Research on state prediction of secondary device in smart substation based on dynamic interval weighted fuzzy c-means algorithm

Chen Zhongyang, Zhang Pengfei, Mahuan *
State Grid Shanghai Electric Power Research Institute, Shanghai, 200437, China
*Corresponding author’s e-mail: 443956240@qq.com

Abstract: A weighted fuzzy clustering algorithm based on dynamic interval is proposed to solve the problem of inaccuracy of missing attributes in clustering. By selecting the index that can reflect the secondary equipment state of intelligent substation comprehensively and combining with the weighted fuzzy clustering algorithm based on dynamic interval, the clustering classification result of the whole secondary equipment state of intelligent substation can be obtained accurately. According to the result, the corresponding maintenance strategy is adopted. The experimental results show that the weighted fuzzy clustering algorithm based on dynamic interval can effectively improve the clustering accuracy and the stability of the convergence, which can be very good to adapt themselves to the secondary device state overhaul.

1. Introduction
Maintenance of the state refers to the safety, reliability, environment, cost as the basis, through the equipment status evaluation, maintenance decision, to achieve safe and reliable operation, maintenance cost is reasonable a maintenance strategy. Compared with the traditional regular maintenance, the state maintenance can reasonably combine with the online monitoring system to conduct overall analysis on the operation status of each equipment in the substation and make an economical and effective maintenance decision. Through status detection and maintenance of the equipment, the running state of the equipment is evaluated, and the corresponding maintenance mode can be selected according to the results [1-3].

At present, a state maintenance of equipment has been carried out in depth. According to the requirements of equipment condition maintenance guidelines, various sensors and detection technologies have technically realized a state maintenance of equipment [4-6]. Compared with primary equipment, due to the low cost of secondary equipment and the complexity of secondary system, the application of state maintenance in relay protection is less. However, with the development of smart grid technology, the scale and number of relay protection equipment have made a leap in development, and the contradiction between the increase of equipment scale and the relatively insufficient maintenance force is increasingly prominent. With the large number of intelligent substations put into operation, the number of intelligent equipment increases sharply, and the coexistence of intelligent equipment and traditional equipment makes there are more secondary equipment in the site, and complex types and information increase the difficulty of correct evaluation of equipment on a large scale. Through modern evaluation algorithm, many researchers in China hope to build a state evaluation model with good effect. Expert evaluation model in the secondary...
equipment status evaluation system is dominant and based on the study of the laws and characteristics of the secondary equipment operation, on the basis of expert guidance planning out the state criteria, relevant staff and W this as the foundation, according to the established evaluation time, corresponding points for operation of the equipment, at the same time with the aid of the expert evaluation analysis model to determine the operation environment of the equipment.

In this paper, the status of relay protection equipment is analyzed by cluster analysis, which is a data mining method. According to the principle of clustering analysis and the idea of data mining, a new classification method for state maintenance of relay protection equipment based on clustering analysis is proposed[7]. The clustering analysis method in this paper is to perform q-type clustering on the repaired relay protection equipment according to the state quantity information of the equipment, and divide each equipment into its corresponding state class.

As the Fuzzy c-means algorithm (FCM), the traditional clustering analysis method, is greatly affected by noise and is very sensitive to the initial clustering center and the input order of sample data, usually individual singular values may lead to the deviation of the clustering center, resulting in strong randomness of the clustering results[8-10]. In this paper, the probability clustering algorithm is combined with the FCM algorithm, and the distance between the sample and the clustering center is given a weight, so that the clustering analysis results can better accord with the actual situation of electric power system, and a dynamic electric power system interval weighted fuzzy mean clustering algorithm is proposed. Compared with FCM algorithm, this algorithm improves the anti-jamming ability of clustering process. In addition, EWPCM algorithm in this paper assigns different weights to different evaluation contents according to the different proportion of inspected equipment in evaluation during state maintenance, which makes the clustering results more reliable.

2. Basic principles

2.1 Fuzzy c-means clustering algorithm

The purpose of fuzzy c-means clustering (FCM) algorithm is to achieve the maximum similarity between objects divided into the same cluster in a certain aspect, and the minimum similarity between objects belonging to different clusters.

Let the sample data set be: \(X=(x_1, x_2, \ldots, x_n)\), where \(X=(x_{k1}, x_{k2}, \ldots, x_{km})\) is the eigenvector of sample \(x_k\), corresponding to an object in the eigenspace. FCM algorithm is to generate a c partition of X on the sample set X.

