New Method for Solving Substation Expansion Planning Problem Using Fuzzy Clustering Algorithms

Zahra Moravej[1], Hossein Kiani Rad[2]
Department of Electrical Engineering, Semnan University, Semnan, Iran
1- Email: Moravej.zahra@gmail.com
2- Email: kiani@semnan.ac.ir

Received: August 2013 Revised: January 2014 Accepted: May 2014

ABSTACT
This paper presents a new method for solving Substation Expansion Planning (SEP) problem using three basic algorithms in fuzzy clustering. Clustering algorithms are mainly associated with distance functions and measure dissimilarities of data set in different clusters. It is equivalent to measure similarities of data in a cluster. That is, a lot of varieties exist to find and create such arranged clusters. The proposed clustering algorithms are Hard C-Means (HCM), Fuzzy C-Means and Possibilistic C-Means. At first, each algorithm is introduced and the differences are characterized. Objective function and optimization procedure of each algorithm are described afterward. Proper evaluation was done by simulating each algorithm. On the other hand, one of the complex and difficult issues in power systems is to find an appropriate response for substation expansion planning. By inspiring from HCM clustering method and by adding some necessary constraints, a new method was developed for solving SEP problem. The proposed method was applied to a typical network and good results were obtained. The results showed that the proposed method was highly effective in dealing with large networks. One of the features of this method is the possibility of introducing the location of new substations during the substation expansion planning. The fast convergence, conformity of solution with engineering perspectives, consideration of real-world networks limitations as problem constraints and simplicity in applying to real networks are the other features of the proposed method.

KEYWORDS: Algorithm, Clustering, Fuzzy, Substation Expansion Planning.

1. INTRODUCTION

The term "clustering" refers to partitioning data set into subsets or separated clusters. So that data contained in a cluster has common features as much as possible and data contained in separate clusters, on the other hand, has different features as much possible [1]. Clustering methods are mainly used to reduce data elements. Instead of dealing with large sets of scattered data, arranged groups can be found by using this method.

Clustering algorithms are mainly associated with distance functions and measure dissimilarities of data set in different clusters. It is equivalent to measure similarities of data in a cluster. That is, a lot of varieties exist to find and create such arranged clusters [1].

One of the complex and difficult issues in power systems is to find an appropriate response for Substation Expansion Planning (SEP) [2]. Generally, SEP can be expressed as an optimization problem. In any optimization problem, the decision variables and the technical and economic constraints must be determined. The number, size and location of new substations, number and expandable capacity of existing substations, number and conductor type of new installed feeders, etc., are the main optimization parameters and decision variables of the problem under consideration [2-5].

So far, different algorithms have been developed by researchers for this purpose. For example, mathematical planning methods, integer programming approach, dynamic programming algorithm, linear programming, mixed integer linear programming, and Branch-and-Bound dynamic programming can be found in the literature [2-12]. Other related references are [13]-[16], which employed the Genetic Algorithm (GA) to the SEP, [5] and [17-19] which considered the effect of uncertainty on some variables such as the amount of loads and [20], which employed the Simulated Annealing (SA) to search for the best solution among all possible solutions. Moreover, in [21] distribution expansion planning considering distributed generations is studied.

The abovementioned methods have some drawbacks, for example, in applying to large scale networks, finding the candidates of installation of new substations, and low speed of algorithm convergence.

Along with studying various algorithms of fuzzy clustering, a new method is developed to find the solutions of SEP in the present work which most of these drawbacks are overcome in the presented method. The main advantage of the presented clustering-based method is to find the location of candidates of new substations automatically that is a major drawback of the other previous methods.

Fuzzy clustering methods and basis of the SEP problem are described in the next sections.

