Asian Journal of Electrical Sciences
ISSN: 2249-6297 (P) Vol.8, No.2, 2019, pp.11-15
© The Research Publication, www.trp.org.in
DOI: https://doi.org/10.51983/ajes-2019.8.2.2365

Application of K-Means Clustering for Designing Power Distribution Network

Shilpa Bansal[1*], Neha Gupta[2] and Shelly Vadhera[3]
[1]PG Student, [2]Assistant Professor, Department of Electrical Engineering,
[1&2]Om Institute of Technology and Management (G.J.U) Haryana, India
[3]Associate Professor, Department of Electrical Engineering, National Institute of Technology, Haryana, India
*Corresponding Author
E-Mail: bansalshilpa123@gmail.com

(Received 14 February 2019; Revised 7 March 2019; Accepted 28 March 2019; Available online 5 April 2019)

Abstract **- Installation of independent power distribution network with central generation using solar power plant is a trending solution to meet the demands of electricity of energy starved rural areas. In this paper, application of K-means clustering is proposed for designing power distribution network which segregates the houses in suitable number of clusters and all the clusters centroids are connected to the central location of solar power plant. Total number of clusters is selected by application of three methods like Elbow method, CH index method and Average Silhouette method. This paper can act as a useful tool for planning and designing the distribution network while optimizing the distribution losses and installation cost of the network.**

Keywords: **Elbow Method, K-Means Clustering, Power Distribution, Rural Electrification**

I. INTRODUCTION

Out of 195 countries, most of the developed countries have full access to electricity while in developing countries like India the electrification is only 84 percent as mentioned in World Bank [1]. Rural areas located at remote locations are not connected to central grid due to unavailability of necessary infrastructure and supply needs. However, continuous works and efforts put by government and private organizations are improving the situation with time. Many exemplary set-ups of independent microgrid like electrification in villages of Karnataka by Selco company [2], electrification in Rajasthan villages mentioned by Ahuja [3] etc. are catering to the day to day electricity needs of the area satisfactorily. [4] Shows that the commercially available solar power plants have maximum efficiency of 15% only. So, this small value of efficiency demands minimization of losses during distribution stage so that maximum amount of power can be harnessed for useful work. The most feasible way of reducing distribution losses is to design a power distribution network which uses minimum length of conductor while supplying electricity to all the houses. For designing a power distribution network, the settlement of houses over the area needs to be analyzed in such a way that the houses can be grouped in number of clusters and each cluster has single distribution centre. Various clustering techniques like K-means clustering, Hierarchical clustering etc are available for data mining. K-means clustering is one of the oldest and popular clustering techniques which have been used in solving large data handling in many fields for years. Clusters are created by K-means clustering in such a way that each point is nearest to its cluster centroid as compared to other centroids. Reference [5] shows the use of K-means clustering in selection of launch locations to reduce delivery time. Useful work in segregating students data based on academic performance parameters was proposed by Oyelade [6]. Reference [7] shows usage of this technique in medical fields, color-based segmentation to track tumor objects in magnetic resonance brain images. Special applications of satellite data like rapid response to disaster warnings produced by image data analysis make use of K-means clustering as mentioned in [8]. Reference [9] shows the usage of this technique in power systems in which detection of abnormal conditions which may lead to blackout was done by creating pattern cluster. In this paper, K-means clustering technique is used in designing a power distribution network which distributes centrally generated power by solar power plant.

The configuration of the network should be designed in such a way that it uses minimum length of conductors. Optimizing the length of conductor will also reduce the copper losses as well as installation cost. This technique groups the data points (house locations) in different clusters and each cluster is given power from the pole located at its centroid. Then, all the poles are connected to central location of solar power plant. The essential parameter, the number of clusters, needs to be provided by the designer to execute K-means clustering algorithm. There is no universal method for deciding the suitable number of clusters. Three direct methods Elbow method, CH index method and Average silhouette method are used for selecting the suitable number of clusters. The results are compared for actual house locations of a rural area, Jaitgarh village, Madhya Pradesh, India. Although this comparison is done for a specific rural area, the proposed approach can be used for designing a power distribution network for any energy starved rural area.

