

Clustering Analysis and its Application in Electrical Distribution System

N.V.V.Karunakar.Jureedi, K.M.Rosalina, N. Prema Kumar

Department of Electrical Engineering, College of Engineering Andhra University, Visakhapatnam, INDIA Email: <u>Karunasmilesin@gmail.com</u>,

mercy345@reddifmail.com,Prem navuri@yaoo.co.in Abstract- Distribution Generation is expected to become more important in future generation system. In distribution system, due to load uncertainties the load exceeds the generating capacity which leads to power loss and unreliable operation of the system. To overcome this problem DG units are incorporated into the distribution system to meet the excess demand which results in power loss minimization, improvement of voltage profile, power quality improvement, reliable operation, etc., clustering techniques are proposed in this paper for optimal placement of DG. Multi DG units have to be placed to optimize the overall performance resulting in technical and economic benefits due to DG installation. In this paper IEEE standard 33 bus radial distribution test system is considered to indicate the efficacy of the proposed hierarchical and k-means algorithms in DG allocation.

Index Terms— Distribution Generation, clustering techniques, multiple DG units.

I. INTRODUCTION

Distributed Generation (DG) in the form of renewable power generation systems is currently preferred for clean power generation. It has a significant impact on the distribution systems. This impact may be either positively or negatively depending on the distribution system, distributed generation (DG) has much potential to improve distribution system performance. DG has an important role in power distribution system planning because of its modularity, small size and low investment cost as well as environmental concerns. In addition, DG also causes considerable reduction in required investment cost for supplying increased load in future. Unlike large central power plants, DG can be installed at or near the consumer load. Generally DG ratings range from 5 kW up to 100 MW. The introduction of DG to distribution system can significantly impact the power flow and voltage profile at consumer's utility equipment. These impacts may be either positively or negatively depending on the distribution system operating characteristics and the DG characteristics. In general, DG refers to small-scale generation connected to Medium Voltage and Low Voltage grids, excluding the large size production plants connected to the High Voltage grid.

Many studies have been indicating that inappropriate selection of location of DG may lead to system losses greater than the losses without DG. Due to high efficiency, small size, low investment cost, modularity in the construction and exploitation of the units, DG has increasingly become an attractive alternative to network reinforcement and expansion. Therefore, the factor of best location is one of the important issues in the implementation of DG sources in the distribution system. Effective integration of DG sources in electrical distribution networks would influence the performance of the distribution system. It enhances the performance in terms of voltage profile improvement, reduction in system losses, improved power quality and reliability of supply, reduction of the electricity cost, and possibility to assist congestion management in transmission lines. So, it is expected that DG sources will have increasing contribution in supplying the electricity demanded by customers in future. Also because of using new and renewable energy in DG units, it can help protect the environment. Interconnecting DG to the distribution feeders can have significant effects on the system in terms of power flow, voltage regulation and reliability. A DG installation changes the traditional characteristics of the distribution system. Determination of the optimal location of DG units in distribution systems with various objective functions have been studied with their respective proposed solutions. Regarding the placement of DG sources, in the literature there are many solutions proposed, namely using fuzzy techniques [4][5], analytical technique [6], heuristic optimization algorithm [7][8], mixed integer nonlinear programming [9], Evolutionary Programming (EP) optimization technique [10], and others.

In this sense, a new approach is proposed in this paper. Using hierarchical and k-means clustering method, the optimal placement of DG into electrical distribution network is determined based on the operational characteristics (the loss sensitivity factor (LSFs) and the voltage magnitude at each node). The use of clustering techniques for representing data, measuring the similarity between elements representative for the group, can be very suggestive and attractive. The k-means clustering method is tested by IEEE 33 bus distribution network to show the effectiveness of the methods in determining the optimal nodes where DG sources have to be placed. Compared with the methods presented in [1][2], and [4], that used the same variables to find the optimal place for DG into a distribution network, the new method based on k means clustering techniques obtains the results very quickly and is particularly effective.

II. LOSS SENSITIVITY FACTORS

The loss sensitivity factor method has been used to solve some of the problems referring to distribution systems, for instance for capacitor location, and can be successfully used to determine the optimal place of DG [1][2][4]. To calculate the loss sensitivity factors into a radial distribution network, it is necessary to know the real and reactive power loads, the real and reactive power losses in network nodes and branches that can be obtained from power flow calculations.



