

Received April 27, 2022, accepted May 10, 2022, date of publication May 23, 2022, date of current version June 7, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3175541

A Weighted Classification Method Based on Adaptive Feature Selection

RUIZHENG NI¹, (Student Member, IEEE), RUICHANG QIU¹, ZHIWEI LUO²,
JIE CHEN¹, (Member, IEEE), ZHEMING JIN¹, (Member, IEEE),
AND ZHIGANG LIU¹

¹School of Electrical Engineering, Beijing Jiaotong University, Beijing 100044, China

²Beijing Institute of Space Launch Technology, Beijing 100076, China

Corresponding author: Ruichang Qiu (rchqiu@bjtu.edu.cn)

This work was supported by the Fundamental Research Funds for the Central Universities under Grant 2018JBZ004.

ABSTRACT This paper proposes a new classification method, which can adaptively select effective features and give different classification weights according to the classification ability of features, this method effectively eliminates the influence of invalid features in classification process. For the classification problem of fault diagnosis, this method judges the fault type from the azimuth relative to the normal state in multi-dimensional space, and judges the fault degree from the ratio of distance, the method can not only give the fault category, but also give the fault probability, fault degree and whether there is a new fault type. In the process of updating the classification center, the classification method proposed in this paper solves the problem that the location of the classification center changes with the sample processing order. The effectiveness of this method is tested by the transformer vibration data (TVIB) and five data sets published in MATLAB and UCI, and compared with K-Mean, self-organizing map (SOM) methods, it is proved that considering the classification weight can effectively improve the classification accuracy, and the proposed method has strong applicability for different classification problems and good processing ability for large capacity and multi feature data.

INDEX TERMS Classification method, classification features, fault diagnosis, transformer vibration.

I. INTRODUCTION

According to whether there is a normal state, the classification problems can be divided into two kinds. The first kind is the classification without a normal state, such as the determination of animal and plant species, the data sets ‘Fisheriris’ (FSR), Wine, Soybean and ‘Human activity’ (HMT) used in this paper belong to the first classification problem, this kind of classification problem does not have a central category that can connect all categories, a data sample can only belong to one category. The second kind refers to the classification problem with normal state or called ideal center point, such as general fault diagnosis problem. In addition to distinguishing categories, this kind of classification problem usually needs to judge the severity of faults and the possibility of other types of faults, because there may be two or more

The associate editor coordinating the review of this manuscript and approving it for publication was Hao Luo¹.

faults at the same time in the actual process, this will greatly increase the difficulty of classification. The fault diagnosis problem based on transformer vibration belongs to the second kind of classification problem, it takes the vibration of the transformer in the normal state as the classification center, the determination of other vibration states is compared with the normal state, and the actual fault diagnosis process is also relative to the normal state.

There are many classification methods, such as KNN [1] (k-nearest neighbor), K-Mean [2]–[5], SVM [6] (support vector machine), DT [7], [8] (decision tree), NB (naive Bayes), etc. KNN is a classification algorithm based on analogy learning, which is simple and has strong anti noise ability. In paper [1], an AdaBoost KNN algorithm based on adaptive feature selection is proposed, which has achieved good results in facial expression recognition. K-Means algorithm is a classical distance based clustering algorithm, its advantage is easy to understand and implement, and

has good clustering effect, the disadvantages are the lack of theoretical basis for the selection of the initial cluster number, and the selection of initial cluster center has a great impact on the clustering results and is sensitive to noise. Many people have made improvements on the k-means algorithm. The hybrid leapfrog algorithm is combined in the K-Means method in paper [2], and the effectiveness of the new method is verified by multiple data sets such as FSR, WINE and GLASS. In paper [3], a pairwise and size constrained K-Means method is proposed to avoid the phenomenon of clustering with few samples. SVM is a data classification method based on statistical theory, which uses hyperplane to separate different types of samples, this method has many applications in transformer fault diagnosis. In [6], SVM is used to classify the frequency response analysis data of several deformation faults of transformer internal winding. In [9], SVM is used to analyze a large number of DGA data of oil immersed transformer. There is also the combination of SVM and other methods, SVM and particle swarm optimization algorithm are combined to improve the accuracy of oil immersed transformer fault classification in [10]. In [11], the SVM method is processed by power average, and the processed algorithm has good classification ability for large capacity and multi feature data. DT is a classification algorithm based on inductive thought, it is expressed as a tree structure, in which each branch structure represents a test output, and each leaf node represents a category label. NB is a learning algorithm based on classical mathematics. It combines priori probability and posteriori probability to improve the accuracy. However, this method requires the independence of each feature, which is difficult to guarantee in the actual process.

The method proposed in this paper is similar to K-Mean method in the first classification problem, both methods divide categories based on distance. The difference is that the method proposed in this paper will make further selection of feature data and allocate different classification weights. The update process of category center is similar to SOM method, it was originally proposed by Kohonen [12], [13]. However, SOM classification method following the rule of winner takes all, and the classification method in this paper is that all categories are possible, it may belong to many categories at the same time for the second classification problem. SOM classification method has the problem that the change of sample order will affect the location of winning neurons [14]–[18], but this problem will be weakened due to the multiple iterations of samples and the introduction of time-dependent variables, the category center updating method (winning neuron in SOM) proposed in this paper overcomes this problem. The advantages of the method proposed in this paper are:

- 1) Adaptive feature selection can eliminate the influence of inferior feature data on classification results, and the role of high-quality feature data can be further increased through the distribution of classification weight of features.

- 2) It is more reasonable to interpret the classification results from the perspective of probability, especially for the data samples in the junction areas of different categories.
- 3) The problem that the final category center changes with the change of sample processing order is solved.
- 4) For the second kind of classification problems such as fault classification, it can judge whether there is outlier data and whether there is a new fault, which has strong practicability in fault classification.

II. PROPOSED CLASSIFICATION METHOD

For data set X , it contains m categories, and the center of each category is $\omega_1, \omega_2, \dots, \omega_m$. Each category contains n_1, n_2, \dots, n_m samples, $n_1 + n_2 + \dots + n_m = N$, and each sample contains p features.

For the first kind of classification problem, Euclidean distance Dt and relative distance Dt_R is used as the classification criteria, as shown in equation (3), the relative distance is calculated by dividing the absolute distance by the distance between the 20% quantile and the 80% quantile of such feature. For the second kind of classification problem, due to the existence of normal center, the second kind of data are processed centrally, that is, all sample data are the result of subtracting the central point data, and then take the normal state as the central point, judge the fault type through the orientation of the sample point relative to the normal state in multi-dimensional space, as shown in equation (2), judge the severity of fault by the distance between the fault point and the normal state, as shown in equation (3).

$$X = \begin{bmatrix} x_1 & x_2 & \dots & x_m \\ x_{11} & x_{12} & \dots & x_{1,n_i} \\ x_{21} & x_{22} & \dots & x_{2,n_i} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \dots & x_{p,n_i} \end{bmatrix} \quad (1)$$

$$\theta = a \cos \left(\frac{x_{11}^T * \omega_1}{\sqrt{x_{11}^T * x_{11}} * \sqrt{\omega_1^T * \omega_1}} \right) \quad (2)$$

$$Dt = (x_{11} - \omega_1)^T (x_{11} - \omega_1) \quad (3)$$

$$Dt_R = \frac{(x_{11} - \omega_1)^T (x_{11} - \omega_1)}{\|x_p\|_{(0.2 \sim 0.8)}}$$