II. K-MEANS CLUSTERING

K-Means clustering partition n data points into k clusters in which each data point belongs to the cluster with the nearest mean as proposed by Kaufman [10]. This method generates k different clusters of greatest possible distinction where number of clusters k needs to be computed from the data. The objective of K-Means clustering is to minimize total intra-cluster variance. The algorithm of K-means clustering algorithm is given below:

1. Clusters the data points into k groups where k is predefined.
2. Select k points at random as cluster centers.
3. Assign data points to their closest cluster center according to the Euclidean distance function.
4. Calculate the centroid or mean of all objects in each cluster.
5. Repeat steps 2, 3 and 4 until the same points are assigned to each cluster in consecutive rounds.

K-Means clustering is relatively an efficient method. The final results obtained by this method are sensitive to initialization. There is no global theoretical method to find the optimal number of clusters. A practical approach is to compare the outcomes of multiple runs with different k and choose the best one based on a predefined criterion. Larger value of k decreases the error but increases the cost involved in installing the network.

III. DETERMINATION OF OPTIMAL NUMBER OF CLUSTERS

Determining the optimal number of clusters in a data set is a fundamental issue in K-means clustering where total number of clusters k needs to be decided. The optimal number of clusters depends on the method used for measuring similarities and the parameters used for partitioning. Three direct methods namely Elbow method, CH index method and Average silhouette method are used to determine value of k which optimize a particular criterion which is within cluster variance, Between cluster variance and the average silhouette width respectively.

A. Elbow Method: Kodinariya and Dan [11] propose a similarity parameter which is used in this method i.e. ‘within-cluster variation’. The K-means algorithm minimizes the within-cluster variation (W) which is given by equation (1).

$$W = \sum_{k=1}^{K} \sum_{C(i)=k} (x_i - x_k)^2 \qquad (1)$$

Where x_k is the average of points in cluster k and is given by

$$x_k = \frac{1}{n} \sum_{i=1}^{n_k} x_i$$

Where
n= total number of data points
n_k= number of data points in k^{th} cluster.
When we plot the variation of W with k, we observe a bend in the graph which is named as Elbow point of the curve. This point is considered as an indicator of the appropriate number of clusters. But implementing this on the real data, gives a continuous curve with no visible elbow. This creates ambiguity in determining the value of k.

B. *CH index method:* Within-cluster variation measures the extent by which the clusters are packed together. As we increase the number of clusters K, this parameter goes down. However, Between-cluster variation measures how spread apart the groups are from each other which is given by equation (2).

$$B = \sum_{k=1}^{K} n * (x_k - x_c)^2 \qquad (2)$$

Where
x_k is the average of points in group k, and x_c is the overall average given by

$$x_k = \frac{1}{n_k} * \sum_{i=1}^{n_k} x_i$$

$$x_c = \frac{1}{n} * \sum_{i=1}^{n} x_i$$

The highest value of B is selected as the clusters should be well separated from each other. But like with-in cluster variance, Between-cluster variance also increases continuously and cannot be the only deciding criteria for choosing k. The method suggested by Caliński is CH index method which uses with-in cluster variance and between cluster variance both [12]. The CH index looks at a ratio of Between-clusters to Within-clusters. The formula is given by equation 3.

$$CH(k) = \frac{B(k)/(k-1)}{W(k)/(n-k)} \qquad (3)$$

For a range of values, that value of k is selected which gives the largest score CH (k), i.e.

$$k = \text{Arg max } CH(k)$$

C. *Average Silhouette Method:* The average silhouette approach determines how well each object lies within its cluster. A high average silhouette width indicates a good clustering. Average silhouette method computes the average silhouette of observations for different values of k. The formula for calculating the silhouette value is given by equation 4.

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (4)$$

Where b(i) is minimum distance between ith element in cluster P and elements of cluster Q (Fig 1) i.e.

$$b(i) = \min_{R \neq P} d(i, R)$$

And a(i) = Average distance between ith element and other elements of the same cluster. i.e.

$$a(i) = \text{avg}_P\, d(i, P)$$

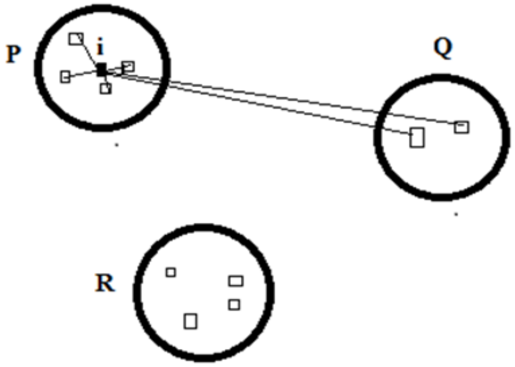

Fig. 1 Illustration of Average Silhouette method

In general,

$$-1 \leq s(i) \leq 1$$

The negative value of s (i) indicates the worst solution. And positive value close to 1 shows the best solution. A value close to zero gives a choice for the data point to be kept in either cluster P or cluster R. The average silhouette width as the average of the s (i) for all objects i belonging to that cluster. The optimal number of clusters k is the one that maximize the average silhouette over a range of possible values for k. Rousseeuw explains the algorithm as follows [13]:

1. Execute k-means clustering for different value of K.
2. For each k, calculate the average silhouette of observations.
3. Plot the curve between Average silhouette Width and number of clusters k.
4. The location of the maximum is considered as the appropriate number of clusters.

