

Extended Single-Iteration Fuzzy C-Means, and Gustafson-Kessel Algorithms for Medium-Sized (10^6) Multisource Weber Problem

Tarik Kucukdeniz, Department of Industrial Engineering, Istanbul University-Cerrahpasa, Istanbul, Turkey

Sakir Esnaf, Department of Industrial Engineering, Istanbul University-Cerrahpasa, Istanbul, Turkey

Engin Bayturk, American College of the Middle East, Eqaila, Kuwait

ABSTRACT

An uncapacitated multisource Weber problem involves finding facility locations for known customers. When this problem is restated as finding locations for additional new facilities, while keeping the current facilities, a new solution approach is needed. In this study, two new and cooperative fuzzy clustering algorithms are developed to solve a variant of the uncapacitated version of a multisource Weber problem (MWP). The first algorithm proposed is the extensive version of the single iteration fuzzy c-means (SIFCM) algorithm. The SIFCM algorithm assigns customers to existing facilities. The new extended SIFCM (ESIFCM), which is first proposed in this study, allocates discrete locations (coordinates) with the SIFCM and locates and allocates continuous locations (coordinates) with the original FCM simultaneously. If the SIFCM and the FCM, show differences between the successive cluster center values are still decreasing, share customer points among facilities. It is simply explained as single-iteration fuzzy c-means with fuzzy c-means. The second algorithm, also proposed here, runs like the ESIFCM. Instead of the FCM, a Gustafson-Kessel (GK) fuzzy clustering algorithm is used under the same framework. This algorithm is based on single-iteration (SIGK) and the GK algorithms. Numerical results are reported using two MWP problems in a class of a medium-size-data (10^6 bytes). Using clustering algorithms to locate and allocate the new facilities while keeping current facilities is a novel approach. When applied to the big problems, the speed of the proposed algorithms enable to find a solution while mathematical programming solution is not doable due to the great computational costs.

KEYWORDS

Big Data Analysis, Multisource Weber problem (MWP), Single-Iteration Fuzzy C-means (SIFCM), Single-Iteration Gustafson-Kessel

INTRODUCTION

Arabani and Farahani (2012) divided application of fuzzy approaches in facility location problems into two categories, which are selection of facility location (a decision-making problem), and the location-allocation problem (an optimization problem). For the first category, multi-criteria decision-making methods such as Fuzzy Analytical Hierarchy Process (AHP), Fuzzy TOPSIS (The Technique for Order Preference by Similarity to Ideal Solution) are applied (Kahraman et al., 2010; Temur et al.,

DOI: 10.4018/IJORIS.2019070101

This article, originally published under IGI Global's copyright on July 1, 2019 will proceed with publication as an Open Access article starting on February 2, 2021 in the gold Open Access journal, International Journal of Operations Research and Information Systems (converted to gold Open Access January 1, 2021), and will be distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

2014). These approaches are fuzzy extension to the well-known classical decision-making methods. A recent study with the classical methods employs DEMATEL-ANP integrated method for location decision problem (Sharma et al., 2018). Problems in the second category can be also categorized as continuous or discrete, capacitated or uncapacitated, and single and multi-facility problems. Mathematical programming, heuristics and metaheuristics are the most common methods used for the second category problems. Statistical methods can also be applied to this category (Kuruvilla et al., 2011). If problem is discrete, facility must be located in a specific point. However, in continuous problems the required number of facilities can be located anywhere in a plane. Metaheuristics, of which handles values in real number space, can be applied to both discrete and continuous problem spaces. Nevertheless, Samanta and Jha (2012) applied genetic and ant colony algorithms, which are combinatorial in their nature, to the continuous search spaces, too.

In this study a new type of facility location problem, which is in between the discrete and the continuous are defined. Here, existing facilities are handled with a discrete fashion, allocation of customers to the existing facilities is under consideration and at the same time a predefined number of new facilities are to be opened on a planar space. This type of problem has not been defined and solved in the literature.

In multisource Weber problems (MWP), also known as location-allocation problem, we are interested in finding the location of p facilities in continuous space in order to serve customers at n fixed points as well as the allocation of each customer to the facilities so that total transportation costs are minimized (Salhi and Gamal, 2003). Solving the MWP and its variants such as uncapacitated or capacitated, weighted or unweighted, continuous or discrete are much studied problem in facility location literature (Brimberg et al., 2008). Optimization skills of exact solution algorithms are not sufficient to find an optimal solution due to NP -hard nature. In real world problems, especially when struggling with very large data (classified by Hathaway and Bezdek (2006) as in Table 1), it is very complicated to find optimum solutions. This evokes hybrid and heuristics-based solution approaches to these types of problems. Li et al. (2012) proposed a two staged approach to solve the p -median problem. The first part uses a greedy search method to find an initial solution; the second part sequentially substitutes medians in the initial solution with additional vertices to reduce the total travel cost. In recent years one may find numerous papers related to the MWP but only few of them involves with large data sets. Taillard (2003) proposes three heuristic algorithms to solve p -median, sum of squares and the MWP problems including PLA85900, largest data set in all classes investigated, consists of 85900 customer points and 15000 facilities but there is no information that which category this dataset falls into according to Table 1. Küçükdeniz and Esnaf (2016) used the same data with 1000 facilities in medium to large data with 1.7×10^6 bytes, to compare new focal particle swarm optimization (FPSO) with K-means, the FCM-based clustering and classical PSO algorithms.

