

© 2007 IEEE. Reprinted, with permission, from Prahastono, I., King, D.J. and Ozveren, C.S. A review of Electricity Load Profile Classification methods. In: *Proceedings of the 42nd Universities Power Engineering Conference, Brighton, 4-6 September 2007*. pp. 1187-1191. 2007.

This material is posted here with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of the University of Abertay Dundee's products or services. Internal or personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution must be obtained from the IEEE by writing to pubs-permissions@ieee.org.

By choosing to view this document, you agree to all provisions of the copyright laws protecting it.

A Review of Electricity Load Profile Classification Methods

Iswan Prahastono, Dr. D. King and Dr. C.S. Özveren

University of Abertay Dundee, Scotland - UK

ABSTRACT

With the electricity market liberalisation in Indonesia, the electricity companies will have the right to develop tariff rates independently. Thus, precise knowledge of load profile classifications of customers will become essential for designing a variety of tariff options, in which the tariff rates are in line with efficient revenue generation and will encourage optimum take up of the available electricity supplies, by various types of customers. Since the early days of the liberalisation of the Electricity Supply Industries (ESI) considerable efforts have been made to investigate methodologies to form optimal tariffs based on customer classes, derived from various clustering and classification techniques. Clustering techniques are analytical processes which are used to develop groups (classes) of customers based on their behaviour and to derive representative sets of load profiles and help build models for daily load shapes. Whereas classification techniques are processes that start by analysing load demand data (LDD) from various customers and then identify the groups that these customers' LDD fall into. In this paper we will review some of the popular clustering algorithms, explain the difference between each method.

Keywords : Electricity Load Profile Classification, Clustering Methods, Hierarchical, K-Means, Follow The Leader, Fuzzy K-Means, Fuzzy Classification.

1 INTRODUCTION

A new electricity act for electricity market liberalisation has been introduced by the Indonesian government, which will allow regions to develop their own electricity systems, including tariff making powers. The clustering and classification of customer load demand profiles becomes important, not only to design tariffs, but also to identify representative sets of standard profiles and to build models of daily load shapes. Various clustering techniques for classifying electricity customers have been identified in the literature [1–12]. The aim of all these clustering techniques is to develop groups of customers based on their behaviour and to achieve load profiling goals by starting with load demand data from these customers and then grouping them into several clusters, which have a similar profile.

In Indonesia, at present, there are no highly developed analytical methods used for electricity customer classification. The current Indonesian Load Profile Classification method uses a historical database of load demand customer profiles, collected over several years, to group customers according to their load pattern, such as industrial, business, public services and residential loads. Because of the geographic layout of Indonesia, which covers thousands of islands with differing socio-economic circumstances, the appropriate method of customer classification may have to vary between different locations.

This paper describes several clustering techniques that could be considered relevant for applying in Indonesia.

These methods are Hierarchical, K-means, Fuzzy K-means, Follow the Leader and Fuzzy Relation clustering techniques.

2 CLUSTERING METHODS

This section describes some of the reported clustering techniques, which have been used to classify electricity consumers and are likely to be more appropriate and applicable to circumstances in Indonesia.

All methods start by deriving a matrix of feature vectors for each load profile data set, followed by the procedures as described below :

2.1 Hierarchical Clustering

Hierarchical clustering groups data, simultaneously over a variety of scales, by creating a cluster tree. The tree is a multilevel hierarchy, where clusters at one level are joined to clusters at the next level.

To perform hierarchical clustering, it is necessary to find the similarity or dissimilarity between every pair of load profiles in the data and then group them into binary clusters based on the previously computed similarity matrix. The process is iteratively repeated by merging the clusters of each level into bigger ones at the upper level until all samples are grouped into expected clusters.

The advantage of this method is that the original data is kept unchanged in the root of the cluster tree [5].

