Proceedings of the 5th WSEAS Int. Conf. on Power Systems and Electromagnetic Compatibility, Corfu, Greece, August 23-25, 2005 (pp212-217)

Determination of Consumers' Load Profiles based on Two-stage Fuzzy C-Means

K.L LO, Z ZAKARIA, M H SOHOD
Power System Research Group,
Department of Electronics & Engineering,
University of Strathclyde,
UNITED KINGDOM

Abstract: - Electricity industries throughout the world have been using load profiles for many years. Previously, in a regulated environment, load profiles have been employed to provide information for forecasting, system planning and demand side planning. However, in the deregulated environment, load profiles have become more significant. The determination of customer load profile may provide utility companies with better marketing strategies and improved efficiency in operating the current facilities.

This paper presents a two-stage clustering technique to determine consumers load profiles on the basis of their electricity behaviour. Fuzzy c-means (FCM) is utilised in this work. The load data used in this work are from actual measurements from different feeders derived from a distribution network. Cluster validity indices will be used to determine the best cluster number.

Keywords: - Load profiling, clustering, fuzzy c-means

1 Introduction

In this current state of electricity deregulation, the dynamic of energy pricing which has a major impact on metering and billing systems has changed. In many countries, consumers now have the flexibility to choose their electricity provider. Detailed knowledge on consumer's load consumption can facilitate distribution companies in determining specific tariff options for different type of consumers.

Ideally, the most efficient method of determining electricity consumption would be the direct monitoring. This can be achieved by installing time interval meters, quarter-hourly, half-hourly or hourly at each point of consumption. However, this approach is cost-prohibitive due to the equipment and processing costs. Furthermore, a significant amount of time would be needed to develop such a system. An alternative to this approach is by determining load profiles for consumers.

The term load profile describes the pattern of electricity usage for a customer or a group of customer over a given period. There are two general load profile models i.e. the area and the category model [1]. The area model includes all those customers that are not metered on time interval basis within the geographic region covered by a network. In this model, the non-metered customers constitute the residual profile, which is an adjusted load-profile for the node under consideration. On the other hand, the category model grouped customers with a similar load pattern into categories. Each individual customer is then associated with a predetermined representative load profile. This model is rather a popular practice, however the precondition is always that sufficient load measurement have been made earlier.

The choice of the most suitable load profiling method for any situation depends on factors including cost, data availability, equipment availability, accuracy requirements, regulatory requirements and the need of the utility distribution company. Generally there are several approaches being used including static load profiling, dynamic modelling and dynamic load profiling [2].

Efforts toward determining load profiles to represent types of consumers have already been made and they are reported in a number of papers. United Kingdom regulatory authorities have came up with a eight generic profiles (two for domestic and six for non-domestic consumers) to represent their consumers which have under 100 kW demand [3]. A rather similar approach together with a comprehensive survey system has been implemented in Taiwan [4, 5]. Interest in using Artificial intelligence application in determining load profiles has also increased. Among the popular method are fuzzy clustering, artificial neural network (ANN) and self-organizing map (SOM) as described in [6-8].

This paper is organised as follows: In section 2, an overview of fuzzy clustering is presented followed by the description of the clustering algorithms employed in this study. Section 3 discusses some of the cluster validity index used to determine the optimal number of clusters. In section 4, we present our case study – test data from actual measurements of daily load curve of 300 feeders in a distribution network. Finally, section V contains the concluding remarks.

2 Fuzzy Clustering

Cluster analysis is the formal study of algorithms and methods for grouping data. It is also a tool for exploring data structure. Therefore, it may reveal relations and structure in data. Cluster analysis has been used in a variety of disciplines such as pattern recognition, image processing, information retrieval, marketing and many more.

Most traditional cluster analysis algorithm is crisp partitioning which means each pattern belongs to one and only cluster. However, most objects have ambiguous attributes and thus method for soft partitioning is required.

Fuzzy set theory proposed by Zadeh in 1965 provides a tool for this purpose. Application of fuzzy set theory in cluster analysis were early proposed in the work of Bellman, Kalaba and Zadeh [9] and Ruspini [10]. In general, there are two categories in using fuzzy theory in cluster analysis [10]. The first category is fuzzy clustering based on relation matrix such as correlation coefficient, equivalence relation, similarity relation and fuzzy relation. On the other hand, the second category is based on objective functions, which include FCM.