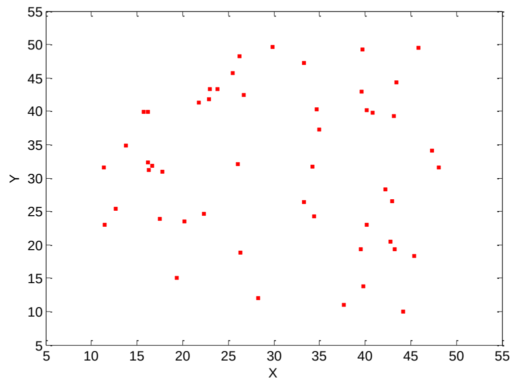
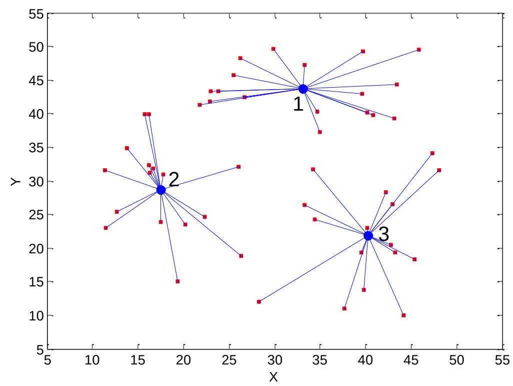
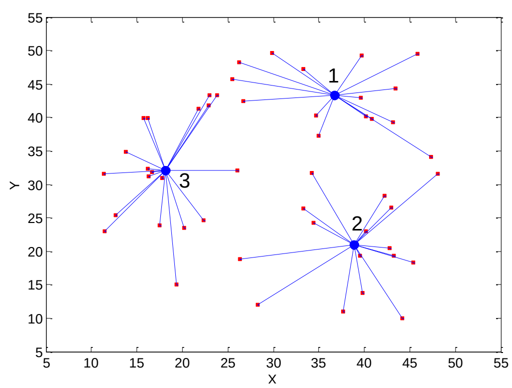


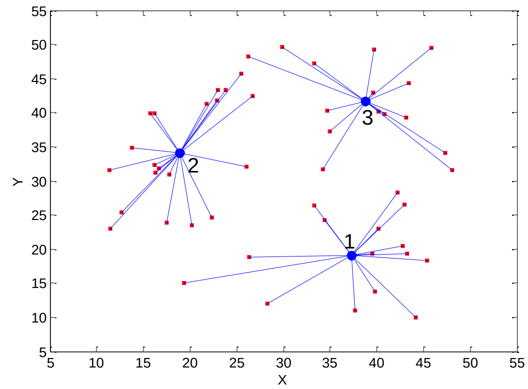
Fig. 1. Set of 100 data points



(a)



(b)



(c)

Fig. 2.a,b,c. Three clustered data set by HCM

2. BASIC CLUSTERING ALGORITHMS

Fuzzy C-Means (FCM) and Possibilistic C-Means (PCM) that derived from Hard C-Means (HCM) are described in this section.

Because the algorithms are based on objective function, after describing objective function of each algorithm the optimization procedure to minimize them is characterized.

A. Hard C-Means

Allocation each data X_j in the data set $X=\{x_1, \ \ldots \ , x_n\}$, $X \subseteq \Re^p$ to each cluster Γ_i, in the HCM algorithm, is unique. Set of clusters $\Gamma=\{\Gamma_1, \ \ldots \ , \Gamma_c\}$ is an extensive partition of data set X in c disjoint subsets Γ_i (1<c<n).

Data partition in HCM is optimal when the sum of squares of the distances between cluster centers and data points assigned to them is minimum [1]. Equation (1) describes the objective function of HCM.

$$J_h(X,U_h,C)=\sum_{i=1}^{c}\sum_{j=1}^{n}u_{ij}d_{ij}^2 \qquad (1)$$

Where $C=\{C_1, \ \ldots \ C_c\}$ is set of clusters prototypes; d_{ij} is distance between X_j and cluster center C_i; U is partition matrix of c × n (binary matrix); and the u_{ij} indicates the assignment of data to clusters.

$$u_{ij} \in \{0,1\} \qquad (2)$$

The following constraint ensures each data point is assigned exactly to one cluster:

$$\sum_{i=1}^{c}u_{ij}=1, \quad \forall j \in \{1, \ \ldots \ n\} \qquad (3)$$

Extensive partitions and avoiding the impracticable solution are resulted from constraint (3). Also, no cluster should remain empty, so:

$$\sum_{j=1}^{n} u_{ij} > 0, \quad \forall i \in \{1, \quad \ldots \quad c\} \tag{4}$$

J_h depends on c and U, which are found by NP-hard to minimize the C-means objective function [3]. Then, using alternating optimization (AO) scheme, the HCM clustering algorithm minimizes J_h [4]. Description of the method is neglected that is beyond the scope of this paper. The complete explanation of the method can be found in [1].

To achieve an appropriate criterion for comparing the algorithms, some examples are simulated for each case. Figure 1 shows a set of 100 completely dispersed data points. In Figure 1 the data is assigned to three clusters. Results of different executions of the three clusters are shown in Figures 2 to 4.

Detail review of Figure 2 shows several different clusters for a set of 100 data points obtained by HCM. This is a feature of HCM to be easily stuck in local minima which makes it necessary to conduct several runs of the algorithm with different initializations. Afterward, best obtained result can be chosen based on the values of J_h [1].