The hierarchical clustering algorithm can be described in the following steps [5–6] :

1. Determine the similarity between every pair of load profile data sets.
2. Group into a pair and create a hierarchical cluster tree based on the determined similarity criterion.
3. Determine the cutting position of the hierarchical cluster tree.

Figure 1 illustrates the hierarchical cluster tree, where the horizontal axis represent the load profile data sets and the vertical axis represent the distance between clusters. A possible cutting position is indicated by a dashed line.

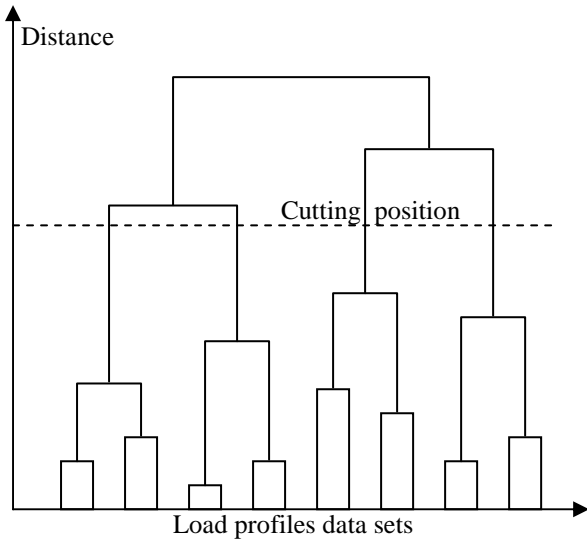


Figure 1 Hierarchical tree

Determining the similarity or the distance between load profiles can be done in various ways such as Euclidean distance, Mahalanobis distance, City Block metric, Minkowski metric and Hamming distance.

Grouping can be processed by linking pairs of load profiles that are in close proximity using linkage criterion such as Shortest distance, Average distance, Centroid distance and Ward distance, which use the previously calculated distance. As objects are paired into binary clusters, the newly formed clusters are grouped into larger clusters until a hierarchical tree is formed.

The number of groups is determined by the cutting position in the binary tree which is chosen either by the maximum distance admissible or by selecting directly the distance corresponding to the desired number of clusters.

The hierarchical clustering method is suitable if the specific number of groups is not predetermined. In fact, the cutting position will determine the number of clusters.

2.2 K-means Clustering

The K-means clustering method groups load profile data by determining a certain number of clusters and a centre point for each cluster. After determining the centre point

of each cluster, each data set should be assigned to the nearest centre point then a recalculation of the new centre point will be done iteratively until the position of the centre point is stable [5].

Assigning data to a centre point which is evaluated using Euclidean, City Block, Cosine, Correlation, or Hamming distance, automatically creates boundaries between each data set. Each particular data set will become a member of the nearest cluster after the first iteration. The next iteration only has the function of updating the centre point position.

The K-means clustering sequence is described by the flowchart in Figure 2.

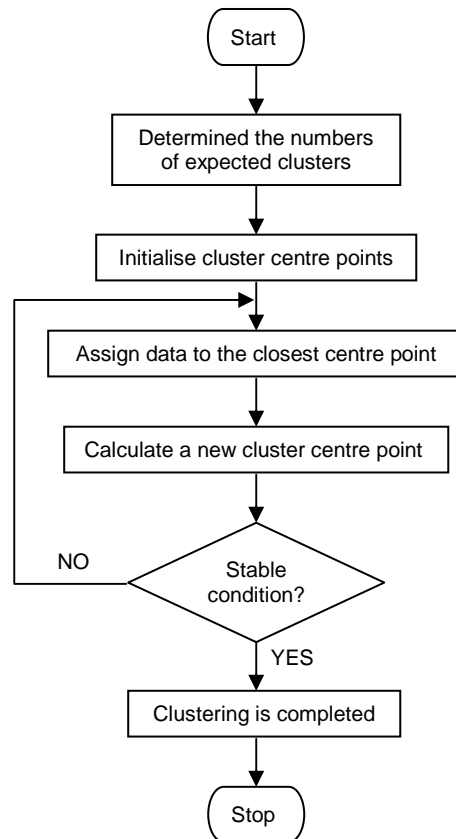


Figure 2 K-Means Clustering

This method does not create a tree structure to describe the groupings of data, but rather creates a single level of clusters.