Through the c partition of data set X, K clusters were set as: \(C=(c_1, c_2, \ldots, c_k)\). The membership relation between samples and clusters is expressed by membership degree \(\mu_{ij}\), and \(\mu_{ik}\) is the characteristic function of this subset, and \(0 \leq \mu_{ik} \leq 1\). Then we get the objective function of FCM algorithm is shown in (1) and (2):

\[
\min J_m = \sum_{i=1}^{N} \sum_{j=1}^{K} \mu_{ij}^m \left\| x_i - c_j \right\|^2
\]

(1)

\[
\sum_{j=1}^{K} \mu_{ij}^m = 1, \mu_{ij}^m \in [0,1]
\]

(2)

FCM algorithm is to evaluate the extreme value of the objective function. By using Lagrange algorithm and combining with the constraints of the above equation, the derivative of the above equation can be obtained is shown in (3) and (4):

\[
\mu_{ij}^m = \frac{1}{\sum_{k=1}^{K} \left( \frac{x_i - c_j^2}{c_i - c_k^2} \right)^{2/(m-1)}}
\]

(3)
3. Weighted fuzzy clustering algorithm

3.1 Electric power system weighted feature improvement fuzzy clustering model
For electric power system, the maintenance purpose of analysis is different, and the importance of each data object to the analyst is also different. For example, in the maintenance of the relay protection status of the electric power system, the relevant intelligent electronic device only needs to feedback the information of the relevant equipment or loop after collecting the status characteristic value of the key monitoring equipment, which greatly improves the operation accuracy and operation speed. Since the secondary device of smart substation is mainly composed of microprocessor, data memory, program memory, software and relay, the range of judging its state is different. Therefore, it is necessary to set interval factors to control the size of the data interval and realize the corresponding data attributes of the dynamic interval. In order to reduce the influence of sample outliers on clustering and improve the clustering accuracy, the interval samples were weighted. The weight is obtained by calculating the density of the sample in the neighboring region. The higher the density, the larger the power value, and the lower the weight value, so as to enhance the influence of the center sample on the selection of clustering center in the iterative process, and weaken the influence of outliers on the selection of clustering center. Finally, the interval weighted clustering analysis is carried out. From the perspective of the management system, the analysis of the data according to the different degrees of importance also meets the requirements of the management system on the work procedures and methods.

Based on the above analysis, we constructed the following model in formula (5):

$$J_m(U,P) = \sum_{k=1}^{n} \left( \sum_{i=1}^{c} (\omega_j d_{ik} - v_{ik})^2 \right)^{-1} + \sum_{i=1}^{n} \left( \sum_{k=1}^{c} (\varphi_k - \mu_{ik})^2 \right)^{-1} + \gamma \sum_{i=1}^{c} \sum_{k=1}^{c} \left( \sum_{j=1}^{s} (\omega_j d_{ij})^2 \right)^{-1}$$

(5)

$$\mu_{ik} > 0, 0 \leq \omega_j \leq 1, j \leq s$$

In this model, $\omega_j$ is the special weight coefficient of the objective function, representing the weight of the j data object. In the electric power system, for the data information obtained from the key equipment, the weight can be increased to make the final result more accurate. $R$ is the total number of eigenvalues of the data object; $d_{ik}$ represents the distance value of the K dimension component of the j data point to the I data point, and $v_{ik}$ represents the interval center between the j data point and the i data point, whose upper and lower bounds are $v_{ik} = [v_{ik}^-, v_{ik}^+, ... v_{nk}^-]$, $v_{ik}^+ = [v_{ik}^+, v_{ik}^+, ... v_{nk}^+]$. Represents the distance value of the k dimension component from the j center to the i center. In this algorithm, in order to distinguish PCM algorithm, highlight the role of eigenvalues and meet the needs of electric power system data analysis, it must be guaranteed $\sum_{i=1}^{n} (\omega_j d_{ik} - v_{ik})^2 \neq 0$.

3.2 Algorithm implementation process
On the basis of the above algorithms and in combination with the optimization of m and c, a complete weighted eigenvalue fuzzy clustering algorithm is proposed. The algorithm steps are as follows:

Step 1: determine the value of the optimal number of clusters c and m, determine the value of the characteristic $\omega$ weight according to the importance of the sample, and determine the accuracy of a calculation;

Step 2: initialize $\mu_{ik}$, $c_i$, and $\gamma$, $i \leq c$, $k \leq n$: 

$$c_j = \frac{\sum_{i=1}^{N} \mu_{ij}^m x_i}{\sum_{i=1}^{N} \mu_{ij}^m}$$

(4)
Step 3: use formula (6) to calculate $\eta_k$

$$\eta_k = \frac{\sum_{i=1}^{n} \mu_{ik}^m \sum_{j=1}^{s} (\omega_j d_{ij} - v_{ilk})^2}{\sum_{k=1}^{n} \mu_{ik}^m}$$

Step 4: use (7) to calculate $\mu_{ik}$

$$\mu_{ik} = \frac{1}{1 + \left( \frac{1}{\eta_i} \sum_{j=1}^{s} (\omega_j d_{ij} - v_{ilk})^2 \right)^{1/m-1}}$$

The membership degree $\mu_{ik}$ is calculated by the formula, where $\mu_{ik}$ is obtained by the extreme value of the objective function in the model.

