IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received August 18, 2020, accepted August 28, 2020, date of publication September 1, 2020, date of current version September 14, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3020906

A Fault Diagnosis Method Based on Improved Adaptive Filtering and Joint Distribution Adaptation

ZHANG XU[1], HUANG DARONG[1], (Member, IEEE), GUOXI SUN[2], AND WANG YONGCHAO[3]

[1]College of Information Science and Engineering, Chongqing Jiaotong University, Chongqing 400074, China
[2]Department of Computer and Electronic Information, Guangdong University of Petrochemical Technology, Maoming 525000, China
[3]Xinjiang Electric Power Research Institute, State Grid Xinjiang Electric Power Company Ltd., Ürümqi 830000, China

Corresponding author: Huang Darong (drhuang@cqjtu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61304104, in part by the Science and Technology Research Project of the Chongqing Municipal Education Commission of China under Grant KJZD-K201800701, in part by the Chongqing Technology Innovation and Application Special Key Project under Grants cstc2019jscx-mbdxX0015, in part by the Key Projects of Guangdong Natural Science Foundation under Grant 2018B030311054, in part by the Innovation Foundation of Chongqing Postgraduate Education under Grant CYS20282 and Grant CYS20285, in part by the Program of Chongqing Innovation and Entrepreneurship for Returned Overseas Scholars of China under Grant cx2018110, and in part by the Science and Technology Project of Power Science Research Institute, State Grid Xinjiang Electric Power Company Ltd., under Grant SGXJDK00JLJS1800161.

ABSTRACT In the real environment of industrial equipment, the vibration signals of essential components show deviations due to the fault and noise. Notably, the noise in the signal will interfere with the diagnosis process of the signal and reduce the accuracy of fault diagnosis. Based on the above problem, adaptive filtering (AF) is used as an excellent method to attenuate noise without specifying the noise type. However, how to define the most appropriate length and type of morphological filter element is the most inherent problem which needs to be solved first. This paper proposed a cooperative diagnosis method of rolling bearings vibration signal based on improved adaptive filtering and joint distribution adaptation (JDA). First, the kurtosis under different element types and lengths is calculated as an index. The structural element corresponding to maximum kurtosis is selected as the most suitable morphological filter element because the different morphological filter elements reflect the effect of feature extraction. Then, JDA aims to improve both the marginal distribution and the conditional distribution to solve the chaotic distribution of time-domain features under variable working conditions. Finally, the improved least squares support vector machine (LSSVM) verified the effectiveness and improvement of the proposed method under bearing acceleration signal. At the same time, the comparative experiment proved that the proposed method not only directly corrects the most appropriate elements greatly optimizes the feature structure, but also enhances the accuracy of fault diagnosis.

INDEX TERMS Adaptive filters, morphological operation, time domain features, transfer learning, joint distribution adaptation, LSSVM, fault diagnosis.

I. INTRODUCTION

In the mechanical industrial system, the safety status of critical parts has always been concerned. During the mechanical operation process, for example, the health state of bearing plays a very significant role in industrial systems. However, bearings in a complex industrial environment may cause some faults during long time operation, such as external cracks, inner cracks, wear and other faults. These faults may not obvious and are difficult to identify at the beginning. But if the fault is not immediately recognized, it will cause serious trouble to the mechanical process. Besides, the complex noise caused by the fault will also affect the accuracy of signal diagnosis.

Acceleration data is usually used as a detection object to verify the validity of the diagnostic model, because acceleration most likely reflects the safety of the bearing. However, the data collected from the actual equipment contains a lot of complex noise, and has a serious impact on the extracted features. Initially, time-domain features are greatly affected

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

by noise, and the extracted features may show the noise features rather than fault features. Unfortunately, the features of bearing without de-noising would lead to reduced accuracy of fault diagnosis. Therefore, most researchers try to de-noising the signal of acceleration data before extracting features.

At the beginning of the fault diagnosis, the first step is to reduce the noise of the acceleration signal, and then the extracted feature would correctly represent the characteristics of fault. However, some de-noising methods have limitations on the type of noise, such as least mean squares (LMS) [1]. Also, in the actual environment, the type of noise is complex and variable [2]. In this paper, an adaptive morphological filter is selected as the signal de-noising method under experiment acceleration data. Specifically, the methods of adaptive morphological filters have many different transformations, and they can be used in many different situations with different operators. The combination of different operators will bring different effects in the de-noising process. Some researchers have constructed some progressive methods of combination morphological filter such as (enhanced difference morphological filter (EDMF), adaptive smoothing multiscale morphological filtering (ASMMF), an adaptive rank-order morphological filter (ARMF), multiscale morphology spectral (MMS), *et al.*) in [3]–[7]. Based on this, an appropriate combination operator can be constructed to accelerate the rolling bearing. Moreover, there is another problem that cannot be ignored is how to select the length and elements of the morphological filter. In order to solve this problem, there are some relevant literature has discussed the methods of defining the appropriate length and elements of the morphological filter. In [8], PSO is proposed to select the best length and elements of the morphological filter. [9] applies different evolutionary algorithms to find optimal parameters. Genetic algorithm is used to optimize the structure of elements in open and closed operators. Reference [10] uses the sensitive index of fault feature ratio (FFR) to determine the optimal result of the element. Although all of the above methods and swarm intelligence optimization algorithm can find the optimal parameters of the denoising algorithm, notably intelligence optimization algorithms will cost more running time of the de-noising model. All in all, when constructing an adaptive de-noising model, it is necessary to find appropriate indexes to evaluate the best choice of the morphological filter type. Especially, kurtosis can reflect the fault characteristics of the signal, the larger kurtosis value is gotten, the more obvious the fault characteristics will be shown. Therefore, kurtosis is used to calculate the optimal parameters of structural elements to ensure the de-noising effect of morphological filtering and reduce the running time of the algorithm.

After describing the signal de-noising method in detail, the next step is to construct the model of signal decomposition and feature extraction. In this paper, Variational Mode Decomposition (VMD) is used to decompose the acceleration of rolling bearings because it has the advantages of high decomposition accuracy and strong noise robustness [11]–[14]. Then, time-domain features are adopted to construct feature sets from signal components [15]–[19].

In multiple fault diagnosis, the classifier cannot classify the feature set of the fault signal because the distribution of the feature set is very confusing. Although the multi-dimensional feature set of time-domain characteristics can describe the state of the signal from different angles. Therefore, the transferability of data can be used to adjust the structure of feature sets under variable working conditions [20]. Based on this, some researchers applied related methods to improve the structure of feature sets; for example, [21]–[24] proposed the methods such as transfer component analysis (TCA), Multi-Domain Semi-Supervised TCA, et.al, which further improve the performance of fault diagnosis. Through the above related method, the transferability of data can enhance the accuracy of fault diagnosis by the TCA algorithm. Unfortunately, TCA cannot solve the problem of marginal distribution and costs too much time in operation. Thus, the advanced method of joint distribution adaptation (JDA) will be applied in this model of fault diagnosis. JDA can improve the distribution of marginal and conditional at the same time [25]. After reducing the dimension of the dataset, a new dataset with a low dimension can enhance the effectiveness and accuracy of bearing diagnosis. Some researchers have applied the JDA to model classification, such as (JDA improved by sparse filtering, joint distribution optimal deep domain adaptation, Deep transfer network with JDA, joint-space force distribution) in [26]–[30]. JDA is mainly used for feature distribution in image processing, but rarely in signal processing. JDA originated from transfer learning and is suitable for signal feature analysis under variable working conditions, it can solve the problem of unbalanced data distribution. At the same time, JDA can adjust the marginal distribution and conditional distribution of signal features to improve the efficiency of the fault diagnosis model.