In this study, it is aimed to work on another variant of the MWP, which assumes new facilities determined from infinite points of plane, to add to the existing locations. For this purpose, two algorithms called extended Single-Iteration Fuzzy C-Means (ESIFCM) and extended Single-Iteration Gustafson-Kessel (ESIGK) deal with existing and new unknown locations of facilities in every clustering iteration are developed. ESIFCM is based on Esnaf et al.'s (2014) Single-Iteration Fuzzy C-Means (SIFCM) algorithm which uses fixed cluster center solution approach to an uncapacitated facility location problem (UFLP). The logic of the SIFCM is copied to the SIGK algorithm. The

Table 1. Size of the data sets after Huber (1996) (Adapted from Hathaway and Bezdek, 2006)

Bytes “sizes”	(10 ²) tiny	(10 ⁴) small	(10 ⁶) medium	(10 ⁸) large
	(10 ¹⁰) huge	(10 ¹²) monster	(10 ^{n>12}) VL(very large)	infinite

ESIFCM, and the ESIGK are cooperative algorithms and based on the SIFCM and the FCM, and ESIGK and GK algorithms respectively. The proposed ESIFCM and ESIGK employ allocation abilities of the SIFCM and SIGK, and location allocation abilities of the FCM, and GK. SIGK allocates customers to existing facilities and GK first locates then allocates customers to required number of new facilities to be opened simultaneously like the SIFCM and FCM does. Unlike other algorithms, demand of a customer should not be satisfied by a single facility. To verify the applicability of the algorithm, first dataset called PLA85900 is taken from TSPLIB compiled by Reinelt (1995). In this dataset locations of 85900 cities are defined. Size of the first data set is 1.285 megabytes. Second problem instance is from a real-world data including 88572 customers and 81 existing and 3 new totally 84 facilities. Size of this data is 1.746 megabytes.

LITERATURE REVIEW

MWP related studies usually ignores to look at existing facilities with new ones. The lack of methods that combine existing facilities with finding new facility locations problem is handled by Mohammadi et al. (2010). They introduce a new variant of capacitated MWP assumed new facility centers are to be located while considering existing centers. All centers in a plane but location of newly offered facilities to be opened selected from a predetermined potential, facility alternatives like p -median problems. Like Mohammadi et al.'s model (2010), Esnaf et al. 's (2014) Single-iteration Fuzzy C-Means (SIFCM) algorithm solves a similar model for uncapacitated facility location problems (UFLPs) like Mixed-integer programming models do.

Kazakovtsev et al.'s (2015) study employs genetic algorithm with greedy heuristic for solving large sized location discrete and continuous p -median problems. They transform continuous p -median problem into a k -means problem by using squared Euclidean distance metric.

In this paper, FCM and GK based fuzzy clustering algorithms are applied to the combined crisp and continuous facility location allocation problems first time. Large problem instances are handled and this size of data is solved first time as a Multisource Weber Problem. Definition of Single Iterated GK algorithm and its application to the facility location and allocation problem is also mentioned first time in this study.

The rest of paper organized as follows. A new variant of the uncapacitated Multisource Weber problem and its non-linear programming model is described in the second section. In the third and fourth sections, extended Single-iteration Fuzzy C-means and extended Single-iteration Gustafson-Kessel algorithms are explained respectively. In the fifth section, proposed algorithms are applied to two medium-sized (10^6) problem. Conclusions are discussed in the sixth and last section of the paper.

PROBLEM DEFINITION

In case of extending the existing locations with new facilities, Multisource Weber problem can be modeled as a non-linear program which is adapted from Salhi and Gamal (2003), Kloze and Drexl (2005), and Mohammadi et al. (2010) and can be formulated as in formula (1), (2) and (3).

$$V(\text{MWP}) = \text{Min} \sum_{k=1}^p \sum_{j=1}^n w_{kj} d(X_k, a_j) z_{kj} + \sum_{i=p+1}^c \sum_{j=1}^n w_{ij} d(x_i, a_j) z_{ij} \quad (1)$$

Subject to

$$\sum_{k=1}^p z_{kj} + \sum_{i=p+1}^c z_{ij} = 1 \quad j = 1, 2, \dots, n \quad (2)$$

$$\sum_{i=p+1}^c z_{ij} = c_{max} \quad (3)$$

$$z_{kj} \geq 0 \quad k = 1, 2, \dots, p; j = 1, 2, \dots, n$$

$$z_{ij} \geq 0 \quad i = p + 1, p + 2, \dots, c; j = 1, 2, \dots, n$$

Where w_{kj} is the quantity assigned from existing facility k to fixed customer point j also denoting the allocation of customers to the existing facilities, $d(X_k, a_j)$ is the Euclidean distance from the discrete, in other words, location (coordinates) of existing facility k , (X_k, Y_k) to the location of a customer at fixed point j , (x_j, y_j) , and w_{ij} is the quantity assigned from facility i to fixed point j also denoting the allocation of customers to the open facilities, $d(x_p, a_j)$ is the Euclidean distance from the continuous, in other words, unknown location (coordinates) of facility i , (X_p, Y_p) to the location of a customer at fixed point j , (x_j, y_j) . This model works under the assumption that there are no capacity constraints at the new facilities hence the customers may split their demand between two or more facilities. z_{kj} is a variable denotes the fraction of demand from existing facility k to customer j , z_{ij} is another variable that denotes the demand fraction of customer j will be served by new and unknown location of facility i . c_{max} is the maximum number of new facilities to be opened. It should be noted that the capacitated model for MWP proposed by Mohammadi et al.'s (2010) is very similar to Esnaf et al.'s (2014) model for UFL. Contrary to the model presented here as (1) to (3), in both models, solutions for new facilities are determined among the predetermined fixed locations of candidate facilities. The objective function (1) minimizes the total demand-weighted distance from existing and new facilities to customers. Constraint (2) ensures that all demands are met by existing or/and new facilities. The equation (3) control that exactly p new facilities to be located. Also, it should be noted that, in mathematical modelling approach demand sharing among multiple facilities is impossible when there is not a capacity constraint. Because the minimization mechanism of the model will always assign customers to a single facility to ensure the minimum total cost. But in some real-life problems, demand sharing between multiple facilities may be a necessity to increase customer service level. Proposed two fuzzy clustering-based approaches in this study enables this requirement by its nature.

