Electrical Power and Energy Systems 118 (2020) 105651

Contents lists available at ScienceDirect

Electrical Power and Energy Systems

journal homepage: www.elsevier.com/locate/ijepes

A novel fault diagnosis method for circuit breakers based on optimized affinity propagation clustering

Yang Lu, Yongli Li*

Key Laboratory of Smart Grid of the Ministry of Education, Tianjin University, Tianjin 300072, China

ARTICLE INFO

Keywords:
Circuit breaker
Trip/close coil current
Clustering
Fault diagnosis

ABSTRACT

Online condition monitoring and fault diagnosis of circuit breakers (CBs) is a significant method to effectively improve the stability and reliability of the power system. However, the currently used fault diagnosis method still have certain defects including the inability to identify unknown faults for training samples. Therefore, this paper proposes an evolving method for fast and accurate online fault diagnosis of CBs. On the basis of collecting samples of CB trip/close coil current (CC) features, an optimized affinity propagation (AP) clustering algorithm to accurately extract the sample clustering exemplars is presented. Additionally, operating state identification and fault diagnosis of CBs is carried out by calculating the similarity coefficient between the new sample and exemplars online. Diagnosis of unknown faults is also achieved by introducing the threshold and comparing it with similarity coefficient results. Simulation results prove that the proposed method can precisely identify various known CBs faults and has the ability to recognizes unknown CBs fault samples even when the number of training samples is small, providing a foundation for CB fault location and condition-based maintenance.

1. Introduction

Circuit breakers (CBs) are recognized as one of the most crucial components to power equipment. They are the key to isolating faulty components driven by protection devices, and play a dual role in the protection and control of the power system [1]. Incorrect operation of CBs can cause a power grid accident or expand the scope of the accident. In severe cases of CB failure, the power system may collapse and cause major economic losses. Therefore, online monitoring and fault diagnosis of CBs have practical significance for enhancing the reliability and stability of the electric power system.

According to the CIGRE surveys, more than 80 per cent of the fault of CBs is caused by operating mechanism and auxiliary control circuits failures. The trip and close coil current (CC) signal is an accessible and noninvasive parameter in CB online condition monitoring. Previous studies have determined that analysis of the CC characteristics can identify effectively many signs of various faults type occurring in control circuits and operating mechanism [2–4]. In recent years, increasingly advanced data-analytics algorithms have been applied to implement the fault diagnosis of various kinds of power equipment [5,6], and also provide a new approach in the fault diagnosis of CBs [7–10]. In Ref. [11], a fault diagnosis method was proposed which has high detection accuracy, and utilizes back-propagation neural network (BPNN) technique. BPNN requires a large number of training samples to ensure an accurate diagnosis. However, it is usually quite difficult to obtain general and sufficient CC fault samples in practical applications, which is because the operating frequency and average failure rate of CBs are relatively low, and different types of the CB may have different CC waveform. In addition, Ref. [12] proposed a fault diagnosis method combining CC characteristics and support vector machine algorithm, which can obtain better diagnostic results compared with the BPNN method in the small training sample case.

However, the methods above are all based on the supervised learning algorithm, which trains a diagnosis model depended on samples combined of feature data and corresponding type label. One of main defects of these conventional methods is that the diagnosis model can only be used to identify the sample of the normal or known fault types of the CB, which are already included in the training set. An unknown fault type sample will be classified arbitrarily into the normal or a known type, which obviously leads to a wrong diagnosis result [13]. In particularly, due to the small number of fault samples in CB historical data, and the difficulty in simulating each type of fault through experiment, it is generally not possible to obtain CC samples of all failure types at the training stage. Therefore, the relevant supervised learning-based methods become difficult to reproduce in application effectively and flexibly.

* Corresponding author.
E-mail address: lyltju@163.com (Y. Li).

https://doi.org/10.1016/j.ijepes.2019.105651
Received 8 January 2019; Received in revised form 18 September 2019; Accepted 24 October 2019
0142-0615/

Therefore, as a common technique of unsupervised learning, cluster analysis is considered as a feasible way to solve the above problem in the paper. The clustering is used to draw inferences from datasets consisting of input without label responses, to find hidden patterns or groupings in data [14]. Cluster analysis algorithm is proven to be an effective theoretical basis for the fault diagnosis method, and has broad application prospects in the field of fault diagnosis for power system equipment [15–17]. In Ref. [16], a method based on kernel fuzzy c-means algorithm (KFCM) was proposed to recognize known and unknown faults in a wind turbine gearbox. Cluster analysis also has a preliminary application in the fault diagnosis of CBs. In Ref. [17], the agglomerative hierarchical clustering (AHC) method is employed to identify CB in a normal or fault state based on the CC features. However, there is currently no relevant research result that can effectively diagnose CB fault types in small sample cases and have the ability to identify unknown faults.

In considering the insufficient current research into CB fault diagnosis, this paper proposes a novel CB fault diagnosis method based on optimized affinity propagation (AP) clustering algorithm. The sample set of CB trip/close CC features collected by historical data or experimental data is first classified using AP clustering. This clustering technique is an advanced unsupervised learning method which has the ability to automatically output high quality cluster numbers and clustering exemplars of a sample set [18]. In addition, this paper selects the appropriate clustering result validity indices to achieve parameter adaptive optimization of the algorithm, which effectively improves the performance of clustering. By calculating the similarity coefficient between new data and known cluster exemplars online, the faults of the known or unknown types can be effectively identified.

The contributions of this research are summarized as follows.

(1) Based on optimized AP clustering, a novel CB fault diagnosis method is designed in the paper, which can accurately identify the CB fault type corresponding to the CC feature sample even when the number of training samples is small.
(2) The optimal similarity matrix is constructed as clustering input based on Euclidean distance function and Gaussian kernel function.
(3) The additional parameter optimization process is added to classify the training samples effectively and output high validity cluster exemplars.
(4) The similarity coefficient based on cluster exemplars is proposed to carried out to effectively recognize unknown faults data of CC online, which will be incorrectly classified by conventional methods.

To highlight the superiority of the proposed method and provide detailed and practical guidelines, the remaining paper is structured as follows: Section 2 analyzes the correlation between the CC features and various categories of CB operating states. Section 3 introduces AP clustering, which is the theoretical basis of the proposed diagnosis method. In Section 4, the improvement of AP clustering algorithm to increase the CBs diagnostic accuracy is discussed, including the construction of a similarity matrix and the optimization of parameters. Section 5 presents the complete diagnostic procedure. In Section 6, comparative simulations are performed to validate the excellent diagnostic accuracy of the proposed method and the feasibility of identifying unknown fault types.