After that, the marginal and conditional distribution of rolling bearing feature sets has been greatly improved. The last step is to diagnose the optimized feature set by the classifier [31]. In general, there are many useful and classic classifiers of machine learning algorithms such as K-Nearest Neighbor (KNN), which can measure the distance between different features for classification purposes. Random forest (RF) integrates all sub-classifiers to selects the classifier with the highest vote [32], [33]. Support Vector Machine (SVM) which uses training data to calculate the hyperplane [34]. Compared with the least squares support vector machine (LSSVM), on the one hand, SVM and LSSVM have different constraints. LSSVM changes the constraints of SVM from inequality to equality. On the other hand, the linear equation of LSSVM can improve the calculation efficiency. Based on this, some scholars applied LSSVM method to recognition research such as (particle swarm optimization LSSVM, genetic algorithm LSSVM, local objective set LSSVM, et.al) in [35]–[38].

Based on this, AF was introduced to weaken the noise in the original signal, while the optimal elements of signal in AF need be calculated first. So, kurtosis is applied to screen the optimal elements for acceleration data and construct the improved AF. The optimized AF can better eliminate the influence of noise on fault diagnosis based on adapting to the signal characteristics. VMD is then used to decompose the de-noising signal into components with different characteristics. In order to construct a feature extraction and distribution optimization model, JDA can shrink the dimension of the time-domain feature set and improve the efficiency of diagnosis. At the same time, because the distribution of the original time-domain feature is chaotic, JDA is introduced to optimize the feature distribution in both marginal and conditional. The compactness between the same states of the features becomes smaller, while the distance between different features becomes larger. To sum up, the JDA can improve fault diagnosis accuracy by optimizing feature distribution. Finally, the improved LSSVM which optimized by particle swarm optimization (PSO) to overcome the local optimization will be applied to verify the accuracy and effectiveness of the proposed algorithm.

The remaining part of this paper is organized as follows: Section 2 will construct the signal preprocessing model under improved AF and VMD. In section 3, the time-domain features will be combined with JDA to construct an integrated feature extraction algorithm. In section 4, LSSVM will be introduced to classify the optimized feature set of rolling bearings. Section 5 will discuss the reliability and effectiveness of the fault diagnosis model under experiment acceleration data. In section 6, the conclusion and expectation will be reached.

II. OPTIMIZED MODEL OF SIGNAL PREPROCESSING BASED ON ADAPTIVE FILTERING AND VARIABLE MODE DECOMPOSITION

Since the selection of morphological filters is significant and complex, the improved method of signal de-noising aims to select the optimal operators, which depends on the kurtosis of acceleration. In order to achieve the adaptive optimization of length and type of elements, kurtosis is an important index to measure the filtering effect. The signal with the most prominent and obvious fault characteristics will also have the largest kurtosis value. Especially, to describe the original signal from different angles of time and frequency domain, VMD decomposes the denoised signal into multidimensional signal components with different characteristics. All in all, the improved preprocessing method can obtain the optimal morphological filters and divide the original signal into several components. Next, the optimized signal preprocessing method will be introduced first in detail.

A. THE BASIC THEORY OF ADAPTIVE FILTERING

The principle of morphological filter is to use the filter window structure to modify or match the noise of the signal to extract the features from the de-noising signal. There are four basic operations in morphological filtering, and they are corrosion, expansion, opening operator, and closing operator. To illustrate the above four operations, $f(n)$ represents the original signal, and $g(m)$ is the structural element. The range of $f(n)$ and $g(m)$ are $F = (0, 1, \cdots N-1)$ and $G = (0, 1, \ldots, M-1)$ respectively. And $N \gg M$. Then, the function of four basic operators is defined as follows:

$$(f\Theta g)(n) = \min[f(n+m) - g(m)] \tag{1}$$

$$(f \oplus g)(n) = \max[f(n-m) + g(m)] \tag{2}$$

$$opf(n) = (f \circ g)(n) = (f\Theta g \oplus g)(n) \tag{3}$$

$$clf(n) = (f \cdot g)(n) = (f \oplus g\Theta g)(n) \tag{4}$$

where Θ is the dilation operator, $\oplus$ is the erosion operator, $\circ g$ is the opening operator, and $\cdot g$ is the closing operator.

Because the effect of basic operators is limited, the basic operator is rarely single used for signal filtering. So, several classic combinatorial filter operators (such as opening-closing (FOC), closing-opening (FCO), combination morphological filter (CMF), morphological gradient (MG), average (AVG), white top-hat (WTH), black top-hat (BTH)) are proposed as follows:

$$FOCf(n) = (f \circ g \cdot g)(n) \tag{5}$$

$$FCOf(n) = (f \cdot g \circ g)(n) \tag{6}$$

$$CMFf(n) = (FOCf(n) + FCOf(n))/2 \tag{7}$$

$$MGf(n) = (f \oplus g) - (f\Theta g)(n) \tag{8}$$

$$AVGf(n) = (f \cdot g + f \circ g)(n)/2 \tag{9}$$

$$WTHf(n) = f(n) - (f \circ g)(n) \tag{10}$$

$$BTHf(n) = (f \cdot g)(n) - f(n) \tag{11}$$

Different combinatorial operators of the irreversible morphological filter have different effects on signal de-noising. In the actual experiment, the real acceleration data has both positive and negative pulses. Therefore, the morphological difference operation (MDO) composed of corrosion and expansion is used to process the positive and negative pulses of the signal. The function of MDO is expressed as follows:

$$mdo(n) = (f \cdot g)(n) - (f \circ g)(n) \tag{12}$$

MDO is constructed to eliminate positive and negative pulses. The next step is to define the specific parameters in the morphological difference operation.

B. THE OPTIMAL SELECTION STRATEGY OF FILTER PARAMETERS

The theory of adaptive morphological filter is first used in image processing. Shape characteristics often accompany the structural elements. The classic one-dimensional structural elements include linear, triangular, circular, et.al. The schematic diagrams of linear, triangular, and circular structural elements are shown as follows.

For example, the linear and triangular structural elements with length 5 and height 1 can be represented as $g = (1, 1, 1, 1, 1)$, $g = (0, 0.5, 1, 0.5, 0)$ respectively.

The parameters of structural elements are length and height. Both determine the effectiveness of structural elements. The length and type of elements mainly influences the filtering effectiveness of structural elements.

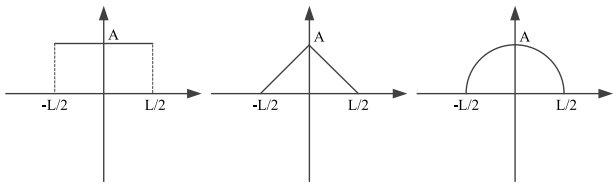

FIGURE 1. Schematic diagram of linear, triangular, and circular structural elements.

Therefore, kurtosis is used to evaluate the filtering effect of different parameters of the original signal. The kurtosis is defined as follows:

$$K = \frac{E(x-\mu)^4}{\sigma^4} \tag{13}$$

where, μ is the average of x, σ is the standard deviation of x.

The height H of the linear and triangular structural elements is 1,3,5,8, and the length L ranges from 1 to 40. After de-noising the signal via MDO, the filtering effect is measured by the kurtosis, and H and L with the maximum kurtosis of the processed signal are adopted. A large number of impact components in the signal will significantly increase the kurtosis of the signal. The impact vibration caused by the fault will cause the vibration of bearing in different frequency bands, the impact component with the largest kurtosis is the most obvious, and the fault is the easiest to extract. Therefore, the value of kurtosis is the principle of optimal parameters, and the length and height of structural elements corresponding to the maximum kurtosis are adopted as the optimal parameters for de-noising. Then the denoised signals are used as research objects of signal decomposition and feature extraction.