K-means clustering uses the actual observations of objects or individual data and therefore is more suitable for clustering large amounts of data.

2.3 Fuzzy K-means Clustering

This method is similar to standard K-means described above, the difference is that each data set has a degree of membership to each initial cluster [5], i.e. each data set belongs to all clusters to some degree.

The degrees of membership for each data set to all clusters should sum to one.

The procedure starts with determining the number of clusters and guessing the cluster centre point (most likely incorrect), which is intended to mark the mean location of each cluster, then assigning every data set a membership grade for each cluster.

The next step is updating each cluster centre point and membership grade iteratively until the position of the centre point is stable. In this step the cluster centre point moves iteratively to the correct position within the data sets.

The Fuzzy K-means clustering technique does not create boundaries between data sets for the first iteration, because the clustering process involves all data. The boundaries will automatically evolve when the clustering process is completed.

Compared with the K-means method, the Fuzzy K-means process is longer, because the iteration process is not only updating the centre point but also the degree of membership of each data set.

The flowchart for this procedure is shown in Figure 3.

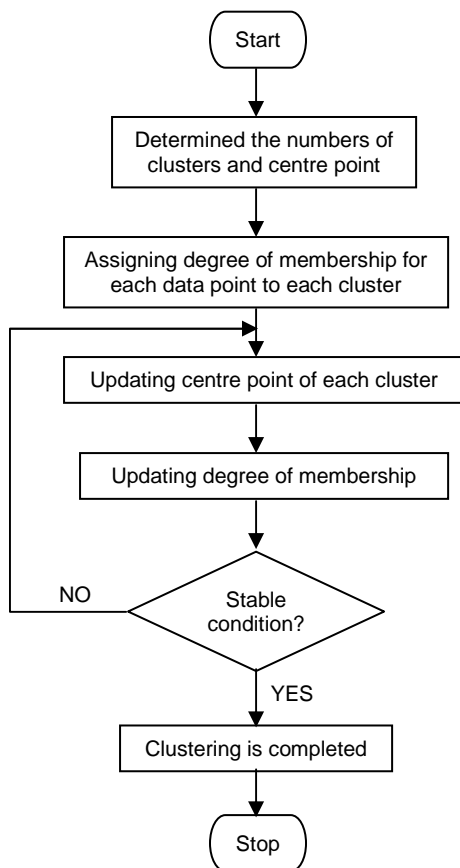


Figure 3 Fuzzy K-Means Clustering

2.4 Follow the leader Clustering

The Follow the Leader clustering method has been described in [7, 8]. This technique uses an iterative process to compute the cluster centres and it is not necessary to predetermine the number of clusters. The

number of clusters is automatically derived from the determination of the distance threshold. The process stops when the cluster centre point is stable.

The function of the first iteration of the algorithm is to determine the number of clusters and their load profile membership. It means that the first iteration also creates boundaries between load profile data sets.

The subsequent iterations adjust the load profile pattern to the nearest cluster and then updates the cluster centre point.

The selection of the distance threshold value should be done by trial and error until the expected number of clusters is produced.

Figure 4 shows the flowchart for the above process as described in [7].