B. Fuzzy C-Means

According to the descriptions in [1], "fuzzy cluster analysis allows gradual memberships of data points to clusters measured as degrees in [0,1]. Aside from assigning a data point to clusters in shares, membership degrees can also express how ambiguously or definitely a data point should belong to a cluster. The concept of these membership degrees is substantiated by the definition and interpretation of fuzzy sets [5]". For more description of the method and their equations go to [1].

To make clear the performance of the FCM, follow the examples. In Figure 3, FCM clustering is performed to partition the data set into only one cluster. Since the data is normally distributed, the optimal point of the cluster is zero. The location of cluster does not change with different repetitions. This characteristic of FCM says that it has no tendency to be stuck into local minima. Another noticeable point regarding HCM is specified from Figure 4. Where the data set is partitioned into five clusters, the fourth and fifth clusters converge to the second and third clusters. This is derived from the fact that in the FCM clustering method the clusters may be overlapped and this makes the clusters to converge to each other [1].

The FCM algorithm is known as a steady and robust classification method. Compared with the HCM, its initialization is absolutely hard and it is not likely to get stuck in an unwanted local minimum of its objective function in practice [6].

C. Possibilistic C-Means

According to [1]: "Fairly high values for the membership of datum in more than one cluster can lead to the impression that the data point is typical for the clusters, but this is not always the case. Consider, for example, the simple case of two clusters shown in Figure 5. Datum x_1 has the same distance to both clusters and thus it is assigned a membership degree of about 0.5. This is plausible. However, the same degrees of membership are assigned to datum x_2 even though this datum is further away from both clusters and should be considered less typical. Because of the normalization, however, the sum of the memberships has to be 1. Consequently x_2 receives fairly high membership degrees to both clusters". Description of the method and their equations can be found in [1].

The results of clustering with the PCM method are illustrated in Figures 6 and 7. In Figure 6, using PCM the data set is partitioned into three clusters that the centers of the clusters are pictured. According to Figure 6 and the dispersion of the centers, it is clear that some numbers of data are unallocated to clusters. Therefore, PCM does not have a comprehensive partition of data set [1].

Because of the nature of the PCM and also separately calculation of the objective function of each cluster, if there exists a global optimum all clusters move toward that point in the data set and coincide. This is evident from Figure 7 that the cluster center 5 moves toward the cluster center 4 to coincide.

Application of HCM method in a newly developed algorithm to solve the SEP problem is discussed in section 3.

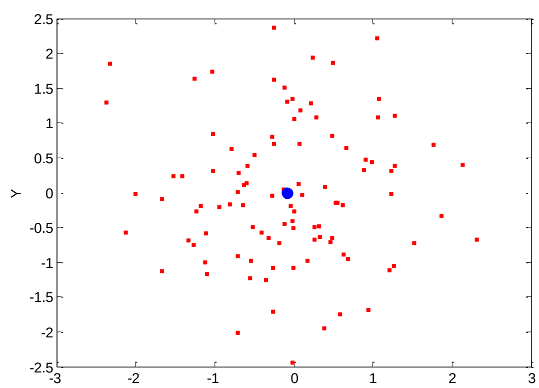


Fig. 3. FCM clustering of data set into one cluster

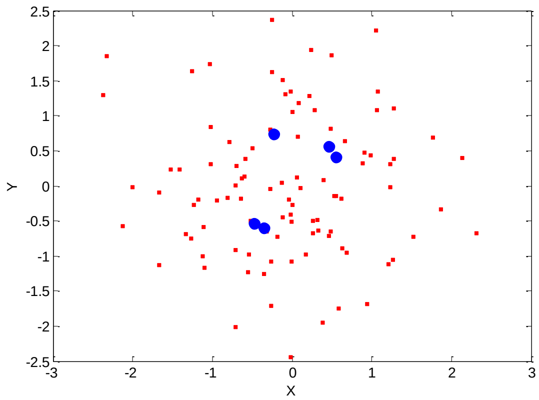


Fig. 4. FCM clustering of data set into five clusters

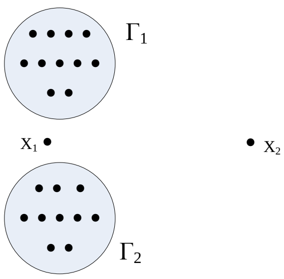


Fig. 5. A situation in which the probabilistic assignment of membership degrees is counterintuitive for datum x2 [1].