2.1 Fuzzy C-Means

This method was originally introduced by Bezdek as an improvement on earlier clustering methods [11]. By using FCM, each data point belongs to a cluster to some degree that is specified by a membership grade. It is based on minimization of the following objective function:

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^m \left\| x_i - c_j \right\|^2 \qquad (3)$$

where

N = number of load profile
C = number of cluster
m = weighting parameter, in general m=2
u_{ij} = is the degree of membership of x_i in the cluster j
x_i = is the profile of ith feeder of measured data,
c_j = jth cluster centre
||*|| = is any norm expressing the similarity between any measured data and the centre

Consider a set of N load profiles $X=\{x_1,x_2,\ldots,x_N\}$ to be clustered into C clusters ($1<C<N$). The steps in this algorithm are as follows:

i) Choose C and m, and initialise the partition matrix $U^{(0)}$.

ii) Calculate the cluster centres.

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^m . x_i}{\sum_{i=1}^{N} u_{ij}^m} \qquad (4)$$

iii) Update the partition matrix for the kth step, $U^{(k)}$ as follows:

$$u_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{\left\| x_i - c_j \right\|^2}{\left\| x_i - c_k \right\|^2} \right)^{\frac{2}{m-1}}} \qquad (5)$$

for $1<i<C$

iv) If $\| U^{(k+1)} - U^{(k)} \| < \varepsilon$ then STOP; otherwise return to step (ii).

3 Cluster Validity

Since clustering algorithms define clusters that are not known a priori, the final partition of the data requires some kind of evaluation. Cluster validity is the term to describe the procedure of this evaluation. One of the most important issues in clustering is to decide the optimal number of clusters that fits a data set. Using a validation index can solve this.

Many different indices have been proposed [12]. In this study, we employed some of the widely used indices as illustrated in Table 1.

Table 1: Clustering validity indices

Functional	Symbol	Authors
Nonfuzziness index	NFI	Roubens (1978)
Partition coefficient	F	Bezdek (1981)
Partition entropy	H	Bezdek (1981)
Minimum hard tendency	Min HT	Rivera (1990)
Mean hard tendency	MeanHT	Rivera (1990)
Separation index	S	Xie-Beni (1991)

4 Case Study

Both algorithms were tested to a set of 300 feeders from a distribution network in Malaysia. The data consists of daily load consumption and measured for every half-hour from 12 midnight that gives 48 values for each feeder.

Since the aim of clustering is to discover the natural group of the data, the algorithm was run several times. Based on the number of main consumers connected to the feeders involved, we fixed 6 as the maximum number of clusters. Therefore the clustering process was repeated from c=2 until c=6. Cluster validity index is calculated at each value of c and this is shown in Table 2.

Table 2: Clustering validity indices for different number of clusters

	F	H	NFI	XB
c = 2	0.7420	0.4061	0.4841	0.0433
c = 3	0.7001	0.5483	0.5501	0.0343
c = 4	0.6182	0.7344	0.4909	0.0291
c = 5	0.5749	0.8664	0.4686	0.0343
c = 6	0.5238	0.9909	0.4285	0.0339

To choose the optimal number of cluster, F, NFI, MinHT and MeanHT should be maximum while H and S minimum. From Table 2, it seems that this can be either c=2, 3 or 4. However, it has been proven that the values of F and H always maximum and minimum respectively at c=2, thus these two indices is not suitable to determine the best cluster number [13].

Furthermore, in this case it is not appropriate to just have two clusters since there is prior knowledge that the feeders supplied more than two different types of consumers. Therefore, in this case, XB index is more favourable since four clusters are more appropriate for electricity feeders. This decision is taken in line with the fact that electricity consumers are normally categorised into residential, commercial, street lighting and industrial even though each of these major categories can be further subdivided. Accordingly, in this case, the feeders' data is clustered into 4 clusters. The other output from clustering is the number of members in each cluster and is depicted in Table 3.