2. CB trip/close CC signals and CB operation

During the trip/close operation of CB, the control unit transforms the CC signal into the mechanical operation of the breaker switching mechanism [19]. The trip/close CC contains crucial information that can be utilized by online monitoring and fault diagnosis of the CBs. This section analyzes the correlation between the CC characteristics and CB operation performance. A typical close CC waveform of normal state CB in the closing process is presented in Fig. 1 (trip CC waveform characteristics are similar) [2–4].

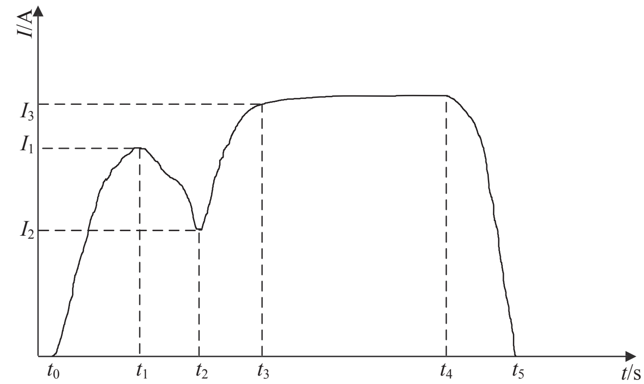

Fig. 1. Illustration of close CC waveform in CB close operation process.

Table 1
Description of CC waveform stages and CC features.

Stages	Stage description	Features	Feature description
$t_0 \sim t_1$	Coil excitation process	t_1	Excitation time
		I_1	First peak current
$t_1 \sim t_2$	Coil plunger motion process	t_2	Latching time
		I_2	Trough current
$t_2 \sim t_4$	Mechanism travel process	t_3	Second peak time
		t_4	Auxiliary contact operation time
		I_3	Second peak current
$t_4 \sim t_5$	Current decay process	t_5	Total energizing time of CC

As seen in Fig. 1, the close CC waveform of a normal state CB contains two peaks and one trough. It has eight features which include three current features and five time features that divide CC waveform into four stages. Table 1 presents the description of waveform stages and CC features, indicating that the CC waveform can be regarded as the dynamic curve of the plunger and mechanism operation process, and therefore contains information such as supply voltage, air gap, electromagnetic force and spring resistance.

Analysis of the CC in various CB operation states demonstrates that as the operation state of the CB is altered, the value of current and time features intuitively changes. The features value of the abnormal state of CB is significantly different from the normal state value, while the same anomaly state has similar feature characteristics, which can be utilized to recognize various CBs states [7–10]. Various CBs failures and their causes are summarized in Ref. [20], and the most affected features of CC can be viewed in Table 2.

Table 2 illustrates that current peaks I_1, I_3, and trough I_2 can reflect information such as supply voltage, auxiliary contact status. The performance of coil excitation, as well as the equivalent inductance and resistance of the coil, can be reflected by the features in stage $t_0 \sim t_1$. In addition, stage $t_1 \sim t_2$ is the coil plunger movement process, which can reveal whether plunger jamming or latch tripping occurs. Finally, the performance of the operating mechanism can be shown by the CC features in stage $t_2 \sim t_4$, in which the spring mechanism drives the moving contact to close. Overall, it is feasible to utilize the value of CC

Table 2
CBs failure types and relevant affected features.

Failure type	Affected features
Voltage supply decreasing	I_1, I_2, I_3
Stiff latch	t_1, t_2, t_3, t_4
Operating mechanism jamming	t_4, t_5
Coil excitation abnormal	t_1
auxiliary contact malfunction	t_4, t_5, I_3

features for distinguishing various fault types of the CB.

3. Theoretical basis of proposed method

Fault diagnosis for CBs utilizing CC features can be classified as a process of data mining, using advanced data-analytics algorithms to locate useful information in a data set [21]. Cluster analysis is a discovery data-mining technique that organizes the research objects into meaningful clusters [22]. This section introduces the advanced AP clustering algorithm [18] employed in this paper, which can be used as the theoretical basis of the proposed diagnostic method, and the cluster external criteria F-measure indices to evaluate the validity of the clustering results[23].

3.1. AP clustering algorithm

This paper employs the AP clustering algorithm, which is an unsupervised learning method, based on the following advantages:

(1) The number of target clusters do not need to be input before the clustering process to carry out CBs fault diagnosis [24].
(2) The algorithm automatically outputs high-quality clustering exemplars, which can be used as a basis for online diagnosis of new data [25].

Assuming that the data set $\boldsymbol{X} = \{x_1, x_2, \cdots\cdots, x_n\}$ contains the size of the data as n, the input of AP is n-order similarity matrix $\mathbf{S}$, and its elements *S(i,j)* can be described as follows:

$$S(i, j) = \begin{cases} S_{ij} & i \neq j \\ p & i = j \end{cases} \quad i, j \in [1, 2, \cdots, n] \tag{1}$$

where S_{ij} is the quantitative similarity between data point x_i and x_j, which can be expressed as a negative value of a type of distance function. The p is called ‘preference’, and its value, which can significantly affect the clustering results, usually can be determined follows:

$$p = \mathrm{median}(S(i, j)) \quad i \neq j \tag{2}$$

The AP algorithm achieves clustering through evidence exchanged between data points. The two kinds of evidence are ‘responsibility’ and ‘availability’, whose update follows Eqs. (3) and (4) during the iteration:

$$R(i, k) = \begin{cases} S(i, k) - \max\limits_{k' \neq k}(A(i, k') + S(i, k')) & i \neq k \\ p(k) - \max\limits_{k' \neq k}(A(k, k') + S(k, k')) & i = k \end{cases} \tag{3}$$

$$A(i, k) = \begin{cases} \min\left(0, R(k, k) + \sum_{i' \notin \{i,k\}} \max(0, R(i', k))\right) & i \neq k \\ \sum_{i' \notin \{i,k\}} \max(0, R(i', k)) & i = k \end{cases} \tag{4}$$

where $S(i, k)$ and $S(i, k')$ are the i row k column element and i row k' column element of the similarity matrix $\mathbf{S}$, *R(i,k)* is the ‘responsibility’ evidence sent from data x_i to candidate exemplar x_k, while *A(i,k)* is the ‘availability’ evidence sent from candidate exemplar x_k to data x_i. The stronger the evidence (the greater the sum of *R(i,k)* and *A(i,k)*), the greater probability that x_k is one of the final exemplars.

AP clustering eliminates some candidate exemplars in each iteration, and finally determines high-quality exemplars $\{o_1, o_2, \cdots\cdots, o_C\}$ for a total of C, which is also the number of clusters, in order to minimize the clustering energy *E(C)*. The *E(C)* can be defined via:

$$E(C) = -\sum_{i=1}^{N} S(i, o_j) \tag{5}$$

where $j \in [1, 2, \cdots, C]$, o_j is the exemplar of cluster C_j, and $i \in c_j$.

