PAPER • OPEN ACCESS

Electrical Load Profile Analysis Using Clustering Techniques

To cite this article: R Damayanti *et al* 2017 *IOP Conf. Ser.: Mater. Sci. Eng.* **180** 012081

Electrical Load Profile Analysis Using Clustering Techniques

R Damayanti[1], A G Abdullah[1*], W Purnama[1], A B D Nandiyanto[2]

[1]Departemen Pendidikan Teknik Elektro, Fakultas Pendidikan Teknologi dan Kejuruan, Universitas Pendidikan Indonesia.

[2]Departemen Pendidikan Kimia, Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam, Universitas Pendidikan Indonesia.

*ade_gaffar@upi.edu

Abstract. Data mining is one of the data processing techniques to collect information from a set of stored data. Every day the consumption of electricity load is recorded by Electrical Company, usually at intervals of 15 or 30 minutes. This paper uses a clustering technique, which is one of data mining techniques to analyse the electrical load profiles during 2014. The three methods of clustering techniques were compared, namely K-Means (KM), Fuzzy C-Means (FCM), and K-Means Harmonics (KHM). The result shows that KHM is the most appropriate method to classify the electrical load profile. The optimum number of clusters is determined using the Davies-Bouldin Index. By grouping the load profile, the demand of variation analysis and estimation of energy loss from the group of load profile with similar pattern can be done. From the group of electric load profile, it can be known cluster load factor and a range of cluster loss factor that can help to find the range of values of coefficients for the estimated loss of energy without performing load flow studies.
Keywords: electric load profile; k-means; fuzzy c-means; k-harmonic means; davies-bouldin index; loss factor; load factor.

1. Introduction

West Java is one of the provinces which consists of various types of loads and electricity consumers, such as households, businesses, industries, and commercial. Electrical power distribution is used as a power source and a wide range of equipment requires electrical power as an energy source. The equipment can generally be light (illumination), load power (for electric motors), heating and electronic equipment resources. A load profile contains information about the amount of energy that burdens a station electricity provider, and this data is usually expressed in kW or MW. This electrical load data are usually made within certain time intervals: every 10, 15 or 30 minutes [1]. Identification and prediction of the load profile are not only the need for power generation and distribution companies, but also useful for the independent generation of small entities, such as micro grid owner who has been there and will continue to grow in the liberal electricity market [2].

Clustering is one method that is included in the field of data mining applications and can be used to classify electrical load profile. However, if the clustering technique is used to obtain electricity customers load pattern, then a suitable clustering algorithm is chosen, and determining the appropriate number of clusters is one of the important things to be considered [3]. In the electricity consumption data, there are various types of load curve shapes, so that the clustering algorithm can be used to identify and classify the load profile similar. Clustering techniques classify the patterns that have much in common in the same group, and this can help to analyze the loss of energy estimates in accordance with different load patterns [4].

Load factor and loss factor of the load profiles that have been classified are used to analyze the pattern of peaks and valleys of the load profile. It aims to identify variations in the demand load on each different group. Cluster load factor and cluster loss factor are used to define load factor and loss factor exclusively for a similar type of load profile [5]. By using the appropriate clustering algorithm, the same load profile on a different day of an area can be grouped. Besides, loss factor is also influenced by the peak load [6].

Based on the background above, paper compared three methods of clustering techniques to get the right method to classify the electrical load profile of West Java during 2014. The methods are K-Means Clustering, Fuzzy C-Means Clustering, and K-Harmonic Means Clustering. To determine the optimal number of clusters, this study uses Davies-Bouldin Index to identify variations in load demand and help to find the range of coefficients to estimate the value of the energy load pattern which is varied in each cluster load profile.

2. Clustering Techniques

Clustering is one method that is included in the field of data mining applications. Clustering is a learning process with the objective of determining the specific attributes of the resulting group. There are various types of algorithms appropriate to the different problems. However, there is no clustering algorithms that are universally recognized as a different algorithm. It shows different functionality in determining the adequacy of measures better [7].

Clustering algorithms can be used in various types of applications, such as image processing, finance, biology, medicine or the optimization of the web [8]. Clustering can be defined as the process of partitioning a large database into groups based on the concept of similarity or proximity between data. Good Grouping is a grouping which has a high similarity to a particular group and has the low similarity among the objects of the other classes [9].