Fig.1 General two bus network

According to the figure supposing a line with impedance of (R + jX) ohms between bus p and bus q, together with the load of $P_{eff} + jQ_{eff}$. The effective active power losses in $k^{\pm h}$ line is as

$$P_{line\ loss} = \llbracket I_k \rrbracket^2 * \llbracket R_k \rrbracket \tag{1}$$

$$I_{k} = \left(\frac{P_{\text{eff}}(q) + jQ_{\text{eff}}(q)}{V_{q}}\right)^{*}$$
(2)

$$= \left(\frac{P_{eff}(q)0 - jQ_{eff}(q)}{V_q^*}\right)$$
(3)

With substituting I_{k} in eq—(3) we get

$$P_{lloss} = \left(\frac{\left[P_{eff}\left(q\right)\right]^{2} + \left[Q_{eff}\left(q\right)\right]^{2}}{\left[V_{q}\right]^{2}}\right) * \left[R_{k}\right]$$
(4)

So the sensitivity analysis factor is derived by derivative of $P_{time \ loss}$ with respective to P_{eff}

$$\frac{\partial P_{line \ loss}}{\partial P_{eff}} = \frac{\left(2P_{eff} \left(q\right) * R_{k}\right)}{\left[V_{q}\right]^{2}} \tag{5}$$

According to equation (5) buses will be ranked and some buses are categorized as the ones which have the most sensitivity for DG placement in order to have best effect on loss reduction.

III. CLUSTER ANALYSIS.

Cluster means a group or a collection of similar objects. Clustering is a procedure to organize data objects into groups or clusters, based on the principle of maximizing the intra-class similarity and minimizing the inter class similarity. Formally clustering task can be defined as "let

 $X \in \mathbb{R}^{mxn}$ a set of data items representing a set of m points X_i in \mathbb{R}^n . The goal is to partition X into K groups C_k such every data that belong to the same group are more "alike" than data in data in different groups. Each of the K groups is called a cluster". The number K might be pre-assigned by the user or it can be an unknown, determined by the algorithm.

Five types of clustering techniques are there in the literature, namely Partitioning, Hierarchical, Density based, Model Based and Grid Based. In this paper partitioning and hierarchical methods are studied, and applied on different radial distribution systems to find zones.

Partitioning Algorithms:

Partitioning techniques are largely similar. They all require some prior information to be specified. The general idea of partitioning strategy is: Given a dataset (D) of n objects, k: the number of clusters to be formed. Then the method will organize the objects into k partitions (k \leq =n) where each partition represents a cluster. The

clusters are formed to optimize the objective partitioning criterion by using iterative relocation technique (objects in a cluster are similar; objects of different clusters are dissimilar). In general, the square-error criterion is used, defined as

$$E = \sum_{i=1}^{k} \sum_{\boldsymbol{p} \in C_i} |\boldsymbol{p} - \boldsymbol{m}_i|^2$$
(6)

Where E is the sum of the square error for all objects in the data set; p is the point in space representing a given object; and m_i is the mean of cluster C_i (both p and m_i are multidimensional). In other words, for each object

in each cluster, the distance from the object to its cluster center is squared, and the distances are summed. This criterion tries to make the resulting k clusters as compact and as separate as possible. For example,

Algorithm: k-means.(The k-means algorithm for partitioning, where each cluster's center is represented by the mean value of the objects in the cluster).

Input:

k: the number of clusters,

D: a data set containing n objects.

Output: A set of k clusters.

Method: arbitrarily choose k objects from D as the initial cluster centers; repeat (re)assign each object to the cluster to which the object is the most similar, based on the mean value of the objects in the cluster; update the cluster means, i.e., calculate the mean value of the objects for each cluster; until no change;



Fig.2 Clustering of a set of objects based on the k-means method. (The mean of each cluster is marked by a "+".)

Hierarchical Algorithms:

Connectivity based clustering, also known as hierarchical clustering, is based on the core idea of objects being more related to nearby objects than to objects farther away. Hierarchical methods are usually classified into Agglomerative and Divisive methods depending on how the hierarchy is formed (bottom-up or top-down). These algorithms connect "objects" to form "clusters" based on their distance. A cluster can be described largely by the maximum distance needed to connect parts of the cluster. At different distances, different clusters will form. These algorithms do not provide a single partitioning of the data set, but instead provide an extensive hierarchy of clusters that merge with each other at certain distances. (Ranjini and Rajalinngum, 2011).

Agglomerative hierarchical clustering: This bottom-up strategy starts by placing each object in its own cluster and then merges these atomic clusters into larger and larger clusters, until all of the objects are in a single cluster or until certain termination conditions are satisfied. Most hierarchical clustering methods belong to this category. They differ only in their definition of inter-cluster similarity.