THE EXTENDED SINGLE-ITERATION FUZZY C-MEANS ALGORITHM

Exact solution algorithms have limited applicability due to the non-polynomial (NP) nature of the problem and when large data sets required to be processed, extended version of the SIFCM algorithm is developed. ESIFCM as a cooperative algorithm utilizes different abilities of the SIFCM and FCM. ESIFCM fulfills allocation function and location and allocation functions are accomplished by FCM simultaneously through sharing customer points in two directions; from SIFCM to FCM and vice versa. In every clustering iteration, the locations or coordinates of the new facilities are updated and for all iterations. If required SIFCM and FCM share the customer points between them with different membership degrees until termination conditions for cluster centers are met. The ESIFCM uses the same objective function called c-means functional adapted by Esnaf and Küçükdeniz (2014):

$$J_p(U, v) = \sum_{k=1}^n \sum_{i=1}^c (u_{ik})^p \|a^k - v_i\|^2 \quad (4)$$

where c is the number of the final clusters, which coincides with the number of rules, $U = \{[u_{ik}], 1 \leq i \leq c, 1 \leq k \leq n\}$ is the partition matrix, $V = \{[v_i], 1 \leq i \leq c\}$ with $v_i \in \mathbb{R}^m$ is the vector of the final cluster prototypes, $a^k (1 \leq k \leq n)$ is the data to be clustered, $p \in (1, \infty)$ is a factor to adjust the membership degree weighting effect; u_{ik} is the membership value of a point belonging to cluster i . Unlike other FCM algorithms V has two subgroups of cluster prototypes such as V_s and V_f . V_s is the set of the cluster centers which have fixed values. The cluster centers of V_s remain same during the clustering iterations. V_f is the set of the cluster prototypes as defined in the original FCM. During the clustering process these prototypes are repeatedly changing until terminated with a stopping criterion.

$$V_s = \{[v_i], 1 \leq i \leq q\} \quad (5)$$

$$V_f = \{[v_i], q+1 \leq i \leq c\} \quad (6)$$

$$V = \{V_s, V_f\} \quad (7)$$

The optimization problem is to minimize $J_p(U, V)$ under the following constraint:

$$\sum_{i=1}^c u_{ik} = 1 \quad \forall k \quad (8)$$

The final prototypes and the respective membership functions that solve this constraint optimization problem are given by the following equations.

$$v_i = \begin{cases} v_i, & i \in V_s \\ \sum_{k=1}^n (u_{ik})^p a^k / \sum_{k=1}^n (u_{ik})^p, & i \in V_f \end{cases} \quad (9)$$

$$u_{ik} = \frac{1}{\sum_{j=1}^c \left(\frac{\|a^k - v_i\|}{\|a^k - v_j\|} \right)^{\frac{2}{p-1}}}, \quad 1 \leq i \leq c, 1 \leq k \leq n \quad (10)$$

Steps of the Extended Single-Iteration Fuzzy C-Means Algorithm

Equations (9) and (10) develop an iterative optimization procedure, which is described by the following steps.

Step 1: Select the number of clusters c , a value for the factor p , fixed values of V_s and initial values for prototypes of V_f .

Step 2: Employ Equation (10) to calculate the membership values u_{ik} ($1 \leq i \leq c, 1 \leq k \leq n$).

Step 3: Calculate the updated cluster center values $v_{q+1}^{new}, v_{q+2}^{new}, \dots, v_c^{new}$ using Eq. (9).

Step 4: If $\max_i \left\{ \|v_i - v_i^{new}\|_{err} \right\} < \varepsilon$ then stop else go to step 2.

V_s and V_f denote the fixed or existing (discrete) locations and unknown (continuous) locations of facilities respectively. Total cost function which is given in formula (1) is computed. Facilities serve customers according to their membership degree values which represent the assignments of demands to facilities using formula (10).

THE EXTENDED SINGLE-ITERATION GUSTAFSON-KESSEL ALGORITHM

Gustafson-Kessel algorithm extends the standard FCM algorithm by introducing an augmented version of the Euclidean distance to be in the form $d_{GK}^2 = (x_k - v_i)^T A_i (x_k - v_i)$ where A_i is calculated using a scaled inverse fuzzy covariance matrix from each cluster (Graves and Pedrycz, 2007). Gustafson-Kessel algorithm is based on an iterative optimization of the objective function of the c-means type (Babuska et al., 2002):

$$J(X; U, V, \{A_i\}) = \sum_{i=1}^c \sum_{k=1}^N (\mu_{ik})^m D_{ikA_i}^2 \quad (11)$$

In equation (11), $U = [\mu_{ik}] \in [0, 1]^{c \times N}$ is the fuzzy partition matrix of the data $X \in R^{n \times N}$, $V = [v_1, v_2, \dots, v_c]$, $v_i \in R^n$ is the cluster prototypes (means) vector and $m \in [1, \infty)$ is the parameter which determines the fuzziness of the resulting clusters. The distance norm D_{ikA_i} can take into account different geometric shapes in one data set and be calculated as follows:

$$D_{ikA}^2 = \|x_k - v_i\|_A^2 = (x_k - v_i)^T A (x_k - v_i) \quad (12)$$

The size of each cluster is defined for the local A_i norm matrix that is used in the formula (11) as one of the optimization variables. This allows the distance norm to adapt to the local topological structure of the data. The minimization of the GK objective functional is achieved by using the alternating optimization method that is suggested by Gustafson Kessel (1979) as based on GK algorithm (Babuska et al., 2002).