To accelerate the convergence process while avoiding numerical oscillations, a damping factor $\eta \in [0, 1)$ is often introduced to scale the evidence *R(i,k)* and *A(i,k)* following the Eqs. (6) and (7).

$$R(i, k)_{new} = (1 - \eta) \times R(i, k) + \eta \times R(i, k)_{old} \tag{6}$$

$$A(i, k)_{new} = (1 - \eta) \times A(i, k) + \eta \times A(i, k)_{old} \tag{7}$$

3.2. F-measure indices

In addition to AP clustering algorithm, it is necessary to quantitatively evaluate the clustering results through the suitable cluster validity assessment method [26]. When the classification label information of the data set is available, the external criteria F-measure indices can be introduced [27].

The F-measure indices (F-M) combine the concepts of ‘precision’ and ‘recall’ in information retrieval for cluster evaluation [23]. The precision and recall of a cluster j and classification i associated with j are defined as:

$$P_{ij} = precision(i, j) = \frac{N_{ij}}{N_i} \tag{8}$$

$$R_{ij} = recall(i, j) = \frac{N_{ij}}{N_j} \tag{9}$$

where N_{ij} is the number of data in classification i for cluster j, N_j is the number of data in cluster j, and N_i is the number of data in classification i. The F-M of classification i. is the harmonic mean between precision and recall and can be determined through:

$$F(i) = \frac{2P_{ij}R_{ij}}{P_{ij} + R_{ij}} \tag{10}$$

For clustering results, the total F-M can be obtained from the weighted average of *F(i)* for each classification:

$$F = \frac{\sum_i [N_i \times F(i)]}{\sum_i N_i} \tag{11}$$

Obviously, the value of F is between 0 and 1, and the larger the value, the better the consistency between the cluster result and classification information. When they are identical, F is 1.

4. Optimized AP clustering algorithm

In order to increase the applicability of the original AP clustering to the sample of CC features and the accuracy of the clustering result in the training stage, availably improvements are carried out in the paper. The effectiveness of the optimized algorithm can be verified by the F-M introduced in Section 3.

4.1. Optimized similarity matrix

The AP clustering input is the similarity matrix $\mathbf{S}$, its non-diagonal element *S(i,j)* can select the corresponding distance function according to the different research objects. Therefore the selection of the function seriously affects the clustering result [18].

According to the characteristics of the CC features sample, The Euclidean distance measure can be employed as a distance function to express (dis)similarity between data points via Eq. (12) [17]:

$$d(x_i, x_j) = \sqrt{\sum_{l=1}^{m} |x_{i_l} - x_{j_l}|^2} = d(x_j, x_i) \tag{12}$$

where x_i and x_j are two samples with the length m. Therefore, according to Eq. (1), the similarity matrix $\mathbf{S}_{old}$ can be constructed as:

$$\mathbf{S}_{old} = \begin{bmatrix} p & -d(x_1, x_2) & \cdots & -d(x_1, x_N) \\ -d(x_2, x_1) & p & \cdots & -d(x_2, x_N) \\ \vdots & \vdots & \vdots & \\ -d(x_N, x_1) & -d(x_N, x_2) & \cdots & p \end{bmatrix} \tag{13}$$

However, chosen $\mathbf{S}_{old}$ as the AP clustering input may cause some contingency. The value of the distance between the features data of various fault types may be obviously different, which can decrease the validity of the clustering result. This is because AP clustering may classify two types of faults with relatively small distance values as one cluster due to the characteristic that automatically outputs a clustering number.

Therefore, based on Euclidean distance, the distance function is improved by following the structure of the Gaussian kernel function [28]. Under the original condition, the difference of distance value between various fault types is reduced. The similarity can be quantified as:

$$S(i, j)_{new} = \exp\left(-\frac{d(x_i, x_j)}{2\sigma^2}\right) - 1 \tag{14}$$

where σ is width coefficient. According to Eq. (14), the value of $S(i,j)_{new}$ ranges from -1 to 0, and the optimized similarity matrix can be derived in Eq. (15), where p is the preference.

It can be seen from Eq. (15) that $\mathbf{S}_{new}$ contains two main parameters σ and p, hence their value will clearly affect the validity of the clustering result.

$$\mathbf{S}_{new} = \begin{bmatrix} p & \exp\left(-\frac{d(x_1,x_2)}{2\sigma^2}\right) - 1 & \cdots & \exp\left(-\frac{d(x_1,x_N)}{2\sigma^2}\right) - 1 \\ \exp\left(-\frac{d(x_2,x_1)}{2\sigma^2}\right) - 1 & p & \cdots & \exp\left(-\frac{d(x_2,x_N)}{2\sigma^2}\right) - 1 \\ \vdots & \vdots & \vdots & \\ \exp\left(-\frac{d(x_N,x_1)}{2\sigma^2}\right) - 1 & \exp\left(-\frac{d(x_N,x_2)}{2\sigma^2}\right) - 1 & \cdots & p \end{bmatrix} \tag{15}$$

4.2. Select optimal parameters

The validity of the AP clustering result depends on the selection of its input parameters σ and p. The optimal values of the parameters for different sample sets are also different [29]. The adaptive optimization of parameters can be carry out by selecting the appropriate cluster internal criteria to evaluate the results of the cluster number space, which are obtained by scanning the given parameter space. The cluster internal criteria differs from external criteria by using only quantities and features inherent to the dataset [27].

In order to select the optimal parameters, this paper introduces two quality evaluation indices: 'cluster compactness' (Cmp) and 'cluster separation' (Sep) [30]. The definitions of these indices are provided below.

The Cmp is based on the generalized definition of the variance of a vector dataset, which can be expressed as:

$$v(X) = \sqrt{\frac{1}{N}\sum_{i=1}^{N} D^2(x_i, \bar{x})} \tag{16}$$

where $\boldsymbol{X} = \{x_1, x_2, \cdots\cdots, x_n\}$ is a dataset, the number of members in $\mathbf{X}$ is N, and $D(x_i, x_j)$ is a distance metric between vector data x_i and x_j, which can be achieved by Eq. (14) in this case. The $\bar{x}$ is the mean of $\mathbf{X}$ as shown in Eq. (17).

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{17}$$

For output clusters c_i ($i \in [1, 2, \cdots, C]$) generated by a system, the Cmp is defined as:

$$Cmp = \frac{1}{C}\sum_{i=1}^{C}\left[\frac{v(c_i)}{v(X)}\right] \tag{18}$$

where C is the cluster number, and $v(c_i)$ is the variance of cluster c_i. The members within each cluster should be as close as possible, so the smaller the Cmp, the more compact the clustering results. The Cmp reaches a minimal value of 0 when every unique input data is encoded into one separate cluster.