As most of the clustering algorithms combine natural clusters in order to reduce the total number of groups to the specified value of k. If k is set too low, the artificial fusions which are formed by merging clusters with larger dissimilarities are yielded with narrow silhouette width and hence are exposed in the plot.

On the other hand if k is set too high, some natural clusters have to be fragmented in an artificial way, in order to conform to the specified number of groups. However, these artificial fragments will also yield narrow silhouette width because for such fragments 'Between cluster' dissimilarities b (i) will become very small, which also results in small silhouette values. This method shows that the silhouettes returns best natural value of k with a larger average silhouette width.

IV. RESULTS AND DISCUSSION

Jaitgarh is a small Village in Silwani Tehsil in Raisen District of Madhya Pradesh State, India. It is located 93 KM from State capital Bhopal. Jaitgarh Village Total population is 338 and number of houses are 65. Reference [14] reports that electricity is not available to this village and to other villages located near it. The house locations of village Jaitgarh village are obtained from Google Earth accessed on April 24th, 2019. Fig. 2 shows the satellite view of location of houses in the village.

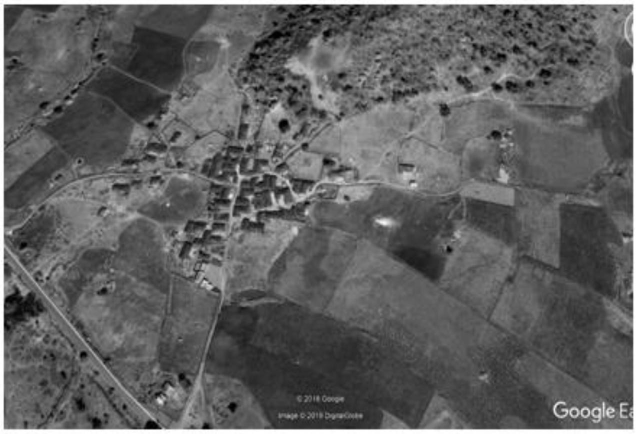

Fig. 2 Satellite view of Jaitgarh village

Application of method 1, Elbow method generates a graph between Within-cluster variations with respect to number of clusters. The resulting graph is shown in Fig.3.

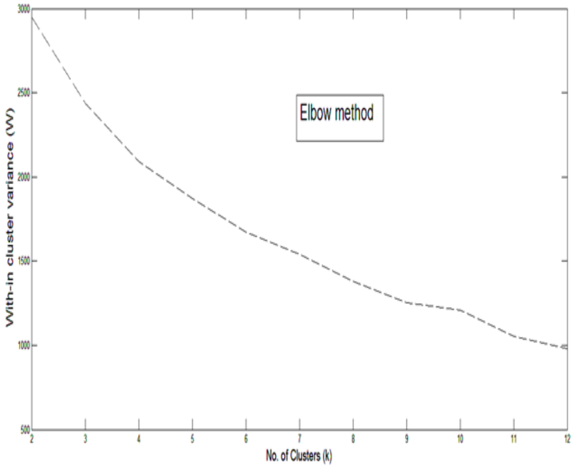

Fig. 3 Graph between Within-cluster variance and K

Fig. 4 shows the variation of Between-cluster variance with different value of k. Fig. 5 shows the resulting graph between CH index versus K as obtained from method 2. Fig.6 shows the variation of Average silhouette width variation with number of clusters as obtained from method 3.

