CONSUMER LOAD PROFILING USING FUZZY CLUSTERING AND STATISTICAL APPROACH

Zuhaina Zakaria[1], Mohd Najib Othman[1] and Mohamad Hadi Sohod[2]
[1]Faculty of Electrical Engineering,
Universiti Teknologi MARA,
Shah Alam, Selangor, Malaysia
[2]Tenaga Nasional Berhad, Malaysia
zuhaina@ieee.org

Abstract - **Load profiling present useful tool for onitoring typical load shape for a group of customers, which can be performed from past or current day data. In a deregulated energy environment, consumers can purchase electricity from any provider regardless of size and location. As a result, there is a growing interest in understanding the nature of variations in consumer's consumption. This information can be used to facilitate electricity supplier in their marketing strategy. Many techniques for load profiling have been reported in the past. The techniques include applications of statistics, unsupervised clustering technique and methods based on frequency domain approach. This paper compares the application of fuzzy clustering with statistical method in load profiling. K-means has been chosen as the statistical approach employed in this study. These two approaches have the same objectives i.e. to recognise similarities, clusters and classify the individual load profiles of different customers to one of the identified categories. The paper evaluates the performance of each method and discusses the strength and weaknesses of both approaches based on the simulated results.**

Keywords: ***load profiling, clustering, fuzzy classification, k-means.***

I. INTRODUCTION

The general aim of load profiling is to produce load profile. Load profile is derived when load curve is produced for a group of consumers, which have common characteristics over a given period. For a long time, power utilities have been using load profiles as the means to facilitate the formulation of retail tariffs. In fact, the idea of using load profiles to bill an individual consumer is well established. Load profiles have also been employed to provide information for forecasting, financial planning, rate design, system planning, demand side planning and regulatory reporting [1]. In addition, load profiles can also be used in evaluating distribution transformer loss-of life [2] and distribution transformer modelling [3].

However, using load profiles in the new electricity market is a new and different development. Information concerning consumer's consumption in the deregulated environment is of primary importance. With more and more pressure from deregulation of the electricity market, the determination of customer load profile will provide utility companies with better marketing strategies and improved efficiency in operating the existent facilities [4].

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. Therefore, load profiling is seen as the alternative solution that would provide a satisfactory, cost-effective approach.

This paper examined the ability of a fuzzy classification method based on fuzzy relations to cluster and classify load data of different feeders in a particular distribution network. The classification produced by this technique was evaluated and compared to the classification made by a more conventional model based on k-means analysis.

The remainder of the paper is organized as follows: Section II focuses on cluster analysis followed by a brief description on both approaches in Section III and IV respectively. Results will be discussed in Section V and finally, the conclusion of the study in Section VI.

II. CLUSTER ANALYSIS

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, which may not known previously, but maybe useful once found. Cluster analysis has been used in a variety of disciplines such as pattern recognition, image processing, information retrieval, marketing and many more.

There are extensive literatures on cluster analysis, resulting in many different type of clustering algorithm. This includes hierarchical clustering, k-means clustering, nearest neighbour clustering, fuzzy clustering and artificial neural network (ANN) application [5]. A brief description on fuzzy clustering and k-means clustering will be presented in this section.

In load profiling, to achieve an effective classification is a very demanding task. There are several factors that determine customer load profiles, representing various technical, commercial and behavioural aspects.

According to [6] the key concepts of classification techniques in electricity customers' are feature selection, time domain approaches (unsupervised clustering algorithm, statistics application, neural networks, fuzzy system) and frequency domain approaches (harmonic analysis, wavelets). Research in these methods were reported in several papers, [7], [8], [9]and [10].

III. FUZZY CLUSTERING

Most traditional cluster analysis algorithm is crisp partitioning which means each pattern belongs to one and only one cluster. Thus, there exist hard boundaries among the clusters. However, most objects have ambiguous attributes and thus method for soft partitioning is required.