The Sep of a clustering system's output is described by:

$$Sep = \frac{1}{C(C-1)}\sum_{i=1}^{C}\sum_{j=1, j\neq i}^{C} \exp\left(-\frac{D^2(o_i, o_j)}{2\sigma_s^2}\right) \tag{19}$$

where o_i is the exemplar of cluster c_i, and $D(o_i, o_j)$ is a distance metric between exemplars o_i and o_j, which can be determined by Eq. (14). The σ_s is the Gaussian constant, and can be taken as $2\sigma_s^2 = 1$ to simplify the calculation. A smaller Sep value indicates a larger dissimilarity among each cluster. If a whole data set is output into one cluster, the Sep has a minimal value of 0.

To comprehensively evaluate the performance of the clustering process and overcome each deficiency, the Cmp and Sep are combined into one indices called the overall cluster quality (Ocq), which is determined through Eq. (20):

$$Ocq = \xi \times Cmp + (1 - \xi) \times Sep \tag{20}$$

where $\xi \in [0, 1]$ is the weight to balance the Cmp and Sep. The smaller value of Ocq, the better performance of the cluster system output.

4.3. Optimized AP clustering algorithm

Based on the description in the first two subsections, the paper optimizes the similarity matrix and parameter selection process in AP clustering, and finally obtains the optimized AP clustering algorithm. The algorithm can output the optimal clustering result and exemplars of the CC features sample set, providing a reliable CB diagnosis basis. The algorithm procedure is outlined as follows:

(1) The sample set $\mathbf{X}$ and parameter space are input.
(2) The similarity matrix $\mathbf{S}_{newi}$ is constructed according to Eq. (15) based on $\mathbf{X}$, p_i and σ_i, which are the value of a step in the scanning parameter space.
(3) AP clustering is performed using $\mathbf{S}_{newi}$ as the input, and the indices Ocq_i are calculated according to Eqs. (18)–(20).
(4) Step 2 and 3 are repeated until all p_i and σ_i in the preset parameter space have been scanned and the cluster number space formed by all possible clustering results is output.
(5) The number of clusters C corresponding to the minimum Ocq in the cluster number space is selected as the optimal result, and exemplars o_i ($i \in [1, 2, \cdots, C]$), optimal p_c and σ_c. are output.

The effectiveness and performance of the proposed optimized AP clustering algorithm is then verified before it is used as the basis for a fault diagnosis method.

4.4. Optimized algorithm verification

Based on external F-M indices introduced in Section 3, this section verifies the validity of the cluster result of the optimized algorithm proposed above for the CC features sample set. As known, the majority of historical CB data samples are those of normal state, the amount of fault samples is relatively small. Moreover, CB cannot repeatedly extract the test samples of various fault types. As a result, the number of available training samples is usually small. Therefore it is also necessary to verify the clustering performance of the proposed algorithm for a small sample set.

Table 3
Parameters and parameter space used for simulations.

Parameters	Value
Feature data dimension m	8
Number of samples N	30
Damping factor η	0.5
AP cluster maximum iterations T	100
Gaussian constant σ_s	0.71
Weight of evaluation indices ζ	0.5
Scanning space of p	$p \in [-5, -0.5]$
Scan step size of p p_{step}	0.05
Scanning space of σ	$\sigma \in [-0.01, 100]$
Scan step size of σ σ_{step}	0.01 ($\sigma < 1$), 1 ($\sigma > 1$)

4.4.1. Select sample set and parameters

This paper take close CC signal as the data source for the clustering algorithm, obtaining a sample set containing 30 samples of CC features, extracted from a high-voltage vacuum CB produced by Sichuan Electric Co., Ltd. Each sample is comprised of all eight feature quantities with a corresponding external fault type label, so that F-M indices calculation can be performed. However, these labels are not included in the cluster input. The CC features data set can reflect normal state and five fault types including decreasing voltage supply, stiff latch, operating mechanism jamming, abnormal coil excitation and auxiliary contact malfunction.

The value of parameters and parameter space used for simulations and verifications are listed respectively in Table 3.

4.4.2. Effectiveness verification of the proposed algorithm

To verify the effectiveness of the algorithm proposed above, all 30 samples are clustered. The parameter adaptive optimization output is: optimal preference p_c of -2, optimal width coefficient σ_c of 0.71, and the cluster results and optimal exemplars are shown respectively in Tables 4 and 5.

In the calculation process, by scanning the given parameter space, the cluster number space can be obtained as $C \in (4, 5, 6, 7, 8)$. The Ocq indices and F-M indices corresponding to each cluster are provided in Fig. 2, and show that when C is 6, the Ocq indices reach a minimum of 3.02, while the F-M indices reach the maximum of 1. This outcome means the clustering result output by the proposed algorithm is completely consistent with the external classification information. Therefore, Fig. 2 demonstrates that the result of the adaptive optimization process is also the highest validity cluster result, verifying the effectiveness of the proposed algorithm in clustering the CC feature samples.

4.4.3. Performance examination of the optimized algorithm

In order to examine the cluster performance of the proposed optimized algorithm in dealing with small sample cases, and highlight the optimization effect compared with the original AP algorithm, comparison simulation is carried out. In the test, the number of training samples is gradually reduced from 30 to 10 while ensuring that 6 operating states are included. Fig. 3 shows the clustering result by the optimization algorithm and the original AP.

It can be seen from Fig. 3 that as the number of samples decrease, the F-M indices of the result output by the original algorithm continues to reduce, illustrating that the clustering validity continues to decline. Meanwhile, the quality of the cluster results of the proposed algorithm are maintained at a high level. Before the set size drops to 12, the quantity of F-M indices can be guaranteed to be 1, demonstrating that the cluster result can remain completely consistent with the external classification label.

In summary, the simulation results adequately verify the effectiveness and performance of the proposed algorithm in dealing with the clustering operation of the CC features even in small sample cases. The results and exemplar output by the proposed algorithm demonstrate a high level of validity.

Table 4
Cluster result using proposed algorithm.

Clusters c_i	Exemplar of c_i o_i	Samples in c_i x_i
1	1	1,2,3,4,5
2	8	6,7,8,9,10
3	14	11,12,13,14
4	16	15,16,17,18,19,20
5	25	21,22,23,24,25
6	29	26,27,28,29,30

Table 5
CC feature quantities of the optimal exemplars.

o_i	I_1(A)	I_2(A)	I_3(A)	t_1(ms)	t_2(ms)	t_3(ms)	t_4(ms)	t_5(ms)
1	2.21	1.62	1.13	24.5	37.8	43.5	46.7	50.3
8	1.81	1.22	0.92	23.76	36.71	42.27	45.77	50.1
14	2.25	1.56	1.16	30.15	43.46	49.13	52.49	56.02
16	2.2	1.61	1.14	24.2	37.5	43.2	49.9	54.1
25	2.19	1.6	1.12	24.1	39.7	42.8	45.9	49.8
29	2.23	1.62	1.14	23.92	37.52	43.32	48.13	52.13

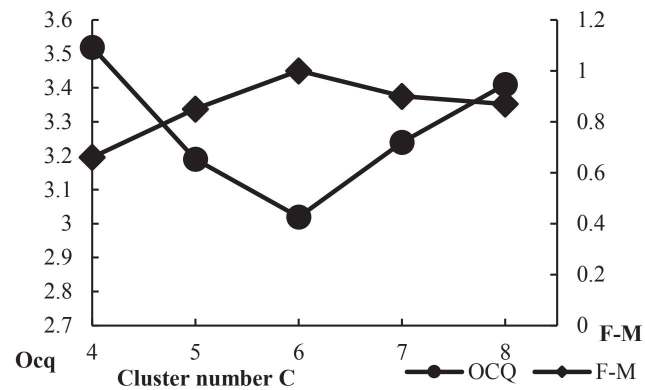

Fig. 2. Ocq and F-M indices quantities of various cluster numbers.