Step 5: use formula (8) to calculate the cluster center $c_i$. In formula (8), we can get $f(c_{ij}^{(p-1)})$ through (9), and get $g(c_{ij}^{(p-1)})$ through (10).

$$c_{ij}^{p} = \frac{f(c_{ij}^{(p-1)})}{g(c_{ij}^{(p-1)})}$$

$$f(c_{ij}^{(p-1)}) = \sum_{k=1}^{n} (\mu_{ik})^m x_{ij} - \gamma \sum_{l=1, l \neq i}^{s} \left( \sum_{j=1}^{s} (\omega_j |c_{ij}^{(p-1)} - c_{ij}^{(p-1)}|)^2 \right)^{-2} c_{ij}^{(p-1)}$$

$$g(c_{ij}^{(p-1)}) = \sum_{k=1}^{n} (\mu_{ik})^m - \gamma \sum_{l=1, l \neq i}^{s} \left( \sum_{j=1}^{s} (\omega_j |c_{ij}^{(p-1)} - c_{ij}^{(p-1)}|)^2 \right)^{-2}$$

$$P$$ is the number of iterations.

Step 6: if $\|c_{ij}^{p} - c_{ij}^{p-1}\| < \epsilon$, the algorithm stops and outputs partition matrix $U$, clustering prototype $P$ and clustering center $c$; otherwise, turn to step 3.

4. Algorithm application analysis

In order to make the evaluation indicator, truly reflect the secondary intelligent equipment running status, and considering the maneuverability of the secondary device condition assessment, in addition to the equipment basic information, build light intensity, power supply module, CPU temperature, supply voltage, SV and GOOSE faults is state parameters of the secondary intelligent device status evaluation system.

The traditional status evaluation of relay protection devices is to compare the status of these devices with the standard limit after collecting the status of these devices. According to the amount of state of the device, the device is divided into three states: normal state, attention state and abnormal state. If the state of the equipment quantity stable and within the standards prescribed by the regulations limit, the equipment is divided into normal state if one or several state close to the standard limit, or the equipment defects generally or important unsolved, or there is a general and important counter measures work has not been completed, but still can continue to run, the equipment is to pay attention to the state; If a certain state quantity of the equipment exceeds the standard limit, or the emergency countermeasures work is completed, or the emergency defects are eliminated, the equipment is classified as an abnormal state, and the equipment in this state can only operate for a short time or stop service immediately. After the status evaluation of the equipment, the maintenance level of the equipment can be determined, including A type of maintenance, B type of maintenance, C type of maintenance and D type of maintenance.
In this paper, the cluster analysis of equipment according to state quantity is carried out to determine the cluster to which the equipment belongs. In order to combine the clustering algorithm, three initial clustering centers should be set according to three device states. Then, each device is assigned to a corresponding cluster through different clustering algorithms. In the process of calculation, the current status of all equipment can be calculated directly and in real time to facilitate timely maintenance decisions.

In the cluster analysis simulation experiment, in order to facilitate the analysis and improve the quality of the data, the obtained state quantity data should be preprocessed first, and the incomplete and noise data in the samples should be removed. Then these data are standardized to remove the dimension of different state variables, so as to avoid the large deviation of clustering center caused by the data of different dimension. After data standardization, cluster analysis can be carried out on data samples. For weighted fuzzy clustering algorithm in this paper, weighted index \( m \) is 2, computing accuracy \( \varepsilon \) is 0.01, and \( \gamma \) is 0.2. In EWPCM algorithm, the evaluation standard of relay protection equipment is added, and the number of clustering \( c \) is set as 3. The three initial clustering centers are:

\[
(4.08, 11.5, 2.4, 9.55, 9.5, 9.55, 9.55, 5.7, 3) ;
(2.86, 8.5, 1.65, 7.55, 7.5, 7.55, 7.55, 4.7, 2) ;
(2, 4.5, 0.85, 3.55, 3.5, 3.55, 3.55, 3.6, 0.8).
\]

According to the requirements of the weighted fuzzy clustering algorithm and the content of the guideline, different evaluation contents in the sample have different importance to the clustering center, so corresponding weights should be assigned to different evaluation contents. The weighted coefficient of sample data constructed in this experiment is:

\[
(0.05, 0.15, 0.15, 0.15, 0.1, 0.1, 0.1, 0.1, 0.1).
\]

The distance between each evaluation content clustering obtained through k-means clustering analysis is shown in table 1:

| Cluster 1 | Cluster 2 | Cluster 3 |
|-----------|-----------|-----------|
| 1 ——— | 10.62 | 9.96 |
| 2 10.62 | ——— | 6.02 |
| 3 9.96 | 6.02 | ——— |

The distance between each evaluation content clustering obtained by weighted clustering algorithm is shown in table 2:

| Cluster 1 | Cluster 2 | Cluster 3 |
|-----------|-----------|-----------|
| 1 ——— | 8.55 | 10.57 |
| 2 8.55 | ——— | 7.82 |
| 3 10.57 | 7.82 | ——— |

By comparing the distance between the two algorithms, it can be seen that, compared with table 1, the differences of each cluster in table 2 are more obvious and the clustering effect is better. Therefore, the weighted fuzzy clustering algorithm can obviously improve the clustering effect.

