# Application of fuzzy c-means clustering to power system coherency identification

SHU-CHEN WANG[1] PEI-HWA HUANG[2]
Department of Electrical Engineering
National Taiwan Ocean University
No. 2, Peining Rd., Keelung 20224
TAIWAN

CHI-JUI WU[3]
Department of Electrical Engineering
National Taiwan University of Science and Technology
No. 43, Sec. 4, Keelung Rd., Taipei 10607
TAIWAN

*Abstract:* -This paper presents the application to the identification of coherent generators in a power system based on the fuzzy c-means clustering. In view of the conceptual appropriateness and computational simplicity, the fuzzy c-means give a fast and flexible method for clustering analysis. At first, the coherency measures are derived from the time-domain responses of generators to reveal the relations between any pair of generators. Then the coherency measures are used as initial membership matrix in the fuzzy c-means clustering, which can let the clustering procedures converge quickly. An example power system is used to show the effectiveness of this method. The schemes of various number of generator clusters can be procured. The approach in this paper needs less iterative times and can directly begin a clustering procedure for any number of clusters.

*Key-Words:* - **Coherency identification, Coherency measures, Cluster analysis, Fuzzy c-means, Power system dynamic equivalent.**

## 1 Introduction

As power systems become larger, their complexity likewise increases and system analysis has to tackle high-order models. However, computation on the entire system model is highly complex and uneconomical while the final analysis results may have unnecessary portions, leading to high investments but with no apparent advantages. So creating a simplified equivalent model usually is the first step. In general, a system could be divided into two parts: study area and external system [1]. The response of study area occupies the significant portion and is directly relative to the system. The external system can be summarized as a dynamic equivalent to reflect the behavior of this part after a system disturbance [2-4]. The generators in the external system must be similar in terminal behaviors. The methods of combining the group of generators in the external system to an equivalent one can be based on many coherency approaches [5-7]. The identification of coherent generators involves obtaining the response curve of each generator during a system transient and then processing those curves by using different technique.

In the literatures, cluster analysis [8-9] can be used to divide complex systems into some smaller and relatively simpler systems. It is to classify non-processed data into certain categories depending on the features considered. Data in each category have the most resemblance while being very dissimilar with data from other categories. Traditional clustering methods strictly assign each studied object into a certain data cluster, called the hard clustering method. In reality, a data point may be recognized to partially belong to multiple clusters. For numerical data, the members of each cluster bear more mathematical similarity but still have certain relative degree with other clusters. So using fuzzy set theory in data clustering can naturally describe the actual classification problem, called the fuzzy clustering method [10-11]. Fuzzy clustering analysis produces a soft partition and provides more flexibility in describing the structure of the data points. The fuzzy c-means (FCM) approach is one of the most popular fuzzy clustering algorithms and has been extensively used in related applications [12]. This method uses the Euclidean distance as a basis for classification, and aims to group data into some cluster centers. Based on the similarities between data, continuous convergences are achieved so that the data can be grouped into clusters. Each of the calculated cluster centers then replaces the original data points around it to become the representative of the cluster. The formation in the fuzzy c-means approach is closer to the real-word problem to have better performance and the formulated problem can be easier to solve computationally.

In this paper, the fuzzy c-means clustering is presented for recognizing coherent groups of generators. The proposed method uses the coherency measures as a basis for classification and aims to group generators under different prescribed number of coherent groups. The values of coherency measures are obtained from the time-domain generator responses of the system after a disturbance. The degree of relation between any pair of generators is described by the values of coherency measures. Then the coherency measures can be used as the initial membership matrix in the fuzzy c-means clustering, which can let the clustering procedures converge quickly. An example power system is used to show the validity and effectiveness of this method. From the simulation results, the proposed method can begin the clustering procedure directly under any prescribed number of clusters. The responses of cluster centers are also given to reveal the equivalent responses of corresponding coherent generators. It also can be found that the generators in a same cluster have strict relation with others. The comparison with the similarity relation method is also given.

## 2 Coherency Measure

In this study, coherency measures, which are derived from time-domain responses, are proposed to evaluate the relative behaviors between any pair of generators. Consider the rotor angle time-domain response $\delta_i(t)$ of generator $i$ during a system transient, which is often referred to as the swing curve. An approximate time derivative of $\delta_i(t)$, denoted as $\omega_i(t)$, can be obtained as

$$\omega_i(t_k) = \frac{\delta_i(t_k) - \delta_i(t_{k-1})}{t_k - t_{k-1}} \quad (1)$$

where $t_k$ is the $k$th sample instant. Note that $\omega_i(t)$ is essentially the change of the swing curve with respect to time and depicts the shape of the swing curve.