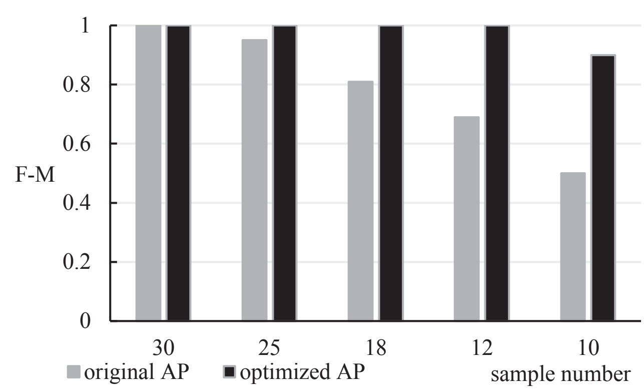

Fig. 3. Comparison results of F-M indices between two algorithms.

5. Fault diagnosis method

Based on the previous description, the optimized AP clustering algorithm can effectively cluster the CC feature sample set consisting of historical or experimental data, and output the optimal cluster result and exemplars. Referring to Table 2, and taking the value of the exemplar features, the corresponding operating state (normal or certain fault type) of each exemplar can be analyzed. In the online diagnosis process, the fault type of the CB can be judged by comparing the quantized similarity between the new data of CC features and the exemplar [25]. However, it is extremely difficult to collect all possible fault type samples through historical data or experimental data. In order to effectively distinguish whether a new sample belongs to a known operating state or a category of unknown fault, an additional threshold is required. This section first defines the similarity coefficient and threshold, then presents a completed fault diagnosis method.

5.1. Similarity coefficient and threshold

To quantify the similarity between the new sample x_{new} and each exemplar o_i, the similarity coefficient ρ_i is defined. This can be expressed as Eq. (21), according to the similarity matrix in the optimized AP cluster and the Eqs. (14) and (15):

$$\rho_i = \exp\left(-\frac{\sqrt{\sum_{l=1}^{m} |x_{newl} - o_{il}|^2}}{2\sigma_c^2}\right) - 1 \quad i \in [1, 2, \cdots, C] \tag{21}$$

where σ_c is the optimal width coefficient output by the proposed clustering algorithm. It can be derived from Eq. (21) that the value of ρ_i range from -1 to 0.

To intuitively identify whether the new sample is a type of unknown fault, a threshold constant λ is introduced based on the similarity coefficient ρ_i. To identify the unknown fault effectively, the value of the threshold needs to be determined with reference to the actual sample data. It must be smaller than the minimum value of the similarity coefficient between training samples in the same fault type and bigger than the maximum value of the similarity coefficient between training samples in different fault types. For a given new sample x_{new}, and the exemplars o_i of the known states, whether x_{new} is an unknown fault can be judged by the following:

$$\begin{cases} \max\limits_{i=1,2,\cdots,C} \{\rho_i\} > \lambda & x_{new} \in normalorknownfault \\ \max\limits_{i=1,2,\cdots,C} \{\rho_i\} \leqslant \lambda & x_{new} \in unknownfault \end{cases} \tag{22}$$

5.2. Procedure of the proposed fault diagnose method

This section describes the procedure of the optimized AP clustering algorithm for CB fault diagnosis, which is illustrated in Fig. 4.

Specific steps are as follows:

(1) Historical data with various fault types is collected, CC features are

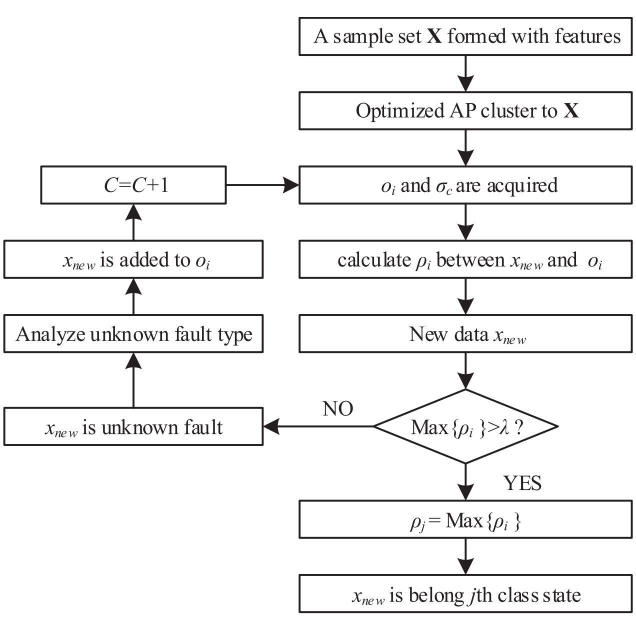

Fig. 4. Procedure of proposed diagnosis method.

extracted, and a sample set **X** is formed.

(2) The optimized AP cluster is applied to **X** and the optimal exemplars o_i and σ_c are acquired. Each exemplar corresponds to an operating state (normal or known fault type).

(3) For the new sample x_{new} to be diagnosed, the similarity ρ_i between x_{new} and each exemplars o_i is calculated using Eq. (21).

(4) Eq. (22) is used to determine whether x_{new} is an unknown fault type.

(5) If x_{new} is not an unknown fault type, it can be judged according to Eq. (23) that x_{new} belongs to the jth class operating state.

$$\rho_j = \max_{i=1,2,\cdots,c} \{\rho_i\} \tag{23}$$

(6) If x_{new} is an unknown fault type, it is defined as a $C + 1$ fault type. After subsequent analysis by the maintenance staff, the cause of the failure is determined. This information is then added to the exemplar set o_{i+1} as the known fault exemplar to achieve the effect of identifying unknown fault type samples.

6. Simulation results

The main aim of this section is to verify the validity and performance of the proposed fault diagnosis method. To prove the outstanding advantages of the proposed method, abundant and essential simulations are completed, including verification of accurate identification of fault types of CC feature sample, verification of the ability to identify unknown fault samples and performance examination in small sample cases.