Table 3: Numbers of feeder in each cluster

Cluster	1	2	3	4
No. of feeders	48	109	58	85

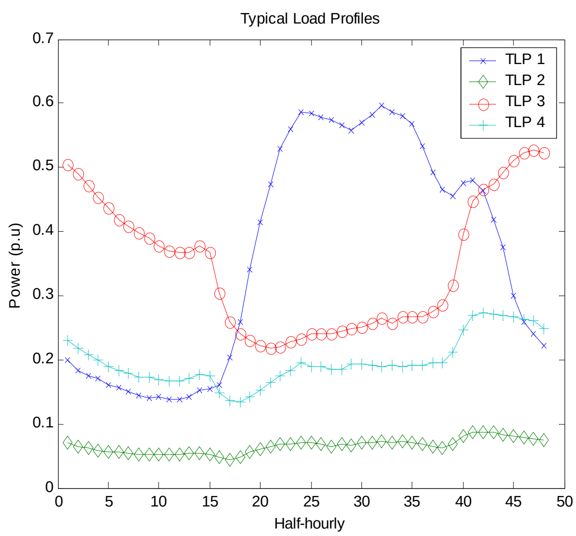

Fig. 1: Four typical load profiles

The typical load profiles (TLPs) for each clusters is obtained from by averaging the load curve of feeders assigned to the same cluster. This is illustrated in Fig.1.

4.1 Re-clustering

From the above results, clustering is seen as a valuable procedure to perform exploratory data analysis and thereby gain some insight into the nature or structure of the data. In this section, a re-clustering process to discover distinct sub clusters for each cluster is proposed. The term *distinct sub clusters* means clusters of patterns whose members are more similar to each other than they are to other patterns.

This approach is useful to discover the true pattern of the TLP especially for cluster 2, which does not reveal significant pattern that can be used in understanding its behaviour. With re-clustering, the feeders in each cluster will be clustered again into a suitable number of clusters. Again, cluster validity indexes are used to determine the optimum number of cluster.

The number of feeders in each cluster is less than before thus the cluster validity indexes are computed for c=2 until c=4. Based on the computed cluster validity indexes, these clusters can be further re-clustered into sub-clusters as shown in Table 4. Then, again TLP for each sub-cluster is established using averaging process. The main clusters are now denoted by C1, C2, C3 and C4. Fig.2 – Fig.5 show the TLP for each sub-cluster after the re-clustering process.

Table 4: Number of sub-clusters for each main cluster

Cluster	**1**	**2**	**3**	**4**
Sub-clusters	2	3	2	2

From the re-clustering process, it can be conclude that C3 consists of the same type of consumer. On the other hand, C1, C2 and C4 consist of mixed consumers. In the initial clustering, these patterns are hidden because only the dominant consumers are detectable. Since there is a prior knowledge about the type of consumers that are connected to some of the feeders, the TLP need to be assigned to this type of consumers.