C. THE SIGNAL DECOMPOSITION ALGORITHM OF VMD

The feature extracted from the one-dimensional signal is limited. Thus, the denoised signal should be decomposed with different characteristics. VMD is an adaptive signal composition method. The variational decomposition of VMD can overcome the lack of modal confusion and insufficient endpoint. Compared with other signal decomposition methods, VMD shows higher decomposition efficiency. The specific method of VMD is as follows:

Step 1: The VMD decomposes the signal into several signal components, which are defined as follows.

$$u_k(t) = A_k(t)\cos\left[\phi_k(t)\right] \tag{14}$$

where the range k is $[1, K]$, K is the number of signal components. $A_k(t)$ is the instantaneous amplitude of $u_k(t)$. The instantaneous frequency of $u_k(t)$ is $w_k(t) = \phi_k'(t)$. f represents the original signal, which is consists of K components of $u_k(t)$. The center frequency of every IMFs is w_k.

Step 2: To calculate the analytic signal of the mode function, the Hilbert transform is applied for every $u_k(t)$ as follows:

$$\left[\sigma(t) + \frac{j}{\pi t}\right] u_k(t) \tag{15}$$

Step 3: The estimated center frequency $e^{-jw_k t}$ of the analytic signal is mixed, and the spectrum of each mode is modulated to the corresponding baseband.

$$\left[\left[\sigma(t) + \frac{j}{\pi t}\right] \times u_k(t)\right] e^{-jw_k t} \tag{16}$$

Step 4: Calculate the square norm of the gradient, and estimate the bandwidth of each modal component. Then, construct a constrain variational model as follows:

$$\min_{\{u_k\},\{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[\left(\sigma(t) + \frac{j}{\pi t} \right) \times u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\}$$
$$s.t. \sum_k u_k(t) = f \tag{17}$$

where $u_k = \{u_1, u_2, \cdots, u_k\}$ represents K components after decomposition and $\omega_k = \{\omega_1, \omega_2, \cdots, \omega_k\}$ represents the center frequency of each component.

Step 5: The secondary penalty factor of α and LaGrange multiplication operator of $\lambda(t)$ are introduced to calculate the above model. The extended function is shown below.

$$\begin{aligned} &L(\{u_k(t)\}, \{\omega_k(t)\}, \lambda) \\ &= \alpha \sum_k \left\| \partial_t \left[\left(\sigma(t) + \frac{j}{\pi t} \right) \times u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \\ &\quad + \left\| f(t) - \sum_k u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_k u_k(t) \right\rangle \end{aligned} \tag{18}$$

Step 6: The center frequency of IMFs is updated by a multiplier alternating direction algorithm. The saddle point of Eq. (17) is the optimal result of the original problem. All the IMFs can be obtained as follows:

$$\widehat{u}_k^{n+1}(\omega) = \frac{\widehat{f}(\omega) - \sum_{i\neq k} \widehat{u}_i(\omega) + \frac{\widehat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \tag{19}$$

where $\hat{u}_k^{n+1}(w)$ is the result of the surplus $\hat{f}(w) - \sum_{i\neq k} \hat{u}_i(w)$ which through the Wiener filtering. The center frequency of each IMFs is updated as follows.

$$\widehat{\omega}_k^{n+1} = \frac{\int_0^\infty \omega \left|\widehat{u}_k(\omega)\right|^2 d\omega}{\int_0^\infty \left|\widehat{u}_k(\omega)\right|^2 d\omega} \tag{20}$$

Based on the above, the process of VMD is thoroughly introduced. In this paper, the optimal value selection of K refers to the latest research about [39], [40]. Through the research on the de-noised signal in the actual experimental environment, under the acceleration signals of five different states, the decomposition number of VMD is defined as

3. After obtaining the de-noising signal components, feature extraction and structure optimization models will be described in the next section.

III. THE COOPERATIVE FEATURE EXTRACTED METHOD BASED ON TIME DOMAIN ANALYSIS AND JDA

The cooperative feature extraction model consists of time-domain features and JDA. In general, the time domain features can describe the state of the fault signal from different angles. However, the multidimensional time-domain features cannot make the fault diagnosis model achieve good results. So, to improve the structure of the multidimensional feature set, JDA is introduced to optimize both the distribution of marginal and conditional of the feature set. The optimized structure of the feature set will significantly enhance the accuracy of fault diagnosis. Based on the above, the feature extracted method will be first expressed.

A. CHARACTERISTIC EXTRACTION BY EQUAL INTERVAL ENERGY PROJECTION

The signal should be denoised before feature extraction; the denoised signal is decomposed into several signal components by VMD. Time-domain features are made up of dimensional indexes and dimensionless indexes. The dimensional indexes include maximum value, minimum value, peak value, average value, root-mean-square value, root amplitude, average amplitude, skewness, kurtosis, variance, peak to peak value, and dimensionless indexes include waveform index, peak index, pulse index, skewness index, and margin index. To further express the time-domain feature method in detail, assuming that the acceleration data is $x_n = \{x_1, x_2, \cdots, x_n\}$, the time-domain characteristics are calculated as follows:

As shown in table 1, the above 16 features are selected to construct the multidimensional feature set of the bearing. Then, the feature structure is optimized under variable working conditions of rolling bearing fault diagnosis.

TABLE 1. Calculation of time domain characteristics.

index	function
maximum	$x_{\max} = \max\limits_{i=1,2,\cdots n}\{x_i\}$
minimum	$x_{\min} = \min\limits_{i=1,2,\cdots n}\{x_i\}$
peak	$x_p = \max\limits_{i=1,2,\cdots n}\lvert x_i\rvert$
average	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
root-mean-square value	$x_{rmsv} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}$
root amplitude	$x_{sra} = \left(\frac{1}{n}\sum_{i=1}^{n}\lvert x_i\rvert^{1/2}\right)^2$
average amplitude	$x_{aa} = \frac{1}{n}\sum_{i=1}^{n}\lvert x_i\rvert$
skewness	$x_{sc} = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(\lvert x_i\rvert - \bar{x}\right)^3}{x_{msv}^3}$
kurtosis	$x_k = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(\lvert x_i\rvert - \bar{x}\right)^4}{x_{msv}^4}$
variance	$\sigma = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$
peak to peak value	$x_{p-p} = x_{\max} - x_{\min}$
waveform index	$x_w = \frac{x_{msa}}{x_{aa}}$
peak index	$x_{pi} = \frac{x_p}{x_{msa}}$
pulse index	$x_{pf} = \frac{x_p}{x_{aa}}$
skewness index	$x_s = \frac{x_{sc}}{\sigma^{3/2}}$
margin index	$x_c = \frac{x_k}{\sigma^2}$

B. THE BASIC THEORY OF JOINT DISTRIBUTION ADAPTATION

The original distribution of time-domain feature sets may not only be chaotic in marginal but also represent the disorder in conditional.

It can be seen from Figure 2 that the distribution of source and target domains is completely different. Marginal distribution adaptation (MDA) can adjust the central distribution of two domains. However, optimization of central distribution cannot satisfy the domain classification tasks. Thus, conditional distribution adaptation (CDA) is introduced to minimize the difference between the labeled source domain feature and the unlabeled target domain feature. JDA is proposed to optimize the distribution of feature sets under variable working conditions in this section.