Then the index $W_{ij}'$ is defined for any possible pair of generators $i$ and $j$ by

$$W_{ij}' = \sum_{t_k \in [0,T]} \left| \omega_i(t_k) - \omega_j(t_k) \right| \quad (2)$$

In (2) the summation is taken for all the samples during the whole study interval $[0,T]$. The index is further normalized to become

$$W_{ij} = W_{ij}' / \max(W_{ij}') \quad (3)$$

The index $W_{ij}'$ calculates the total distance between the swing curves of generator $i$ and generator $j$, and the index $W_{ij}$ measures the degree of difference in shapes.

Finally, the coherency measure $C_{ij}$ is obtained by

$$C_{ij} = 1 - W_{ij} \quad (4)$$

Obviously, $0 \le C_{ij} \le 1$, $C_{ii} = 1$, and $C_{ij} = C_{ji}$. The association between the pair of generators $i$ and $j$ can be evaluated by the value of $C_{ij}$. A larger $C_{ij}$ indicates that generator $i$ and generator $j$ are more similar in the time domain. Therefore, the index $C_{ij}$ does concern the degree of similarity in shapes. This justifies the values of coherency measures to express the level of similarity between any pair of generators in the external system.

## 3 Fuzzy c-means Cluster Analysis

Cluster analysis summarizes a large amount of data into a small number of groups. Fuzzy cluster analysis [10-11] refers to partition the studied objects into a fuzzy classification notion so that objects in a group are similar and the other groups are not similar. Such kind of clustering method provides more flexibility in describing the structure of the data points under study. The fuzzy c-means (FCM) algorithm [12] uses concepts of $n$ -dimensional Euclidean space to determine the geometric closeness of data points by assigning them to various clusters and then determining the distance between the clusters. The distance between points in the same cluster will be considerably less than the distance between points in different clusters. The number of clusters can be assigned arbitrarily for the study purpose.

The most widely used objective function for fuzzy clustering is the weighted sum of squared errors within groups. The objective function $J_m$ is defined as

$$J_m(U, c_i; X) = \sum_{i=1}^{c} \sum_{j=1}^{n} \mu_{ij}^m d_{ij}^2 \quad (5)$$

where:

$X$ : data space of generator time-domain responses whose elements are $\{x_j\}$

$c$ : number of clusters

$n$ : number of generators

$c_i$ : center of cluster $i$

$U$ : membership matrix whose elements are $\{\mu_{ij}\}$

$\mu_{ij}$ : degree of relation of generator $j$ to cluster $i$

$m$ : exponent on $\mu_{ij}$, weighting coefficient

$d_{ij}$ : distance from $x_j$ to $c_i$, that is, $d_{ij} = |c_i - x_j|$

It is worth noting that the cluster center $c_i$ of cluster $i$ is referred to as the prototype of the cluster and can be considered as the representative of that cluster.

The fuzzy c-means clustering algorithm is essentially an iterative procedure. In the conventional fuzzy c-means method, the initial partition set is selected randomly, so that the iterative time would be large. In this section, the coherency measures are used as the initial fuzzy partition. It is hoped that the iterative times could be less. The clustering procedures of using the coherency measures as initial values in the fuzzy c-means method can be formulated as the following steps. The iterative procedure is taken at each sampling instant.

1. Begin the procedure at the sampling instant $t_0$. Construct an $n \times n$ fuzzy relation matrix $R$ for an $n$-generator system with the coherency measures $C_{ij},\ i = 1, \cdots, n,\ j = 1, \cdots, n$, obtained from the time-domain responses.

