Identification of coherent generators based on clustering methods

Wu Peng, Xing Lin, Zhang Ji, Xi Peng
Power Grid Planning Research Center, State Grid Hebei Economic Research Institute, Shijiazhuang, China
wupeng11@126.com

Abstract—The identification of coherent generators is very important for the stability analysis and control of power system. So far, there are three main identification methods of coherent generators which have certain disadvantages. In this paper, two new algorithms based on clustering are proposed for the identification of coherent generators. Algorithm 1 is the optimal classification of generators based on given swing curves. By introducing the validity index into the algorithm, the optimal classification of the given generator swing curves can be obtained. Algorithm 2 is identification of coherent generators under specific faults. This algorithm firstly calculates the critical clear time of the system. Since the swing curves obtained at the critical clear time(critical curves) contain the complete information reflecting the swing and oscillation between generators, they can be used as reliable samples for clustering. Examples show that the two algorithms are effective.

1. INTRODUCTION

When one group of generators is out of synchronization with the other group, the power system often fails to stabilize. In the transient process, a group of generators which can keep synchronous is called coherent generators. The identification of coherent generators is very important for the stability analysis and control of power system.

So far, there are three main identification methods of coherent generators.

1.1 The identification of coherent generators based on geographical division.

Because different regions of China's power system is connected by a few of long-distance transmission lines, the connection between regions is relatively weak. Therefore, generators in different regions are classified into different groups[1][2]. Generally speaking, identification of coherent generators based on geographical division is simple and convenient. The disadvantage of this method is that it can't give a more reasonable results according to the fault location, and the error is large[3].

1.2 The identification of coherent generators based on Slow coherency theory.

Slow coherency theory is based on the following knowledge. Generator groups with relatively weak electrical connections oscillate at a relatively slow frequency. The electrical connection between generators in the same group is relatively strong, and they oscillate at a faster frequency. Based on this theory, the coherent generator group is obtained by analyzing the eigenvalues and eigenvectors of the linear system after the linearization of the dynamic equation of the generators[4][5]. Reference[8] uses the generalized eigenvalues and eigenvectors of the differential algebraic equations describing the dynamic power system, combined with the Krylov-subspace method to identify the coherent generators.
1.3 Classification of generators swing curve based on Time-frequency characteristics and regulations. Classification of generators swing curve based on Time-frequency characteristics and regulations\(^{[9,10]}\). In reference\(^{[9]}\), the time-domain, frequency-domain and wavelet analysis are combined with rule base to extract the characteristics of swing curves. Identify the coherent group of stable multi swing curves in a limited period of time, and separate all potential instability groups. In reference\(^{[10]}\), the pattern method is used to classification swing curves. First, the fast Fourier transform is used to preprocess the swing curves, and then the feature set is obtained. Finally, a rule-based expert system is used to classify the feature set to obtain the organic group. A rule-based expert system is used to classify the feature set to get the generator groups. How to establish a complete and reasonable rule base is the key to whether the method is correct.

In this paper, two new methods based on fuzzy clustering for identification of coherent generators are proposed. The generators with the same time-domain characteristics of swing curve are put into a group by clustering method. The results of several examples show that the methods are effective.

2. Fuzzy C-means clustering algorithm (FCM) and validity check
In the clustering methods, the things that are clustered are usually called samples, and a group of things that are clustered are called sample sets. The idea of fuzzy C-means clustering(FCM) is to maximize the similarity of samples that are divided into the same cluster, while the similarity of samples between different clusters is the smallest.

2.1 FCM algorithm
Given sample set \(X = \{x_1, x_2, ..., x_n\}\) is a Set with \(n\) elements. The \(j\)-th element in data set \(x\) is an \(s\)-dimensional vector: \(x_j = \{x_{j1}, x_{j2}, ..., x_{js}\}\). Fuzzy clustering is to divide \(X\) into \(C\) categories \((2 \leq C \leq n)\), where \(v = \{v_1, v_2, ..., v_C\}\) is the cluster center. In fuzzy clustering, each sample is not strictly divided into a certain class, but belongs to a certain class with a certain degree of membership. Let \(u_{ij}\) denote the membership degree of the \(j\)-th sample belonging to the \(i\)-th category, and the membership matrix and clustering center are respectively expressed as \(U = \{u_{ij}\}\) and \(V = \{v_i\}\), and also \(u_{ij} \in [0,1], \sum_{i=1}^{C} u_{ij} = 1\).