A. UPDATE OF CATEGORY CENTER

The general rules for updating winning neurons in SOM classification method are as follows:

$$\omega(k + 1) = \omega(k) + \eta(x - \omega(k)) \quad (4)$$

Update learning rate η take a fixed value or take a variable related to time, several typical examples are given in equation (5)

$$\begin{aligned} \eta(t) &= \alpha \\ \eta(t) &= \frac{1}{t} \\ \eta(t) &= \alpha * \exp\left(-\frac{d^2}{2\rho(t)}\right) \end{aligned} \quad (5)$$

There will be a problem that the category center will change due to the change of sample processing order, for ease of understanding, η is taken as 0.5, suppose for samples $x_{11}, x_{12}, \dots, x_{1n}$, ω_1 is the winning category center, x_{11} is the initial category center in the first category, $\omega_{1(1)}$ equals x_{11} , for x_{12}, \dots, x_{1n} , the update order of ω_1 is

$$\omega_{1(2)} = \frac{\omega_{1(2)} + x_{12}}{2} = \frac{x_{11} + x_{12}}{2} \quad (6)$$

Similarly, $\omega_{1(3)}, \omega_{1(4)}$ is

$$\begin{aligned} \omega_{1(3)} &= \frac{\omega_{1(2)} + x_{13}}{2} = \frac{x_{11}}{4} + \frac{x_{12}}{4} + \frac{x_{13}}{2} \\ \omega_{1(4)} &= \frac{\omega_{1(3)} + x_{14}}{2} = \frac{x_{11}}{8} + \frac{x_{12}}{8} + \frac{x_{13}}{4} + \frac{x_{14}}{2} \end{aligned} \quad (7)$$

It can be seen that the influence of the original samples x_{11} and x_{12} is declining step by step, the proportion of new samples will always account for half. This means that the change of sample processing order will change the position of the last category center position. When η taken as a time-related variable, the influence of new samples can be reduced with the increase of time t , and the multiple iterations of samples in the classification process will also weaken the influence of sample order on the position of category center, but this does not guarantee that all sample data will occupy the same proportion when determining the final category center. The correct update method should be that all samples occupy the same proportion, regardless of the order in which they appear. The method proposed in this paper solves this problem, as shown in equation (8).

$$\omega_{m(k+1)} = \frac{x_j + \omega_{m(k)} * k}{k + 1} \quad (8)$$

k is the number of wins of categories center ω_m , this method considers all samples belonging to this category equally, and will not change the last category center position due to the sample processing order.

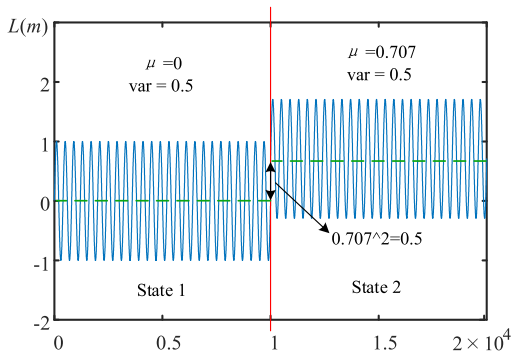


FIGURE 1. Selection criteria of effective classification features.

B. CALCULATION OF FEATURES WEIGHT

The classification ability of different features is different. In particular, there are 60 features in the ‘‘HMT’’ data set, if these features are not selected and all features are given the same classification ability, the classification result will

be very poor. Fig. 1 shows the distribution of a classification feature in two different states. When the sample in the two states overlap more than equation (9), it is considered that this kind of feature does not have effective classification ability in this paper, and the actual sample distribution will be more chaotic than that in the figure, so the degree of overlap will be more serious.

$$\begin{aligned} \text{var}_i &< |\mu_i - \mu_{\text{other}}|^2 \\ wn &= \begin{cases} wn + 1 & \text{if } \text{var}_i < |\mu_i - \mu_{\text{other}}|^2 \\ wn - 1 & \text{if } \text{var}_i \geq |\mu_i - \mu_{\text{other}}|^2 \end{cases} \end{aligned} \quad (9)$$

When the target data set is difficult to classify and the classification ability of the basic feature data is poor, the features can be combined to form new features. For the data set with m features, when the combination of two feature is adopted, $C(m, 2)$ new features can be generated, $C(m, 2) = m*(m-1)/2$, and a total of 2^{m-1} features can be generated. If the features are still insufficient, the feature data can be added in combination with the positive and negative changes of the feature data, but too much feature data will cause computational burden and may also affect the classification results. The method of generating combined features is very helpful for the classification of Pima Indians data set, and it has practical significance, that is, when some features of a patient are abnormal at the same time, it is judged that the patient has been ill. The addition and subtraction combination of features can increase the discrimination between different categories and help to improve the classification effect.

The classification ability of different feature data can be evaluated by equation (9). The best result of each feature data is that it can meet equation (9) for any two categories, then the classification ability of this kind of feature data is assigned combination number $C(m, 2)$, and the worst result is that this kind of data has no classification ability for all categories, then the value is 0. The smaller the data difference within the same category, the greater the difference between different categories, and the stronger the classification ability of this kind of data. Weight allocation is to reasonably allocate the weight according to the classification ability of different features. The weight distribution is calculated by equation (10).

$$\begin{aligned} wt_{m,p} &= \frac{\text{var}(X_p)}{\text{var}(x_{mp})} \\ wt_p &= \frac{\text{var}(X_p)}{\sum_{i=1}^m \text{var}(x_{mp}) / m} \\ WT_m &= \frac{wt_{m,p} * wn}{\sum wt_{m,p} * wn} \\ WT &= \frac{wt_p * wn}{\sum wt_p * wn} \end{aligned} \quad (10)$$

wt_p is the classification ability of each feature data. It is obtained by dividing the overall variance of this

feature data by the mean value of the variance of each category. $w_{t,m,p}$ is the classification ability of each feature data for each category. It is obtained by dividing the overall variance of such feature data by the variance of each category. WT is the classification weight of this feature of data in the classification algorithm. WT_m is the classification weight of this category of data of this feature in the classification algorithm. The weight coefficients of each category WT_m are used to determine the results of each category. When the results determined by the weight coefficients of different categories conflict, the overall weight coefficients WT are used to determine the category of problem data.

C. DETERMINATION OF OUTLIERS AND NEW CATEGORIES

Outlier samples refer to data that are far away from each category center. If they are forced to update the category center, it may lead to the offset of some category centers. Therefore, the category center will not be updated for outlier samples. The outlier is determined by the angle between the sample point and the center of the winning category ($\theta(x, \omega)$), when it is greater than the minimum value of the angle between the centers of two categories, that is, the angle between the new sample point and the nearest category center is greater than the angle between the two category centers, then the sample point is determined as an outlier, as shown in equation (11).

$$\theta(x, \omega) > \min(\theta(\omega_i, \omega_j)) \quad 1 \leq i, j \leq m \quad (11)$$

Whether to establish a new category center depends on the number of outliers and the aggregation degree of outliers. If the average number of samples of each category in the original data is n , when the number of outliers is greater than $n/2$ and the aggregation degree of these outliers is greater than that of each type of samples, this can be compared with the variance of the Euclidean distance between each type of data and its category center.

$$\text{var}(x') \leq \min(\text{var}(x_1), \text{var}(x_2), \dots, \text{var}(x_m)) \quad (12)$$