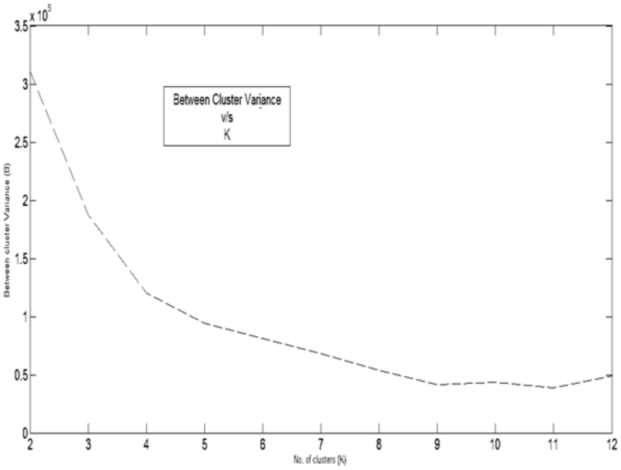

Fig. 4 Graph Between cluster variance and k

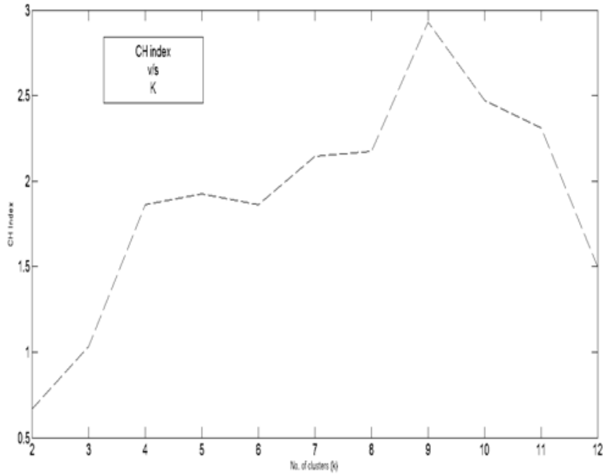

Fig. 5 Graph between CH index and k

The number of clusters suggested by elbow method is 4, 9. The ambiguity in selection of the elbow point of the curve is there due to no clearly visible single elbow point. CH index shows its highest value at K=9 however local maxima are observed at K=4, 7 where values of the index is 1.7, 2.3 respectively as compared to value at k=9 i.e. 2.8. S it can be seen that first two method suggest value of K= 4, 7, 9 the best value is chosen out of these three options using average Silhouette width. Table I shows the Average Silhouette width for shortlisted value of K.

TABLE I AVERAGE SILHOUETTE WIDTH FOR VALUE OF K

No. of Clusters	Average Silhouette Width
4	0.58264
7	0.55938
9	0.52414

The highest silhouette width obtained for k= 4. It means that naturally 4 clusters exist in the data set whereas 7 or 9 clusters are created by artificial fragmentation for increasing the number of clusters. So, the optimum number of clusters for Jaitgarh village is 4.

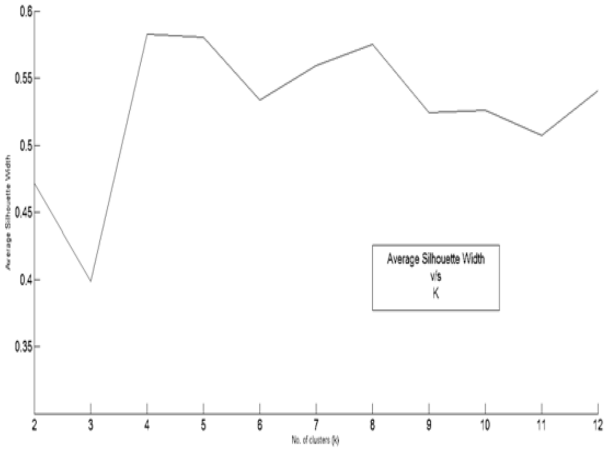

Fig. 6 Graph between Average Silhouette Width and k

V. CONCLUSION

The cluster diagram obtained from K-means clustering for K=4 are shown in Fig.7. All the data points are connected to their respective cluster centroid (◊). Then all the centroids are connected to the central location which is the proposed location of solar power plant for the village. The total length of conductor used is 2527.8 meter. Through this paper, the optimum number of clusters has been found by comparing results of 3 methods and the most suitable configuration for power distribution network is suggested by K-means clustering technique.