In this clustering algorithm, data samples are small or data in a cluster are linearly related to each other, covariance matrix may become singular. To solve this problem in a simple and effective way, Babuska et al. (2002) modified the GK algorithm, as given in the following details:

For the given X data set, $c, m, \varepsilon, \rho_i$ standard parameters, β threshold numbers of conditions and γ weight parameter is chosen. Initial values of partition matrix are determined and covariance matrix F_0 is calculated for all data set.

Repeat for $l = 1, 2, \dots$

Step 1: Compute cluster prototypes (means):

$$V_i = \frac{\sum_{k=1}^N \left(\mu_{ik}^{(l-1)} \right)^m x_k}{\sum_{k=1}^N \left(\mu_{i,k}^{(l-1)} \right)^m} \quad 1 \leq i \leq c \quad (13)$$

Step 2: Compute the cluster covariance matrices:

$$F_i^{(l)} = \frac{\sum_{k=1}^N \left(\mu_{ik}^{(l-1)} \right)^m \left(x_k - v_i^{(l)} \right) \left(x_k - v_i^{(l)} \right)^T}{\sum_{k=1}^N \left(\mu_{ik}^{(l-1)} \right)^m} \quad 1 \leq i \leq c \quad (14)$$

Add a scaled identity matrix:

$$F_i = (1 - \gamma) F_i + \gamma (F_0)^{1/n} I \quad (15)$$

Extract eigenvalue λ_{ij} and Φ_{ij} from F_i

Find $\lambda_{i \max} = \max_j \lambda_{ij}$ and set

$$\lambda_{ij} = \lambda_{i \max} / \beta \quad \forall j \text{ for which } \lambda_{i \max} / \lambda_{ij} > \beta$$

Reconstruct F_i by

$$F_i = [\Phi_{i1} \dots \Phi_{in}] \text{diag}(\lambda_{i1}, \dots, \lambda_{in}) [\Phi_{i1} \dots \Phi_{in}]^{-1} \quad (16)$$

Step 3: Compute the distances:

$$D_{ikA_i}^2 = \left(x_k - v_i^{(l)} \right)^T \left(\rho_i \det(F_i) \right)^{1/n} F_i^{-1} \left(x_k - v_i^{(l)} \right) \quad (17)$$

$$1 \leq i \leq c, 1 \leq k \leq N$$

Step 4: Update the partition matrix:

$$\text{For } 1 \leq k \leq N$$

If $D_{ikA_i} > 0$ for $1 \leq i \leq c$,

$$\mu_{i,k}^{(l)} = \frac{1}{\sum_{j=1}^n \left(D_{ikA_i} / D_{jkA_j} \right)^{2/(m-1)}} \quad (18)$$

Otherwise

$$\mu_{i,k}^{(l)} = 0 \text{ if } D_{ikA_i} > 0 \text{ and } \mu_{i,k}^{(l)} \in [0,1]$$

With $\sum_{i=1}^K \mu_{i,k}^{(l)} = 1$ otherwise.

Run on until $\|U^{(l)} - U^{(l-1)}\| < \varepsilon$ (Babuska et al., 2002)

As in the ESIFCM algorithm, extended single-iteration Gustafson-Kessel Algorithm (ESIGK) employs current locations as static cluster centers and applies Gustafson-Kessel fuzzy clustering algorithm to find additional cluster centers by making reallocations in the solution space.

Conversion of the model inputs and variables for a new variant of MWP and proposed ESIFCM/ESIGK algorithms can be seen in Table 2.