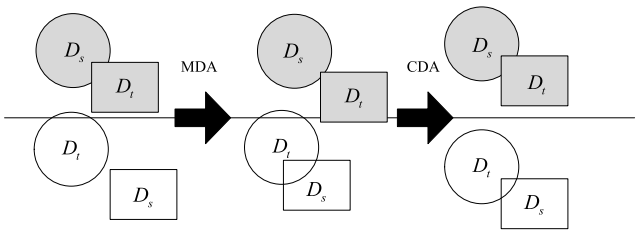

FIGURE 2. Illustration of MDA and CDA.

To introduce the JDA in detail, $X_s = \{x_i^s, y_i^s\}_{i=1}^{n_s}$ represents the source dataset, y_i^s is the label, $X_t = \{x_i^t\}_{i=1}^{n_t}$ represents the target dataset. Especially, the maximum mean discrepancy (MMD) is used to measure the difference between two

distributions.

$$MMD_H^2(X_s, X_t) = \left\| \frac{1}{n_s}\sum_{i=1}^{n_s}\phi\left(x_i^s\right) - \frac{1}{n_t}\sum_{i=1}^{n_t}\phi\left(x_i^t\right)\right\|_H^2 \quad (21)$$

JDA has four main optimization objects: principal component analysis (PCA), MDA, CDA, and JDA.

Step 1: PCA aims to maximize the variances between the orthogonal components to optimize the input features. Input data can be represented as $X = \left[x^i, \cdots, x^n\right]$, where $x^i \in R^{N_{in}\times 1}$. The purpose of PCA is presented as follows.

$$\max_{A^TA=I} tr\left(A^TXHX^TA\right) \quad (22)$$

where the $tr(\cdot)$ denotes the matrix trace. The output feature is $Z = \left[z^1, \cdots, z^n\right] = A^TK$.

Step 2: MDA. The purpose of MDA is to minimize the marginal distributions of the source and target domain. Therefore, applying MMD to MDA to calculate the result is as follows:

$$L_0 = \left\| \frac{1}{n_s}\sum_{i=1}^{n_s}A^Tx_s^i - \frac{1}{n_t}\sum_{j=1}^{n_t}A^Tx_t^i\right\|^2 = tr\left(A^TXM_0X^TA\right) \quad (23)$$

$$M_0 = \begin{cases} \frac{1}{n_sn_s}, x_s^i, x_s^j \in D_s \\ \frac{1}{n_tn_t}, x_t^i, x_t^j \in D_t \\ \frac{1}{n_tn_t}, otherwise \end{cases} \quad (24)$$

where M_0 is defined as Eq. (24). And the Eq. (23) can minimize the discrepancy distributions between source and target domain.

Step 3: CDA. CDA is essential for distribution and adaptation. However, there is a problem that the label of the target domain is unknown. Thus, the technique of using pseudo labels of the target domain in CDA is applied to unsupervised domain adaptation. The conditional distribution $Q(x_s|y_s) = Q(x_s|\widehat{y_t})$ is an alternative of the posterior probability $Q(y_s|x_s) = Q(\widehat{y_t}|x_s)$. So, under the labeled source domain and unlabeled target domain, the distance between $Q(x_s|y_s = \mathrm{c})$ and $Q(x_s|\widehat{y_t} = \mathrm{c})$ can be mapped together. The number of datasets is C, where $\mathrm{c} \in \{1, \cdots C\}$. MMD is rewritten to adapt to the conditional distribution, as shown below.

$$\begin{aligned} L_c &= \sum_c^C \left\| \frac{1}{n_s^C}\sum_{x_s\in D_s^c}A^Tx^i - \frac{1}{n_t^C}A^Tx^j\right\|^2 \\ &= \sum_{c=1}^C tr\left(A^TXM_cX^TA\right) \end{aligned} \quad (25)$$

where the input data $D_s^c = \left\{x^i : x^i \in D_s \cap y\left(x^i\right) = c\right\}$ and $D_t^c = \left\{x^i : x^i \in D_t \cap \widehat{y}\left(x^i\right) = c\right\}$ are in source and target domain, respectively. $y\left(x^i\right)$ and $\widehat{y}\left(x^i\right)$ are the true label of x^i, x^j is the pseudo label of the target domain. n_s^c and n_t^c is the number of samples of source and target domain. M_c is the MMD matrix of CDA can be rewritten as follows.

$$M_c = \begin{cases} \frac{1}{n_s^cn_s^c}, x^i, x^j \in D_s^c \\ \frac{1}{n_t^cn_t^c}, x^i, x^j \in D_t^c \\ \frac{-1}{n_s^cn_t^c}, \begin{cases} x^i \in D_s^c, x^j \in D_t^c \\ x^i \in D_t^c, x^j \in D_s^c \end{cases} \\ 0, otherwise \end{cases} \quad (26)$$

Step 4: JDA. The above method of CDA and MDA should be combined simultaneously. Therefore, the function combined with Eq. (23) and Eq. (25) is as follows:

$$\min_{AXHX^TA=I}\sum_{c=0}^C tr(A^TXM_cX^TA) + \lambda\left\|A\right\|_F^2 \quad (27)$$

Based on the theory of constrained optimization, the Lagrange multiplier can be represented $\Phi = \mathrm{diag}(\varphi_1, \cdots \varphi_k) \in R^{k\times k}$, and then, Lagrange function is defined as follows.

$$L_Z = tr\left(^T\left(X\sum_{c=0}^C M_cX^T + \lambda I\right)A\right) + tr\left(\left(I - A^TXHX^T\right)\Phi\right) \quad (28)$$

$$\left(X\sum_{c=0}^C M_cX^T + \lambda I\right)A = XHX^TA\Phi \quad (29)$$

At last, the problem of calculating the result of the matrix A is transformed into solving the Eq. (29), and the k smallest eigenvectors are obtained. The classifier is used to predict the pseudo labels, and then the pseudo labels can be obtained from the feature extracted in the target dataset.

The advantage of time domain feature set is that it can describe fault characteristics of signals from multiple angles to distinguish different fault signals. But multi-dimensional features will cause feature redundancy and affect the efficiency of diagnosis. Therefore, the purpose of JDA is to realize the dimensional-reduction and distribution optimization.

IV. FAULT DIAGNOSIS OF CLASSIFICATION BASED ON LEAST-SQUARES SUPPORT VECTOR MACHINE

The LSSVM algorithm is an excellent algorithm for fault diagnosis and classification. However, the two significant parameters, penalty factor and kernel function, greatly influence the effectiveness of LSSVM. Therefore, particle swarm optimization is applied to optimize the two parameters in LSSVM with the fitness of training accuracy.

A. A THE COOPERATIVE CLASSIFICATION METHOD BASED ON PSO AND LSSVM

The LSSVM used in this paper is widely used due to its high accuracy and precision in nonlinear signal processing. Compared with artificial neural networks, LSSVM can overcome the problems of long training time, random training results, and the lack of learning. Moreover, previous studies

have proposed a method to simplify the calculations, thereby greatly improving efficiency. However, there are two important parameters of kernel function and the penalty factor in LSSVM. These two parameters determine the calculation results of the LSSVM classification model. So, the LSSVM optimized by PSO is adopted in this paper, and the training accuracy is taken as fitness to find the optimal parameters of LSSVM. Then, the process of optimized LSSVM is shown in Figure 3.