D. CALCULATION OF FAILURE PROBABILITY AND FAILURE DEGREE

For the first classification problem, the distance between the sample and the category center is used to calculate the probability, as shown in equation (13); for the second classification problem, the angle between the sample and the category center is used to calculate the probability, as shown in equation (14).

$$P = \frac{\ln(Dt / \sum Dt)}{\sum \ln(Dt / \sum Dt)} \quad (13)$$

$$P = \frac{\ln(\theta / 180)}{\sum \ln(\theta / 180)} \quad (14)$$

The fault degree is calculated by dividing the distance between the sample point and normal category center by the

distance between the fault category center and the normal category center, as shown in equation(15), $x^T x$ represents the distance of the sample point from the normal category center. $\omega^T \omega$ represents the distance between the category center of the sample and the normal category center.

$$FR = \frac{x^T x}{\omega^T \omega} \quad (15)$$

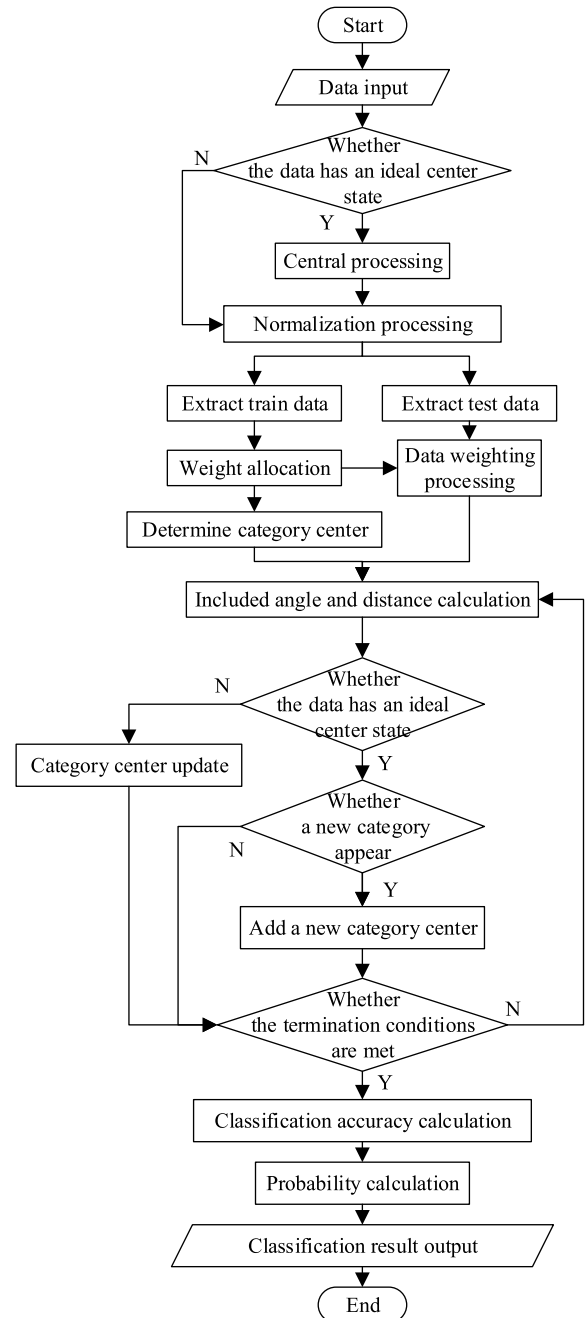


FIGURE 2. Algorithm flow chart.

E. DETERMINATION CALCULATION FLOW OF NEW ALGORITHM

Fig. 2 shows the flow chart of the classification method proposed in this paper. For the first classification problem,

we use algorithm 1, and for the second classification problem, we use algorithm 2. The classification algorithm adopts the method of adaptive feature selection. For different categories, the unique weight calculation of the category is adopted, while for conflicting data points, the overall weight calculation is adopted. For the first classification problem, the category is determined based on distance; for the second classification problem, the category is determined based on angle.

Algorithm 1 Classification Method for the First Type of Data

Input: Sample data and labels
 Output: Sample category and probability
 Initialize: Number of categories m and categories center ω
 n (Maximum number of iterations)
 N (Number of samples)
 k (Number of wins of categories center ω_m)
 Adaptively select classification features and assign classification weight to them
 for $i = 1 \dots n$ do
 for $j = 1 \dots N$ do
 Step1 Calculate Euclidean distance

$$Dt_m = (x_j - \omega_m)^T (x_j - \omega_m)$$

$$Dt_R = \frac{(x_j - \omega_m)^T (x_j - \omega_m)}{\|x_p\|_{(0.2-0.8)}}$$

 Step2 Determine category and update categories center

$$\omega_{m(k+1)} = \frac{x_j + \omega_{m(k)} * k}{k + 1}$$

 end for
 end for
 Step3 Calculate the probability

$$P_m = \frac{\ln(Dt_m / \sum Dt_m)}{\sum \ln(Dt_m / \sum Dt_m)}$$

For the second kind of classification problem, because there is the possibility that one sample belongs to multiple categories at the same time, we should update multiple category centers at the same time based on the probability that the target sample belongs to different categories. For the update of different category centers, corresponding intervention coefficients can be added, such as adjusting the update effect of category centers in the form of logarithm, square or specified probability range, which can make different choices according to different applications. In this paper, the logarithmic form is selected, such as equation (13) and equation (14). The category center update equation in algorithms 1 and 2 can be rewritten as equation (16), for the latter part, multiply the corresponding probability P_m to obtain the update method of each category center, as shown in equation (17). When P_m is 1, only the current winning category center will be updated. When P_m is 0, the current

Algorithm 2 Classification Method for the Second Type of Data

Input: Sample data and labels
 Output: Sample category, fault probability, fault degree and whether to establish a new category
 Initialize: Number of categories m and categories center ω
 n (Maximum number of iterations)
 N (Number of samples)
 k (Number of wins of categories center ω_m)
 Centralized processing of data samples
 Adaptively select classification features and assign classification weight to them
 for $i = 1 \dots n$ do
 for $j = 1 \dots N$ do
 Step1 Calculate Euclidean distance and included angle

$$\theta_j = a \cos \left(\frac{x_j^T * \omega_m}{\sqrt{x_j^T * x_j * \omega_m^T * \omega_m}} \right)$$

$$Dt_m = (x_j - \omega_m)^T (x_j - \omega_m)$$

$$Dt_R = \frac{(x_j - \omega_m)^T (x_j - \omega_m)}{\|x_p\|_{(0.2-0.8)}}$$

 Step2 Calculate the probability

$$P_m = \frac{\ln(\theta_m / 180)}{\sum \ln(\theta_m / 180)}$$

 Step3 if x_j not an outlier, then update category center
 Step4 Determine category and update categories center

$$\omega_{m(k+1)} = \omega_{m(k)} + \frac{P_m}{k + 1} (x_j - \omega_{m(k)})$$

 Step5 Determine whether to establish a new category
 end for
 end for
 Step6 Calculate the fault degree

$$FR = \frac{x_j^T x_j}{\omega_p^T \omega_p}$$

category center will not be updated.

$$\omega_{m(k+1)} = \omega_{m(k)} + \frac{1}{k + 1} (x_j - \omega_{m(k)}) \tag{16}$$

$$\omega_{m(k+1)} = \omega_{m(k)} + \frac{P_m}{k + 1} (x_j - \omega_{m(k)}) \tag{17}$$

III. DATA SET VALIDATION

In this paper, the effectiveness of the proposed classification method is verified by six data sets, FSR, TVIB, HMT, Wine, Pima Indians and Soybean. The FSR, Wine, Pima Indians (PimaInd) and Soybean data set are four commonly used classification data set; HMT data set has large sample size and many kinds of feature data, which can test the classification ability of the algorithm for complex data; TVIB is extracted from transformer vibration data, it is a

TABLE 1. Comparison of classification accuracy of various methods.