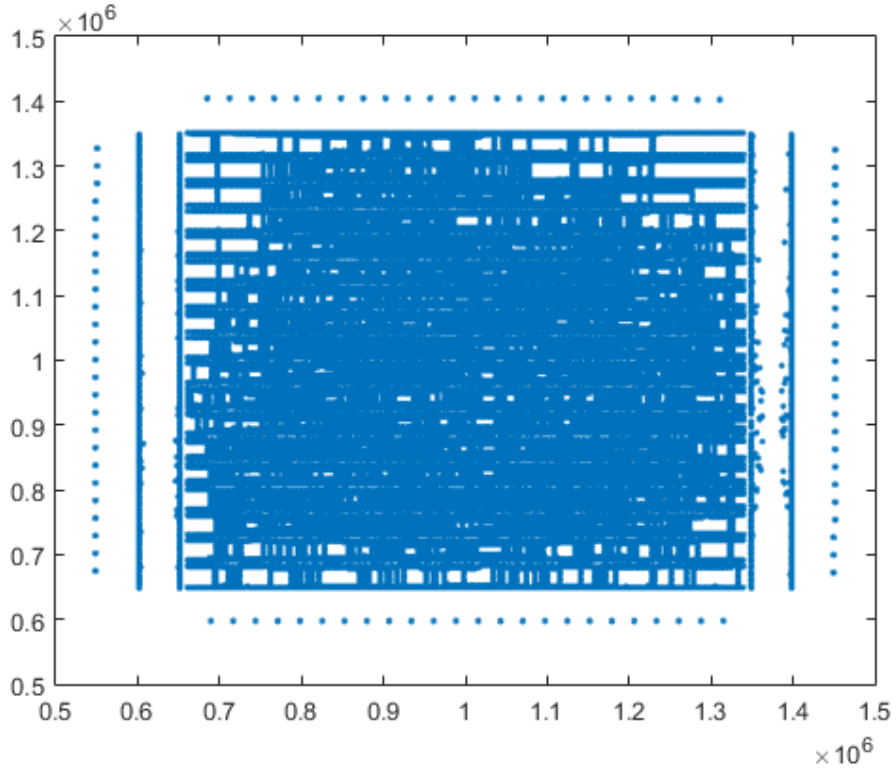
NUMERICAL STUDY

Proposed methods are applied to two big datasets. The first dataset, PLA85900, (Figure 1) can be found in the TSPLIB compiled by Reinelt (1995). This dataset consist of the locations of 85900 cities.

Table 2. Conversion of the model inputs and variables for a new variant of MWP and proposed ESIFCM/ESIGK algorithms

Definition	New variant of MWP	ESIFCM/ESIGK
Demand	w_{ij}	Computed by fuzzy membership values: u_{ik} or μ_{ik}
Distance from existing facilities to the customers	$d(X_k, a_j)$	$a^k - v_i^2, i \in V_s$
Distance from new facilities to the customers	$d(X_i, a_j)$	$a^k - v_i^2, i \in V_f$
Location of existing facilities	(X_k, Y_k)	$v_i, i \in V_s$
Location of new facilities	(X_p, Y_l)	$v_i, i \in V_f$
The fraction of demand from existing facilities to customers	z_{kj}	u_{ik} or $\mu_{ik}, i \in V_s$
The demand fraction of customers will be served by new and unknown facilities	z_{ij}	u_{ik} or $\mu_{ik}, i \in V_f$

Figure 1. PLA85900 data set



On the PLA85900 dataset, a numerical study is conducted by first selecting 100 random locations from PLA85900 as the current facility locations for each dataset. Then the proposed algorithm is executed to find five additional facility locations. The analysis has conducted ten times while choosing different random locations for 100 existing facilities at each run. Total calculated costs and CPU times of the proposed algorithms can be seen in Table 3.

As it can be seen in Table 3, ESIGK algorithm reaches better total cost values than the ESIFCM algorithm in 9 out of ten trials. However, in case of computational cost ESIFCM is faster than the ESIGK algorithm.

The second dataset (Figure 2) is provided from a Turkish distribution company and used after cleaned. There are 81 distribution centers of a company with 88,572 franchisees. In the solution, there will be X, Y coordinates of 3 new distribution centers to be determined, and the problem of assigning customers to the 81 existing and 3 new distribution centers with the closest distance will be obtained.

The number of variables or size of the solution matrix solved by the proposed algorithm in this problem is $88572 \times 84 = 7440048$. In other words, we have 7440048 assignment integer variables to be clustered. The size of the MWP problem in our case is 1746035 bytes or 1.746 megabytes. This size according to Hathaway and Bezdek (2006), which is given in Table 1, makes data set of our case, is medium (10^6). Location (coordinates) of new distribution centers will be added to existing facilities are computed and given in Table 4 and Table 5.

Figure 3 and Figure 4 shows the cluster maps of ESIFCM and ESIGK algorithms for 81 existing and 3 new distribution centers to be opened respectively. Table 6 shows the cost of obtained solutions for each proposed method. In this dataset, ESIFCM is better than the ESIGK algorithm in terms of both total cost and computational time.

Table 3. Total costs and CPU times of the proposed algorithms

Problem set	ESIFCM		ESIGK	
	Total costs	CPU times	Total costs	CPU times
1	476,174,966,351	81.71	479,362,394,562	266.87
2	509,029,777,806	48.02	482,510,805,574	120.32
3	494,970,552,617	50.84	472,091,842,269	521.33
4	497,350,046,883	38.95	486,355,740,125	151.70
5	505,417,101,144	45.61	495,309,825,273	286.44
6	500,845,455,841	49.73	486,658,662,123	329.36
7	498,136,631,650	45.87	483,319,144,004	363.40
8	492,177,407,834	41.78	484,000,595,263	256.58
9	495,066,828,549	44.80	482,510,758,541	239.84
10	507,629,786,355	34.04	507,160,156,554	142.44
Average	497,679,855,503	48.07	485,927,992,429	267.83