Fuzzy set theory proposed by Zadeh in 1965 introduced the idea of uncertainty of belonging, which was described by a membership function that provides a tool for this purpose. Applications of fuzzy set theory in cluster analysis were early proposed in [11] and [12]. According to [13], there are two categories where fuzzy set theory applied in cluster analysis. The first category is fuzzy clustering based on relation matrix which includes correlation coefficient, equivalence relation, similarity relation and fuzzy relation. The second category is based on objective functions such as fuzzy c-means (FCM). The focus of this paper will be the application of fuzzy clustering based on fuzzy relation.

A. Fuzzy clustering based on fuzzy relation (FCR)

This method was initially proposed by Zadeh in 1971 [14], followed by Bezdek and Harris later in 1978 [15]. The traditional crisp relation based on the concept that everything is either has complete relationship or no relationship at all. Thus, a crisp relation maps or relates elements of two or more sets. Fuzzy relations also map elements of one universe to another universe, through the Cartesian product of the two universes. However, the 'strength' of the relation between the two universes is measured with a membership function that expressed various 'degrees' of strength of the relation on the unit interval [0,1].

There are several ways to develop the numerical values that characterize a relation. One of the popular methods is to use a family of procedures termed similarity methods to determine the value assignments for relations [16]. The algorithm can be summarised as below:

- i Produce the fuzzy relations (r_{ij}) matrix using the *cosine amplitude method.*
- ii Produce the fuzzy equivalence matrix using *max-min composition.*
- iii Obtain the clusters of data (crisp-equivalence matrix) using the classification by equivalence relations method using the *lambda-cut method.*

B. Cosine Amplitude

This method makes use of a collection of data samples, *n*. If these data samples are collected, they form a data array, X.

$$X = \{x_1, x_2, \ldots, x_n\}$$

Each of the elements, x_i in the data array X is itself a vector of length *m,* that is

$$x_i = \{x_{i1}, x_{i2}, \ldots, x_{im}\}$$

Each element of r_{ij} results from a pair wise comparison of two data samples, say x_i and x_j where the strength of the relationship between them is given by the membership value that is $0 \leq r_{ij} \leq 1$. The cosine amplitude method calculates r_{ij} in the following manner:

$$r_{ij} = \frac{\left|\sum_{k=1}^{m} x_{ik}, x_{jk}\right|}{\sqrt{\left(\sum_{k=1}^{m} x_{ik}^2\right)\left(\sum_{k=1}^{m} x_{jk}^2\right)}} \qquad \text{Equation (1)}$$

where i,j = 1,2…n

The relation matrix will be of size *n* x *n* and will be reflexive and symmetric – hence a tolerance relation.

C. Max-min Composition

This method is used to reform a tolerance relation into a fuzzy equivalence relation by employing the familiar operators of fuzzy set, max ($\vee$) and min ($\wedge$).The max-min composition is defined by the set-theoretic and membership function-theoretic expressions,

$$\mu_{\underset{\sim}{\in}}\left(x_i, x_j\right) = \underset{x \in X}{\vee}\left\{\mu_{\underset{\sim}{R}}\left(x_i, x_j\right) \wedge \mu_{\underset{\sim}{R}}\left(x_i, x_j\right)\right\} \qquad \text{Equation (2)}$$

D. Lambda-Cut for Fuzzy Relations

Lambda-cut is one of the methods to defuzzify fuzzy results to crisp results. In classification of data, this method was used to cluster the data. This involved taking the fuzzy equivalence matrix and setting every element above the cut-off point, λ (between 0 and 1) to a '1' and every element below it to a '0'. The crisp equivalence matrix, which has a certain number of classes, will be produced. The numbers of classes depend upon the value of λ. As this value gets larger, more classes will be obtained. The whole process for fuzzy clustering in load profiling is shown in Figure 1.