Clustering Data aims to find the structure in a heterogeneous collection of data. These structures define groups or classes of data together in a group and different from the other group. The end result of the clustering algorithm or methodology linkage relationships depends on classification criteria that are used to separate the same data and different [8].

Cluster analysis is also a tool to explore the data structures. Therefore, cluster analysis may declare relationships and structures in the data, which may not be known in advance, and may be particularly useful when it is found. Cluster analysis has been used in various disciplines such as pattern recognition, image processing, information retrieval, marketing and much more [10].

2.1. K-means clustering

K-means clustering is a method of classifying the load profile data to determine the number of clusters and a central point for each cluster. After determining the center point of each cluster, each data set should be placed closest to the center point and then calculated back from the center point of the new iteration until a stable position of the center point [11].

Assigning the data to a central point which is evaluated using the Euclidean, City block, Cosine, Correlation, or Hamming distance, automatically makes the boundary between each set of data. Each specific set of data will become a member of a nearby cluster after the first iteration. The next iteration only has the function of updating the position of the center point.

This method does not create a tree structure to describe the data grouping, but to create a single level of the cluster. K-means uses actual observations of the object or individual data and therefore more suitable for classifying large amounts of data.

2.2. Fuzzy c-means clustering

This method is similar to a standard K-means described above, but the difference is that each set of data has a degree of membership for each initial cluster, i.e. each set of data included in all of the group to some degree. Degree of membership for each set of data for all cluster must be equal to one [11].

The procedure begins by determining the number of clusters and guessing the focal point of the group (probably not true), which is intended to mark the location of the average of each cluster, and then assign each data set class membership for each cluster. The next step is to update each cluster and the central point class membership iteratively until a stable position of the center point. In this step iterative cluster center point is moved to the correct position in the data set.

Fuzzy C-means clustering does not make the boundaries between sets of data for the first iteration, because the process involves grouping all the data. The limits will automatically develop when the grouping process is completed. Compared with the method of K-means clustering, fuzzy c-means has a longer process, because the iteration process updates not only the central point but also the degree of membership of each data set.

2.3. K-harmonic means clustering

K-Harmonic Means Clustering (MIC) is a method introduced by Zhang, Hsu, and Dayal made to overcome the problems that exist in the K-means clustering [12]. MIC is one example of center-based cluster and is a method in which the groups are formed by iterative refinement based on the location of the center point of each cluster. In KHM, objective function value is generated by finding the average total harmonic of all the data points to the distance between each data point to all existing cluster center point [13].

This is different from the K-means method wherein the objective function is obtained from the total distance of all data to a central point group. In a harmonic function, if there is one member of the small value, the average value of any harmonic is small, but if no member is small then the value is too large [14]. Average harmonic is very sensitive to the situation where there are two or more adjacent center point. This method naturally puts one or more areas to the central point of data points that are far from the central points that existed previously. This will make the objective function will be smaller [15].

3. Methods

The first step to do in this study is to collect the daily load profile data for the year 2014. Once the data is obtained, the process of grouping the load profile was conducted by using three pieces of clustering methods and broken into several classes, namely two to ten classes.

After the process of grouping, further evaluation of clustering results of the three methods was conducted. The evaluation is done by comparing the sum of squares within-group and the sum of squares between-group. The best method is a method that has the smallest ratio of standard deviation [16]. The standard deviation in the group can be calculated by equation (3.1).

$$S_w = \frac{1}{c}\sum_{i=1}^{c} S_i \tag{3.1}$$

Similarly sum of squares between-group can be calculated using equation (3.2).

$$S_b = \left[(c-1)^{-1}\sum_{i=1}^{c}(\bar{x}_i - \bar{x})^2\right]^{\frac{1}{2}} \tag{3.2}$$

The next stage determines the optimal number of clusters based on indicators of the validity of clustering, Davies-Bouldin index which can be expressed in the equation (3.3) as follows [17]:

$$D = \frac{1}{K}\sum_{i=1}^{K} m_{i \neq j} \left(R_{i,j}\right) \quad (3.3)$$

$R_{i,j}$ is a measure of how well the ratio of the value of the ratio between cluster i-th and j-th cluster. Its value is obtained from a component of cohesion and separation that exist in equation (2.7) and (2.8). Best Cluster is a cluster that has the smallest possible cohesion and separation as possible. $R_{i,j}$ is formulated by equation (3.4).

$$R_{i,j} = \frac{SS_i + SS_j}{S_{i,j}} \quad (3.4)$$

The criteria for the optimum number of groups given by the smallest value of DBI (non − negative ≥ 0).