Method	FSR	HMT	TVIB	Wine	PimaInd	Soybean
SOM	88.67%	57.09%	80%	85.07%	75.35%	79.30%
K-Mean	89.33%	33.56%	82.50%	70.22%	23.70%	59.84%
TC	82%	36.81%	88.33%	83.30%	64.94%	63.30%
NWC	90.67%	92.11%	95.83%	90.74%	68.40%	89.89%
AWC	100%	92.76%	100%	98.15%	78.81%	93.09%

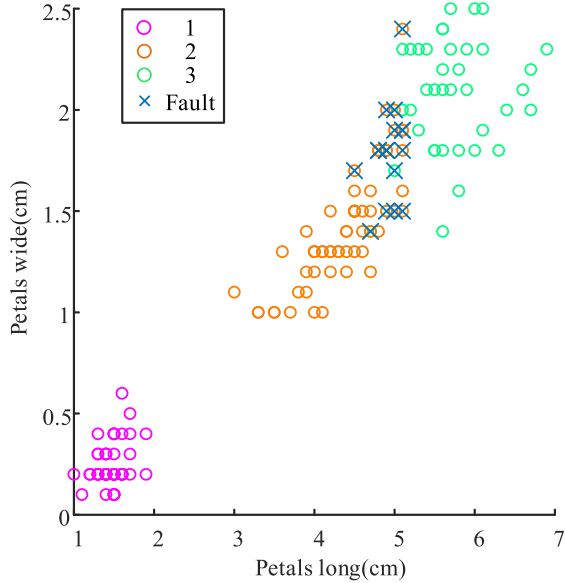


FIGURE 3. SOM classification results of FSR data set.

typical application of fault diagnosis, the features of time domain, frequency domain and correlation are used as the diagnosis basis, and other features can be referred to [19]. These six data sets are labeled data, the category of each sample is known. This part compares the classification ability of the six data sets under different classification methods, which are “SOM”, “K-Mean”, “traditional category center update method (TC)”, “proposed method does not consider the classification weights (NWC)” and “proposed weighted classification method based on adaptive feature selection (AWC)” respectively. Due to the small amount of data in some data sets, the proportion of training data in all data sets is 70%.

A. FSR DATA SET

The FSR data set includes three categories of iris data, namely setosa, versicolor and virginica. Each category of iris has 50 samples, a total of 150 samples. Each sample has four features: calyx length, calyx width, petal length and petal width.

Fig. 3 show the classification results of SOM classification method. Fig. 4 show the classification results of K-Mean classification method, the classification results of the two classifiers for FSR data sets are very similar. The classification difficulty of the data set mainly lies in distinguishing the latter two categories of iris, that is, the part of yellow circle and green circle in the figure. The circles of different colors in the figure represent different categories, and those with

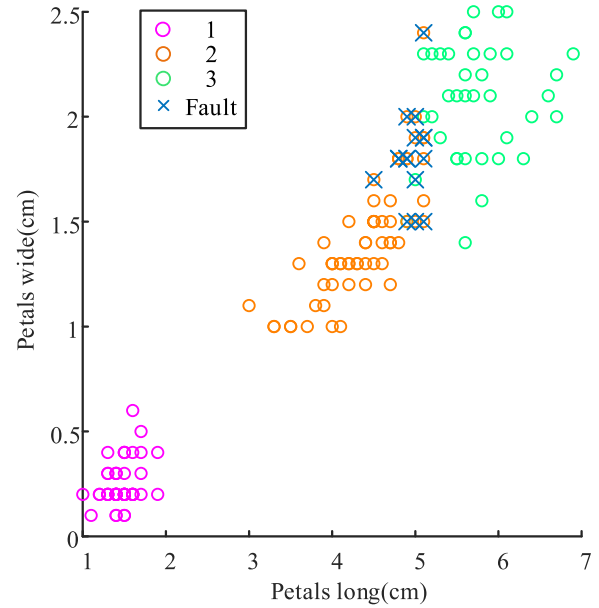


FIGURE 4. K-Mean classification results of FSR data set.

‘x’ on the circle represent samples with wrong classification. Fig.5, Fig.6 and Fig.7 are the classification results of FSR data sets under the three conditions of TC, NWC and AWC. The comparison of classification accuracy is shown in Table 1.

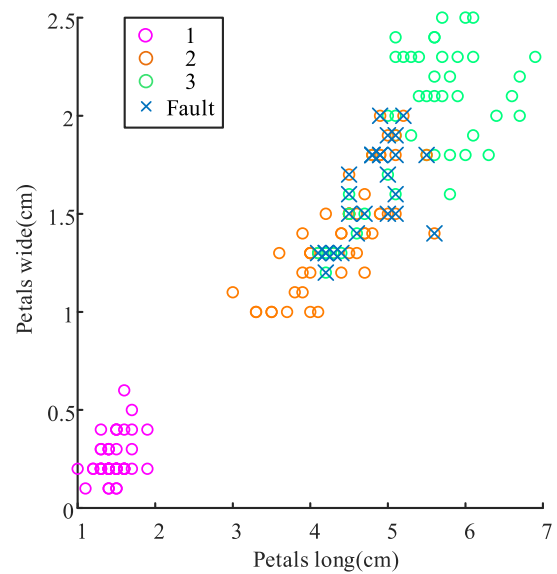


FIGURE 5. TC classification results of FSR data set.

For FSR data set, it can be seen that the TC method has the lowest classification accuracy. If the weight distribution is not considered, the four features give the same classification weight, the classification accuracy is 90.67%, after considering the weight distribution, the classification accuracy of AWC is improved to 100%. As can be seen from Fig. 7. The difference between the three categories centers of FSR data set based on AWC method are shown in FIGURE 8.

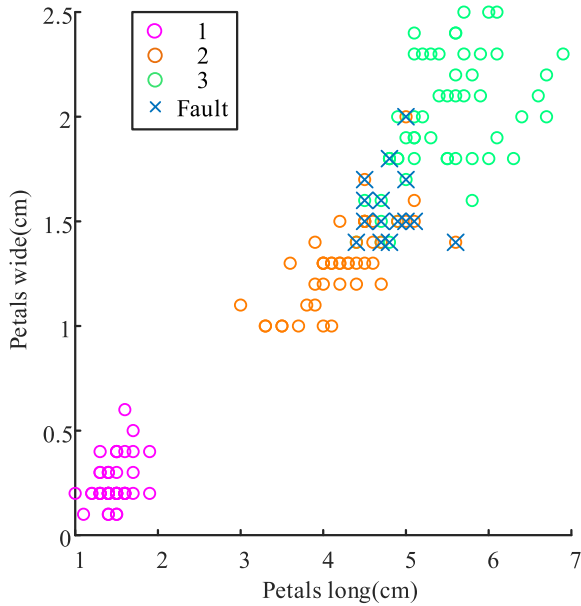


FIGURE 6. NWC classification results of FSR data set.