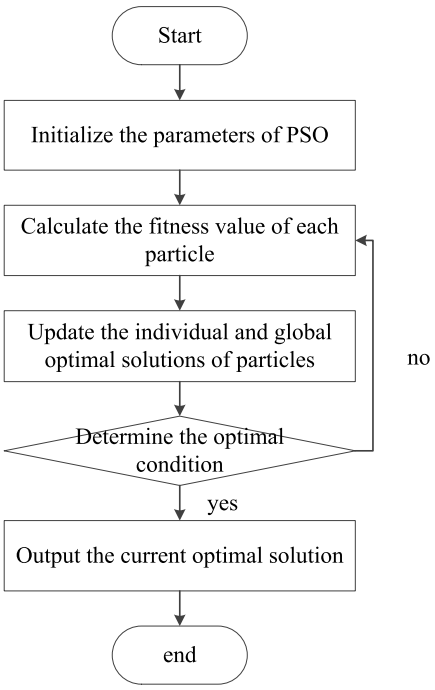

FIGURE 3. The process of PSO-LSSVM.

Based on Figure 3, the improved LSSVM could be described as follows.

Step 1: Input the optimized feature set into the classifier to train the optimal parameters.

Step 2: Initialize the parameters of the particle swarm optimization.

Step 3: Set the training accuracy as the fitness, and calculate the speed and position of each particle.

Step 4: Update the optimal solutions according to the fitness of the particle.

Step 5: Determine whether the results of the algorithm reach the optimal condition, and save the parameters under optimal condition.

Step 6: Apply the optimal parameters to classifier, and obtain training and test accuracy.

According to the above algorithm, the cooperative fault diagnosis method has been completely shown in figure 4. The flow of the whole diagnosis method is expressed as follows.

Step 1: The acceleration signal data is collected from the sensor.

Step 2: Use different structural elements to filter the acceleration and calculate the kurtosis of denoised.

Step 3: Find the optimal element type corresponding to the maximum kurtosis.

Step 4: Use the optimal structural element to filter the signal, and then decompose the denoised signal into several components by VMD.

Step 5: Extract the time-domain features from the signal components, and construct multidimensional feature sets.

Step 6: Divide the time-domain features sets into two datasets, the source domain and the target domain, both of which are input into kernel space and JDA.

Step 7: Use PSO-LSSVM as the classifier to calculate the accuracy of the optimized feature set.

V. SIMULATION AND EXPERIMENT ANALYSIS

To verify the reliability and effectiveness of the proposed diagnosis model, the experimental data is collected from the key laboratory of Guangdong Petrochemical Equipment. Figure 5 shows the fault diagnosis platform of the extruder and expansion dryer of the 100,000-ton butadiene unit. The rolling bearing has been tested on the extruder and expansion dryer and used as the fault diagnosis system in Maoming Branch of China Petrochemical Company.

Figure 6 shows the phenomenon of three different bearing faults. The inner crack and outer crack of the bearing are mainly caused by the shock or overloading of the rolling bearing during the operations. Moreover, the inner crack and outer crack, high temperature and overload working environment will also cause the ball bearing abrasion. The inner and outer crack of bearings in pressures deformation or small clearance will also produce fault phenomenon of rolling ball missing.

Table 2 illustrates the acceleration data under five different working conditions in key laboratory of Guangdong Petrochemical Equipment. Among them, the working conditions of A, B, C are constructed as the source domain dataset. D and F are constructed as the target domain dataset.

According to the algorithm proposed in this paper, next section will show the experiment results in stages.

TABLE 2. Bearing details under different working conditions.

Condition	Length(mm)	Speed(r/min)
A	1150	2870
B	1730	2980
C	1800	2970
D	2450	2980
E	2200	4800

Besides, the fault diagnosis method in this paper can also apply to multiple fault diagnosis. So, we obtain another experimental data of multiple faults, which were also collected from the key laboratory of Guangdong Petrochemical Equipment. Multiple fault diagnosis system is based on bearing and gearbox, and the multiple faults consist of bearing fault and gearbox fault. The specific construction of multiple faults is shown in Table 3.

In this paper, the experimental data collected from the sensor on fault diagnosis platform includes a normal state and

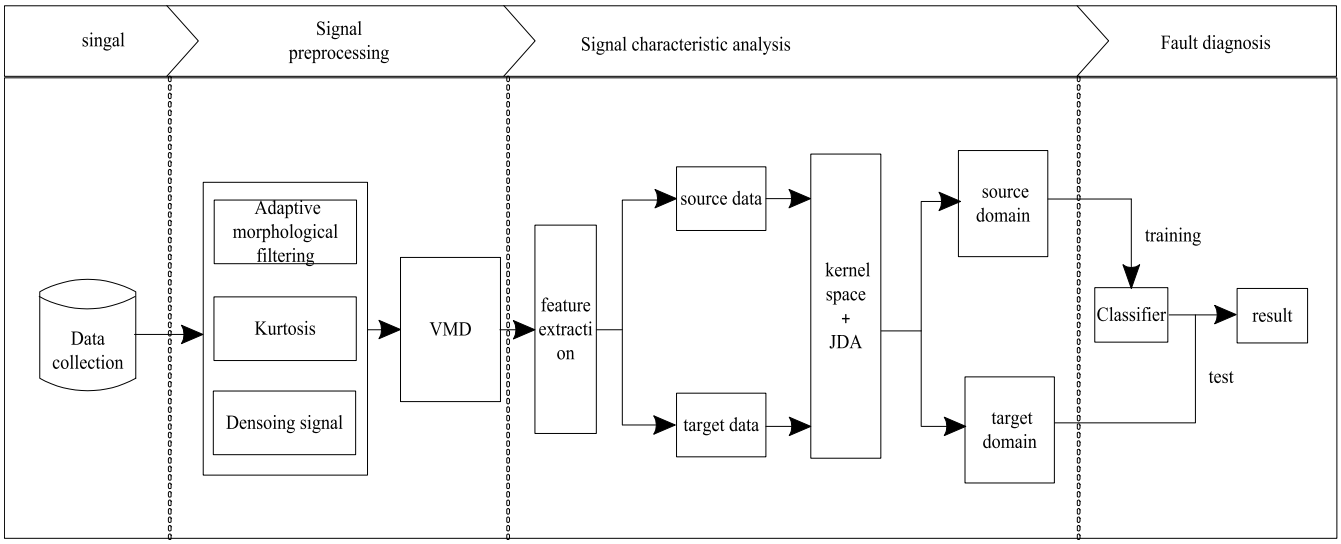

FIGURE 4. The flow of the fault diagnosis method.

FIGURE 5. The fault diagnosis system of extruder and expansion dryer of the 100,000-ton butadiene unit.

TABLE 3. Specific construction of multiple faults.

Condition	Bearing	Gearbox
1	Outer crack on the right side	Missing teeth
2	Missing bolls on the left side	Missing teeth
3	Outer crack on the left side	Missing teeth
4	Inner crack on the right side	Missing teeth

four failure states. The experimental signals in each state have the same number of samples and no missing sampling points.

A. THE SIGNAL PREPROCESSING BASED ON IMPROVED ADAPTIVE FILTERING AND VMD

Morphological difference operation is used to de-noising the signal. To select the optimal parameters of structural elements in morphology filtering, the height H of linear and triangular structural elements is 1,3,5,8, and the length L ranges from 1 to 40. The de-noising signals kurtosis under different structural elements types are calculated, and element types corresponding to the maximum kurtosis are selected as the optimal parameters. The kurtosis results are shown in Table 4 and Table 5.

TABLE 4. The optimal parameter of linear element under different bearing states.

State of fault bearing	height	length	kurtosis
normal	8	6	5286.907
Outer	8	5	49.821
inner	8	4	469.698
wearing	8	6	359.395
missing	8	5	119.222

TABLE 5. The optimal parameter of triangle element under different bearing states.