Having in mind that the optimum number of groups, load profile analyzes performed on each group by observing the peak load, so it can be loaded and loss factor of each group. Until finally, from this study we can conclude and give advice, both for further research and application of the results obtained in the study.

4. Findings and discussion

4.1. Grouping electrical load profile using clustering techniques

After grouping using all three methods, further evaluation of the results of clustering considers the value of the ratio of the average Sum of squares within-group and the sum of squares between-group.

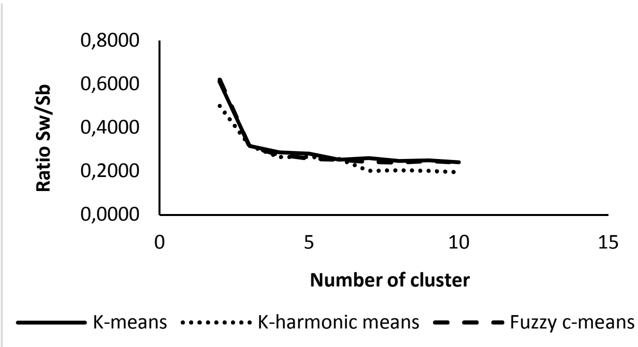

Figure 1. Ratio of Sw / Sb of the three methods

Figure 1 shows the ratio Sum of squares within-group and the sum of squares between-group of the three methods used to classify electrical load profile. K-means clustering method has a standard deviation of the ratio of the highest compared with Fuzzy c-means and K-harmonic means, which is 0.3046. Fuzzy c-means has an average value ratio Sw / Sb at 0.2987. While the lowest ratio of Sw / Sb shown by method K-harmonic means. It can be concluded that in this case the most appropriate method to the average ratio of standard deviation in the group and the lowest standard deviation among the groups is the method of K-harmonic means, which is 0.2676.

4.2. Determine the optimum number of clusters

Having in mind that K-harmonic means clustering method is the most appropriate method, then the determination of the optimum number of clusters is performed. Cluster validity index which is used to determine the optimum number of clusters is Davies-Bouldin Index, where the optimum cluster number is given on the value of the smallest DBI. Figure 2 is DBI value of K-harmonic means clustering.

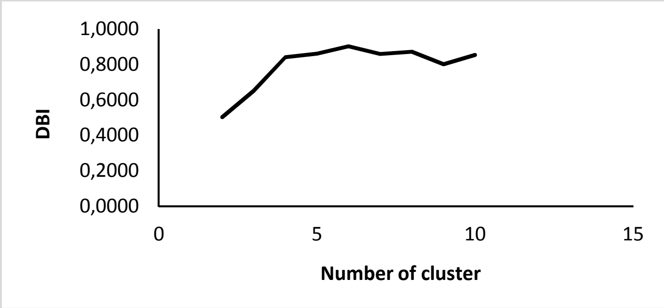

Figure 2. Davies-Bouldin Index

Based on Figure 2, it can be seen that the smallest value of DBI is when the number of clusters is equal to 2. This indicates that the optimum number of clusters to classify electrical load profiles using the K-harmonic means is 2.

4.3. Electrical load profile analysis

Figure 3 shows the results of clustering using K-harmonic means with the number of cluster is 2, so that it is known that West Java has two patterns of consumption average load difference. The 1st cluster consists of 28 days, in which 22 days are a holiday and the rest are weekdays. The pattern of the load profile in this cluster tend to be irregular and the range of maximum load is only about of 4000 kW. While in 2nd cluster, the pattern of consumption load profiles describes the load on weekdays, where the amount of activity that caused the load demand is higher than the first cluster.