6.1. Select sample set and parameters

The source of the CC feature data is the same as the related description in Section 4, and the parameters of the proposed method are shown individually in Table 3, where the threshold constant λ is -0.5.

To verify the validity of the proposed method and its ability to identify unknown faults, 11 samples containing four operating states are extracted from the entire set of 30 samples for training. At the same time, another 10 samples are taken, including five operating states, as a set of test samples. The training sample set and test sample set are listed in the Appendix A. The label of categories in the table are used as external classification information to evaluate the accuracy of the method, and are not used as a proposed method input in the training process of unsupervised learning. As shown in Appendix A, Type N represents normal, and F1, F2, and F3 in the training set are obviously the three known fault types of voltage supply decreasing, stiff latch and operating mechanism jamming, respectively. The F4 is abnormal coil excitation, which is not contained in the training samples. Therefore the samples of F4 can be treated as unknown faults for the training stage.

6.2. Validity verification of the proposed diagnosis method

To verify the effectiveness of the proposed method and highlight diagnostic accuracy, the test samples are diagnosed online using the proposed method and the original AP clustering method with Euclidean distance as the similarity coefficient, respectively. The results of the contrasting simulation are detailed as follows:

Clustering process: the optimized AP accurately divides the training samples into four clusters. The output optimal parameters are p_c at -1.1, σ_c at 0.71, and the exemplars o_i are samples numbered with 1, 4, 7 and 10, whose features are listed in Table 6. In comparison, the original AP comparatively output five clusters and exemplars o_i including 1, 4, 7, 8 and 10. The calculation result by the external classification label shows that the F-M indices of the cluster result output by the proposed AP algorithm is 1, which ensures training samples are completely and precisely classified, while the original AP result F-M indices is only 0.91.

Table 6
Exemplars and their features quantities output by optimized AP algorithm.

	I_1(A)	I_2(A)	I_3(A)	t_1(ms)	t_2(ms)	t_3(ms)	t_4(ms)	t_5(ms)	Types
1	2.21	1.62	1.13	24.5	37.8	43.5	46.7	50.3	N
4	1.81	1.23	0.91	23.8	36.7	42.3	45.8	50.1	F1
7	2.23	1.61	1.11	30.1	43.5	49.1	52.4	56.1	F2
10	2.18	1.6	1.13	24.19	37.54	43.26	49.87	54.07	F3

Diagnosis process: The similarity coefficient results and diagnostic results of the test samples are calculated using Eqs. (21)–(23) are presented intuitively in Table 7.

As illustrated in Table 7, the maximum similarity coefficient ρ_i can effectively reflect the operating state of test samples in the process of executing the proposed method. Furthermore, when the similarity coefficients are all smaller than λ, the unknown fault category can be precisely identified. However, the original AP clustering results are not accurate enough, thus the diagnosis results of the test samples are not completely correct.

Table 8 details the training time and online diagnosis time of the two methods. Compared with the original method, in order to guarantee a higher diagnostic accuracy, the proposed method introduces a parameter optimization link, which prolongs training time. However, the training process is offline and does not affect the operating time of online diagnosis. In addition, the diagnostic process calculation time of the proposed method is only around 50 ms, which is short enough to satisfy the practical application requirements of online diagnosis.

6.3. Verification of the ability to identify unknown fault samples of the proposed diagnosis method

To further highlight the ability of the proposed method to identify unknown type faults, BPNN algorithm as a supervised learning algorithm is used to train the same training sample set while additionally attaching classification labels for fault diagnosis of the test samples. The comparison diagnosis results between the proposed method and BP neural network are shown in Table 9.

As detailed in Table 9, the BP neural network algorithm can accurately recognize the test samples of fault categories contained in the training set with numbers from 1 to 8. However, BP classifies test samples of unknown fault type with numbers 9 and 10 into normal categories, which is inconsistent with the actual results. This occurs because the supervised learning algorithm only memorizes the types in the training samples and gives a false diagnosis to unknown fault samples based on this account. Comparatively, as for the proposed method, not only known fault samples from N, F1, F2 and F3 are accurately classified to each type, but samples 9 and 10 are also explicitly identified as not belonging to any types of the training samples and recognized as a novel fault type F4. The results indicate that the proposed method can diagnoses samples of both known and unknown fault types accurately.

Table 8
Training time and online diagnosis time of two methods.

Method	The proposed method			Method based on original AP		
Time	Longest	Shortest	Average	Longest	Shortest	Average
Single cluster (ms)	124.8	78.0	104.5	124.8	78.0	104.5
Training(s)	–	–	153.9	–	–	0.1
Diagnosis(ms)	78.0	15.6	52.4	75.3	18.3	51.8

Table 9
Comparison diagnosis results between the proposed method and BP neural network.

Test samples	Types	Proposed method	BP neural network
1	N	N	N
2	N	N	N
3	F1	F1	F1
4	F1	F1	F1
5	F2	F2	F2
6	F2	F2	F2
7	F3	F3	F3
8	F3	F3	F3
9	F4	F4	N
10	F4	F4	N

6.4. Diagnostic performance examination of the proposed method with a small set of samples

As described in Section 1, in practical applications, the number of available training samples for fault diagnosis methods is usually small. To test the performance of the proposed method in the case of a small set of samples, first, the number of training samples is reduced, but all six types are included, which means that the test samples do not contain unknown faults. The remaining samples in the set of 30 samples are

Table 7
Diagnostic results and similarity coefficient results of the test samples using two methods.

Samples	Types	Proposed method				Results	Original method					Results
		1	4	7	10		1	4	7	8	10	
1	N	**−0.1571**	−0.8804	−1.0000	−0.9937	N	**−0.1709**	−2.1237	−12.6676	−12.6766	−5.0664	N
2	N	**−0.2757**	−0.8839	−1.0000	−0.9938	N	**−0.3225**	−2.1531	−12.7782	−12.7849	−5.0776	N
3	F1	−0.8987	**−0.2096**	−1.0000	−0.9973	F1	−2.2896	**−0.2352**	−14.7124	−14.7271	−5.9234	F1
4	F1	−0.8798	**−0.1174**	−1.0000	−0.9974	F1	−2.1182	**−0.1249**	−14.6189	−14.6327	−5.9685	F1
5	F2	−1.0000	−1.0000	**−0.1617**	−1.0000	F2	−12.6489	−14.5078	−0.1764	**−0.1741**	−10.6994	F3
6	F2	−1.0000	−1.0000	**−0.1460**	−1.0000	F2	−12.7241	−14.5889	−0.1578	**−0.1025**	−10.7930	F3
7	F3	−0.9935	−0.9974	−1.0000	**−0.0840**	F3	−5.0305	−5.9519	−10.7073	−10.7335	**−0.0877**	F4
8	F3	−0.9931	−0.9972	−1.0000	**−0.0478**	F3	−4.9779	−5.8758	−10.7774	−10.8044	**−0.0490**	F4
9	F4	−0.8959	−0.9578	−1.0000	−0.9981	F4	−2.3014	−3.1858	−13.0888	−13.1050	−6.3124	F5
10	F4	−0.8944	−0.9556	−1.0000	−0.9982	F4	−2.2623	−3.1662	−13.0470	−13.0628	−6.2546	F5