State of fault bearing	height	length	kurtosis
normal	8	6	4040.964
outer	8	7	96.190
inner	8	5	720.722
wearing	8	7	559.408
missing	8	7	220.294

According to Table 4-5, the kurtosis corresponding to triangular elements is larger than linear elements. Therefore, the results in Table 4 are taken as the optimal parameters in adaptive filtering. Next, the morphological difference operation of the structural element with the optimal parameter is used to de-noise the original signal. Taking the normal acceleration data as an example, Figure 7 shows the time domain figure of the original signal and the denoised signal, and it can be seen from the time domain diagram that the positive and negative pulses in the signal can be effectively attenuated. Figure 8 is the frequency domain of the original signal and the denoised signal of normal and external faults in two different states. From figure 8, we can see that the frequency fault features are relatively clear and orderly reduced. After signal adaptive de-noising, the denoised signal will be decomposed into several components by VMD, where $K = 3$.

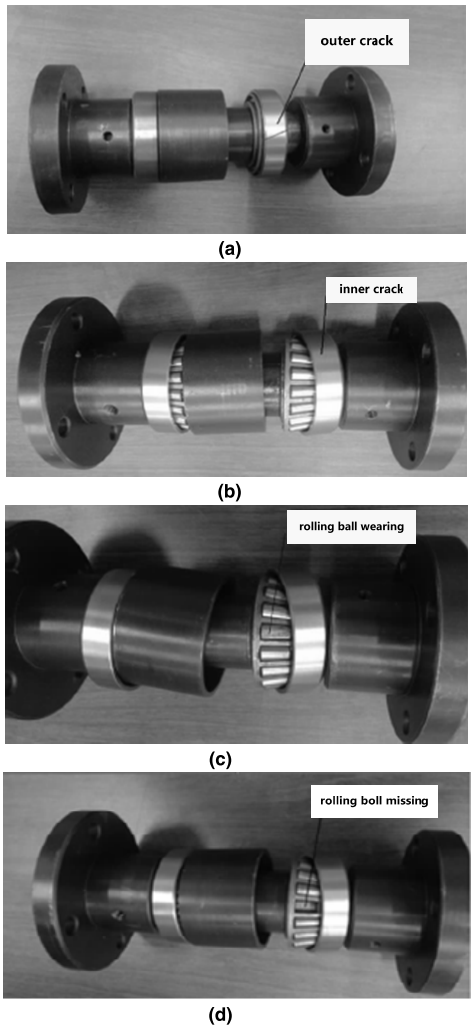

FIGURE 6. (a) Fault phenomenon of outer crack. (b) Fault phenomenon of inner crack. (c) Fault phenomenon of rolling ball wearing. (d) Fault phenomenon of rolling ball missing.

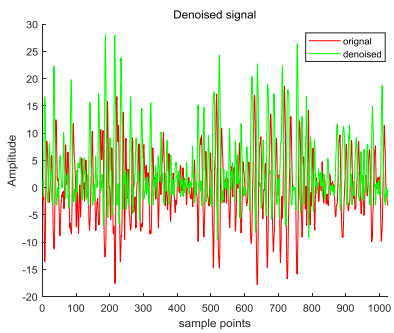

FIGURE 7. Time domain figure of original signal and denoised signal.

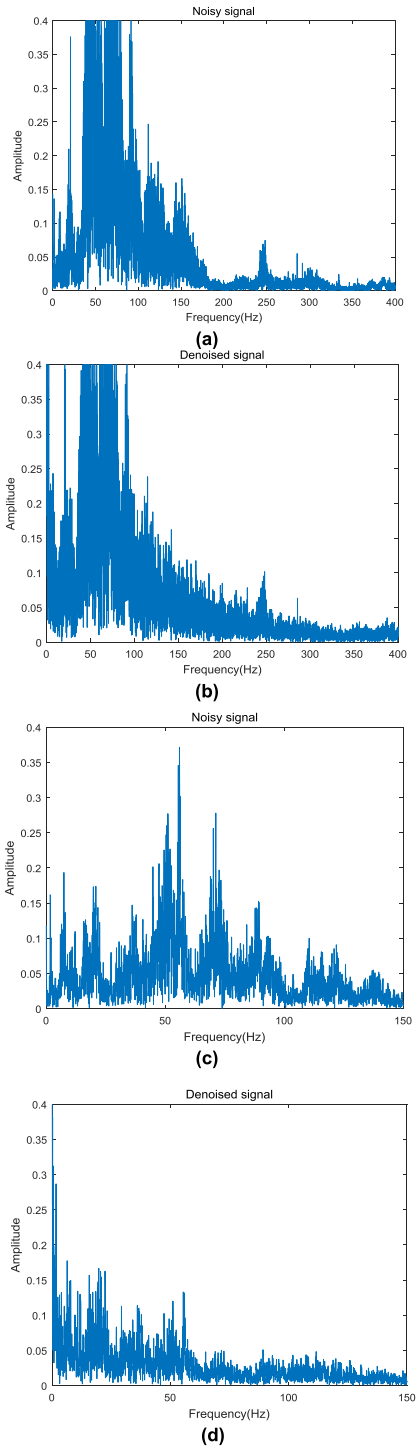

FIGURE 8. (a) Spectrum diagram of the normal signal. (b) Spectrum of the denoised normal signal. (c) Spectrum diagram of the external fault signal. (d) Spectrum of the external fault denoised signal.

B. THE FEATURE EXTRACTION AND OPTIMIZATION

To verify the effectiveness of feature extraction and structure optimization, time-domain features are extracted from each component and feature sets are directly constructed. Then, the feature set is divided into the source domain dataset of 900 features and the target domain dataset of 600 features. After using JDA algorithm to optimize the features, it will greatly improve the structure of the feature set.

From figure 9 and 10, it is hard to diagnose the bearing fault because the original feature distribution is so messy. But after optimizing the feature distribution by JDA, the discrimination (different states) and compactness (same states) are becoming

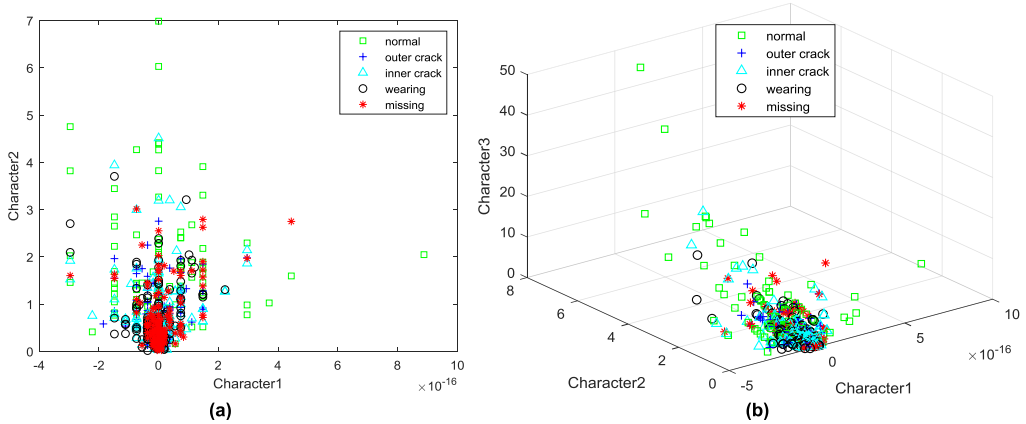

FIGURE 9. (a) Two-dimensional original feature distribution. (b) Three-dimensional original feature distribution.