$$R = \left[ C_{ij} \right], \quad 0 \le C_{ij} \le 1 \tag{6}$$

   where $C_{ij}$ indicates the degree of association between generator $i$ and generator $j$.
2. Select $c$, the number of clusters. It wants to partition $n$ generators into $c$ clusters. Let $l$ denote the iterative time. Initialize the $c \times n$ membership matrix $U$ with a sub-matrix of $R$, that is, $U^{(1)} = [C_{ij}],\ i = 1, \cdots, c,\ j = 1, \cdots, n$.
3. Begin a new iterative procedure at the sampling instant $t_k$.
4. At the $l$ th iteration, calculate the cluster center $c_i^{(l)}$ of cluster $i$:

$$c_i^{(l)} = \frac{\sum_{j=1}^{n} (\mu_{ij}^{(l)})^m x_j}{\sum_{j=1}^{n} (\mu_{ij}^{(l)})^m}, \quad i = 1, \cdots, c \tag{7}$$

   where $x_j$ is the time-domain response value of the generator $j$ at this sample instant, and $\mu_{ij}$ is the element of $U$ and expresses the degree of membership of generator $j$ to cluster $i$. Note that the value of $m$ normally falls in th range of $1.5 \le m \le 3$.
5. Compute the distance between generator $j$ and cluster center $i$:

$$d_{ij}^{(l)} = \left| c_i^{(l)} - x_j \right| \tag{8}$$

6. Update the membership matrix $U^{(l+1)}$ by

$$\mu_{ij}^{(l+1)} = \frac{1}{\sum_{k=1}^{c} (\frac{d_{ij}^{(l)}}{d_{kj}^{(l)}})^{2/(m-1)}}\ , i = 1, \cdots, c, \quad j = 1, \cdots, n \tag{9}$$

   The newest degree of belonging of generator $j$ to cluster $i$ is given by $\mu_{ij}^{(l+1)}$. Then all degrees of belonging of all generators to all clusters are updated.
7. Check if $\left\| U^{(l+1)} - U^{(l)} \right\| \le \varepsilon$, that is, $\left| \mu_{ij}^{(l+1)} - \mu_{ij}^{(l)} \right| \le \varepsilon,\ , i = 1, \cdots, c, \quad j = 1, \cdots, n$, where the $\varepsilon$ is the convergent tolerance, or a predefined number of iteration is reached, then stop; otherwise, set $l = l + 1$ and go to Step 4.
8. Use the convergent $U$ at the sampling instant $t_k$ as the initial membership matrix to begin a new iterative procedure for the next sampling instant $t_{k+1}$ and go to Step 3 until the final sampling instant.
9. Defuzzify the convergent $U$ of the final sampling instant. The defuzzification is called the maximum membership method for hardening the fuzzy classification matrix that is required to assign data into hard partitions.

The above steps perform the fuzzy clustering analysis for power system coherency identification. It is noted that in the above algorithm, the cluster center of each cluster is referred to as the prototype of that cluster and can be considered as the representative of that cluster.

## 4 Example

In order to show the proposed fuzzy c-means clustering scheme, the New England 39-bus test system as shown in Figure 1 is used. The detailed system data and initial operating conditions can be found in [2]. Note that generator 10 is equivalent for representing a large system, and generators 6 and 7 are belonging to the study area; therefore, they will not be considered in the coherency analysis. The other generators are seen as the external system and

Table 1 Coherency measures between generators

| Generator | 1 | 2 | 3 | 4 | 5 | 8 | 9 |
|---|---|---|---|---|---|---|---|
| 1 | 1.0000 | 0.7482 | 0.6814 | 0.7176 | 0.5355 | 0.8115 | 0.6231 |
| 2 | 0.7482 | 1.0000 | 0.8191 | 0.6602 | 0.3814 | 0.6860 | 0.5465 |
| 3 | 0.6814 | 0.8191 | 1.0000 | 0.6498 | 0.3962 | 0.6538 | 0.5595 |
| 4 | 0.7176 | 0.6602 | 0.6498 | 1.0000 | 0.6997 | 0.7483 | 0.7498 |
| 5 | 0.5355 | 0.3814 | 0.3962 | 0.6997 | 1.0000 | 0.5592 | 0.6121 |
| 8 | 0.8115 | 0.6860 | 0.6538 | 0.7483 | 0.5592 | 1.0000 | 0.6627 |
| 9 | 0.6231 | 0.5465 | 0.5595 | 0.7498 | 0.6121 | 0.6627 | 1.0000 |