Finally, the three final clustering centers obtained by the dynamic interval weighted fuzzy c-means algorithm are shown in table 3.

| Element | Cluster1 | Cluster2 | Cluster3 |
|---------|----------|----------|----------|
| Operating environment | 1.83 | 2.67 | 3.09 |
| Mean time between failure | 2.78 | 4.15 | 6.30 |
| Family mean time between failure | 3.32 | 8.55 | 9.64 |
| Light intensity in the module | 3.66 | 6.49 | 6.74 |
The power supply temperature 2.45 4.49 5.41
CPU temperature 8.54 6.34 2.50
Mains input 9.57 10.98 7.33
SV alarm 4.66 3.56 4.33
GOOSE alarm 2.49 2.30 5.15

The algorithm is combined to obtain each cluster member and the number of devices in each cluster, as shown in table 4 and table 5:

Table 4 Relationship between devices and clusters

| No | Subordinate to the clustering | Distance from center |
|----|-------------------------------|---------------------|
| 1  | 3                             | 5.01                |
| 2  | 1                             | 5.10                |
| 3  | 3                             | 5.73                |
| 4  | 3                             | 5.45                |
| 5  | 3                             | 7.89                |
| 6  | 3                             | 5.92                |
| 7  | 3                             | 4.47                |
| 8  | 2                             | 5.64                |
| 9  | 1                             | 5.11                |
| 10 | 3                             | 6.22                |
| 11 | 2                             | 5.20                |
| 12 | 3                             | 7.82                |
| 13 | 2                             | 6.40                |
| 14 | 2                             | 4.83                |
| 15 | 3                             | 7.04                |

Table 5 Number of devices of every cluster

| Cluster | Devices Number |
|---------|----------------|
| 1       | 2              |
| 2       | 4              |
| 3       | 9              |

According to the clustering results in table 3 and table 4, the status evaluation results of each device can be clearly obtained. Through these results, the corresponding maintenance decisions can be made.

5. Conclusion

The dynamic interval weighted fuzzy c-mean algorithm proposed in this paper, based on the requirements of secondary device state maintenance of smart substation, not only provides a new method of state evaluation in secondary device state maintenance of smart substation, but also improves the clustering result by combining the content of state evaluation and the right value. Therefore, the status maintenance of secondary device in smart substation has the characteristics of real-time and effective.

References

[1] Shi Y., Jiang Z., An C. (2016) Research on practical state evaluation of protection device in smart substation, 44(10):119-125.
[2] Xiao F., Wang Z., Zhang Z., et al. (2018) Study on maintenance strategy of relay protection system based on condition monitoring, 46(06):74-83.
[3] Ji G., Wu W., Zhang B., et al. (2013) A time-varying component outage model used in electric power system condition-based maintenance[J]. Proceedings of the CSEE, 33(25): 139-146.
[4] Zhang H., Wei B., Li K., et al. (2016) Research on risk assessment technology based on Markov state evaluation model for power transformer and entropy weighted fuzzy comprehensive evaluation[J]. Electric power system Protection and Control, 44(5): 134-140.

[5] Li Ming, Han Xueshan, Wang Yong, et al. Decision making model and solution of condition-based maintenance for substation[J]. Proceedings of the CSEE, 2012, 32(25): 196-202.

[6] Lu Z., Liu K. (2019) The method of fuzzy cluster analysis and its application[J]. Mathematics in Practice and Theory, 49(06): 288-291.

[7] Yue S., Huang T., Wang P. (2014) Matrix Eigenvalue Analysis-Based Clustering Validity Index[J]. Journal of Tianjin University (Science and Technology), 47(08): 689-696.

[8] Ge W., Wang Z., Gu H., et al. (2011) The information fusion of condition based maintenance system for primary and secondary power equipments and its application research in PMS system[J]. Electric power system Protection and Control, 39(21): 150-154.

[9] Chen Y., Li S., Zhang L., et al. (2019) Association analysis for defect data of secondary device in smart substations based on improved Apriori algorithm[J]. Electric power system Protection and Control, 47(20): 135-141.

[10] Guo S., Tong X., Wang H. (2019) Smart substation protection system reliability assessment method considering the aging of equipment and planned maintenance[J]. Electric power system Protection and Control, 47(09): 97-104.