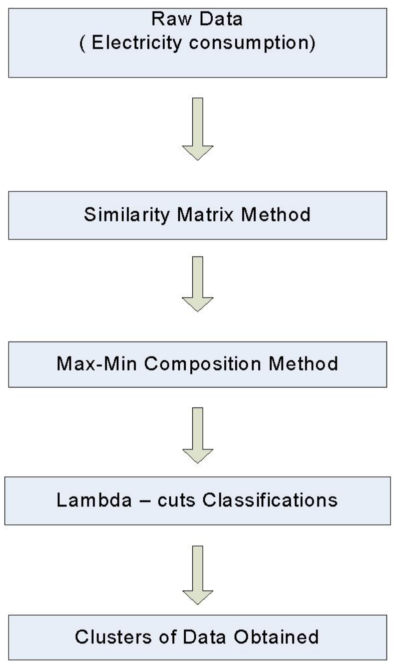

Fig. 1: Overall process in fuzzy clustering for load profiling

IV. K- MEANS ANALYSIS

K-means is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters, k fixed a priory. The main idea is to define k centroids, one for each cluster. However, the algorithm requires the number of clusters to be specified.

A. Classification Procedures

This procedure attempts to identify relatively homogeneous groups of cases based on selected characteristics, using an algorithm that can handle large numbers of cases. This algorithm aims at minimizing an *objective function*, in this case a squared error function. The objective function

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2 \qquad \text{Equation (3)}$$

where $\left\| x_i^{(j)} - c_j \right\|^2$ is a chosen distance measure between $x_i^{(j)}$ a data point and the cluster centre c_j, is an indicator of the distance of the n data points from their respective cluster centres.

The classification function can be used after value of centroid (cluster centre) and distance are obtained.

B. Classification Process

The K-Means Cluster Analysis procedure begins with the construction of initial cluster centres. Then, cases are assigned to clusters based on distance from the cluster centres. Finally, the location of cluster centres based on the mean values of cases in each cluster is updated by performing iterations.

V. RESULTS AND DISCUSSION

Both classification techniques were tested to a set of load data of different feeders. The data is the electricity consumption usage taken from 30 different feeders derived from a distribution network in Malaysia. These feeders are connected to four types of consumers i.e. domestic (D), commercial (C), industrial (I) and mixed consumers (D & C).

The time interval of sampling load curve data is 30 minutes and taken from 12 midnight until 11.30 pm the following day. Therefore the load profile is represented by 48 load values throughout the day. Table 1 illustrates the summary of the load data.

TABLE 1
FEEDERS DATA DESCRIPTION

Feeder No	Original Group	Type of consumer	Approximate number of consumers		
			D	C	I
1 - 13	1	D	177	-	-
14 - 19	2	C	-	21	-
20 - 22	3	I	-	-	3
23 - 30	4	D & C	106	17	-

A. Fuzzy Classification

As mentioned earlier, lambda-cut approach is used to cluster the data. Therefore, a suitable values need to be chosen to obtained an appropriate number of clusters. From the simulation, a lambda value of 0.873 is found as a suitable value because it gives a good classification results compared with other value. Using this value, the algorithm classified the feeder data into six major clusters. The load curves in each clusters is shown in Figure 2 to 6.

Figure 2 show that Cluster 1 consists most of the domestic feeders (10 out 13 feeders), four commercial and five D&C feeders. However, there are only three feeders in cluster 2 i.e. combination of domestic and commercial feeders. On the other hand, most of the industrial feeders are grouped in cluster 3 as illustrated in Figure 4.