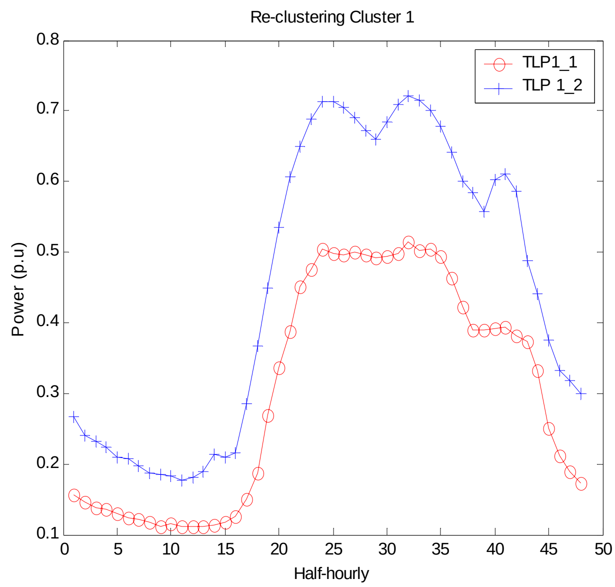

Fig. 2: TLPs for C1 when re-clustered into 2 sub-clusters

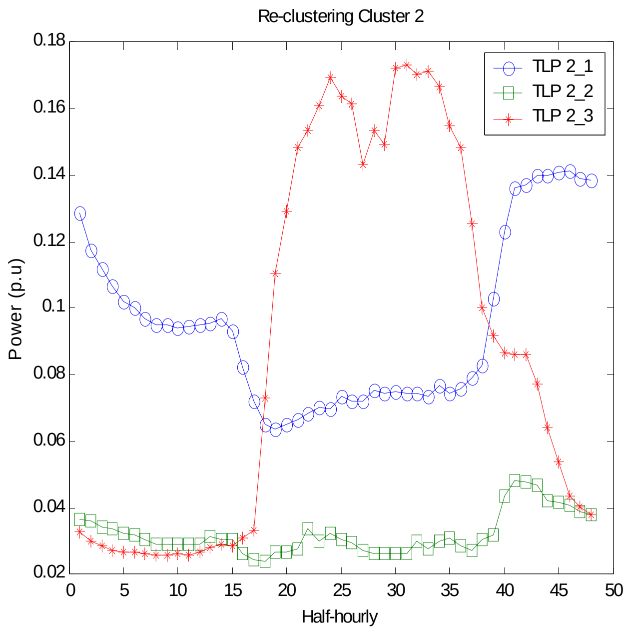

Fig. 3: TLPs for C2 when re-clustered into 3 sub clusters

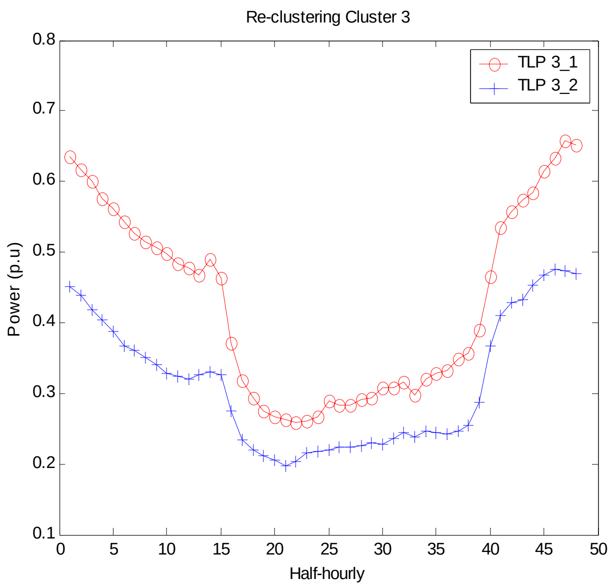

Fig. 4: TLPs for C3 when re-clustered into 2 sub-clusters

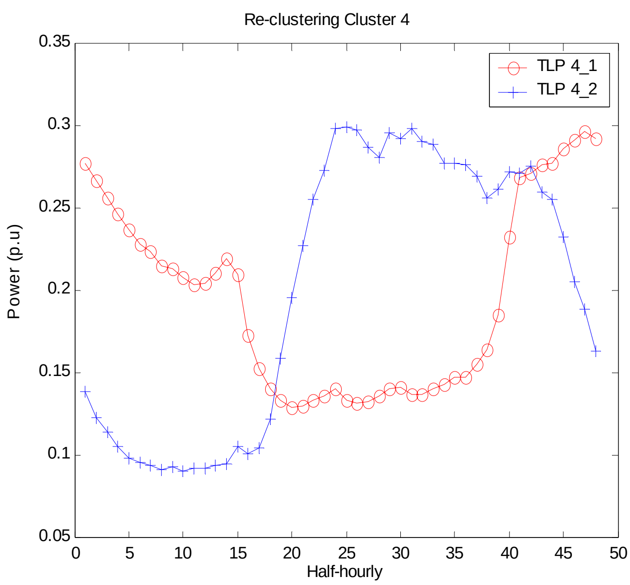

Fig. 5: Typical load profiles for C4 when re-clustered into 2 sub clusters

Comparing the pattern of the TLP to the specific type of consumer will help to visualise the situation if it fits to any type of consumer. From the utility database, it is known that the feeders are connected to the following type of consumer:

i) Domestic
ii) Commercial
iii) Small Scale Industrial
iv) Mixed Load (Domestic, Commercial and Small Scale Industrial)
v) Mixed Load (Domestic and Commercial)
vi) Mixed Load (Commercial and Small Scale Industrial)

Based on the comparison with the available feeders' load curve, the TLP for each cluster can be assigned as follows:

- TLP for cluster 1: Mixed load consist of domestic, commercial and industrial consumers
- TLP for cluster 2: Mixed load of domestic and commercial consumers which has lower load value
- TLP for cluster 3: Domestic consumers
- TLP for cluster 4: Two categories of commercial consumers