Table 2 Clustering results of fuzzy c-means algorithm

| Method | Number of clusters | Generators | Average iterative times |
|---|---|---|---|
| With coherency measures | 6 | (2, 3), (1), (4), (5), (8), (9) | 2.24 |
| | 5 | (2, 3), (1, 8), (4), (5), (9) | 1.98 |
| | 4 | (2, 3), (1, 8), (4, 9), (5) | 2.05 |
| | 3 | (2, 3), (1, 4, 8, 9), (5) | 1.92 |
| | 2 | (1, 2, 3, 4, 8, 9), (5) | 1.85 |
| | 1 | (1, 2, 3, 4, 5, 8, 9) | 2.12 |
| Without coherency measures | 6 | (2, 3), (1), (4), (5), (8), (9) | 3.91 |
| | 5 | (2, 3), (1, 8), (4), (5), (9) | 3.28 |
| | 4 | (2, 3), (1, 8), (4, 9), (5) | 3.47 |
| | 3 | (2, 3), (1, 4, 8, 9), (5) | 3.35 |
| | 2 | (1, 2, 3, 4, 8, 9), (5) | 3.29 |
| | 1 | (1, 2, 3, 4, 5, 8, 9) | 3.51 |

Table 3 Comparison of the fuzzy c-means clustering and the similarity relation method [11] in the 3-cluster scheme

| Method | Computation times |
|---|---|
| Fuzzy c-means clustering | 333 |
| Similarity relation method | 668 |

Table 4 Elements of convergent membership matrix at the final sampling instant of the 3-cluster scheme

| Generator | | 1 | 2 | 3 | 4 | 5 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|
| Cluster group | 1 | 0.0024 | 0.9938 | 0.9943 | 0.0024 | 0.0004 | 0.0001 | 0.0078 |
| | 2 | 0.9971 | 0.0043 | 0.0045 | 0.9966 | 0.0019 | 0.9998 | 0.9917 |
| | 3 | 0.0005 | 0.0019 | 0.0012 | 0.0010 | 0.9977 | 0.0001 | 0.0005 |

the dynamics equivalent based on coherency is required.

A three-phase short circuit fault at bus 16 at $t = 0.1$ sec and then clearing the fault after three cycles by opening the line connecting bus 16 and bus 21 is used to obtain the swing curves. The response data from 0 sec to 2.5 sec are used for calculation of coherency measures. The sample period is 0.01 sec, so that the iterative procedures are done at all 250 sample instants. The swing curves of generators 1, 2, 3, 4, 5, 8, and 9 following the disturbance are shown in Figure 2. Table 1 tabulates the 7x7 coherency measure matrix for those generators from the swing curves. The index $C_{ij}$ reveals the degree of similarity in shapes between the swing curves of generator $i$ and generator $j$. For example, $C_{18}$ is the largest coherency measure value for generator 1, which means that generator 1 is more similar with generator 8.

In Table 2, results of the fuzzy c-means clustering approach with and without using the coherency measures as initial values of the membership matrix are given. Table 2 also gives the average iterative times of different clustering schemes. In the computation, the convergent tolerance is given as $\varepsilon = 0.01$ and the maximal iterative times is 50. Both methods have the same clustering results. Obviously, when the coherency measures are used as initial values, the average iterative times are smaller. The use of coherency measures in the fuzzy c-means clustering can let clustering procedure converge rapidly. The iterative times of the 3-cluster scheme of all sampling instants are given in Figure 3. It can be found that when the coherency measures are used, the clustering reaches at a stable state about at t=1 second, while that without coherency measures at t=1.5 seconds. In Table 2, the clustering result of the 3-cluster scheme is (2, 3), (1, 4, 8, 9), and (5). This means that generator 2 and generator 3 could be equivalent to be generator (2, 3). The generator (1, 4, 8, 9) is the equivalent of four generators. And generator 5 alone is in the third cluster. The main advantage of the proposed fuzzy c-means clustering technique lies in its capability of creating objective clustering of generators at any prescribed number of coherent groups. Any appropriate number of coherent groups can be determined according to the need for model reduction.

The comparison result of the proposed method with the similarity relation method [11] is shown in Table 3. The similarity relation method is described in Appendix. The method of similarity relation needs to begin the clustering from a maximal number of clusters, and then the next smaller number of clusters.