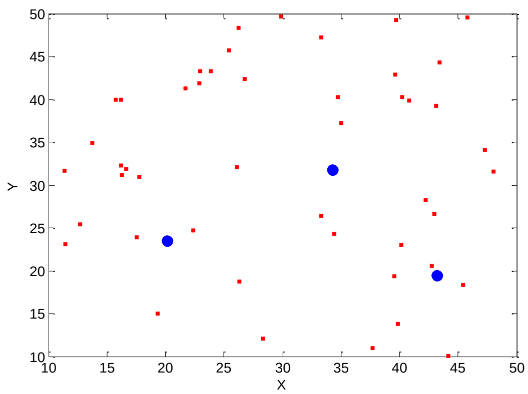


Fig. 6. The PCM clustering of data set in 3 clusters

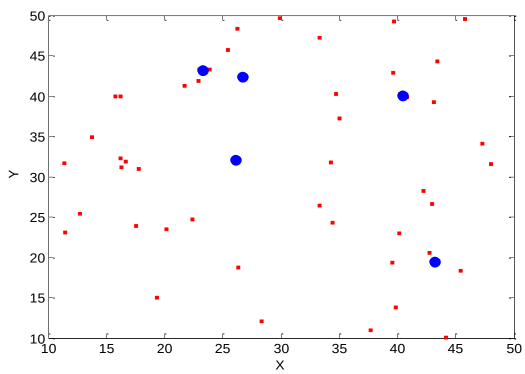


Fig. 7. The PCM clustering of data set in 5 clusters

3. Substation Expansion Planning by Clustering Method

Due to the nature of substation expansion planning HCM method was used to develop an applicable algorithm to solve SEP. For detailed discussion of SEP see [22].

A. Substation Expansion Planning

Substation expansion planning is implemented in transmission, sub-transmission and distribution levels independently. Location of distribution substations depends on the way consumers are distributed geographically, and the location of sub-transmission substations depends on the location of distribution substations. Similarly, the location of transmission substations depends on the way sub-transmission substations are located. To find the configuration and location of substations in a level, therefore, either geographical distribution of loads and their quantity or geographical distribution of downward substations may be used along with their loads. Considering geographical distribution of the loads it is obvious that the connection among the downward substations and studied substations are not determined, while their primary connection is not a question in the study.

Rated capacity of each substation (S) can be obtained as follows:

S = minimum capacity of (transformers, disconnectors, etc.) (5)

Total loads fed by each substation must be less than installed capacity of substation. Equation (6) expresses this constraint:

$$(1-k)\times S \le SL \qquad (6)$$

Where k is the capacity of each substation as reserve ranging between (0, 1). This constraint, like LOLP (Loss of Load Probability) in generation expansion planning or single contingency in transmission expansion planning, models the reliability in the study. In addition, SL is total loads fed by the corresponding substation (MVA).

As mentioned earlier, the loads are fed through the output feeders of substations. Each feeder, regarding its conductor type, is able to transfer a certain amount of power. It is the first constraint of feeders expressed in (7):

$$LOAD_a \le S_f \qquad (7)$$

Where, $LOAD_a$ is the amount of load fed by feeder (MVA) and S_f is the rated capacity of feeders (MVA).

Taking into account maximum permissible voltage drop, a feeder is able of transferring a certain amount of power for a certain distance. This is the second constraint of feeders. To express this constraint, a certain MVA×L value is attributed to each particular feeder. This value can be calculated with regards to resistance and reactance of feeders per length. Thus, Equation (8) is also satisfied:

$$L_a \times LOAD_a \leq MVA \times L \quad (8)$$

Where, La is distance between load point and substation under study (length of feeder) and MVA×L is the product of the capacity and the length of feeders.

The loads or downward substations must be supplied with the lowest possible cost. There are four components in the cost function. First component is installation cost of new substations, the second one includes the cost of feeders that feeds the loads, the third one is the cost of expansion of the existing substations, and finally the last component represents the cost of losses in feeders and substations (no-load losses of transformers). Therefore, the objective function to be minimized in this study is given as:

Objective Function = min {inst_cost + feeder_cost + exp_cost + loss_cost} (9)

Where, inst_cost is cost of installing a new substation ($), feeder_cost is cost of installing a new feeder ($), exp_cost is cost of expanding an existing substation ($), and loss_cost is cost of power losses in feeder and substation ($)

The clustering method, as a powerful tool for solving the above complicated optimization problem, will be discussed in what follows.

B. The Proposed Clustering Algorithm

Cost of feeding of each load from all substations is calculated in the proposed algorithm, and each load is attributed to the substation with the lowest feeding cost. Then, by subtracting the sum of the allocated loads of each substation and its permitted capacity (considering the reserve coefficient), an index is calculated for all substations. Negative index implies that sum of the loads attributed to the substation is higher than its capacity.

Furthermore, other indices are calculated by subtracting the cost of feeding each load from a substation with second priority and the lowest feeding cost of each load for all the loads attributed to each substation. Among the attributed loads of a substation, the largest index means that feeding cost of the load from other substations is higher than other loads of the substations. Thus, to reach the minimum objective function, loads arranged in descending order are fed by the substation with the lowest index respectively.