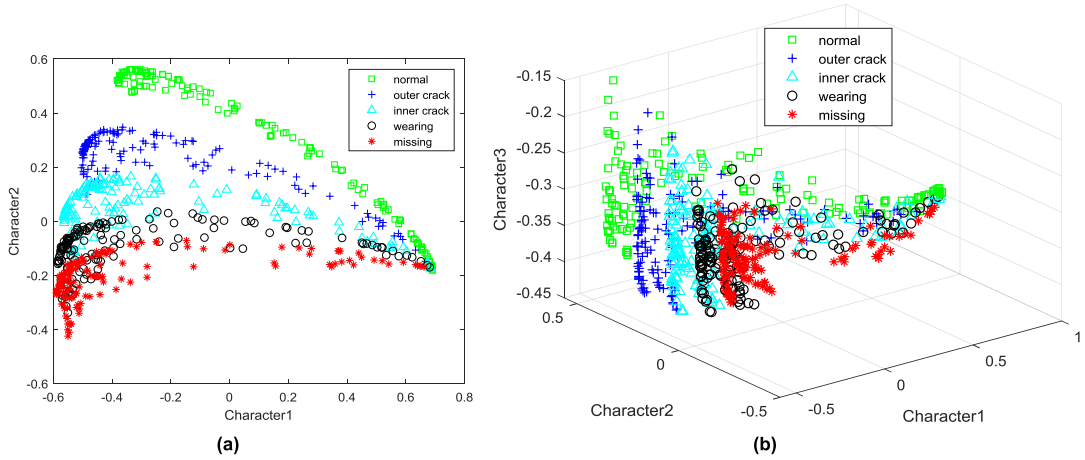

FIGURE 10. (a) Two-dimensional feature distribution after JDA. (b) Three-dimensional feature distribution after JDA.

more and more apparent. The figure of feature distribution can further prove the effectiveness of JDA.

In addition, some parameters of JDA need to be set in the proposed method, the regular parameter of λ is 0.95, the number of iterations is 10, the kernel type is linear, and the dimensions after adaptation is 10.

Compared with Figure 9 and Figure 10, it can be observed that the structure of the feature set will be greatly improved through the optimization of the JDA algorithm. Then, the optimized bearing feature data is input to the classifier for diagnosis.

C. THE COMPARISON ANALYSIS OF DIAGNOSIS METHOD

To verify the superiority of the signal preprocessing method, the denoised signal is compared with the original signal under the same signal decomposition and feature optimization experiment environment of the. Then the KNN is used as a classifier to diagnose the feature set of rolling bearings. The results of diagnostic accuracy based on KNN are as follows: the accuracy of the original signal is 91.0%; the accuracy of the denoised signal is 93.0%.

TABLE 6. The effectiveness of diagnosis model under different feature analyzation.

fault category	Fault analysis method	Training (%)	Test (%)
Single fault	TCA	99.5	96.2
	SSTCA	100	96.0
	JDA	100	97.8
Multiple fault	TCA	98.5	97.5
	SSTCA	100	99.5
	JDA	100	100

Especially, the PSO-LSSVM is used as a classifier in the comparison experiment, and the signal of each comparison experiment is decomposed into three components by VMD. The source domain dataset includes 900 features under three different working conditions, and the target domain dataset

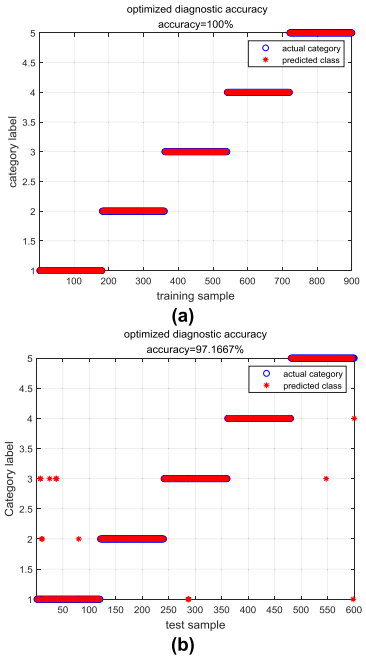

FIGURE 11. (a) Training accuracy of fault diagnosis model. (b) Test accuracy of fault diagnosis model.

contains 600 features under two different working conditions. Figure 11 shows the training and test fault diagnosis accuracy of the proposed method. All 900 features of the training experiment were correctly classified, while in the test experiment, 583 features were correctly classified and 17 features were misclassified.

TABLE 7. The superiority of the proposed diagnosis model.

fault category	Fault analysis method	Training	Test
Single fault	Original method	30.6%	18.0%
	General signal denoised	40.8%	18.8%
	Improve signal denoised	41.7%	20.0%
	Signal denoised and JDA	99.5%	96.2%
	Original signal and JDA	99.4%	95.8%
	Proposed method	100%	97.6%
Multiple fault	Original method	41.75%	16.5%
	General signal denoised	58.25	15.5%
	Improve signal denoised	58.75	20.0%
	Signal denoised and JDA	100%	99.5%
	Original signal and JDA	99.8%	99.5%
	Proposed method	100%	100%

Therefore, to analyze the specific superiority of improved adaptive filtering and JDA algorithm in the fault diagnosis model, some comparative experiments are carried out in terms of the signal de-noising and feature structure optimization in this paper. Table 6 and Table 7 show the results of comparative experiments. Among them, Table 6 analyzes the diagnosis efficiency of fault diagnosis models under different feature analysis models. Compared with the other feature distribution optimization algorithms, JDA can optimize the feature distribution effectively and has high accuracy in both training and test results.

Table 7 illustrates the necessity of the improved method. From the experimental results of two fault conditions, the accuracy of the fault diagnosis model is much improved after optimizing the feature structure by JDA. So, the effectiveness of JDA can be proved once again. Besides, the diagnostic accuracy of the de-noising signal directly confirmed that the improved signal de-noising method could effectively eliminate the influence of noise on fault diagnosis accuracy.

The same acceleration data collected from the Key Laboratory of Guangdong Petrochemical Equipment were used to prove the superiority of the diagnosis method proposed in this paper. Table 8 shows the accuracy of different methods. Where, a variety of different signal decomposition algorithms, signal denoising algorithms, and classification algorithms are adopted to compare with the method in this paper. It is evident that the classification and decomposition algorithm in this paper has substantial advantages. The results show that the proposed method has higher accuracy and efficiency than the previous methods.

TABLE 8. The superiority of the proposed diagnosis model.

Fault analysis method	Accuracy (%)
EMD	66.7
EEMD	86.6
BP	84.2
LMS	75.0
SVM	83.3
KNN	81.7
LSSVM	89.6

VI. CONCLUSION

In summary, this paper proposed a cooperative method for bearing fault diagnosis, which aims to explore the optimal parameters in morphology filtering and optimize the structure of the feature set. Kurtosis is applied to adaptive filtering for parameter optimization, and then JDA optimizes the marginal and conditional distribution of the bearing dataset. Through the algorithm discrimination and comparison experiment, the main conclusions are as follows.

(1) Because of the difficulty in selecting the width and height of structural elements, kurtosis is used as the index to measure the superiority of parameters of length and height in structural elements. The experimental results show that the superiority of the improved adaptive filtering lies in

the greatly improved diagnosis accuracy of the signal after de-noising the signal through the improved adaptive filtering.

(2) JDA is applied to carry out transfer learning on features under different working conditions and optimize the condition distribution and marginal distribution of the features. Moreover, compared with TCA and SSTCA, the comparative experiments show that the diagnosis accuracy of the feature set, optimized by JDA is higher than the above two algorithms.