Table 5 Coherency measures of seven real generators and two virtual generators in the 3-cluster scheme

| Generator | 1 | 2 | 3 | 4 | 5 | 8 | 9 |
|---|---|---|---|---|---|---|---|
| Virtual generator (2, 3) | 0.7286 | 0.9095 | 0.9095 | 0.6681 | 0.3938 | 0.6804 | 0.5613 |
| Virtual generator (1, 4, 8, 9) | 0.8187 | 0.6866 | 0.6648 | 0.8507 | 0.6402 | 0.8531 | 0.7856 |

Table 6 Clustering results of the 3-cluster scheme using different time interval data

| Time interval (sec-sec) | Generators |
|---|---|
| 0-0.5 | (1), (2), (3, 4, 5, 8, 9) |
| 0-1.0 | (1, 2, 3, 8, 9), (4), (5) |
| 0-1.5 | (2, 3), (1, 4, 8, 9), (5) |
| 0-2.0 | (2, 3), (1, 4, 8, 9), (5) |
| 0-2.5 | (2, 3), (1, 4, 8, 9), (5) |
| 0-3.0 | (2, 3), (1, 4, 8, 9), (5) |

For example, if the 4-cluster scheme is needed, the clustering procedure should begin from the 6-cluster scheme, and then the 5-cluster scheme. But the fuzzy c-means clustering can begin the 4-cluster scheme directly. However, the results of the fuzzy c-means clustering are the same with the similarity relation method, but the former needs less computation times.

The final convergent membership matrix of the 3-cluster scheme is revealed in Table 4. In the cluster group 1, the elements of generator 2 and generator 3 are, respectively, 0.9938 and 0.9943, and are very close to 1, which proves that they form this cluster.

If generator (2, 3) and generator (1, 4, 8, 9) are used as the virtual (equivalent) generators, their cluster center values can be seen as the equivalent dynamic responses of those generators. Figure 4 shows the swing curves of generator (5) and the two equivalent generators. And the coherency measures of the seven real generators and two virtual generators can also be given in Table 5. It can be found that the coherency measures of virtual generator (2, 3) to real generator 2 and generator 3 are, respectively, 0.9095 and 0.9095. Those values are very close to 1. The coherency measures of generator (2, 3) to other generators are smaller. This proves that generator (2, 3) could be used to represent generator 2 and generator 3 and seen as an equivalent generator.

It is required to examine that whether using different time interval data would affect the clustering results. Table 6 shows the clustering results of the 3-cluster scheme using different time interval data. It can be found that if the time interval is equal of large than 2 sec, a stable clustering result can be obtained. It is because that the oscillation periods of those generators in Figure 2 are near 2 sec. So the data in the interval longer than the oscillation periods are required to determine the coherent generators.