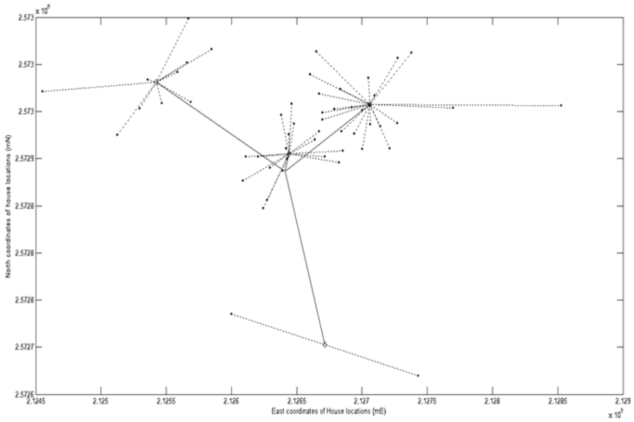

Fig. 7 Resulting cluster diagram for Jaitgarh Village with 4 clusters

The total length of conductor is dependent on the distribution of data points in the clusters. And K-means clustering lacks consistency and generates a number of different distributions for the same data because of its dependency on choice of initial centroids. Therefore, the clustering technique is executed 50 times to find the best solution. The configuration obtained from the technique facilitates the central generation of power with a choice between distributed maintenance at various centroids or centralized maintenance services at the plant itself. Downsizing the within-cluster variance by K-means also reduces the length of conductors in the same proportion. Further, the savings in conductor lead to proportional savings in installation cost, distribution losses and conductor replacement cost.

REFERENCES

[1] World Bank, “Sustainable Energy for All (SE4ALL)”, database from the SE4ALL Global Tracking Framework led jointly by the World Bank, *International Energy Agency, and the Energy Sector Management Assistance Program*, [Online] Available: https://data. world bank.org/indicator/eg.elc.accs.zs

[2] S. Mukherji and U. Jumani, SELCO “Solar Lighting for the Poor - Growing Inclusive Markets”, 2011, [Online] Available: http://www. growinginclusivemarkets.org/media/cases/India_SELCO_2011.pdf.

[3] M. Ahuja “Solar power lights up remote villages in Rajasthan” in The Hindustan times, 2017, [Online] Available: https://www. hindustantimes.com/jaipur/solar-power-lights-up-remote-villages-in-rajasthan/ story-WJSQE0dBdhEba1E1UcbpuK.html

[4] B. D. Sharma, “Performance of solar power plants in India” by Central Electricity Regulatory Commission, New Delhi, Jan 2011, [Online] Available: http://www.cercind.gov.in/2011/Whats-New/ Performance %20of%20solar%20power%20planTS.pdf

[5] S. M. Ferrandez, H. Timothy, T. Weber, R. Sturges and R. Rich, “Optimization of a truck-drone in tandem delivery network using K-means and genetic algorithm” in *Journal of Industrial Engineering and Management*, Vol. 9 No. 2, pp. 374-387, 2016

[6] J. Oyelade, O. Olufunke and I. Obagbuwa, “Application of K-means Clustering algorithm for prediction of Students Academic Performance” *in International Journal of Computer Science and Information Security*, Vol. 7, No. 1, pp. 292-295, 2010.

[7] M. Wu, C. Lin and C. Chang, “Brain Tumor Detection Using Color-Based K-Means Clustering Segmentation”, in *Third International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pp. 245-250, 2007.

[8] M. Yang, H. Mei and D. Huang “An effective detection of satellite images via K-means clustering on Hadoop system” *in International Journal of Innovative Computing, Information and Control*, Vol. 13, pp. 1037-1046.

[9] O. Ozgonenel, D. W. P. Thomas, T. Yalcin and I. N. Bertizlioglu, “Detection of blackouts by using K-means clustering in a power system” in *11th IET International Conference on Developments in Power Systems Protection (DPSP 2012),* Birmingham, UK, pp. 1-6, 2012.

[10] L. Kaufman and P. Rousseeuw, “*Finding Groups in Data: An Introduction to Cluster Analysis*” by Wiley Series in Probability and Statistics, 1990.

[11] T. Kodinariya, M. P. R. Dan, “Review on Determining of Cluster in K-means Clustering” *in International Journal of Advance Research in Computer Science and Management Studies*, Vol. 1, pp. 90-95, 2010.

[12] T. Caliński and J. Harabasz, “A dendrite method for cluster analysis”, in *Communications in Statistics*, Vol. 3, No. 1, pp. 1-27.

[13] P. J. Rousseeuw, “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis” in *Journal of Computational and Applied Mathematics*, Vol. 20, pp. 53-65, 1987.

[14] Banerjee and R. K. Singh “Modi claims all Indian villages have electricity but India Today report says otherwise” *India Today*, April 30, 2018. [Online] Available: https://www.indiatoday.in/india/ story/pm-modi-claims-all-indian-villages-have-access-to-electricity-india-today-ground-report-says-otherwise-1223078-2018-04-29.