The objective function of FCM analysis is:

\[
J(U,V) = \sum_{i=1}^{C} \sum_{j=1}^{n} u_{ij}^m d_{ij}^2
\]

where \(d_{ij} = \|x_j - v_i\| = \left[\sum_{k=1}^{s} (x_{jk} - v_{ik})^2\right]^{1/2}\)

\(d_{ij}\) is the Euclidean distance between the sample \(x_j\) and the cluster center \(v_i\), \(m \geq 1\) is the fuzzy weighted index, which indicates the fuzzy degree of the membership matrix \(U\). The larger \(m\) is, the higher the fuzzy degree of classification. Therefore, the FCM classification algorithm is to find a membership matrix \(U\) that make \(J(U,V)\) the minimum value under the constraint conditions \(\sum_{i=1}^{C} u_{ij} = 1\).

The constrained optimization problem composed of equations (1)-(2) can be solved by the Lagrange method. The algorithm flow is:
1) Initialization: Set the classification number \( C(2 \leq C \leq n) \) and give the fuzzy weighted index \( m=2 \). The number of samples is \( n \), and the iteration stop threshold \( \varepsilon \) is given. Given the initial classification matrix \( \mathbf{V}^{(0)} = \{v_1, v_2, \ldots, v_C\} \), let the number of iterations \( k=0 \).

2) Use the following formula (3) to calculate the membership matrix \( U^{(k)} \) at the \( k \)-th time:
   For any \( i (1 \leq i \leq C) \) and \( j (1 \leq j \leq n) \):
   \[
   u_{ij}^{(k)} = \frac{1}{\sum_{r=1}^{C} \left( \frac{d_{ij}^{(k)}}{d_{ij}^{(r)}} \right)^{\frac{1}{m-1}}} , \quad \text{if} \quad d_{ij}^{(k)} > 0 , \\
   u_{ij}^{(k)} = 1 , \quad \text{if} \quad d_{ij}^{(k)} = 0 . \quad \text{Also when} \quad r \neq i , \quad u_{ij}^{(k)} = 0 . \quad (3)
   \]

3) Update the cluster center \( V^{(k+1)} \) with the following formula (4):
   \[
   v_{ij}^{(k+1)} = \frac{\sum_{r=1}^{C} \left( u_{ij}^{(k)} \right)^m x_j}{\sum_{j=1}^{n} \left( u_{ij}^{(k)} \right)^m} 
   \]  
   \[
   (4)
   \]

4) If \( \| V^{(k+1)} - V^{(k)} \| < \varepsilon \), the algorithm ends, otherwise \( k=k+1 \), and return to step 2).

   It can be seen from the above algorithm flow that the FCM algorithm is a dynamic iterative process, and the cluster center matrix \( V \) and membership matrix \( U \) are continuously revised in each step.

2.2 Validity function in fuzzy clustering

Clustering validity function \( V_{\text{rie}}(c) \) can be used to measure the quality of clustering results. A good validity function requires that on the one hand, the class should be as compact as possible, that is, the smaller the value, the better; on the other hand, the distance between classes should be as far as possible, that is, the larger the value, the better. In this paper, the Xie-Beni validity function \( V_{\text{rie}} \) is used as a measure of the clustering algorithm.

The \( V_{\text{rie}}(U, V, c) \) validity function is:
   \[
   V_{\text{rie}}(U, V, c) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{C} u_{ij}^m \| v_i - x_j \|^2 \\
   \min_{i \neq j} \| v_i - v_j \|^2 
   \]

   Equation (2.5) represents the comparison between intra-class compactness and inter-class separation. The function numerator is used to measure the compactness within the class. The smaller the value, the more compact. The function denominator is used to measure the separation between the classes. The bigger, the better the separation. Therefore, for a cluster, the smaller the \( V_{\text{rie}}(U, V, c) \), the better.

   Let the average vector of all samples be \( \bar{x} \).
   \[
   \bar{x} = \frac{\sum_{i=1}^{C} \sum_{j=1}^{n} u_{ij}^m x_j}{n} 
   \]

   Then the clustering validity function is:
The specific process of the algorithm is as follows.

Initialization: Set \( u^{(k)}_0 = 0 \), \( C=2 \), the validity index is represented by \( L \), and the fuzzy weighted index \( m=2 \).

Starting from the initial classification number \( C=2 \), the validity index \( L(C) \) is calculated after each clustering (2.7). The cluster number \( C \) with the largest corresponding value of \( L(C) \) is the optimal cluster number. In actual situation, the number of clusters won’t be large, so the upper limit of \( C \) is 5 \( [12] \).

The specific process of the algorithm (Algorithm 1) is as follows.

Initialization: Set \( u^{(k)}_0 = 0 \), \( C=2 \), the validity index is represented by \( L \), and the fuzzy weighted index \( m=2 \).
1) The number of samples in the set is \( n \), given an iteration stop threshold \( \varepsilon \), the initial classification matrix is \( V^{(0)} = \{x_1, x_2, \ldots, x_n\} \), and the number of iterations \( k=0 \).