Loads are eliminated from the cluster by being fed through related substation and this process is repeated until the substation reaches its nominal capacity. When this process is over, related substation is eliminated from the substations list and the process is repeated from the beginning for the other existing substations from substations list.

If all of the loads are fed from the existing substations, there will be no need to install a new substation. Otherwise, there are two possible solutions. If it is possible and economic, numbers of substations will be expanded. Otherwise, new substations should be installed. To do this, the gravity center of the largest load and the nearest load to it is the candidate to install a new substation. When the voltage drop constraint allows, the old gravity center is replaced by the calculated one which is the gravity center of the former center and the nearest load to it. While the exact location of the new substation is not characterized, the process keeps running.

Capacity of new substation is specified according to the total of the loads allocated to the substation when its location is determined. Cost of installing the substation and feeders are two to be considered in calculating the objective function. By installing the new substation, the algorithm is repeated all over until all the loads are supplied.

When this process is over, all of the loads are supplied and the location and capacity of new installed substations are characterized. Furthermore, the expanded substations and capacity of expansion of each substation are specified. As a result, the network is expanded and will be adequate in the horizon year. Figure 8-a shows part 1 and Figure 8-b shows part 2 of the flowchart of the proposed algorithm.

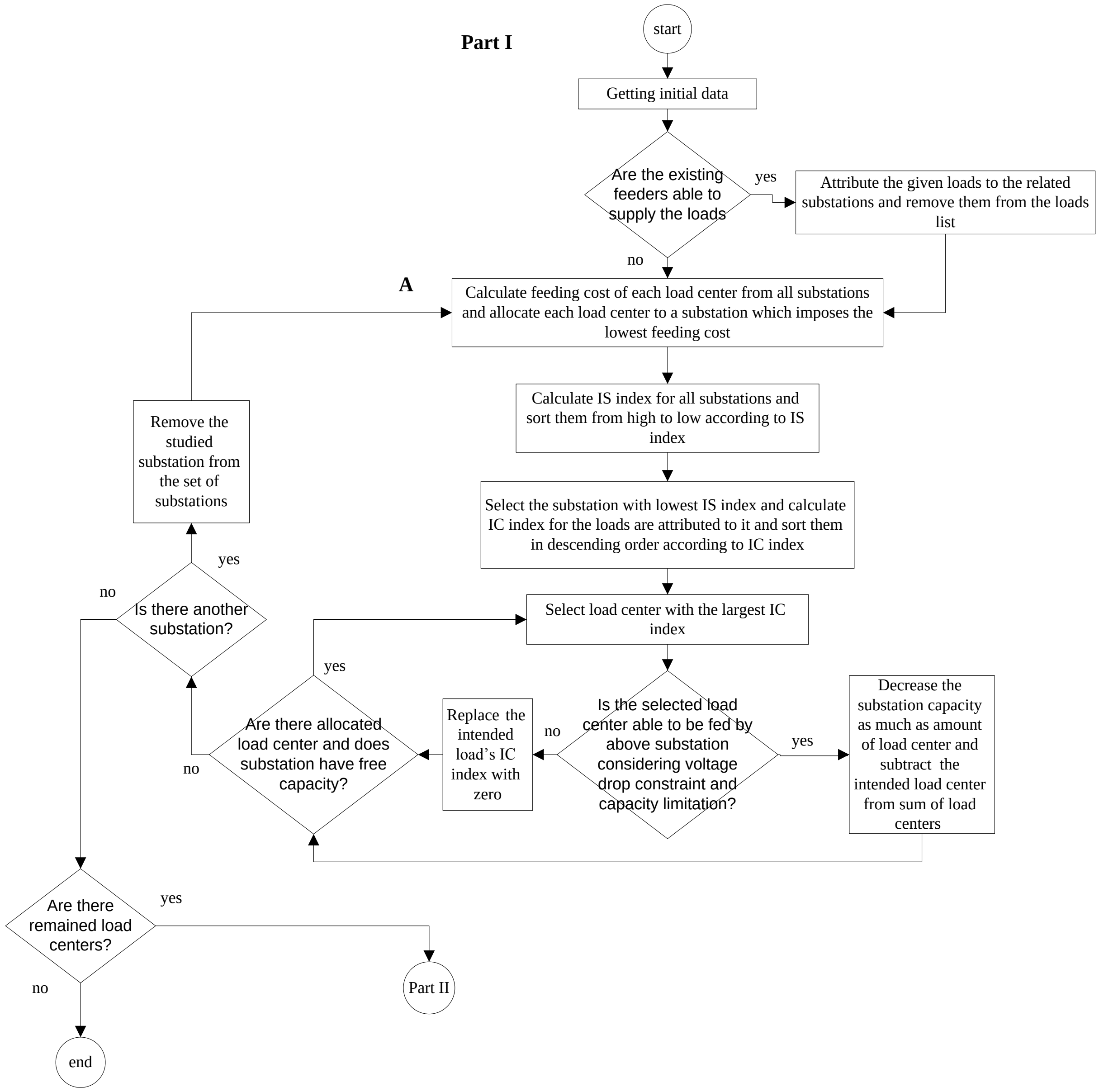


Fig. 8 .(a). Flowchart of the proposed algorithm for SEP

4. THE RESULTS APPLIED ON A TYPICAL NETWORK

In this section, a typical network is considered that is consists of 31 load centers and 5 existing substations. Parameters of the network including profiles of the feeders, cost of installation and expansion of the substations, the parameters of the loads and substations are given in [22].

Figure 9 shows SEP response to the proposed network. Three substations are installed with capacities of 15, 30 and 30 MW and some of the substations are expanded. Details of the software output are listed in Table I.