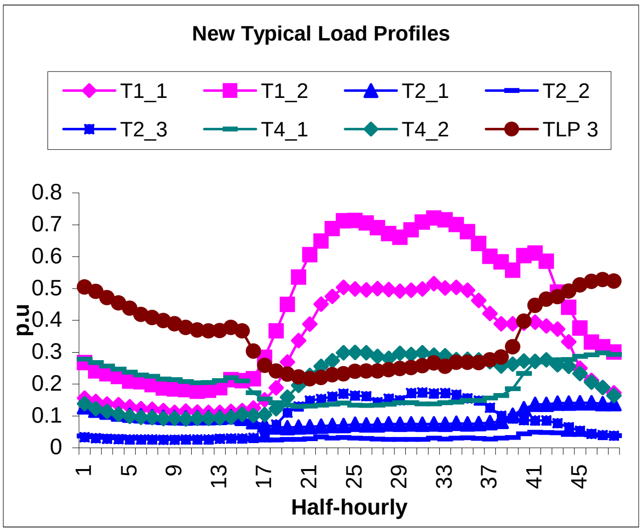

Fig. 6: New typical load profiles

After re-clustering process, it is found that the initial TLP has now increased. All clusters except Cluster 3 consist of a mixed consumer. Therefore, the new number of TLP for this case study has turn into 8 as shown in Fig.6 instead of 4 which is chosen initially. Each TLP established in the re-clustering process are selected as the new TLP. Since re-clustering for cluster 3 provides similar profiles, initial TLP for this cluster is retained.

5 Conclusion

The paper presents the procedure to determine the typical load profiles based on fuzzy clustering. Fuzzy c-means technique is used in this work and tested on daily load data of 300 feeders. Cluster validity indices were used to determine the optimal number of clusters. Typical load profile is obtained by averaging the number of patterns in each cluster. A re-clustering process is also proposed in this paper. Results from the re-clustering process have demonstrated that this technique is useful to discover the true pattern of the typical load profiles produced. Furthermore, the findings show that the energy consumption can be clustered not only based on the load pattern but also load value. In conclusion, typical load profiles established in this chapter has great potential for further analysis in distribution system applications.

References:

[1] System Tariff Issues Working Group, "Metering, Load Profiles and Settlement in Deregulated Markets," Union of Electricity Industry 2000-220-0004, 2000.

[2] "Load Profiling: An Innovate Alternative to Direct Metering When Implementing Customer Choice in the Electric Utility Industry," AEIC Load Research Committee July 1998.

[3] S.V Allera and A.G Horsburgh, "Load Profling for Energy Trading and Settlements in the Uk Electricity Markets," presented at DistribuTECH Europe DA/DSM Conference, London,vol, pp. 1998.

[4] C.S Chen, J.C Hwang, C.W Huang, Y.M Tzeng, and M.Y Chu, "Determination of Customer Load Characteristics by Load Survey System at Taipower," *IEEE Transaction on Power Delivery*, vol. Vol.11, 1996.

[5] C.S Chen, J.C Hwang, and C.W Huang, "Application of Load Survey Systems to Proper Tariff Design," *IEEE Transaction in Power System*, vol. Vol.12, 1997.

[6] K.L Lo and Zuhaina Zakaria, "Electricity Consumer Classification Using Artificial Intelligence," presented at University Power Engineering Conference, Bristol, U.K,vol 1, pp. 443-447, 2004.

[7] D Gerbec, S Gasperic, and F Gubina, "Comparison of Different Classification Methods for the Consumers' Load Profile Determination," presented at 17th International Conference on Electricity Distribution, CIRED, Barcelona,vol Session 6, pp. 2003.

[8] Mihai Gavrilas, Viovel Calin Sfintes, and Marius Nelu Filimon, "Identifying Typical Load Profiles Using Neural Fuzzy Model," presented at IEEE Power Engineering Society Transmission and Distribution Conference,vol 1, pp. 421-426, 2001.

[9] R Bellman, R Kalaba, and L Zadeh, "Abstraction and Pattern Classification," *Journal of Math Analysis and Applications*, vol. 13, pp. 1-7, 1966.

[10] Miin-Shen Yang, "A Survey of Fuzzy Clustering," *Math. Comput. Modelling*, vol. 18, pp. 1-6, 1993.

[11] James C Bezdek and J Douglas Harris, "Fuzzy Partitions and Relations; an Axiomatic Basis for Clustering," *Fuzzy Sets and Systems*, vol. 1, pp. 111-127, 1978.

[12] Hyun-Sook Rhee and Kyung-WhanOh, "A Validity Measure for Fuzzy Clustering and Its Use in Selecting Optimal Number of Clusters," presented at 5th IEEE International Conference on Fuzzy Systems, Seoul, Korea,vol 2, pp. 1020-1025, 1996.

[13] Pei Jihong and Yang Xuan, "Study of Clustering Validity Based on Fuzzy Similarity," presented at 3rd World Congress on Intelligent Control and Automation, China,vol, pp. 2444-2447, 2000.