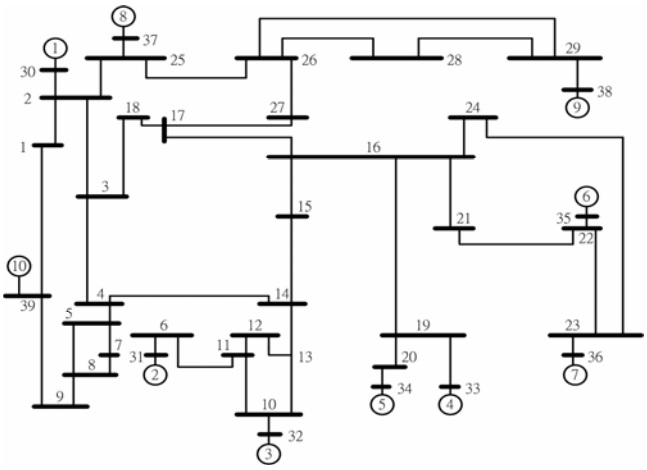

Figure. 1 One-line diagram of the New England test system

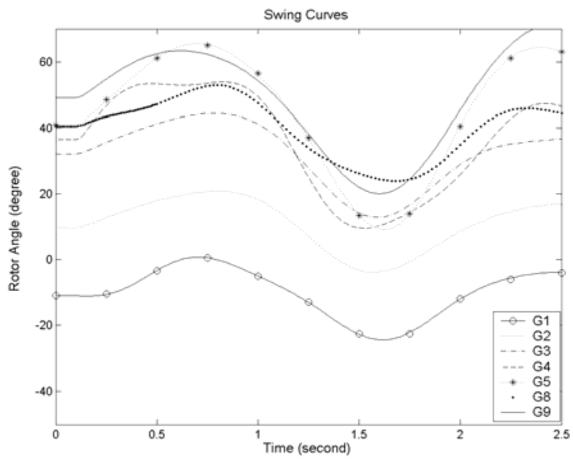


Figure. 2 Swing curves of all generators in the external system

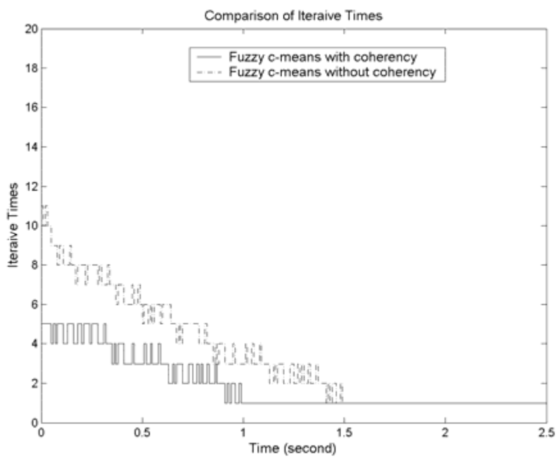


Figure. 3 Iterative times in the clustering procedure of the 3-cluster

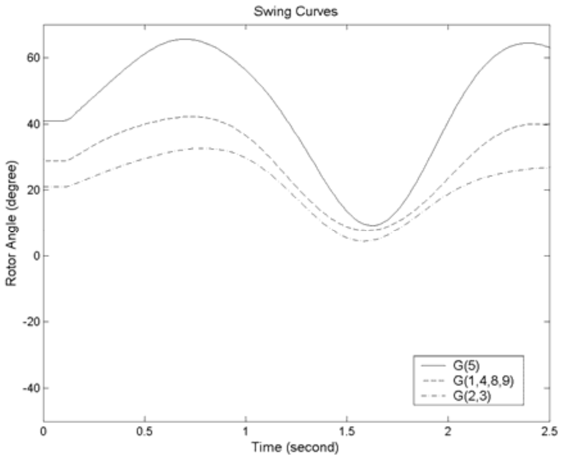


Figure. 4 Responses of three cluster centers in the 3-cluster scheme: generators (5), (1, 4, 8, 9), (2, 3)

## 5 Conclusion

A new method for the study of power system coherency, which comprises a coherency measure and a fuzzy c-means clustering approach, has been presented in this paper. The coherency measures are computed from the system time-domain dynamic responses. The fuzzy c-means clustering approach is used for processing the computed coherency measures and the clustering of generator groups at different prescribed number of coherent groups can be obtained. Application results of a sample test system are presented to show the validity and effectiveness of the proposed approach. By comparing to the similarity relation method, it can be found that the proposed method is flexible in determining the coherency groups. The iterative times are smaller. The responses of cluster centers for representing the coherent generators are also given. This approach is suitable for the selection of an equivalent system in power system stability analysis and is suggested for the future study.

*References:*


[1] O. Badeeb and G. Hazza, Application of the slow coherency decomposition method to the Yemeni network, *International Journal of Electrical Engineering Education*, Vol. 41, No. 1, 2004, pp. 56-63.

[2] R. Podmore and A. Germond, Development of Dynamic Equivalents for Transient Stability Studies, *EPRI Report EL-456*, 1977.

[3] J. Chow, *Time-Scale Modeling of Dynamic Networks with Applications to Power Systems*, Springer-Verlag, 1982.

[4] W. Price, A. Hargrave, B. Hurysz, J. Chow, and P. Hirsch, Large-scale system testing of a power system dynamic equivalencing program, *IEEE Trans Power Systems*, Vol. 13, No. 3, 1998, pp. 768-774.

[5] E. Armaca and N. Serifoglu, A rank correlation-based method for power network reduction, *Electrical Engineering*, Vol. 85, 2003, pp. 211-218.

[6] M. Jonsson, M. Begovic, and J. Daalder, A new method suitable for real-time generator coherency determination, *IEEE Trans Power Systems*, Vol. 19, No. 3, 2004, pp. 1473–1482.

[7] M. Wang and H. Chang, Novel clustering method for coherency identification using and artificial neural network, *IEEE Trans Power Systems*, Vol. 9, No. 4, 1994, pp. 2056-2062.

[8] M. Aldenderfer and R. Blashfield, *Cluster analysis*, Sage Publications, Beverly Hills, 1984.

[9] L. Kaufman and P. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*, John Wiley & Sons, 1990.

[10] F. Hoppner, F. Klawonn, R. Kruse, and T. Runkler, *Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition*, John Wiley & Sons, 1999.

[11] J. Bezdek, J. Keller, R. Krisnapuram, and M. Pal, *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*, Kluwer Academic, 1999.

[12] N. Pal, K. Pal, J. Keller, and J. Bezdek, A possibilistic fuzzy c-means clustering algorithm, *IEEE Trans Fuzzy Systems*, Vol. 13, No. 4, 2005, pp. 517–530.