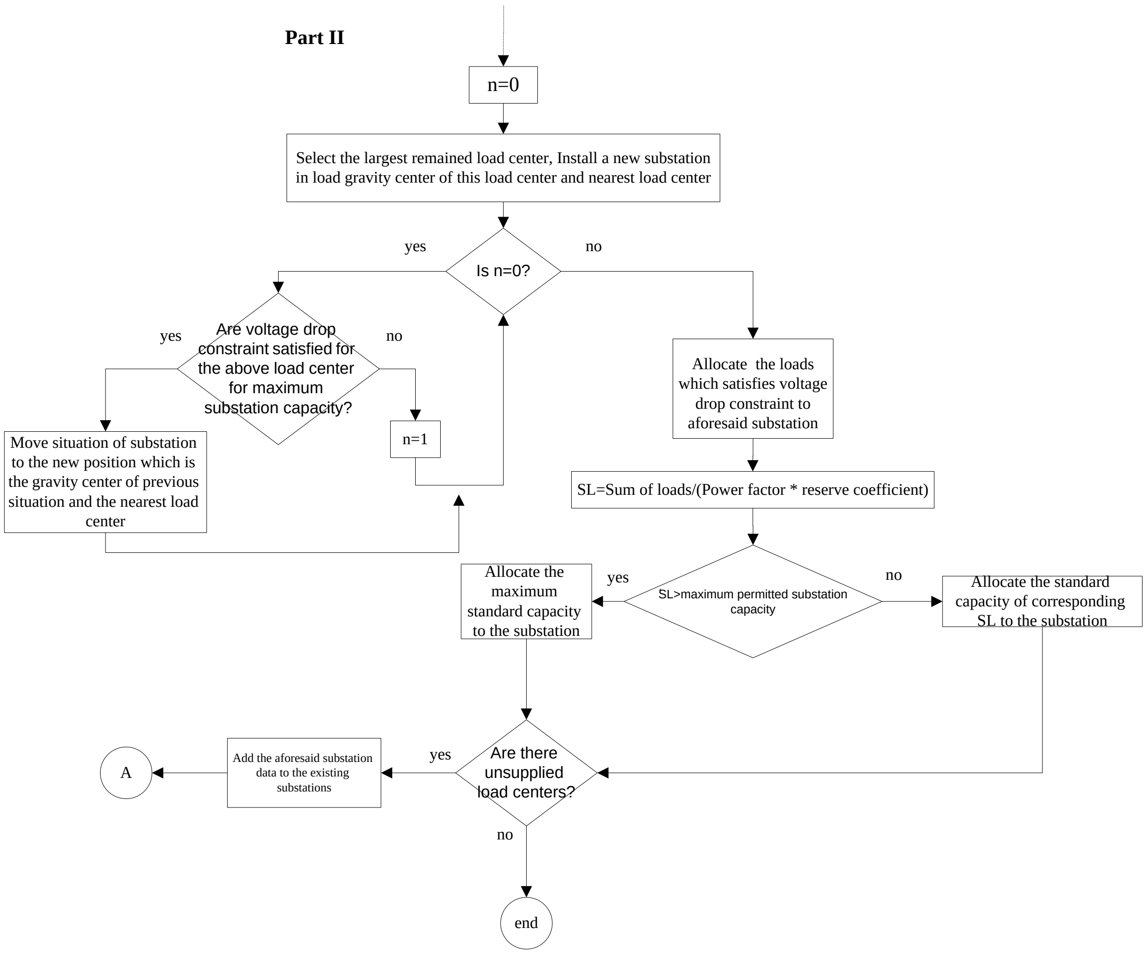


Fig. 8 (b). Flowchart of the proposed algorithm for SEP

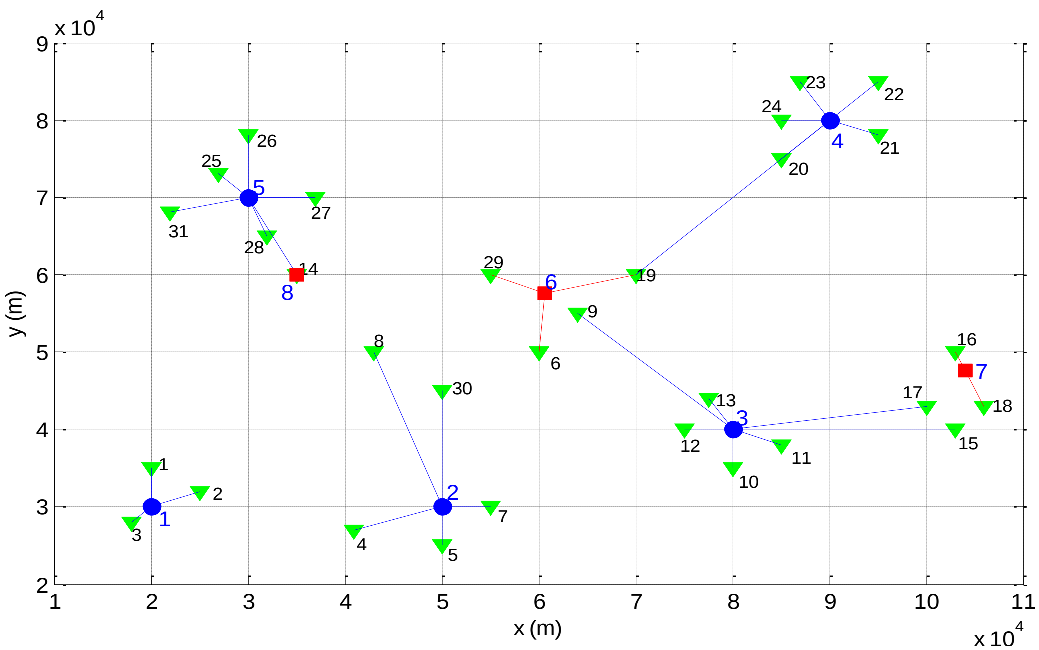


Fig. 9. The output of the clustering approach to SEP for the given network

Table 1. Results of applying the method to the typical network

Number of new installed substations	3
Sum of loads	152 (MW)
Sum of the new substations capacity	75 (MVA)
Installation cost of the new substations	7,500,000($)
Expanded substations	1,2,3,4
Sum of the old substations expansion capacity	150 (MVA)
Expansion cost of the old substations	3,850,000($)
Cost of feeders	6,463,000 ($)
Total cost	17,813,000($)

Compared to the results of the other methods – for example the GA results in [22] - the main advantages of the proposed algorithm are: ability to choose the new location of substations automatically, applicability on networks with large size, high speed of convergence, conformity of the presented answers with the engineering mentality, considering the real network constraints as the constraints of SEP problem and having no complexity of the method in applying to real networks.

5. CONCLUSION

The main methods of clustering were discussed in this paper. HCM uses a binary system (0, 1) for data assignment while FCM and PCM use membership degrees. The HCM is often stuck in local minima while this is not the case for PCM and FCM. For FCM, it was found capable of clustering overlaps while PCM was capable of clustering coincidences.

Application of fuzzy clustering in SEP was also addressed in this paper and an algorithm was developed to solve SEP problem by observing the limitations on substation capacities, feeder capacities, and voltage regulations. The results of applying the proposed algorithm on a typical network were presented. The ability of the clustering based algorithm was proved as the results showed.

REFERENCES

[1] J., Valente de Oliveira, W., Pedrycz, **"Advances in Fuzzy Clustering and its Applications",** *John Wiley*, 2007.

[2] X., Wang , J.R., McDonald, **"Modern Power System Planning"**, *MCGRAW-HILL Publication*, 1994.

[3] M.S., Sepasian, H., Seifi, A., Akbari, S.H., Hosseini and E., Mohseni Kabir, **"A New Approach for Substation Expansion Planning"**, *IEEE Transactions on Power Systems*, Vol.21, pp. 997-1004, 2006.