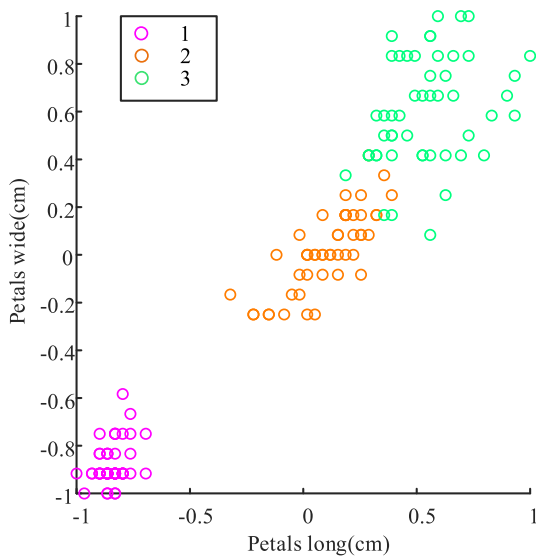


FIGURE 7. AWC classification results of FSR data set.

The difference between the first category and the other two categories is large, in which the difference between the first category and the third category is the largest. The difference between the second category and the third category is the smallest.

The classification ability of the first and second features in FSR data set is poor, especially the classification ability of the second feature is very poor. Therefore, the classification results in Fig. 3 to Fig. 7 are drawn using the latter two feature data, which is more conducive to observation. The classification weight of each feature of FSR data sets is shown in Table 2. After adaptive selection of features, the second features will be abandoned and only the other three features are used for classification. The sum of the latter two features weights accounts for more than 90%, which indicates

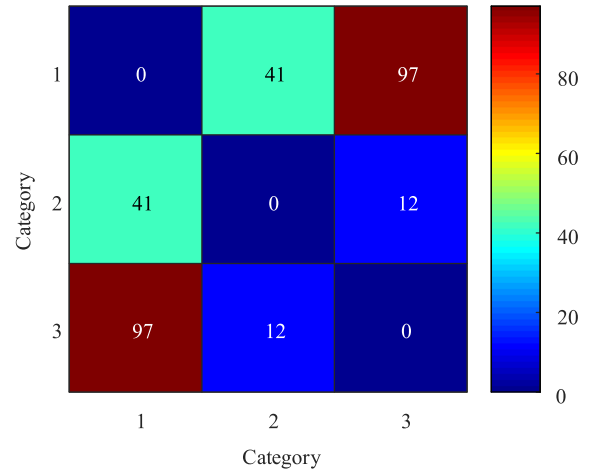


FIGURE 8. The distance between the three category centers of FSR data set (AWC).

TABLE 2. FSR data set weight allocation result.

Feature name	Weight
calyx length	0.077
calyx width	0
petal length	0.5055
petal width	0.4167

that the difference between the three iris mainly lies in the petal.

B. HMT DATA SET

The HMT data set includes five categories of human activities data, namely sitting, standing, walking, running and dancing. The data set has 24075 samples, and each sample has 60 features. After screening by equation (9), 30 features are selected as the classification basis. The comparison of classification accuracy of various classification methods is shown in Table 1. The classification accuracy of the TC method for HMT data set is only 36.81%. Without considering the weight distribution, the classification accuracy of HMT data set is 92.11%, after considering the weight distribution, the classification accuracy can be improved to 92.76%. The differences between the five category centers of HMT data set based on AWC method are shown in Fig. 9.

The classification weight of each feature of HMT data set is shown in Table 3. The classification accuracy of HMT data set under NWC and AWC classification methods has little difference, this is because the classification ability of various feature data in HMT data set has little difference, therefore, whether to consider the distribution of classification weight has no obvious impact on the results.

C. TVIB DATA SET

As shown in Fig. 10, the vibration test transformer has 14 fastening bolts, 6 transverse bolts (A-F) and 8 longitudinal bolts (1-8). There are five vibration sensor mounting locations on the transformer. TVIB data set includes the

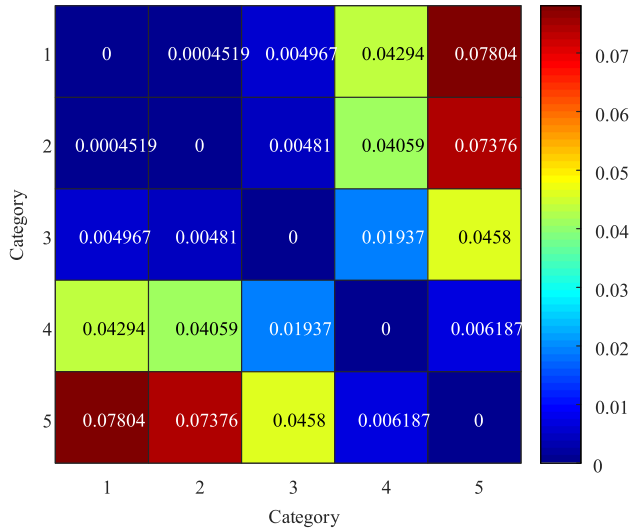


FIGURE 9. The distance between the five category centers of HMT data set (AWC).

TABLE 3. HMT data set weight allocation result.

Feature number	4	5	6	7	10	13
	24	25	26	37	38	39
	47	48	49	50	51	53
	55	56	59			
Weight	0.1242	0.1891	0.0728	0.0551	0.0726	0.0314
	0.0300	0.0289	0.0274	0.0210	0.0226	0.0228
	0.0179	0.0252	0.0257	0.0256	0.0276	0.0547
	0.0228	0.0719	0.0308			

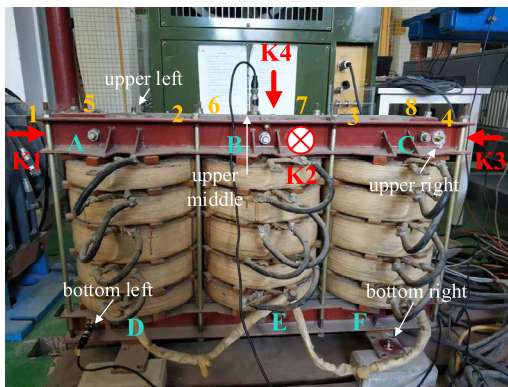


FIGURE 10. Vibration test transformer.

vibration data of transformer under 12 states. The fault types studied in this paper include different degrees of looseness of bolts in different directions of transformer and short-circuit fault of some windings. Table 4 lists the types of transformer faults. Each fault type has 10 samples, a total of 120 samples, and each sample has twelve features, as shown in Table 5.

Fig. 11 shows the angle difference of 12 transformer states. There is a great difference between the transformer short circuit and the other 11 states, and the difference between horizontal bolt looseness and longitudinal bolt looseness. The effect of loose end bolts (1, 4, 5, 8, A, C, D, F) is greater than that of other internal bolts. In the

TABLE 4. Transformer status setting.

Number	Transformer States
1	Initial state
2	1-6 with 40Nm tightening force, 7,8 loosen to 0Nm
3	1-6, 8 with 40Nm tightening force, 7 loosen to 0Nm
4	1-8 with 40Nm tightening force
5	A-F loosen to 0Nm
6	A, D-F with 70Nm tightening force, B, C loosen to 0Nm
7	A, C, D-F with 70Nm tightening force, B loosen to 0Nm
8	A, B, D-F with 70Nm tightening force, C loosen to 50Nm
9	A-F with 70Nm tightening force
10	1-8 with 40Nm tightening force, A, C-F with 90Nm, B with 70Nm
11	1-8 with 60Nm tightening force, A-F with 90Nm
12	Phase B partial winding short circuit

TABLE 5. Transformer vibration feature data.