Divisive hierarchical clustering: This top-down strategy does the reverse of agglomerative hierarchical clustering by starting with all objects in one cluster. It subdivides the cluster into smaller and smaller pieces, until each object forms a cluster on its own or until it satisfies certain termination conditions, such as a desired number of clusters is obtained or the diameter of each cluster is within a certain threshold.

The Following figure shows how Agglomerative and divisive hierarchical clustering on dataset $\{a, b, c, d, e\}$ happens



Fig:3 Agglomerative and divisive hierarchical clustering

In either agglomerative or divisive hierarchical clustering, the user can specify the desired number of clusters as a termination condition.

Dendrogram

A tree structure called a dendrogram is commonly used to represent the process of hierarchical clustering. It shows how objects are grouped together step by step. Figure shows a dendrogram for the five objects presented in Figure 3.



Fig.4: Dendrogram

There are different hierarchical clustering methods; they can be distinguished depending up on the definition of distance (similarity) measure. Several distance measures are used to merge the clusters in agglomerative clustering techniques.

Minimum Distance: $d_{\min}(C_i, C_j) = \min_{p \in G_i, p' \in C_j} \left| p - p' \right|$ Maximum Distance: (7)

$$d_{\max}(C_i, C_j) = \max_{p \in C_i, p' \in C_j} \left| p - p' \right|$$
(8)

$$d_{mean}(C_i,C_j) = \left| m_i - m_j \right|_{(9)}$$

Mean Distance: Mean Distance:

$$d_{avg}(C_{i},C_{j}) = \frac{1}{n_{i}n_{j}} \sum_{p \in C_{i}} \sum_{p' \in C_{j}} |p-p'|$$
(10)

When an algorithm uses the minimum distance, dmin (Ci, Cj), to measure the distance between clusters, it is sometimes called a nearest-neighbor clustering algorithm. Moreover, if the clustering process is terminated when the distance between nearest clusters exceeds an arbitrary threshold, it is called a single-linkage algorithm.

When an algorithm uses the maximum distance, dmax (Ci, Cj), to measure the distance between clusters, it is called a farthest-neighbor clustering algorithm. If the clustering process is terminated when the maximum distance between nearest clusters exceeds an arbitrary threshold, it is called a complete-linkage algorithm.

The minimum and maximum measures represent two extremes in measuring the distance between clusters. The use of mean or average distance is a compromise between the minimum and maximum distances and overcomes the outlier sensitivity problem. Whereas the mean distance is the simplest to compute, the average distance is advantageous in that it can handle categorical as well as numeric data.

IV. DG ALLOCATION METHOD WITH CLUSTERING TECHNIQUES

Regarding the placement of DG sources, in the literature there are many solution proposed, namely using fuzzy and

real coded genetic algorithm, using particle swarm optimization (PSO) and others.

To determine the optimal placement of DG sources into an electrical distribution system proposed solution has the following steps:

In this step from distribution load flow analysis we obtain the results of real and reactive power losses at each node and branch and the voltage at each node.

The second step it to determine the loss sensitivity factors (LSFs) in nodes of distribution system analyzed. The normalized LSFs and voltages are then computed. The k-means clustering method is used for grouping the nodes from the viewpoint of the operation characteristics (LSF and V(p.u)). Based on the clusters obtained in the previous step, the reference node where DG sources can be located is recognized for each cluster.

Flow chart :



Fig.5 Flow chart for k-means clustering

V. CASE STUDY

To show the capability of the proposed method to solve the problem of the optimal placement of DG sources, IEEE 33 bus distribution network has been considered (Fig.6).



Fig. 6IEEE 33 bus radial distribution system

For this network we have information about the line parameters (i.e line and load data), information which have led to performing the Load flow analysis obtain the required parameters obtained, the loss sensitivity factors (LSFs) were determined to be used in the process of clustering.

Table 1 The Loss Sensitivity Factors (LSF) and Voltage at each node (V(p.u)) for IEEE 33 bus distribution network

Node	V(p.u)	LSF	Node	V(p.u)	LSF
1	1	0	18	0.9038	0.0034
2	0.997	0.0003	19	0.9965	0.0006
3	0.9829	0.0019	20	0.9929	0.0057
4	0.9754	0.0014	21	0.9922	0.0016
5	0.968	0.0015	22	0.9916	0.0027
6	0.9495	0.0034	23	0.9793	0.0018
7	0.946	0.0008	24	0.9726	0.0036
8	0.9323	0.0074	25	0.9693	0.0036
9	0.926	0.0045	26	0.9476	0.0008
10	0.9201	0.0046	27	0.945	0.0012
11	0.9193	0.0009	28	0.9336	0.0045
12	0.9177	0.0017	29	0.9254	0.0035
13	0.9116	0.0066	30	0.9218	0.0022
14	0.9093	0.0025	31	0.9176	0.0043
15	0.9079	0.0027	32	0.9167	0.0014
16	0.9065	0.0034	33	0.9164	0.0015
17	0.9044	0.0159			