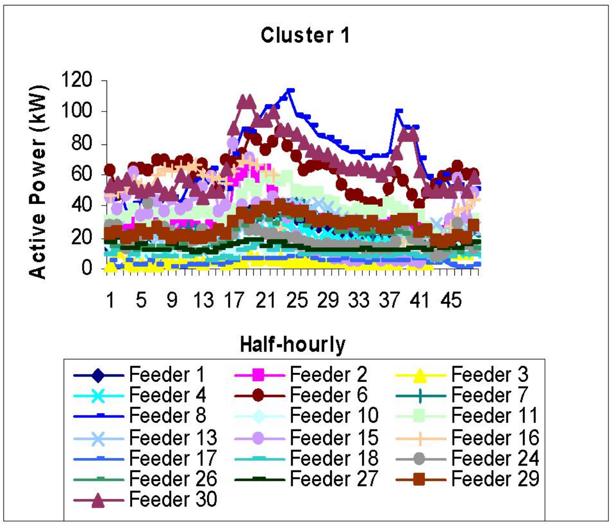

Fig. 2: Cluster 1 (FCR)

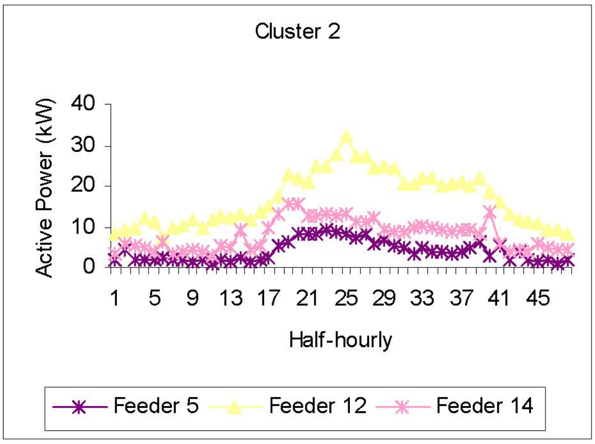

Fig. 3: Cluster 2 (FCR)

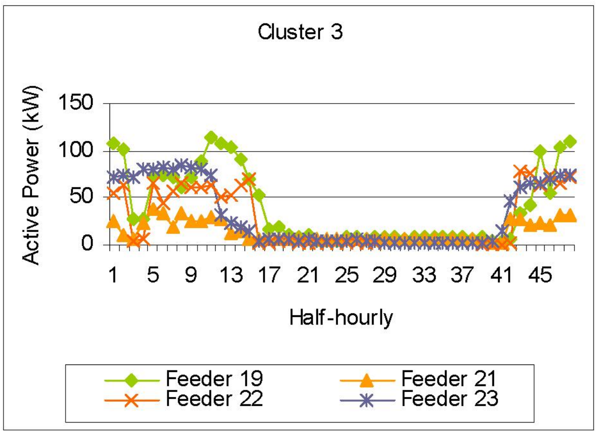

Fig. 4: Cluster 3 (FCR)

Cluster 4 consists of two feeders which are domestic and commercial feeders. There are also several load patterns that could not be categorized into their original group. This is shown in Figure 6 where cluster 5 and 6 has only a single feeder. Feeder 20 which is an industrial feeder was clustered in cluster 5 while and feeder 28 (D&C) in cluster 6.

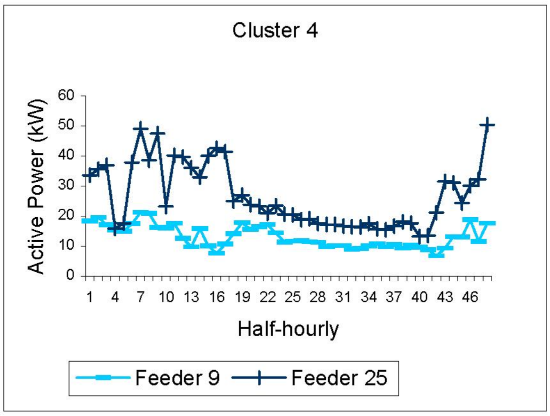

Fig. 5: Cluster 4 (FCR)