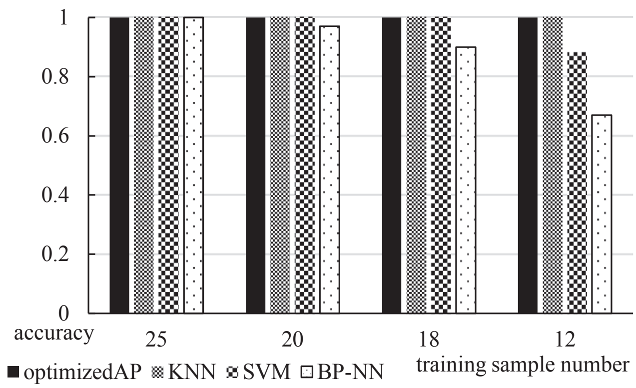

Fig. 5. Comparison diagnosis results of the proposed method and other supervised learning methods in no unknown fault case.

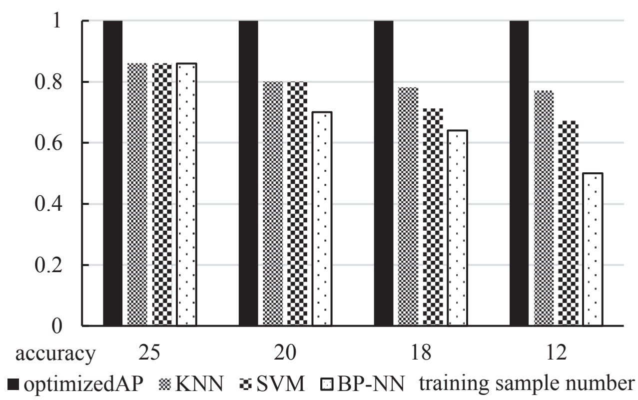

Fig. 6. Comparison diagnosis results of the proposed method and other supervised learning methods with unknown fault samples.

used as test samples. The proposed method and several common supervised learning methods are used to diagnose the same test samples, and comparison results are depicted in Fig. 5.

Several common supervised learning algorithms are selected including: k-nearest neighbor (KNN), support vector machine (SVM), and BPNN. The distance function utilized by KNN is Euclidean distance. Kernel function in SVM is chosen as a Gaussian kernel function with the Gaussian constant of 1.4. The number of neurons in the input layer, hidden layer and output layer of the BP neural network are 8, 15, and 6, respectively, and the transfer functions are tan-sigmoid and linear function. It can be seen from Fig. 5 that compared with the supervised learning algorithm, the proposed method can maintain high diagnostic accuracy in small sample cases. When the number of training sample is only 12, the diagnostic accuracy is still 100%. In the same circumstance, the accuracy of the SVM and the BPNN begin to decline.

Moreover, to test the ability of the proposed method to identify unknown faults with a small sample set, another simulation has been made. In that case, samples of the F2 fault type are not included in the training samples and are considered as a type of unknown fault only included in the test samples. And like the former simulation, the number of training samples is reduced, but all five types are included except F2 type. The remaining samples of unknown types in the set of 30 samples and F2 type samples are used as test samples. The proposed method and above-mentioned supervised learning methods are used to diagnose the same test samples, and comparison results are depicted in Fig. 6.

It can be seen from Fig. 6 that compared to the results in Fig. 5, the diagnostic accuracy of supervised learning methods are greatly reduced under different training sample numbers. That is because supervised learning methods cannot accurately identify unknown faults (F2), but classifies unknown fault samples into a type of faults included in training samples. However, when novel fault data not belonging in the existing training samples arrive, the proposed method based on unsupervised learning algorithmic will find these hidden patterns and update for the new condition. Therefore, as shown in Fig. 6, the proposed method can effectively identify unknown faults (F2), and maintain 100% diagnostic accuracy until the sample number is 12.

In general, the simulation results adequately verity that the proposed method carry out the diagnosis of various CBs faults accurately, and has the ability to identify unknown fault type. In addition, dealing with small sample set cases, the proposed method guarantees a relatively high diagnostic accuracy rate compared with the supervised learning method.

7. Conclusion

This paper proposed a novel CB fault diagnosis method based on optimized AP clustering with CC feature to quickly and accurately identify the operating state of CBs online, as well as effectively distinguish fault types. Through theoretical analysis and simulation results, the following conclusions can be drawn:

(1) The proposed improved for the original AP clustering algorithm, including the construction of the optimal similarity matrix and adaptive optimization of clustering parameters, effectively increases the clustering accuracy and helps to find the optimal exemplars.
(2) The CB diagnosis procedure designed in the paper based on optimal exemplars and the similarity coefficient can effectively and efficiently identify the CC samples of both known and unknown fault types online, which is difficult to achieve by other conventional data-driven methods.
(3) The simulation verification proves that the proposed method can identify various CB fault types more accurately compared with the original AP clustering. In addition, the proposed method has a better performance for CB fault diagnosis in small sample cases.

Therefore the proposed method is more suitable for fault diagnosis of high voltage CBs and has greater application prospects.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported in part by National Key Research and Development Program of China (2016YFB0900603) and Technology Projects of State Grid Corporation of China (52094017000W).

Appendix A

See Tables 10 and 11.

Table 10
Training sample set used for simulations.

	I_1(A)	I_2(A)	I_3(A)	t_1(ms)	t_2(ms)	t_3(ms)	t_4(ms)	t_5(ms)	*Types*
1	2.21	1.62	1.13	24.5	37.8	43.5	46.7	50.3	N
2	2.2	1.63	1.15	24.59	37.76	43.57	46.54	50.35	N
3	2.18	1.62	1.17	24.34	37.88	43.38	46.59	50.21	N
4	1.81	1.23	0.91	23.8	36.7	42.3	45.8	50.1	F1
5	1.82	1.25	0.96	23.82	36.84	42.29	45.81	49.97	F1
6	1.81	1.22	0.92	23.76	36.71	42.27	45.77	50.1	F1
7	2.23	1.61	1.11	30.1	43.5	49.1	52.4	56.1	F2
8	2.26	1.62	1.09	30.17	43.47	49.11	52.48	55.99	F2
9	2.21	1.62	1.17	24.21	37.57	43.25	50.02	54.09	F3
10	2.18	1.6	1.13	24.19	37.54	43.26	49.87	54.07	F3
11	2.2	1.61	1.14	24.2	37.5	43.2	49.9	54.1	F3

Table 11
Test sample set used for simulations.