Fig.7 Dendrogram for the IEEE 33 bus system



Fig. 8 The schematic diagram clusters obtained for the IEEE 33 bus system



Fig.9 Hierarchical cluster formation for IEEE 33-bus



Fig.10 k-means cluster formation for IEEE 33-bus

The k-means clustering is to identify the reference nodes in which the DG have to be placed. In the k-means clustering process, the best result was obtained for three clusters (Figure 8). On the basis of the results obtained in the clustering process, two have been considered to locate the distributed generation sources: Considering the nodes located in a single cluster (the cluster with high LSF and low voltage) resulting after the clustering process, the DG sources are located in these nodes and the optimal sizes in these nodes are determined through an exhaustive search method. Considering the clusters obtained after applying the k-means clustering technique, and identified as the representative node as 8 and 17 of the clusters.

VI. CONCLUSION

This paper presents novel technique of identifying optimal location for the installation of DG units in Radial Distribution Systems using cluster base analytical technique. Distribution load flow studies performed on standard IEEE 33 bus system. Load flow parameters Loss sensitivity factor (LSF) and Voltage at each node (V(p.u)) are assigned as a attributes for the selection of clusters. The Hierarchical and k-means clustering techniques are used for the application. The results are competent and proved the superiority of clustering techniques.

REFERENCES

- Pavel Berkhin, Accrue Software, 1045 Forest Knoll Dr., San Jose, CA, 95129; e-mail: pavelb@accrue.com."Survey of Clustering Data Mining Techniques"
- [2] Florina Scarlatache*, Gheorghe Grigoraş*, Gianfranco Chicco**, Gheorghe Cârţină* *Technical University "Gheorghe Asachi" of Iasi, Romania **Politecnico di Torino, Italy E-mail: flr_rotaru@yahoo. "Using k-Means Clustering Method in Determination of the Optimal Placement of Distributed Generation Sources in Electrical Distribution Systems,"
- [3] FLORINA SCARLATACHE and GHEORGHE GRIGORAȘ* "AN APPROACH REGARDING THE PLACEMENT OF DISTRIBUTED GENERATION SOURCES IN ELECTRIC DISTRIBUTION SYSTEMS USING HURWITZ CRITERION"
- [4] R.Varikuti and M. Damodar Reddy, "Optimal placement of DG units using fuzzy and real coded genetic algorithm," Journal of Theoretical and Applied Information Technology, pp. 145-151. Vol. 7, No. 2, 2009.
- [5] M. T. Ameli, V. Shokri and S. Shokri, "Using fuzzy logic & full search for distributed generation

allocation to reduce losses and improve voltage profile", International Conference on Computer Information Systems and Industrial Management Applications (CISIM), pp. 626- 630, 2010.

- [6] C. Wang and M. H. Nehrir, "Analytical approaches for optimal placement of distributed generation sources in power systems," IEEE Transactions on Power Systems, Vol. 19, No. 4, pp. 2068–2076, Nov. 2004.
- [7] Y. M. Atwa, E. F. El-Saadany, M. M. A. Salama and R. Seethapathy, "Optimal renewable resources mix distribution system energy loss minimization", IEEE Transactions on Power Systems, Vol. 25, No. 1, February 2010.
- [8] S. R. A. Rahim, T. K. A. Rahman, I. Musirin, S. A. Azmi, M. F. Mohammed, M. H. Hussain and M. Faridun, "Comparing the network performance between the installation of DG and compensating capacitor using EP," in International Journal of Power, Energy and Artificial Intelligence, Vol. 1, pp. 14-20, August 2008.
- [9] Fl. Rotaru G. Chicco, G. Grigoras and G. Cartina, "Two-stage distributed generation optimal sizing with clustering-based node selection", Electrical Power Systems & Energy Systems, in press.
- [10] G. Grigoras, G. Cartina and Fl. Rotaru, "Using k-Means Clustering Method in Determination of the Energy Losses Levels from Electric Distribution Systems", World Scientific and Engineering Academy and Society (WSEAS), Section: Mathematical Methods and Computational Techniques in Electrical Engineering, pp. 52 - 56, ISSN: 1792-5967, ISBN: 978-960-474-238-7, Timisoara, 2010.
- [11] S. Ray and R.H. Turi, "Determination of Number of Clusters in K-Means Clustering and Application in Colour Image Segmentation", Available:www.csse.monash.edu.a u/~roset/papers/cal99.pdf.

~~