Figure 2. The data of Turkish distribution company

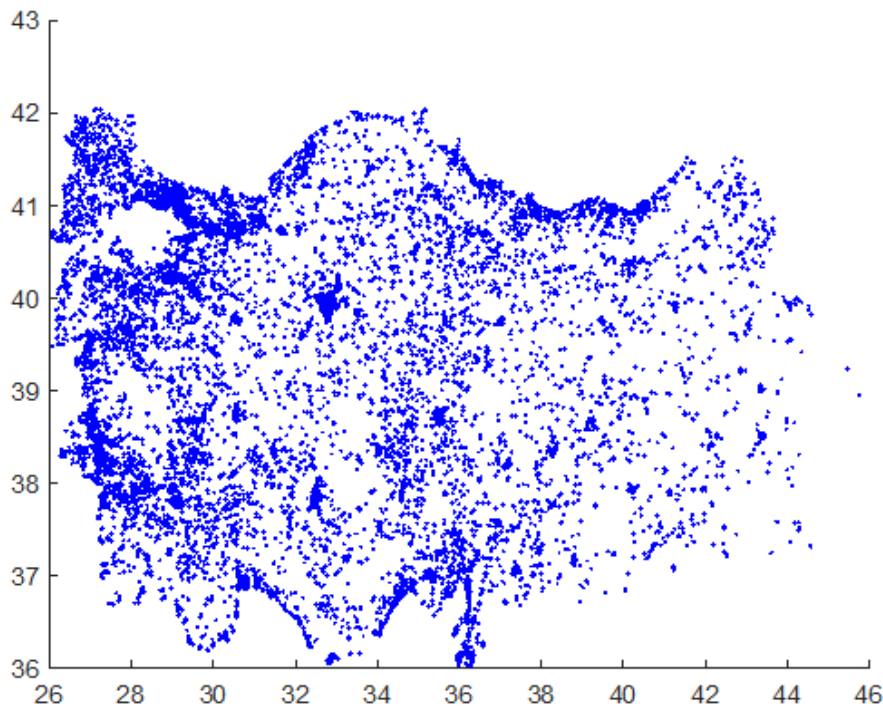


Table 4. Location (coordinates) of new distribution centers for ESIFCM

	X	Y
New Distribution Center 1 (82)	44.19	37.63
New Distribution Center 2 (83)	31.49	38.27
New Distribution Center 3 (84)	42.71	41.11

Table 5. Location (coordinates) of new distribution centers for ESIGK

	X	Y
New Distribution Center 1 (82)	35.15	36.95
New Distribution Center 2 (83)	31.08	38.28
New Distribution Center 3 (84)	42.73	41.10

Figure 3. Cluster maps of ESIFCM algorithm for 81 existing and 3 new distribution centers

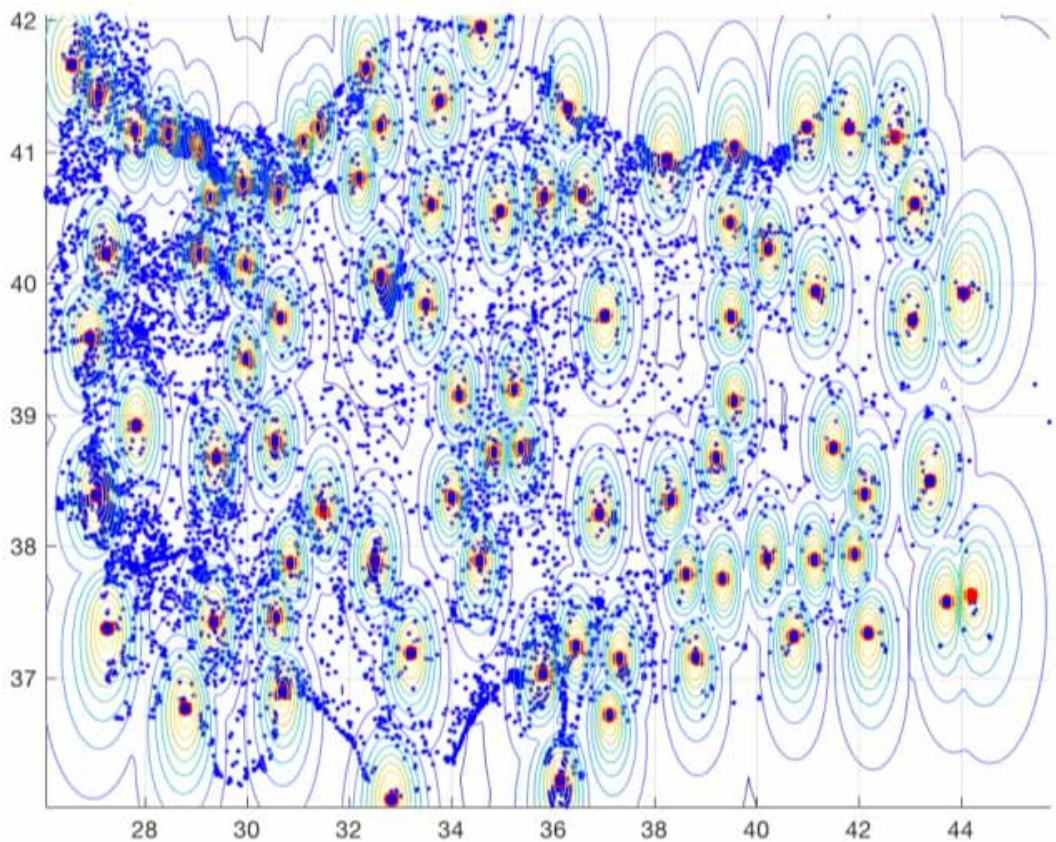


Figure 4. Cluster maps of ESIGK algorithm for 81 existing and 3 new distribution centers