	I_1(A)	I_2(A)	I_3(A)	t_1(ms)	t_2(ms)	t_3(ms)	t_4(ms)	t_5(ms)	*Types*
1	2.2	1.58	1.12	24.56	37.93	43.48	46.62	50.29	N
2	2.28	1.61	1.21	24.36	37.87	43.49	46.84	50.08	N
3	1.78	1.19	0.88	23.7	36.57	42.18	45.75	50.19	F1
4	1.85	1.3	0.93	23.8	36.69	42.29	45.71	50.08	F1
5	2.26	1.61	1.13	30.17	43.35	49.06	52.42	56.08	F2
6	2.25	1.56	1.16	30.15	43.46	49.13	52.49	56.02	F2
7	2.27	1.63	1.17	24.23	37.5	43.25	49.95	54.11	F3
8	2.21	1.6	1.16	24.19	37.5	43.2	49.86	54.11	F3
9	2.2	1.61	1.13	24.09	39.75	42.86	45.89	49.79	F4
10	2.22	1.64	1.13	24.15	39.73	42.82	45.94	49.82	F4

References

[1] Vakilian, et al. Priority assessment of online monitoring investment for power system; circuit breakers-Part I: qualitative-quantitative approach. IEEE Trans Power Deliv 2013;28(2):928–38.
[2] Peilei R, Jian H, Xiaoguang H, Jin X. Testing of circuit breakers using coil current characteristics analysis. In: Proc. IEEE Control Autom. Conf.; 2009-Dec. p. 185–9.
[3] Johal H, Mousavi MJ. Coil current analysis method for predictive maintenance of circuit breakers. In: Proc. Transm. Distrib. Conf. Expo. Conf.; 2008-Apr. p. 1–7.
[4] Xiaoguang H, Chao L. Research on the condition parameter tester of high voltage circuit breakers. In: Proc. IEEE Ind. Electron. Appl. Conf.; 2008-Jun. p. 2389–93.
[5] Hinow Martin, Mevissen M. Substation maintenance strategy adaptation for life-cycle cost reduction using genetic algorithm. IEEE Trans Power Deliv 2010;26(1):197–204.
[6] Su Zuqiang, Tang Baoping, Lei Deng, et al. Fault diagnosis method using supervised extended local tangent space alignment for dimension reduction. Measurement 2015;62:1–14.
[7] Natti S, Kezunovic M. Assessing circuit breaker performance using condition-based data and Bayesian approach. Electr Power Syst Res 2011;81(9):1796–804.
[8] Kezunovic M, Natti S. Risk-based maintenance approach: a case of circuit breaker condition based monitoring. In: 3rd international CIGRE workshop on liberalization and modernization of power systems, Irkutsk, Russia, August; 2006.
[9] Zhong Jun, et al. A RankBoost based data-driven method to determine maintenance priority of circuit breakers. IEEE Trans Power Deliv 2017;99. 1–1.
[10] Zhong Jun, et al. Incorporating a condition monitoring based aging failure model of a circuit breaker in substation reliability assessment. IEEE Trans Power Syst 2015;30(6):3407–15.
[11] Shi Liu, et al. Study of PSO-BP neural networks application in high-voltage circuit breakers mechanical fault diagnosis. In: International conference on electricity distribution (CICED), 2016 China. IEEE; 2016.
[12] Geng Pulong, et al. Fault pattern recognition method for the high voltage circuit breaker based on the incremental learning algorithms for SVM. In: International conference on condition monitoring and diagnosis (CMD), 2016. IEEE; 2016.
[13] Object detection in optical remote sensing images based on weakly supervised learning and high-level feature learning. IEEE Trans Geosci Remote Sensing 2015;53(6):3325–37.
[14] Tu Zhuowen, et al. Unsupervised object class discovery via saliency-guided multiple class learning. IEEE Trans Pattern Anal Mach Intell 2015;37(4):862.
[15] Zhang Zhiqiang, Zhang A. A novel strategy for fault diagnosis of analog circuit online based modified kernel fuzzy C-means. In: IEEE international conference on industrial technology. IEEE; 2016.
[16] Zhuang Li, et al. Fault diagnosis of wind turbine gearbox based on kernel fuzzy c-means clustering. IET Conference Proceedings. The Institution of Engineering & Technology; 2015.
[17] Razi-Kazemi Ali Asghar, et al. Data mining of online diagnosed waveforms for probabilistic condition assessment of sf $ _ 6 $ circuit breakers. IEEE Trans Power Deliv 2015;30(3):1354–62.
[18] Frey J Brendan, Delbert Dueck. Clustering by passing messages between data points. Science 2007;315(5814):972–6.
[19] Razi-Kazemi Ali Asghar. Circuit breaker condition assessment through a fuzzy-probabilistic analysis of actuating coil's current. IET Gener Transm Distrib 2016;10(1):48–56.
[20] Razi-Kazemi Ali Asghar, et al. Circuit-breaker automated failure tracking based on coil current signature. IEEE Trans Power Deliv 2014;29(1):283–90.
[21] Han Jiawei, Pei Jian, Kamber Micheline. Data mining: concepts and techniques. Elsevier; 2011.
[22] Ferrari Daniel Gomes, Castro Leandro Nunes De. Clustering algorithm selection by meta-learning systems: a new distance-based problem characterization and ranking combination methods. Inf Sci 2015;301:181–94.
[23] Han Junwei, et al. Background prior-based salient object detection via deep re-construction residual. IEEE Trans Circuits Syst Video Technol 2015;25(8):1309–21.
[24] Wei Zexian, et al. A novel intelligent method for bearing fault diagnosis based on affinity propagation clustering and adaptive feature selection. Knowl-Based Syst 2017;116:1–12.
[25] Hu Di, Sarosh Ali, Dong Yun-Feng. A novel KFCM based fault diagnosis method for unknown faults in satellite reaction wheels. ISA Trans 2012;51(2):309–16.
[26] Wu Chih-Hung, et al. A new fuzzy clustering validity index with a median factor for centroid-based clustering. IEEE Trans Fuzzy Syst 2015;23(3):701–18.
[27] Halkidi Maria, Batistakis Yannis, Vazirgiannis Michalis. On clustering validation techniques. J Intelligent Inform Syst 2001;17(2-3):107–45.
[28] Ding Meng, Fan Guoliang. Articulated and generalized gaussian kernel correlation for human pose estimation. IEEE Trans Image Process 2016;25(2):776–89.
[29] Jia Hongjie, et al. A density-adaptive affinity propagation clustering algorithm based on spectral dimension reduction. Neural Comput Appl 2014;25(7-8):1557–67.
[30] He Ji, Tan Ah-Hwee, Tan Chew-Lim. Modified ART 2A growing network capable of generating a fixed number of nodes. IEEE Trans Neural Networks 2004;15(3):728–37.