Feature type	Number	Feature name
Time domain	1	mean square value
	2	the correlation between two channels
	3	peak value
	4	tolerance index
	5	impulsion index
	6	Skewness
	7	kurtosis index
Frequency domain	8	mean value
	9	the correlation with the normal conditions
	10	the correlation between two channels
	11	Mean square frequency
	12	Frequency variance

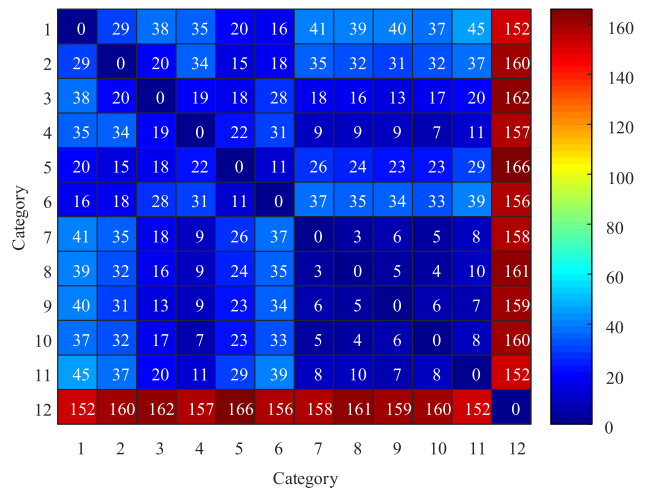


FIGURE 11. Angle difference of 12 transformer states (AWC).

early stage of bolt loosening, it is difficult to identify, for example, the difference between 4, 7, 8, 9, 10 and

TABLE 6. TVIB data set weight allocation result.

Feature number	Weight
1	0.0960
2	0.1154
3	0.0822
4	0
5	0
6	0.0213
7	0.0343
8	0.1294
9	0.2785
10	0.1719
11	0.0488
12	0.0221

11 fault states are small, and it is difficult to distinguish these states.

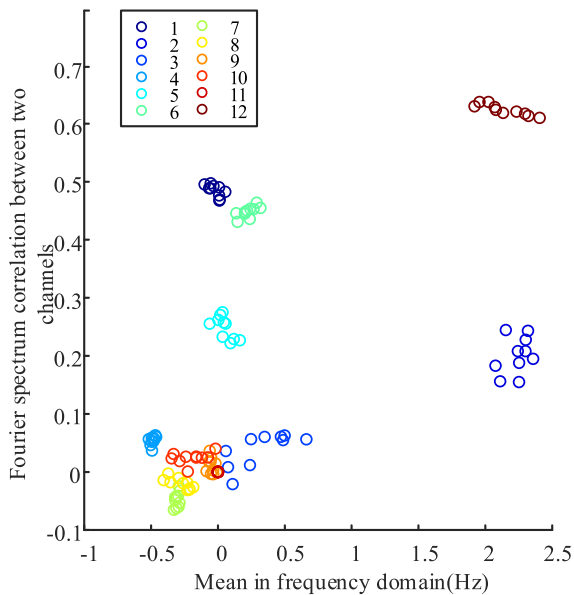


FIGURE 12. Classification results considering the classification weights (AWC).

The weight allocation result of TVIB data set is shown in Table 6, and the weight of the feature 9 is the largest, in other words, the correlation with the frequency spectrum under normal conditions is the best feature to identify the transformer fault state. When considering the classification weights, the classification accuracy reaches 100%, as shown in Fig. 12. For newly measured data or data of uncertain categories, the angle and distance between the new data sample and the center of each category can be calculated to determine whether the new data belongs to the original fault type or the new fault type, the new fault type is determined according to equations (11) and (12). Fig. 13 shows the classification results when a new fault type occurs, the new category is not divided into the original category, the data marked with diamond is the data of the new category. After the classification, the new fault type should also be added to the original fault categories to enrich the transformer fault diagnosis database.

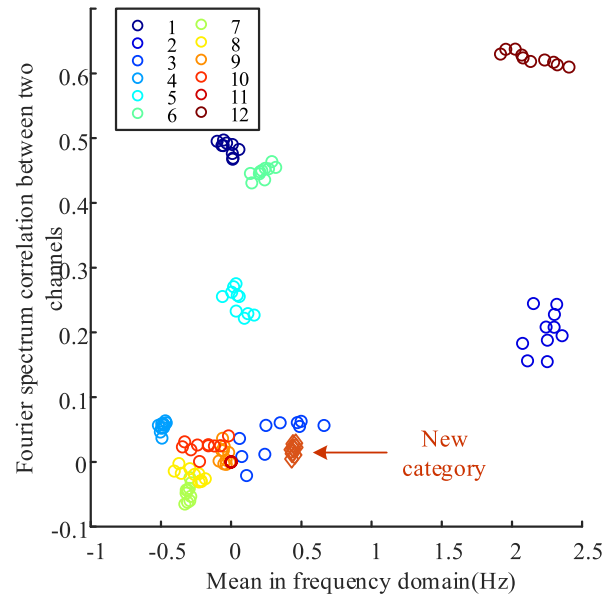


FIGURE 13. Classification results when a new category appears.

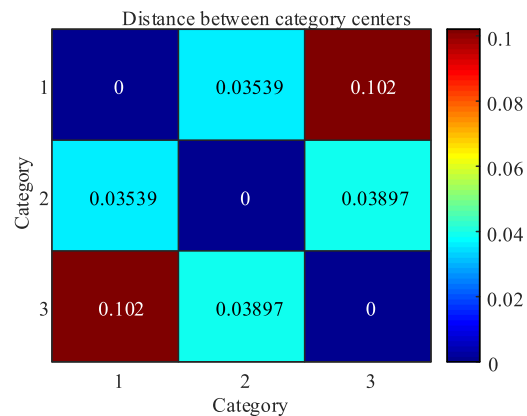


FIGURE 14. The distance between the three category centers of wine data set (AWC).

D. WINE DATA SET

The wine data set comes from UCI database, which records the chemical components of three different varieties of wine in the same region of Italy. The data set has 178 samples, and each sample contains 13 features. The difference between the three categories centers of wine data set based on AWC method are shown in Fig. 14. The difference between the first category and the third category is the largest. After adaptive feature selection, 7 features are selected as the basis for classification. In the classification results of wine dataset, the classification results of one sample are wrong, because the data of the sample deviate greatly, as shown in Fig.15.

E. PIMA INDIANS DATA SET

Pima Indians data set also comes from UCI database, which records the medical data of patients and whether they have diabetes in 5 years. It is a Binary classification problem.

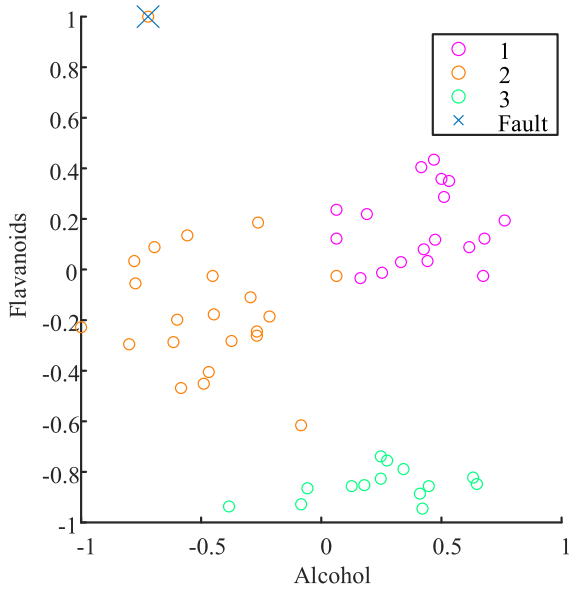


FIGURE 15. Classification results considering the classification weights (AWC).