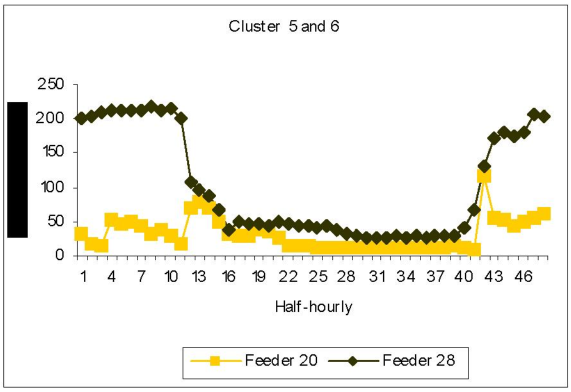

Fig. 6: Cluster 5 and 6 (FCR)

B. K- Means Analysis

Similarly, the number of clusters must be specified before performing K-Means analysis. Since the optimal number of clusters in FCR is 6, therefore the same number of clusters is used here. In this manner, the results can be compared more clearly. Figure 7 - 12 shows the load patterns of the feeders in each cluster.

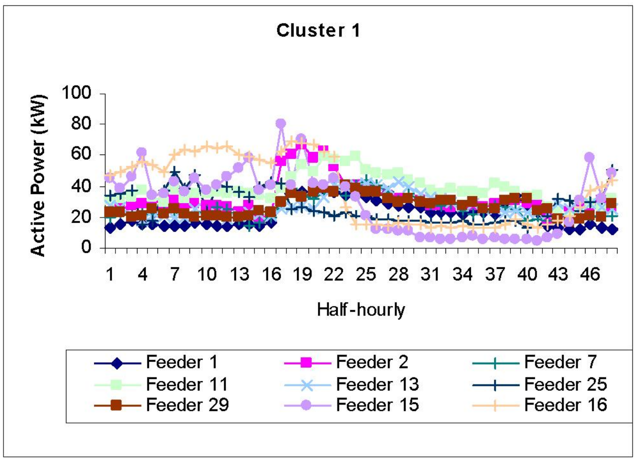

Fig. 7: Cluster 1 (K-means)

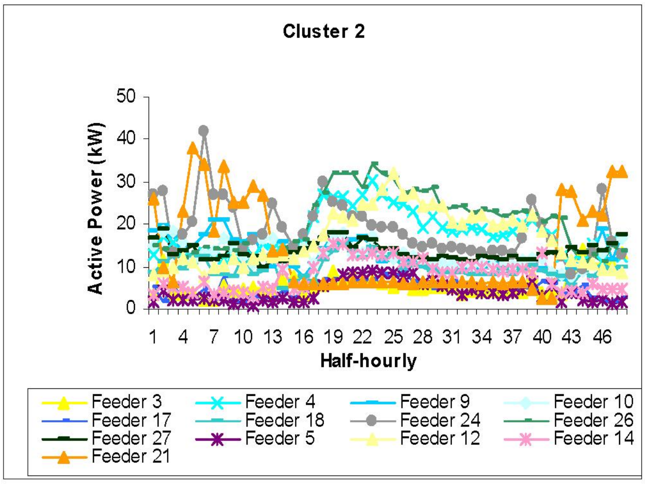

Fig. 8: Cluster 2 (K-means)

As shown in Figure 8 and 9, both cluster 1 and cluster 2 consists of high number of feeders. Cluster 1 contains various types of feeders except industrial feeder while cluster 2 comprises of all feeder types including one industrial feeder. On the other hand, there are only three feeders in cluster 3 i.e. two domestic feeders and one mixed feeder (D&C).