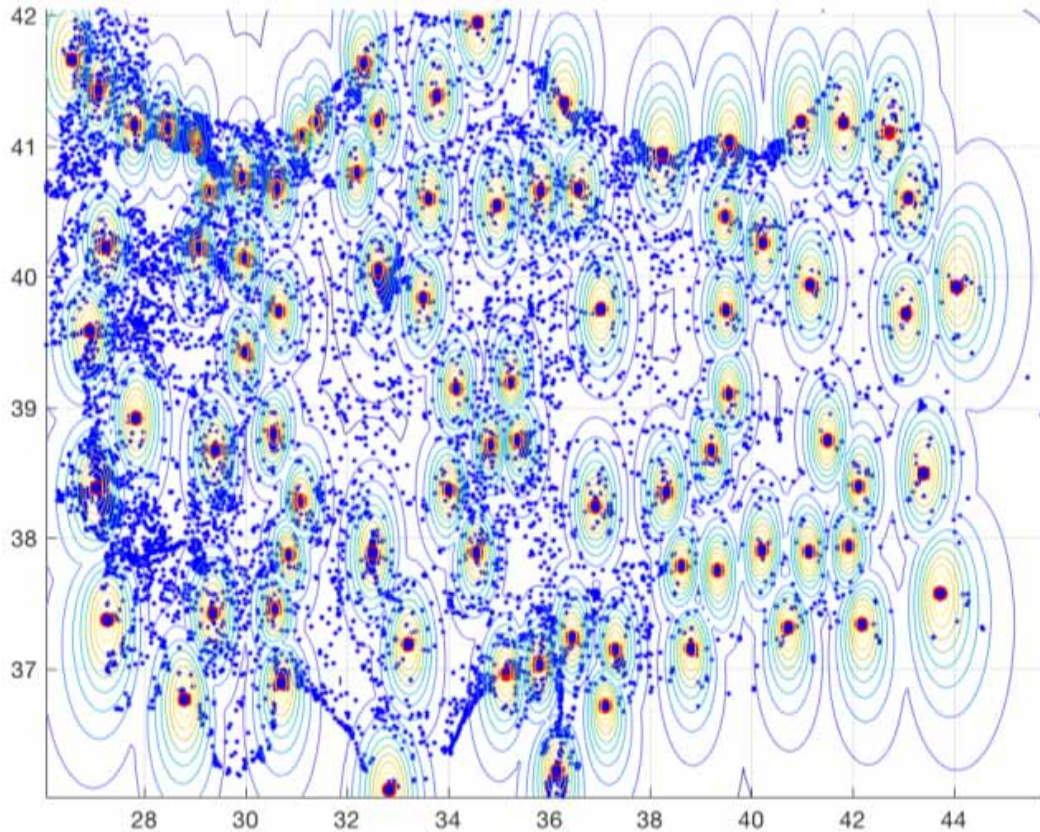


Table 6. Total cost of proposed algorithms

Method	Total Cost	CPU Time
ESIFCM	74572.82	23.19
ESIGK	80006.54	231.73

The computer that was used during test runs in section 5 has the following configuration; Intel CPU at 2.4 GHz with 16 GB of RAM. Codes of the ESIFCM and ESIGK algorithms were developed and executed by MATLAB R2017a. The main inputs for the ESIFCM and the ESIGK algorithms are X and Y coordinates of customer points and existing facilities. The ESIFCM and the ESIGK are applied to an uncapacitated new variant of MWP using weighting exponent, p , is taken as 2 and termination tolerance is determined as 10^{-6} .

CONCLUSION

In this study, two new and cooperative algorithms, the ESIFCM and ESIGK, are developed for a new variant of multisource Weber problem. The new variant of the MWP assumes new facility centers are to be located in a plane and changing during clustering while existing facilities having fixed locations are taken into consideration. This combined approach to the facility location allocation problem is

solved first time by using the FCM and GK based fuzzy clustering algorithms. Large problem instances are handled and this size of data is solved first time as a Multisource Weber Problem.

EISFCM uses different features of the SIFCM; and the FCM and ESIGK uses different features of the SIGK and GK, concurrently. Customer coordinates are already known a priori. The sizes of the data sets experimented here were 1.285 and 1.746 megabytes, which are in the medium (10^6 bytes) size category. One of the datasets is gathered from a distribution company in Turkey. ESIFCM and ESIGK are applied to this data set and the coordinates of the new locations are computed.

When the results on experimental studies analyzed, in the first dataset the ESIGK algorithm was better than the ESIFCM with a 2.42% lower total transportation cost. However, in the second dataset ESIFCM algorithm has reached 6.79% lower total transportation cost than ESIGK. These results show that in large problems, locations of the customers affect the solution performance because of the different distance metrics employed by different clustering methods. A good approach for the location and allocation problem would be trying different clustering algorithms, which have different distance measures, on the data and then comparing the results.

Adding new facilities, whose planar locations are not known priori, to a supply chain while maintaining the existing facilities is a special kind of supply chain management problems. The proposed algorithms are capable of solving these types of problems in real life we face with such as locating spare part logistics, relay stations, recycling centers, disaster relief stations, search and rescue headquarters, healthcare waste disposal facilities, cell towers, renewable energy systems, and medical emergency units etc. Due to the high number of decision variables, big problems are difficult and computationally expensive to solve with mathematical programming techniques. In this study clustering analysis-based solution methods are proposed which are combining existing locations with the process of finding new locations for the new facilities. As the experimental studies show that proposed algorithms are suitable for the stated business problem. Future research will be focused on much larger data sets sized of large (10^8 bytes), huge (10^{10} bytes), and monster (10^{12} bytes).