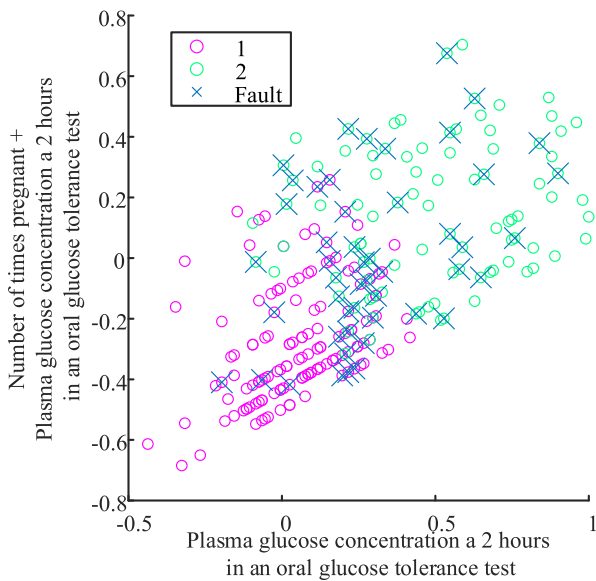


FIGURE 16. Classification results considering the classification weights (AWC).

As can be seen from Fig. 16, there is a large overlap area between the diseased and normal data, and the boundary between the two is not obvious, which indicates that the data characteristics are not good enough and that this kind of classification problem is difficult. In the various deformation algorithms based on multilayer perceptron and deep neural network, the classification accuracy of the data ranges between 65%-80% [20]. The method of generating combined features is used in the Pima Indians data set. Its practical significance is that when some indicators of the patient are high or low at the same time, the patient can be judged to be ill. The classification algorithm proposed in this paper can achieve 78.81% classification accuracy after adaptive

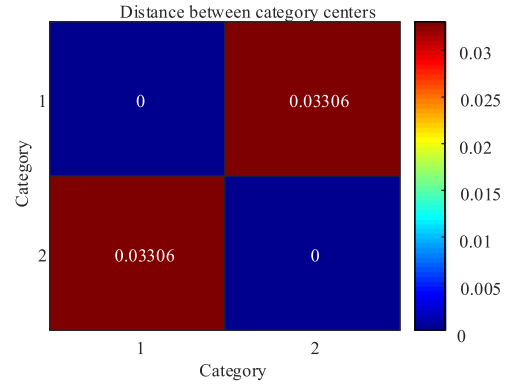


FIGURE 17. The distance between the two category centers of Pima Indians data set (AWC).

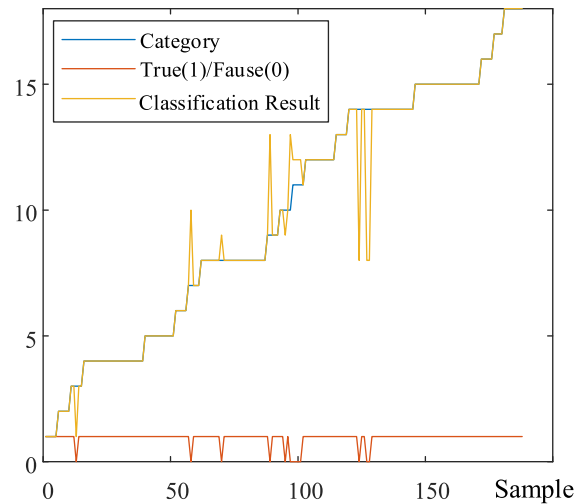


FIGURE 18. Classification results considering the classification weights (AWC).

feature selection. The difference between the three categories centers of Wine data set based on AWC method are shown in Fig. 17.

F. SOYBEAN DATA SET

The soybean dataset is also from the UCI database, which records the surrounding environment and characteristics of 19 different soybean varieties. Fig. 18 and Fig. 19 show the classification results of the soybean dataset. Because most of the data in the dataset are integers such as 0, 1 and 2, and the data coincidence degree is very high, so the classification results are displayed in the way of Fig. 18. During the data set processing, attention should be paid to the processing of the value 0 to prevent the denominator from being zero. Fig. 19 shows the distance distribution between 19 categories in the soybean dataset. Among the various classification methods in paper [21], the highest classification accuracy of this data set is 90,87%. The classification accuracy of the classification method proposed in this paper can reach 93.09%

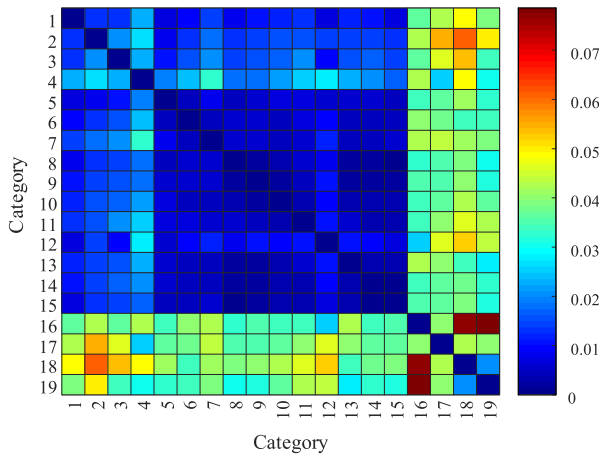


FIGURE 19. The distance between the nineteen category centers of Soybean data set (AWC).

IV. CONCLUSION

This paper proposes a classification method, which will not change the influence of each sample in determining the category center because of the processing order of sample data. This method can select effective features and give different classification weights according to the classification ability of different features. The classification method has been tested on six data sets and achieved good classification results.

For the first kind of simple classification problem, this method can not only give the specific classification results, but also give the corresponding probabilities belonging to different categories, which is more in line with the actual situation. For the application of fault diagnosis, this method can give the classification results, the corresponding fault probability and fault degree. More importantly, this method can judge whether there is a new fault type, as shown in Fig. 13, this is beneficial to early warning of faults and it is very useful for fault diagnosis.

The classification accuracy of the six data sets is shown in Table 1. The traditional category update method has poor classification ability, especially for the data set with large sample size, such as HMT data set. The new method significantly improves the classification accuracy, and considering the classification weights can further improve the classification accuracy. It should also be noted that the classification accuracy of the TVIB data set has always been high, because the features of the data set are carefully selected, so the reasonable selection of evaluation features is an important factor to ensure the classification accuracy.

REFERENCES

- [1] L. Chen, M. Li, W. Su, M. Wu, K. Hirota, and W. Pedrycz, "Adaptive feature selection-based AdaBoost-KNN with direct optimization for dynamic emotion recognition in human-robot interaction," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 5, no. 2, pp. 205–213, Apr. 2021.
- [2] W. Ding, Y. Zhang, Y. Sun, and T. Qin, "An improved SFLA-kmeans algorithm based on approximate backbone and its application in retinal fundus image," *IEEE Access*, vol. 9, pp. 72259–72268, 2021.