[4] S.M., Mazhari, H., Monsef, **"Dynamic Sub-Transmission Substation Expansion Planning Using Learning Automata"**, Electric Power System Research, Vol.96, pp. 255-266, 2013.

[5] T., Akbari, M., Heidarizadeh, M., Abdisiab and M., Abroshan, **"Towards Integrated Planning: Simultaneous Transmission and Substation Expansion Planning"**, *Electric Power System Research*, Vol. 86, pp. 131-139, 2012.

[6] M., Ponnaviakko, K.S., Prakasa Rao and S.S., Venkata , **"Distribution System Planning through a Quadratic Mixed Integer Programming Approach"**, *IEEE Transactions on PWRD*, Vol, 7, pp. 1157-1163, 1987.

[7] U.G.W., Knight, **"The Logical Design of Electrical Networks using Linear Programming Methods"**, *Proceedings of the IEE - Part A: Power Engineering*, Vol.107, pp. 306-316, 1960.

[8] J.V., Oldfield, M.A., Lang**," Dynamic Programming Network Flow Procedure for Distribution System Planning"**, *Proceeding Power Industry Computer Applications Conference* ,1965.

[9] T., Belgin, **"Distribution System Planning using Mixed Integer Programming"**, *ELEKTR-IK*, Vol.6, pp. 37-48, 1998.

[10] G.L., Thompson, D.L., Wall, **"A Branch and Bound Model for Choosing Optimal Substation Locations"**, *IEEE Transactions on PAS*, Vol.100, pp. 2683-2688, 1981.

[11] M.T., Ameli, M., Shivaie, **"A Multi-Objective Framework for Transmission Expansion Planning in Electricity Market"**, *International Review of Electrical Engineering (IREE)*, Vol.6, No.4, pp. 2036-2047, August 2011.

[12] E.C., Yeh, S.S., Venkata and Z., Sumic, **"Improved Distribution System Planning using Computational Evolution"**, *IEEE Transactions on Power Systems*, Vol.11, pp.668-674, 1996.

[13] R., Saberi, M., Khosravi, H., Falaghi, T., Vahidi, M., Ramezani and M., Dadgar**,"Optimal Planning of Medium Voltage Distribution Networks in the Presence of wind Generators"**, *Integration of Renewables into the Distribution Grid, CIRED 2012 workshop*, pp.1-4, 2012.

[14] I.J., Ramirez-Rosado, J.L., Bernal-Agustin**, "Optimization of Power Distribution Network Design by Application of Genetic Algorithm",** *International journal of power & energy systems*, Vol.15, pp.104-110, 1995.

[15] S.H., Hosseini, S., Jalilzadeh, A., Jalilvand, and A., Bagheri, **"A New Approach for Sub-transmission System Expansion Planning Using Genetic Algorithm"**, *International Review of Electrical Engineering (IREE)*, Vol.5, No.5, pp.2343-2349, October 2010.

[16] A., Bagheri, S.H., Hosseini, S., Jalilzadeh, A., Jalilvand, and M., Parsa Moghaddam, **"Sub-transmission System Expansion Planning Including Distributed Generation and Considering the Losses by Genetic Algorithm"**, *International Review of Electrical Engineering (IREE)*, Vol.5, No. 4, pp. 1655-1663, August 2010.

[17] M.R., Haghifam, M., Shahabi, "Optimal Location and Sizing of HV/MV Substation in Uncertainty Load Environment Using Genetic Algorithms", *International Journal of Electric Power Systems Research*, Vol.63, pp.37-50, 2002.

[18] P.M.S., Carvalho, L.A.F., Ferreira., **"Optimal Distribution Network Expansion Planning under Uncertainty by Evolutionary Decision Convergence "**, *International Journal of Electric Power Energy Systems*, Vol.20, pp.125-129, 1998.

[19] S., Lopez, A., Aguilera and G., Blanco**," Transmission Expansion Planning Under Uncertainty: An Approach Based on Real Option and Game Theory Against Nature"**, *Latin America Transactions, IEEE*, Vol.11, pp. 566-571, 2013.

[20] J.M., Nahman., D.M., Peric, **"Optimal Planning of Radial Distribution Networks by Simulated Annealing Technique"**, *IEEE Transactions on Power Systems*, Vol.23, pp.790-795, 2008.

[21] H., Shayeghi, A., Bagheri, **"Dynamic Sub-Transmission System Expansion Planning Incorporating Distributed Generation Using Hybrid DCGA and LP Technique"**, *International Journal of Electrical Power & Energy Systems*, Vol.48, pp. 111-122, 2013.

[22] H., Kord, **"Substation expansion planning including load uncertainties"** *MS.C. Dissertation, Univ. Zanjan, College electrical engineering*, 2010.