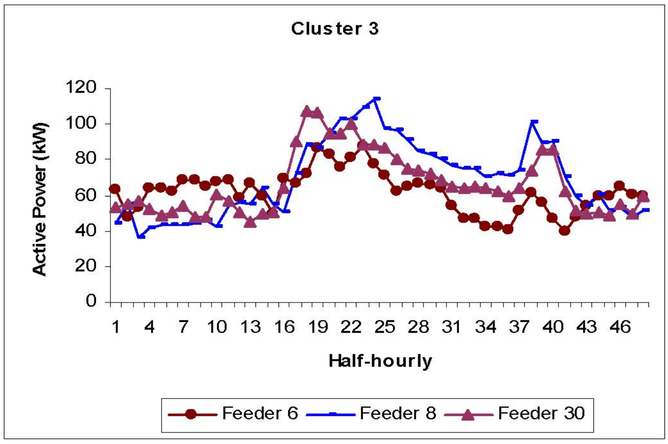

Fig. 9: Cluster 3 (K-means)

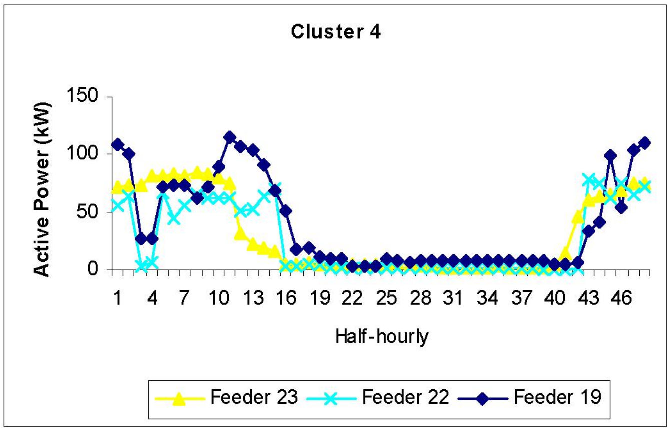

Fig. 10: Cluster 4 (K-means)

Cluster 4 comprises of three different feeders i.e. commercial, industrial and combination of domestic and commercial. Finally, cluster 5 and 6 consists of a unique load pattern that could not be grouped with other feeders.

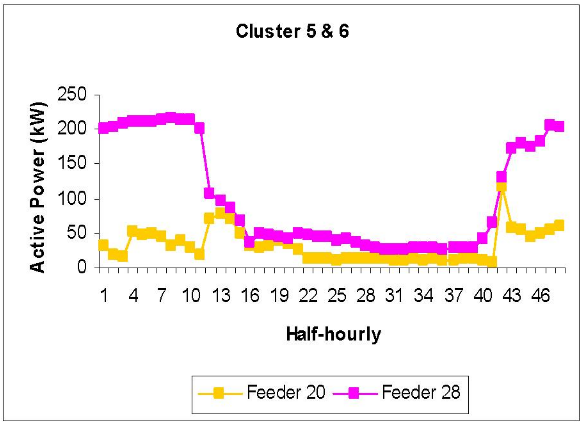

Fig. 11: Cluster 5 and Cluster 6

C. Analysis from both approaches

As illustrated in Table 1, there are two clusters produced from K-means which has a high number of feeders. Unlike clusters produced from FCR, where the number of feeders in each cluster excluding cluster 1, consists of several feeders only. The reason is probably due to the nature of FCR that has more sensitivity to pattern, which are so distinct from the others.

TABLE 2
NO OF FEEDERS IN EACH CLUSTERS PRODUCED BY FCM AND ANN

Cluster	FCR	K-means
1	19	9
2	3	13
3	4	3
4	2	3
5	1	1
6	1	1

In FCR, most of the domestic feeders have been successfully grouped together in cluster 1. On the other hand, in K-means approach, domestic feeders were clustered in two main clusters i.e. cluster 1 and 2. FCR has also managed to group most of the industrial feeders in a cluster while in K-means analysis, industrial feeders are grouped with other type of feeders. From these observations, FCR is seen as more accurate than K-means analysis.

However, from both analyses, the results show that most of the commercial load patterns were grouped together with the domestic pattern. This may be due to the fact that these commercial consumers consist of small businesses which operated similar to domestic usage style but with higher loads.