- [3] L. Han, S. Luo, H. Wang, L. Pan, X. Ma, and T. Zhang, "An intelligible risk stratification model based on pairwise and size constrained kmeans," *IEEE J. Biomed. Health Informat.*, vol. 21, no. 5, pp. 1288–1296, Sep. 2017.
- [4] M. Li, H. Wang, H. Long, J. Xiang, B. Wang, J. Xu, and J. Yang, "Community detection and visualization in complex network by the density-canopy-Kmeans algorithm and MDS embedding," *IEEE Access*, vol. 7, pp. 120616–120625, 2019.
- [5] K. Peng, V. C. M. Leung, and Q. Huang, "Clustering approach based on mini batch Kmeans for intrusion detection system over big data," *IEEE Access*, vol. 6, pp. 11897–11906, 2018.
- [6] J. Liu, Z. Zhao, C. Tang, C. Yao, C. Li, and S. Islam, "Classifying transformer winding deformation fault types and degrees using FRA based on support vector machine," *IEEE Access*, vol. 7, pp. 112494–112504, 2019.
- [7] Y. Li, M. Dong, and R. Kothari, "Classifiability-based omnivariate decision trees," *IEEE Trans. Neural Netw.*, vol. 16, no. 6, pp. 1547–1560, Nov. 2005.
- [8] A. Suárez and J. F. Lutsko, "Globally optimal fuzzy decision trees for classification and regression," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 12, pp. 1297–1311, Dec. 1999.
- [9] U. M. Rao, I. Fofana, K. N. V. P. S. Rajesh, and P. Picher, "Identification and application of machine learning algorithms for transformer dissolved gas analysis," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 28, no. 5, pp. 1828–1835, Oct. 2021.
- [10] J. Li, G. Li, C. Hai, and M. Guo, "Transformer fault diagnosis based on multi-class AdaBoost algorithm," *IEEE Access*, vol. 10, pp. 1522–1532, 2022.
- [11] J. Wu and H. Yang, "Linear regression-based efficient SVM learning for large-scale classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 10, pp. 2357–2369, Oct. 2015.
- [12] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biol. Cybern.*, vol. 43, no. 1, pp. 59–69, 1982.
- [13] T. Kohonen, "Self-organizing neural projections," *Neural Netw.*, vol. 19, nos. 6–7, pp. 723–733, Jul./Aug. 2006.
- [14] C. Yang, S. Su, X. Ju, and J. Song, "A mobile sensors dispatch scheme based on improved SOM algorithm for coverage hole healing," *IEEE Sensors J.*, vol. 21, no. 18, pp. 21080–21089, Sep. 2021.
- [15] N. Sinclair, D. Harle, I. A. Glover, J. Irvine, and R. C. Atkinson, "An advanced SOM algorithm applied to handover management within LTE," *IEEE Trans. Veh. Technol.*, vol. 62, no. 5, pp. 1883–1894, Jun. 2013.
- [16] Y. Zhang, M. Chen, D. Tian, and L. Ding, "Biomimetic SLAM algorithm based on growing self-organizing map," *IEEE Access*, vol. 9, pp. 134660–134671, 2021.
- [17] J. Fu and G. Che, "Fusion fault diagnosis model for six-rotor UAVs based on conformal Fourier transform and improved self-organizing feature map," *IEEE Access*, vol. 9, pp. 14422–14436, 2021.
- [18] S. Misbahulmunir, V. K. Ramachandaramurthy, and Y. H. M. Thayoob, "Improved self-organizing map clustering of power transformer dissolved gas analysis using inputs pre-processing," *IEEE Access*, vol. 8, pp. 71798–71811, 2020.
- [19] Y. Li, S. Jiao, and B. Geng, "A comparative study of four multi-scale entropies combined with grey relational degree in classification of ship-radiated noise," *Appl. Acoust.*, vol. 176, May 2021, Art. no. 107865.
- [20] W. Zhenyuan, *Deep Learning, Python Practice Based on Keras*. Beijing, China: Electronic Industry Press, 2018.
- [21] Y. Jiang and Z. Zhou, "Editing training data for kNN classifiers with neural network ensemble," in *Advances in Neural Networks—ISNN 2004*. Berlin, Germany: Springer, 2004.



RUIZHENG NI (Student Member, IEEE) was born in Shanxi, China. He received the B.S. degree in electrical engineering and automation from Shanxi Agriculture University, Shanxi, in 2016. Since 2016, he has been a Postdoctoral Researcher with Beijing Jiaotong University, Beijing, China. His current research interests include transformer life-time prediction, electric traction, and transmission control.



RUICHANG QIU received the B.S. degree in electrical engineering and automation from Beijing Jiaotong University, Beijing, China, in 1993. From 1990 to 2005, he worked as a Lecturer with the School of Electrical Engineering, Beijing Jiaotong University, where he is currently an Associate Professor. His current research interests include intelligent detection technology, power electronics and power conversion technology, electric traction, and transmission control.



ZHIWEI LUO working with the Beijing Aerospace Launch Technology Research Institute, has long been committed to the scientific research of power electronics and motor drive. He has successively presided over and completed a number of major research projects in the field of power supply and power drive. His research results have been widely used in the field of special vehicles and industrial equipment. He is responsible for preparing a number of national military standards, industrial standards, and enterprise standards related to power supply and electric drive. He has compiled two books, published more than 20 papers in domestic core journals, industry journals, and academic conferences, and obtained more than 40 invention patents.



JIE CHEN (Member, IEEE) received the B.S. and Ph.D. degrees in electrical engineering and automation from Beijing Jiaotong University, Beijing, China, in 2008 and 2013, respectively. Since 2013, he has been a Postdoctoral Researcher with Beijing Jiaotong University and the Institute of Electrical Engineering, Chinese Academy of Sciences, Beijing. He was a Visiting Scholar of WEMPEC with the University of Madison. He is currently an Associate Professor with Beijing Jiaotong University. His research interests include variable frequency drive, rail transportation traction control, and inverter parallel control.



ZHEMING JIN (Member, IEEE) was born in Jilin, China. He received the B.S. degree in electrical engineering and the M.S. degree in power electronics and AC drives from Beijing Jiaotong University, Beijing, China, in 2013 and 2015, respectively, and the Ph.D. degree in power electronic systems from the Department of Energy Technology, Aalborg University, Aalborg, Denmark, in 2018. In 2019, he joined the Department of Energy Technology, Aalborg University, as a Postdoctoral Researcher, and was subsequently appointed as an Assistant Professor. In 2021, he joined Beijing Jiaotong University, where he is currently an Associate Professor. His research interests include control of power electronic converters, stability of power electronic based power systems, energy storage, DC microgrids, and their applications in transportation electrification.



ZHIGANG LIU received the bachelor's, master's, and Ph.D. degrees in electric drive for locomotives from Beijing Jiaotong University, Beijing, China, in 1986, 1990, and 1994, respectively. He is currently a Full Professor with Beijing Jiaotong University. In recent years, he presided over a number of national key scientist research projects and achieved fruitful results in the field of rail transit power supply, traction control, safety prediction, and control. He has published a book on power electronics and spent several months as a Visiting Scholar in U.S. and Canada. His teaching activities and research interests include power electronics circuit and systems, rail transportation traction control, and safety. He is the Vice Chairperson of the China Electro-Technical Society Rail Transport Electrical Technical Committee, and he is also an evaluation expert of several national key plans.

...