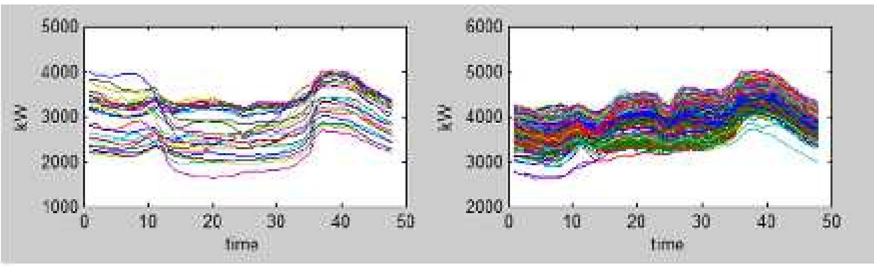

Figure 3. Clustering results of 02 clusters with KHM

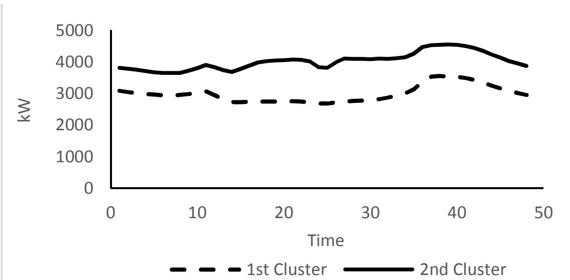

Figure 4. Typical load profiles of 02 clusters

Based on Figure 4, it can be seen from both of typical load profiles that there are differences in load consumption patterns. The 1st cluster is marked with dashed lines indicating that at 07:00 until 17:00 the load demand tends to be stable, ranging in 2770 kW. Meanwhile, the 2nd cluster is shown in black line showing the load demand 12.00 until 13:00 and indicating a decrease load demand. Overall, from 00.30 to 24.00 the 2nd cluster has a higher value load demand than the 1st cluster.

From both clusters, the load factor of each cluster is known using equation (4.1):

$$L_f = \frac{L_a}{L_m} \tag{4.1}$$

With the load factor, the interval of loss factor can be known using equation (4.2):

$$L_f^{2} \leq L_s \leq L_f \tag{4.2}$$

Where:

L_f	=	Load factor
L_a	=	Average Load demand
L_m	=	Maximum load demand
L_s	=	Loss factor

Table 1. Information of Cluster load, cluster loss factors, and maximum demand

Cluster No. (no. of days)	Cluster LF	Range of Cluster Loss Factor	Max. Demand (kW)	Day & Time of Max. Demand
1 (28)	0.7407	0.5487 - 0.7407	4027.380	23 Feb (19.00)
2 (333)	0.7982	0.6371 - 0.7982	5037.260	22 Oct (20.00)

Table I shows variations of the load demand and value ranges for estimated energy loss coefficients of each cluster. The 1st cluster has a variation of load demand which is smaller than the 2nd cluster, which is 0.7407 with a range of cluster loss factor 0.5487 - 0.7407. This is because the demand load in the 1st cluster is a holiday, where the load is dominated by a group of domestic and commercial sectors which tend to have capacitive load. Maximum demand in the 1st cluster occurred at 19:00 on Sunday that on 23 February. In the 2nd cluster, load factor value reached 0.7982 with range of estimates of the energy loss coefficient of 0.6371 - 0.7982, which was higher than in the 1st cluster. This is because the second cluster consists of working days, which dominates with inductive load due to many industries operating. The maximum load demand in the 2nd cluster occurred at 20:00 on Wednesday, October 22.

The load factor shows the variation of load demand to the feeder while the loss factor can help to estimate the average energy lost in the distribution system without load flow studies. By knowing and understanding the load of consumption patterns can help to regulate the load demand. The electricity provider can predict and anticipate load demand variations that might occur on certain days.

5. Conclusions and recommendations

Clustering electrical load profiles can be done using three methods of clustering techniques, namely K-means clustering, fuzzy c-means and K-harmonic means. From the three methods, K-harmonic means is the best method to classify the electrical load profile. This is indicated by the ratio of sum of squares within-group and the sum of squares between-group has the lowest value compared with the method of K-means and Fuzzy c-means.