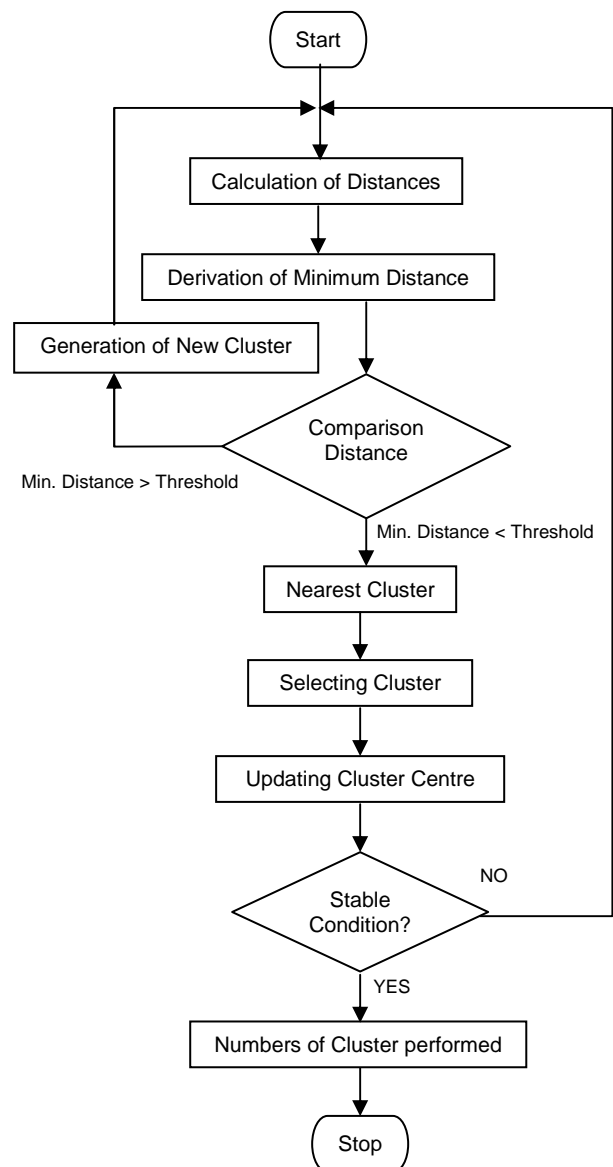


Figure 4 Follow the Leader Clustering

2.5 Fuzzy Relation Clustering

Another technique to classify electricity load demand profiles uses Fuzzy Relation clustering as describe in [9]. Actually, this is a complex iterative process that can be simplified into the following steps :

1. Determine the similarity of load profile data sets using Cosine Amplitude Method.
2. Group the load profile data sets by using Max-Min Composition Method.
3. Determine the number of clusters data by using the Lambda-Cuts for fuzzy relation Method.
4. Obtain number of cluster.

Within the above process, there is an assignment of the degree of membership for each data set and an adjustment to the cluster centre position after determining the value of the threshold in the Lambda-Cuts process.

The number of clusters is dependent on the determination of the threshold value in the Lambda-Cuts process. This value can be determined by trial and error to produce the required number of clusters.

This technique uses fuzzy relation to evaluate the similarities and to group the data sets. Therefore, the boundaries between data sets is created in the final iteration after the threshold value is determined.

This technique is suitable for handling large sets of fuzzy data with complex interactions. The numbers of customer classes being decided by the Lambda-Cuts.

3 SUMMARY

The above clustering methods are summarised below :

The Hierarchical method is more suitable for developing new electricity customer classes, because the groupings can be decided by analysing the cluster tree.

Alternatively, the Fuzzy Relation classification technique can be used, if there is a large set of fuzzy data. Customer classes can be decided by choosing an appropriate threshold value using the Lambda-Cuts.

The Follow the Leader method seems appropriate for the cases where the approximate number of classes is known before hand. In this case the Fuzzy Relation technique is also applicable.

If the number of customer classes has been decided and is relatively constant for the future, the K-means or the Fuzzy K-means clustering method could be applied depending on the spread of data and the results needed.

In case of the data sets containing fuzzy data, both Fuzzy K-means and Fuzzy Relation clustering method are better than Hierarchical and K-means.

The differences between each method is summarised in Table 1 below.