REFERENCES

- Arabani, A. B., & Farahani, R. Z. (2012). Facility location dynamics: An overview of classifications and applications. *Computers & Industrial Engineering*, 62(1), 408–420.
- Babuska, R., Van Der Veen, P. J., & Kaymak, U. (2002). Improved covariance estimation for Gustafson Kessel clustering. In *Proceedings of the IEEE International Conference on Fuzzy Systems*, Honolulu, HI (pp 1081-1085).
- Balasko, B., Aboyni, J., & Feil, B. (2005). Fuzzy clustering and data analysis toolbox. Retrieved from <http://www.fmt.vein.hu/softcomp/fclusttoolbox>.
- Brimberg, J., Hansen, P., Mladenovic, N., & Salhi, S. (2008). A survey of solution methods for the continuous location-allocation problem. *International Journal of Operations Research*, 5(1), 1–12.
- Esnaf, Ş., Küçükdeniz, T., & Tunçbilek, N. (2014). Fuzzy C-Means Algorithm with Fixed Cluster Centers for Uncapacitated Facility Location Problems: Turkish Case Study. In C. Kahraman & B. Öztayşi (Eds.), *Supply Chain Management Under Fuzziness*. Berlin, Heidelberg: Springer.
- Graves, D., & Pedrycz, W. (2007). Fuzzy c-means, gustafson-kessel fcm, and kernel-based fcm: A comparative study. In *Analysis and Design of Intelligent Systems using Soft Computing Techniques* (pp. 140–149).
- Gustafson, D. E., & Kessel, W. C. (1979). Fuzzy clustering with fuzzy covariance matrix. In *Proceedings of the IEEE* (pp. 761-766). San Diego, CA: CDC.
- Hathaway, R. J., & Bezdek, J. C. (2006). Extending fuzzy and probabilistic clustering to very large data sets. *Computational Statistics & Data Analysis*, 51, 215–234.
- Kahraman, C., Cebi, S., & Tuysuz, F. (2010). Fuzzy Location Selection Techniques. In C. Kahraman & M. Yavuz (Eds.), *Production Engineering and Management under Fuzziness*. Springer.
- Kazakovtsev, L. A., Orlov, V., Stupina, A. A., & Kazakovtsev, V. (2015). Modified Genetic Algorithm with Greedy Heuristic for Continuous and Discrete p-Median Problems. *Facta Universitatis, Series Mathematics and Informatics*, 30(1), 89–106.
- Klose, A., & Drexel, A. (2005). Facility location models for distribution system design. *European Journal of Operational Research*, 162, 4–29.
- Kuruvilla, A., Alexander, S. M., & Li, X. (2011). A Study of the Cascading Effects of Ambulance Diversion among Hospitals. *International Journal of Information Systems in the Service Sector*, 3(3), 60–70.
- Küçükdeniz, T., & Esnaf, Ş. (2016). Heuristic optimization-based clustering solution for large facility location problems. In *Intelligent Techniques for Data Analysis in Diverse Settings* (pp. 171-191). Hershey, PA: IGI Global.
- Li, X., Claramunt, C., Zhang, X., & Huang, Y. (2012). A fast and deterministic approach to a near optimal solution for the p-median problem. *International Journal of Operations Research and Information Systems*, 3(3), 1–14.
- Mohammadi, N., Mohammed, R. M., & Alesheikh, A. A. (2010). A new GA based solution for capacitated multi-source Weber problem. *International Journal of Computational Intelligence Systems*, 3(5), 514–521.
- Reinelt, G. (1995). *Tsplib95. Interdisziplinäres Zentrum für Wissenschaftliches Rechnen*. Heidelberg: IWR.
- Salhi, S., & Gamal, M. D. H. (2003). A Genetic algorithm based approach for the uncapacitated continuous location-allocation problem. *Annals of Operations Research*, 123, 203–222.
- Samanta, S., & Jha, M. K. (2012). Applicability of genetic and ant algorithms in highway alignment and rail transit station location optimization. *International Journal of Operations Research and Information Systems*, 3(1), 13–36.
- Sharma, S., Routroy, S., & Desai, R. (2018). Retail Location Decision Using an Integrated DEMATEL-ANP Method. *International Journal of Operations Research and Information Systems*, 9(1), 51–65.
- Taillard, É. D. (2003). Heuristic methods for large centroid clustering problems. *Journal of Heuristics*, 9(1), 51–73.
- Temur, G. T., Kaya, T., & Kahraman, C. (2014). Facility Location Selection in Reverse Logistics Using A Type-2 Fuzzy Decision Aid Method. In *Supply Chain Management under Fuzziness* (pp. 639-654). Springer-Verlag.

Tarik Kucukdeniz is an Associate Professor of Industrial Engineering in the Engineering Faculty of Istanbul University-Cerrahpasa. He received his M.Sc. and Ph.D. degrees from the Industrial Engineering Department of Istanbul University. His current research topics include optimization methods in supply chain problems and machine learning algorithms for time series prediction and regression.

Sakir Esnaf is a Professor of Department of Industrial Engineering at Istanbul University-Cerrahpasa. He graduated from Istanbul Technical University, Department of Industrial Engineering in 1986. He received his M.Sc and Ph.D degrees in Production Management from Istanbul University. He worked as a postdoctoral research fellow at the Intelligent Fuzzy Systems Laboratory, Department of Mechanical and Industrial Engineering at University of Toronto, Canada. His current research topics are facility location, fuzzy clustering, metaheuristics, big data analytics and revenue management. He has several national and international papers published.

Engin Bayturk is currently a PhD candidate in Industrial Engineering at Istanbul University-Cerrahpasa, Turkey. He is working as a Teaching Instructor at American College of the Middle East, Kuwait. He received his bachelor's degree in Industrial Engineering from Istanbul Kultur University, Turkey in 2012 and his Master's degree in Industrial Engineering from Istanbul University, Turkey in 2016. His research interests include optimization, stochastic processes, fuzzy modelling and artificial intelligence.