2) Calculate the membership matrix \( U^{(k)} \) at the k-th time.

   For any \( i(1 \leq i \leq C) \), \( j(1 \leq j \leq n) \):

   \[
   u^{(k)}_{ij} = \frac{1}{\sum_{r=1}^{C} \left( \frac{d^{(k)}_{ij}}{d^{(k)}_{ir}} \right)^{2}} \text{, if } d^{(k)}_{ij} > 0 , \\
   u^{(k)}_{ij} = 1 \text{, if } d^{(k)}_{ij} = 0 . \text{Also when } r \neq i \text{, } u^{(k)}_{ij} = 0 . \tag{9}
   \]

   Where \( d_{ij} = \|x_j - v_i\| = \left( \sum_{k=1}^{m} (x_{ik} - v_{ik})^2 \right)^{1/2} \) is the Euclidean distance between the sample and the cluster center.

3) Calculate and update the cluster center value \( V^{(k+1)} \).

   \[
   V^{(k+1)} = \frac{\sum_{j=1}^{n} (u^{(k)}_{ij})^m x_j}{\sum_{j=1}^{n} (u^{(k)}_{ij})^m} 
   \]

4) If \( \|V^{(k+1)} - V^{(k)}\| < \varepsilon \), go to step 5), Otherwise, let \( k=k+1 \), Go back to step 2).

5) Use formula (5) to calculate the validity function \( V_{sic} \), if \( V_{sic} > L \), update \( L \) to \( V_{sic} \), let \( C = C+1 \),turn to step 1). If \( C=5 \), the algorithm ends.

6) According to the matrix \( U \) after calculation stop, Determine the clustering result. Find the membership vector \( \{u_{i1}, u_{i2}, \ldots, u_{in}\} \) with each generator \( x_i \). If the largest element is \( u_{ij} \), then \( k \) is the group number of the generator \( x_i \).After each generator finds its corresponding group number, the generators group is obtained.

3.2 Identification of coherent generators under specific faults (Algorithm 2)

The method for calculating the critical clear time (CCT) based on the potential energy boundary (PEBS) is as follows\(^{[11]}\):

1) Make the fault continue and simulate along the continuous fault curve, using the energy function to calculate the potential energy \( V_{pe} (t) \) and kinetic energy \( V_{ke} (t) \) of the system at every moment.

2) The change trend of potential energy \( V_{pe} \) is to increase first and then decrease. Record the maximum value of \( V_{pe} \), record as \( V_{crit} \). \( V_{crit} \) is the critical energy.

3) The \( t \) which makes \( V_{crit} = V_{ke} (t) + V_{pe} (t) \),is the requested CCT.

Since the swing curves obtained at the critical clear time(named critical curves in this paper) contain the complete information reflecting the swing and oscillation between generators, they can be used as reliable samples for clustering .In this paper, identification of coherent generators under specific faults based on critical curves clustering is called algorithm 2.The process is as followed.

1) Calculate the critical clear time of the specified fault.

2) Using CCT as the fault removal time, the system is numerically simulated to obtain the generator swing curves.
3) Treat all generator rotor angle curves as a total sample \( X = \{x_1, x_2, \ldots, x_n\} \), \( n \) is the number of generators, and \( x_i \) represents the swing curve of the \( i \)-th generator.

4) initialization: Set the Number of clusters \( C=2 \). The number of data samples is \( n_\varepsilon \), given the iteration stop threshold \( \varepsilon \), the initial classification matrix \( V^{(0)} \) is selected as \( V^{(0)} = \{x_1, x_2, \ldots, x_C\} \), and the number of iterations \( k=0 \).

5) Use the following formula (10) to calculate the membership matrix \( U^{(k)} \):

For any \( i (1 \leq i \leq C) \), \( j (1 \leq j \leq n) \):

\[
u^{(k)}_{ij} (\text{if } d^{(k)}_{ij} > 0) = \frac{1}{\sum_{r=1}^{n} \left( \frac{d^{(k)}_{ij}}{d^{(k)}_{rj}} \right)^m} \]

\[
u^{(k)}_{ij} = 1, \text{ if } d^{(k)}_{ij} = 0 \]

(10)

where, \( d_{ij} = \|x_i - v_j\| = \left[ \sum_{j=1}^{n} (x_{ij} - v_{ij})^2 \right]^{1/2} \).

6) Use the following formula to calculate and update the cluster center value \( V^{(k+1)} \).

\[
u^{(k+1)}_{j} = \frac{\sum_{i=1}^{n} \left( \nu^{(k)}_{ij} \right)^m x_j}{\sum_{i=1}^{n} \left( \nu^{(k)}_{ij} \right)^m} \]

(11)

7) If \( \|V^{(k+1)} - V^{(k)}\| < \varepsilon \), the algorithm stops. Otherwise, let \( k = k+1 \), Go back to step 5).