Methods	Number of clusters is predetermined	Creates boundaries between data sets
Hierarchical	No	Yes
K-means	Yes	Yes
Fuzzy K-means	Yes	No
Follow the leader	If necessary	Yes
Fuzzy relation	If necessary	No
Methods	Requires Iterative process	Trial and error approach
Hierarchical	No	No
K-means	Yes	No
Fuzzy K-means	Yes	No
Follow the leader	Yes	Yes
Fuzzy relation	Yes	Yes

Table 1 Method Comparison

4 CONCLUSIONS

Results of the above initial investigations can be generalised that each clustering method has characteristics. Therefore, to justify the appropriate method which match to the circumstances of Indonesia is necessary further investigation to determine criteria for each clustering techniques.

ACKNOWLEDGEMENTS

The authors would like to thank to PT PLN (Persero), Indonesia for providing the funding, information and opportunity.

REFERENCES

1. Werner, T.G., *Load Profiling in Germany*, 1st Discussion Forum: Technology Evolution and Future European Electricity Markets, 2-4 September 2002, London, UK.
2. Zakaria, Z., and Lo, K.L., *Load Profiling in the New Electricity Market*, Research and Development, 2002. SCORed 2002. Student Conference on 16-17 July 2002, Syah Alam, Malaysia, IEEE 2002, pp.278 – 281
3. Chang, R.F., and Lu, C.N., *Load Profiling and Its Applications in Power Market*, Power Engineering Society General Meeting, 2003, IEEE Volume 2, 13-17 July 2003, pp. 978.
4. Stephenson, P. and Paun, M., *Consumer advantages from half-hourly metering and load profiles in the UK competitive electricity market*, in Proc. Int. Conf. Elect. Utility Deregulation Restructuring Power Technologies 2000. London, U.K.: City Univ., Apr. 4-7, 2000, pp. 35-40.

5. Chicco, G., Napoli, R. and Piglione, F, *Application of Clustering Algorithms and Self Organising Maps to Classify Electricity Customers*, Proc.IEEE Bologna Power Tech 2003, Bologna, Italy, June 23-26, 2003.
6. Gerbec, D., Gasperic, S., Smon, I., Gubina, F., *An approach to customers daily load profile determination*, Power Engineering Society Summer Meeting, 2002 IEEE, Volume 1, 2002, pp. 587 – 591.
7. Yu, I. H., Lee, J. K., Ko, J. M., and Kim, S. I., *A Method for Classification of Electricity Demands using Load Profile Data*, Computer and Information Science, 2005. Fourth Annual ACIS International Conference on 2005, pp. 164 – 168.
8. Chicco, G., Napoli, R., Postolache, P., Scutariu, M., and Toader, C., *Customer Characterization Options for Improving the Tariff Order*, IEEE Trans. Power System, vol. 18, February 2003, pp.381 – 387.
9. Ozveren, C. S, Birch, A, P, & Vechakanjana, C, *Fuzzy Classification of Electrical Load Demand Profiles – A Case Study*, Power Management and Control, 17-19 April 2002, Conference Publication no. 488, IEEE 2002.
10. Pitt, B. D., and Kirschen, D. S., *Application of data mining techniques to load profiling*, in Proc. IEEE PICA, Santa Clara, CA, May 16–21, 1999, pp. 131–136.
11. Carpaneto, E., Chicco, G., Napoli, R., and Scutariu, M., *Customer classification by means of harmonic representation of distinguishing features*, Power Tech Conference Proceedings, 2003 IEEE Bologna, Volume: 3 23-26 June 2003, pp. 7.
12. D. Gerbec, S. Gasperic, and F. Gubina, *Determination and Allocation of Typical Load Profile to Eligible Consumers*, Paper accepted for presentation at 2003 IEEE Bologna Power Tech Conference, June 23-26, Bologna, Italy.

AUTHOR'S ADDRESS

The first author can be contacted at

School of Computing and Creative Technology
University of Abertay Dundee
Kydd Building
Dundee - DD1 1HG
Scotland - UK
email 0604721@abertay.ac.uk