To determine the optimum number of clusters, the indicator of the validity of clustering, the Davies-Bouldin Index, is used. Criteria for the optimum number of groups are indicated by the value of the smallest DBI contained in cluster totaling 2.

From both clusters, electrical load profile is generated. It can be seen that each group has a different load consumption patterns, as shown by the typical load profile of each group. The 1st cluster patterns tend to be irregular load profile which is dominated by the holidays, where most office and industrial are closed. While in the 2nd cluster, the pattern of consumption load profiles describe the burden on weekdays, where the number of offices and industrial activities that caused the load demand are higher than the 1st cluster, but in this cluster there are a few days off the entrance into it. This shows that the demand pattern of irregular loads occurs not just on holidays, but the irregular load demand on normal days can also occur, and vice versa.

6. Recommendations

Data mining is the process of extracting information that can be done in any field of science. Electric load profile is a set of data that can be used to gather information that could be useful in the future. One data mining techniques in the form of a partition-based clustering techniques have been used in this study. The authors suggest that this study was developed by trying to classify electrical load profile based on the type of customers, so the information obtained from the analysis of the electrical load profile can be more beneficial for utilities and power providers.

References

[1] Keka I, Hamiti M. Load Profile Analyses Using R Language. Int Conf Inf Technol Interfaces. 2013;245–50.
[2] Zhu Y, Lu S. Load profile disaggregation by Blind Source Separation : a Wavelets-assisted Independent Component Analysis Approach. 2014;
[3] Zhang T, Zhang G, Lu J, Feng X, Yang W. A New Index and Classification Approach for Load Pattern Analysis of Large Electricity Customers. IEEE Trans Power Syst. 2012;27(1):153–60.
[4] Xu R, Member S, Ii DW. Survey of Clustering Algorithms. IEEE Trans NEURAL NETWORKS. 2005;16(3):645–78.
[5] Sharma DD, Singh SN. Electrical Load Profile Analysis and Peak Load Assessment using Clustering Technique. IEEE. 2014;
[6] Mikic OM. Variance-Based Energy Loss Computation in Low Voltage Distribution Networks. IEEE Trans Power Syst. 2007;22(1):179–87.
[7] Panapakidis IP, Christoforidis GC. Modifications of the Clustering Validity Indicators for the Assessment of the Load Profiling Procedure. 4th Int Conf Power Eng Energy Electr Drives. 2013;(May):1253–8.
[8] Gavrilas M, Gavrilas G. Application of Honey Bee Mating Optimization Algorithm to Load Profile Clustering. 2010;
[9] Fayyad U, Piatetsky-shapiro G, Smyth P. From Data Mining to Knowledge Discovery in. American Association for Artificial Intelligence. 1996;17(3):37–54.
[10] Othman MN, Sohod MH, Mara UT. Consumer Load Profiling using Fuzzy Clustering and Statistical Approach. Student Conf Res Dev. 2006;(SCOReD):27–8.
[11] Chicco G, Napoli R, Piglione F. Application of Clustering Algorithms and Self Organising Maps to ClassifL Electricity Customers. IEEE Bol PowerTech Conf. 2003;
[12] Zhang B, Hsu M, Dayal U. Clustering Algorithm K -Harmonic Means -A Data Clustering Algorithm. 1999;
[13] Tian Y, Liu D. K-harmonic means data clustering with Differential Evolution. Int Conf Futur Biomed Inf Eng. 2009;369–72.
[14] Mathematics A. K -harmonic means data clustering with simulated annealing heuristic. Appl Math Comput. 2007;184:199–209.
[15] Yang F, Sun T, Zhang C. An efficient hybrid data clustering method based on K-harmonic means and Particle Swarm Optimization. Expert Syst Appl [Internet]. Elsevier Ltd; 2009;36(6):9847–52. Available from: http://dx.doi.org/10.1016/j.eswa.2009.02.003
[16] Bunkers MJ, Miller JR. Definition of Climate Region in the Northern Plains Using an Objective Cluster Modification Technique.pdf. J Clim. 1995;9:130–46.
[17] Davies DL, Bouldin DW. A Cluster Separation Measure. IEEE Trans Pattern Anal Mach Intell. 1979;(2):224–7.