8) According to the matrix \( U \) after calculation stop, Determine the clustering result. Find the membership vector \( \{u_1, u_2, \ldots, u_n\} \) with each generator \( x_i \). If the largest element is \( u_{i_k} \), then \( k \) is the group number of the generator \( x_i \). After each generator finds its corresponding group number, the generators group is obtained.

4. EXAMPLE ANALYSIS

Take IEEE-39 node system as an example, as shown in Figure 1.

![Figure 1 IEEE-39 node system](image-url)
Example 1: The fault line is 10-11. A three-phase short-circuit fault occurs at the near end of the bus 10, and the fault duration is 0.3 seconds. Then remove the fault and reclose. The swing curves of the corresponding 10 generators are obtained.

Algorithm 1 is used to cluster the swing curve of the generators. The result is that when the generators are divided into three groups, the validity index is the largest. The clustering result is: \{10\}, \{2,3\}, \{1,5,6,7,8,9\}.

To verify the results, the generator rotor angle curve is drawn in Figure 2 as followed.

![Generator rotor angle in example 1](image1)

It can be seen from Figure 2 that the result of clustering obtained by algorithm 1 is exactly the same as the distribution of the 10 curves.

Example 2: The fault line is 10-11. A three-phase short-circuit fault occurs at the near end of the bus 10. Algorithm 2 is used to cluster the swing curve of the generators. Firstly, the critical cut-off time is calculated as 0.25 seconds. Simulate the system at the critical time to get the generators’ swing curves. The result of clustering is: \{10\}, \{1,2,3,4,5,6,7,8,9\}.

To verify the results, the generator rotor angle curve is drawn in Figure 3 as followed.

![Generator rotor angle in example 2](image2)

It can be seen from Figure 3 that the result of clustering obtained by algorithm 2 is exactly the same as the distribution of the 10 curves, which proves that the result of algorithm 2 is correct.

5. CONCLUSION
In this paper, two new algorithms based on clustering are proposed for the identification of coherent generators. Algorithm 1 is the optimal classification of generators based on given swing curves. By introducing the effectiveness index into the algorithm, the optimal classification of the given generator swing curve can be obtained. Algorithm 2 is identification of coherent generators under specific faults. This algorithm firstly calculates the critical clear time of the system. Since the swing curves (critical curve) obtained at the critical clear time contain the complete information reflecting the swing and
oscillation between generators, they can be used as reliable samples for clustering. Examples show that the two algorithms are effective.

REFERENCES
[1] Zong Hongliang, Sun Guanghui, Liu Zhi, Wang Rong. Coordination of Out-of-step Protection Equipment in Large Power Systems[J]. Automation of Electric Power Systems. 2003, 27(22): 72-75
[2] LIU Yuanqi, LIU Yutian. An Islanding Cutset Searching Approach Based on Dispatching Area. Automation of Electric Power Systems. 2008, 32(11): 20-24
[3] ZHANG Baohui, WANG Chenggen, HAO Zhiguo. Problems and solutions of power system out-of-step islanding[J]. Electric Power Automation Equipment. 2010,30(10): 1-5
[4] You H, Vittal V, Wang X. Slow coherency based islanding. IEEE Transactions on power systems. 2004, 19(1): 483-491
[5] Wang Xiaoming, Vittal Vijay. System islanding using minimal cutsets with minimum net flow. Proceeding of IEEE Power Systems Conference and Exposition, New York, 2004
[6] Wang X, Vittal V, Heydt G T. Tracing Generator Coherency Indices Using the Continuation Method: A Novel Approach. IEEE Transactions on Power Systems. 2005,20(3):1510-1518
[7] QIAO Ying, SHEN Chen, LU Qiang. Islanding Decision Space Minimization and Quick Search in Case of Large-scale Grids. Proceedings of the CSEE2008, 28(22):23-28.
[8] Najafi S, Hosseinian S H, Abedi M. Proper Splitting of Interconnected Power System. IEEJ Transactions on Electrical and Electronic Engineering.2010, 5:211-220
[9] Dai Chensong, Xue Feng, Xue Yusheng. Classification of Disturbed Trajectories[J]. Automation of Electric Power Systems. 2001,1(10):13-16
[10] Zhang Jun, Chen Yi, Xue Yusheng. Classification of Swing Curves in Transient Stability Analysis[J]. Automation of Electric Power Systems. 1997,21(8):1-3
[11] Ni Yixin, Chen Shousun, Zhang Baolin, Theory and analysis of dynamic power system. Beijing, Tsinghua University Press, 2002
[12] Pal N. R, Bezdek J.C. On Clustering for the Fuzzy C-means Model[J].IEEE Trans FS,1995,3(3):370-379.