWFCM-COG
6	2000	150	150134	146435.9	146749.1	146124.5
7	2500	200	36004.31	35123.45	34905.97	34611.42
8	2800	100	58190.81	57684.55	57464.25	57239.2
9	5000	50	117988.4	117812.9	117900.7	117798.6
9	5000	100	82070.04	81472.6	81587.69	81373.4
10	10000	50	272124.3	271137.8	268174.6	267673
10	10000	100	188013.1	187222.2	186299.8	185868.1
10	10000	200	131749.9	130582.1	130894.4	130449.8
11	20208	800	553645.9	527133.3	522225.2	519814
12	50111	200	2884595	2850871	2838437	2825892
12	50111	400	2014010	1986422	1986339	1964001
13	100000	100	9304092	9329676	9360277	9322158
13	100000	1000	1021425	1006551	959266	945384.6

5 Conclusion

In this paper, Probabilistic FCM, Probabilistic RWFCM, and hybrid algorithms with the center of gravity methods are proposed and compared with each other using 13 different datasets and 21 different trials. As a result, in terms of total cost minimization, the Probabilistic RWFCM-COG algorithm performs better than other proposed algorithms.

In future studies, different probability distributions of demand and customer positions can be applied in various data sets. Instead of probabilistic customer location and demand, stochastic versions of proposed algorithms will be used to solve the problem and can be compared in terms of CPU time and total transportation cost. As a fine-tuning method, the Nelder-Mead simplex search which is a derivative optimization algorithm will be combined with a new RWFCM algorithm in both probabilistic and stochastic telecommunication big data. Both the combination of the two methods and its area applied will be novel if it is successful. Also, this approach can be used for simulating the mobility of people infected by Covid-19 through their mobile phone-data in future research.

References

1. Altinel, I.K., Durmaz, E., Aras, N., Ozkisacik, K.C.: A location-allocation heuristic for the capacitated multi-facility Weber problem with probabilistic customer locations. *Eur. J. Oper. Res.* **198**, 790–799 (2009)
2. Bezdek, J.C.: *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York (1981)

3. Cooper, L.: Location-allocation problems. *Oper. Res.* **11**, 331–343 (1963)
4. Diabat, A., Dehghani, E., Jabbarzadeh, A.: Incorporating location and inventory decisions into a supply chain design problem with uncertain demands and lead times. *J. Manuf. Syst.* **43**, 139–149 (2017)
5. Esnaf, S., Kucukdeniz, T.: Solving uncapacitated planar multi-facility location problems by a revised weighted fuzzy c-means clustering algorithm. *J. Multiple Val. Logic Soft Comput.* **21**, 147–164 (2013)
6. Esnaf, S., Kucukdeniz, T.: A fuzzy clustering-based hybrid method for a multifacility location problem. *J. Intell. Manuf.* **20**(2), 259–265 (2009)
7. Gao, Y.: Shortest path problem with uncertain arc lengths. *Comput. Math Appl.* **62**, 2591–2600 (2011)
8. Gao, Y.: Uncertain models for single facility location problems on networks. *Appl. Math. Model.* **36**, 2592–2599 (2012)
9. Hansen, P., Jaumard, B., Taillard, E.: Heuristic solution of the multisource weber problem as a p-median problem. *Oper. Res. Lett.* **22**, 55–62 (1998)
10. Johnson, R.A., Wichern, D.W.: *Applied Multivariate Statistical Analysis*, 6th edn. Pearson, New York (2007)
11. Kuenne, R.E., Soland, R.M.: Exact and approximate solutions to the multisource Weber problem. *Math. Program.* **3**, 193–209 (1972)
12. Kucukdeniz, T., Baray, A., Ecerkale, K., Esnaf, S.: Integrated use of fuzzy c-means and convex programming for capacitated multi-facility location problem. *Expert Syst. Appl.* **39**(4), 4306–4314 (2012)
13. Markovic, N., Ryzhov, I.O., Schonfield, P.: Evasive flow capture: a multi-period probabilistic facility location problem with independent demand. *Eur. J. Oper. Res.* **257**, 687–703 (2017)
14. Megiddo, N., Supowit, K.J.: On the complexity of some common geometric location problems. *SIAM J. Comput.* **13**, 182–196 (1984)
15. Mousavi, S.M., Niaki, S.T.A., Mehdizadeh, E., Tavarroth, M.R.: Capacitated multi-facility location-allocation problem with probabilistic customer location and demand: two hybrid meta-heuristic algorithms. *Int. J. Syst. Sci.* **44**, 1897–1912 (2012)
16. Murray, A.T., Church, R.L.: Applying simulated annealing to location-planning models. *J. Heuristics* **2**, 31–53 (1996)
17. Murtagh, B.A., Niwattisayawong, S.R.: Efficient method for the multi-depot location dash allocation problem. *J. Oper. Res. Soc.* **33**, 629–634 (1982)
18. Ohlemuller, M.: Tabu search for large location–allocation problems. *J. Oper. Res. Soc.* **48**(7), 745–750 (1997)
19. Peidroa, D., Mula, J., Poler, R., Verdegay, J.L.: Fuzzy optimization for supply chain planning under supply, demand and process uncertainties. *Fuzzy Sets Syst.* **160**, 2640–2657 (2009)
20. Wang, S., Watada, J.: A hybrid modified PSO approach to VaR-based facility location problems with variable capacity in fuzzy random uncertainty. *Inf. Sci.* **192**, 3–18 (2012)
21. Wen, M., Iwamura, K.: Fuzzy facility location-allocation problem under the Hurwicz criterion. *Eur. J. Oper. Res.* **184**, 627–635 (2008)
22. Wen, M., Kang, R.: Some optimal models for facility location allocation problem with random fuzzy demands. *Appl. Soft Comput.* **11**, 1201–1207 (2011)

23. Wen, M., Qin, Z., Kang, R.: The α -cost minimization model for capacitated facility location-allocation problem with uncertain demands. *Fuzzy Optim. Decis. Making* **13**, 345–356 (2014)
24. Wen, M., Qin, Z., Kang, R., Yang, Y.: The capacitated facility location-allocation problem under uncertain environment. *J. Intell. Fuzzy Syst.* **29**, 2217–2226 (2015)
25. Yang, L.X., Ji, X.Y., Gao, Z.Y., Li, P.K.: Logistics distribution centers location problem and algorithm under fuzzy environment. *J. Comput. Appl. Math.* **208**, 303–315 (2007)
26. Zhou, J., Liu, B.: New probabilistic models for capacitated location-allocation problem. *Comput. Ind. Eng.* **45**, 111–125 (2003)
27. Zhou, J., Liu, B.: Modeling capacitated location-allocation problem with fuzzy demands. *Comput. Ind. Eng.* **53**, 454–468 (2007)