Another interesting result is that most of the mixed feeder (D &C) was grouped together with domestic feeders. One reason for this result is that the number of commercial consumers in this group is relatively small compared with the domestic users. Thus, the domestic load pattern has suppressed the commercial load pattern.

VI. CONCLUSION

This study evaluated the ability of fuzzy classification in classifying electricity consumers based on their energy consumption compared with statistical approach. Fuzzy clustering based on relation and K-means analysis were chosen for this purpose.

Both methods were tested on 30 different load profiles which comprises domestic, commercial, industrial and mixed feeders. The results from fuzzy clustering shows that the measured load profiles can be grouped into 6 clusters. The same number of clusters is then applied to K-means analysis.

From the clusters obtained, fuzzy clustering has demonstrated its ability in clustering load profiles into their natural groups. This approach is seen as more accurate than K-means analysis. In addition, fuzzy clustering is more flexible and sensitive in selecting the cluster boundary. These characteristics allowed it to single out cases that have unique or slightly different pattern and attributes.

VII. REFERENCES

[1] C. W. Gellings, "The Value of Load Research," in *Public Utilities Fortnightly*, 1988, pp. 32-40.
[2] J. A. Jardini, "Daily Load Profile for Residential, Commercial and Industrial Low Voltage Consumers," *IEEE Transaction on Power Delivery*, vol. Vol.15, pp. 375-380, 2000.
[3] R.-F. Chang, R.-C. Leon, and C.-N. Lu, "Distribution transformer load modeling using load research data," *IEEE Transaction on Power Delivery*, vol. 17, pp. 655-661, 2002.
[4] M. Scutariu, C. Toader, and P. Postolache, "Investigations on Daily Load Curve Characterisation," presented at 2nd Euroconference on Risk Management in Power System Planning & Operation in Market Environment, Porto, 2001.
[5] A. K. Jain and J. Mao, "Artificial Neural Networks: A Tutorial," in *Computer*, 1996, pp. 31-44.
[6] G. Chicco, R. Napoli, P.Postolache, M.Scutariu, and C.Toader, "Electric Energy Customer Characterisation for Developing Dedicated Market Strategies," presented at Power Tech, Porto, 2001.
[7] H. Muller, "Classification of Daily Load by Cluster Analysis," presented at 8th Power System Computation Conference, 1984.
[8] A. P. Birch, C. S. Ozveren, and A. T. Sapeluk, "A Generic Load Profiling Technique using Fuzzy Classification," presented at 8th International Conference on Metering & Tariff for Energy Supply, 1996.
[9] C. S. Ozveren, L. Fyall, and A. P. Birch, "A Fuzzy clustering and classification technique for customer profiling," presented at University Power Engineering Conference, 1997.
[10] B. D. Pitt, "Application of Data Mining Techniques to Load Profiling," UMIST, 2000.
[11] G. Xinbo and X. Weixin, "Advances in theory and applications of fuzzy clustering," in *Chinese Science Bulletin*, vol. 45, 2000, pp. 961-970.
[12] R. Bellman, R. Kalaba, and L. Zadeh, "Abstraction and Pattern Classification," *Journal of Math Analysis and Applications*, vol. 13, pp. 1-7, 1966.
[13] M.-S. Yang and H.-M. Shih, "Cluster analysis based on fuzzy relations," *Fuzzy Sets and Systems*, vol. 120, pp. 197-212, 2001.
[14] L. Zadeh, "Similarity relations and fuzzy orderings," *Information Sciences*, vol. 3, pp. 177-200, 1971.
[15] J. C. Bezdek and J. D. Harris, "Fuzzy partitions and relations; an axiomatic basis for clustering," *Fuzzy Sets and Systems*, vol. 1, pp. 111-127, 1978.
[16] T. J. Ross, *Fuzzy Logic with Engineering Applications*: McGraw Hill International Editions, 1997.