(3) Compared with other classical diagnostic models, the diagnostic model in this paper can make full use of the features under different working conditions, and it can also solve the problems of lacking sample and unbalance distribution. In addition, the distribution of an optimized feature set can improve the diagnosis accuracy of the rolling bearings. The training and test accuracy of the proposed method is 100% and 97.8%, respectively. Therefore, the effectiveness and superiority of the proposed fault diagnosis method are proved for rolling bearings.

The method proposed in this paper discussed the problem of parameter optimization in adaptive filtering and feature optimization. Although the proposed method has greatly improved the accuracy of rolling bearings fault diagnosis, two serious problems need to be solved immediately. On the one hand, in feature extraction, it is found that there are some redundant and unnecessary features in the time domain feature set, which affects the efficiency of the diagnostic model. On the other hand, JDA has significantly improved the distribution of datasets. However, it also takes more time. Therefore, the following research is based on how to apply the feature selection method and improve the efficiency of JDA. In addition, the cause of the bearing fault from its fault location or the variation trend of fault degree should be considered in the next step.

ACKNOWLEDGMENT

The authors would like to thank the Guangdong Institute of Petrochemical Technology for providing the rolling bearing data in this article.

REFERENCES

[1] D. Huang, L. Ke, B. Mi, G. Wei, J. Wang, and S. Wan, "A cooperative denoising algorithm with interactive dynamic adjustment function for security of stacker in industrial Internet of Things," *Secur. Commun. Netw.*, vol. 2019, Feb. 2019, Art. no. 4049765.

[2] N. Liu, Y. Sun, Y. Wang, H. Guo, B. Gao, T. Feng, and P. Sun, "Active control for vehicle interior noise using the improved iterative variable step-size and variable tap-length LMS algorithms," *Noise Control Eng. J.*, vol. 67, no. 6, pp. 405–414, Nov. 2019.

[3] J. Yu, T. Hu, and H. Liu, "A new morphological filter for fault feature extraction of vibration signals," *IEEE Access*, vol. 7, pp. 53743–53753, May 2019.

[4] G. J. Tang, X. L. Yan, and X. L. Wang, "Chaotic signal denoising based on adaptive smoothing multiscale morphological filtering," *Complexity*, vol. 2020, Feb. 2020, Art. no. 7242943.

[5] Y. Xu, C. Fan, Q.-X. Zhu, A. Rajabifard, N. Chen, Y. Chen, and Y.-L. He, "Novel pattern-matching integrated KCVA with adaptive rank-order morphological filter and its application to fault diagnosis," *Ind. Eng. Chem. Res.*, vol. 59, no. 4, pp. 1619–1630, Jan. 2020.

[6] H. Zhang and Q. He, "Tacholess bearing fault detection based on adaptive impulse extraction in the time domain under fluctuant speed," *Meas. Sci. Technol.*, vol. 31, no. 7, Jul. 2020, Art. no. 074004.

[7] B. Sun, Y. Li, Z. Wang, Y. Ren, Q. Feng, D. Yang, M. Lu, and X. Chen, "Remaining useful life prediction of aviation circular electrical connectors using vibration-induced physical model and particle filtering method," *Microelectron. Rel.*, vol. 92, pp. 114–122, Jan. 2019.

[8] Y. Huang, J. Lin, Z. Liu, and C. Huang, "A morphological filtering method based on particle swarm optimization for railway vehicle bearing fault diagnosis," *Shock Vib.*, vol. 2019, Dec. 2019, Art. no. 2593673.

[9] J. Oh and H. Hwang, "Feature enhancement of medical images using morphology-based homomorphic filter and differential evolution algorithm," *Int. J. Control, Autom. Syst.*, vol. 8, no. 4, pp. 857–861, Aug. 2010.

[10] M.-L. Tian and J.-M. Yang, "Optimization of structural element in froth image denoising of coal flotation," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 30, no. 4, May 2016, Art. no. 1654002.

[11] W. Fu, K. Shao, J. Tan, and K. Wang, "Fault diagnosis for rolling bearings based on composite multiscale fine-sorted dispersion entropy and SVM with hybrid mutation SCA-HHO algorithm optimization," *IEEE Access*, vol. 8, pp. 13086–13104, Jan. 2020.

[12] X. Chen, Y. Yang, Z. Cui, and J. Shen, "Wavelet denoising for the vibration signals of wind turbines based on variational mode decomposition and multiscale permutation entropy," *IEEE Access*, vol. 8, pp. 40347–40356, Mar. 2020.

[13] Y. Miao, M. Zhao, Y. Yi, and J. Lin, "Application of sparsity-oriented VMD for gearbox fault diagnosis based on built-in encoder information," *ISA Trans.*, vol. 99, pp. 496–504, Apr. 2020.

[14] W. Liu, W. Chen, and Z. Zhang, "A novel fault diagnosis approach for rolling bearing based on high-order synchrosqueezing transform and detrended fluctuation analysis," *IEEE Access*, vol. 8, pp. 12533–12541, Jan. 2020.

[15] O. W. Samuel, H. Zhou, X. Li, H. Wang, H. Zhang, A. K. Sangaiah, and G. Li, "Pattern recognition of electromyography signals based on novel time domain features for amputees' limb motion classification," *Comput. Electr. Eng.*, vol. 67, pp. 646–655, Apr. 2018.

[16] Z. Huang, J. Zhu, J. Lei, X. Li, and F. Tian, "Tool wear predicting based on multi-domain feature fusion by deep convolutional neural network in milling operations," *J. Intell. Manuf.*, vol. 31, no. 4, pp. 953–966, Apr. 2020.

[17] K. Kaplan, Y. Kaya, M. Kuncan, M. R. Minaz, and H. M. Ertunç, "An improved feature extraction method using texture analysis with LBP for bearing fault diagnosis," *Appl. Soft Comput.*, vol. 87, Feb. 2020, Art. no. 106019.

[18] J. Tian, Y.-T. Ai, C.-W. Fei, F.-L. Zhang, and Y.-S. Choy, "Dynamic modeling and simulation of inter-shaft bearings with localized defects excited by time-varying displacement," *J. Vib. Control*, vol. 25, no. 8, pp. 1436–1446, Apr. 2019.

[19] C.-W. Fei, Y.-S. Choy, G.-C. Bai, and W.-Z. Tang, "Multi-feature entropy distance approach with vibration and acoustic emission signals for process feature recognition of rolling element bearing faults," *Struct. Health Monitor.-Int. J.*, vol. 17, no. 2, pp. 156–168, Mar. 2018.

[20] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Trans. Neural Netw.*, vol. 22, no. 2, pp. 199–210, Feb. 2011.

[21] T. Grubinger, A. Birlutiu, H. Schöner, T. Natschläger, and T. Heskes, "Multi-domain transfer component analysis for domain generalization," *Neural Process. Lett.*, vol. 46, no. 3, pp. 845–855, Dec. 2017.

[22] X. Gu, F. Deng, X. Gao, and R. Zhou, "An improved sensor fault diagnosis scheme based on TA-LSSVM and ECOC-SVM," *J. Syst. Sci. Complex.*, vol. 31, no. 2, pp. 372–384, Apr. 2018.

[23] W. Fu, K. Wang, C. Zhang, and J. Tan, "A hybrid approach for measuring the vibrational trend of hydroelectric unit with enhanced multi-scale chaotic series analysis and optimized least squares support vector machine," *Trans. Inst. Meas. Control*, vol. 41, no. 15, pp. 4436–4449, Nov. 2019.

[24] Y. Liu, H. Xu, Z. Xia, and Z. Gong, "Multi-spectrometer calibration transfer based on independent component analysis," *Analyst*, vol. 143, no. 5, pp. 1274–1280, 2018.

[25] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer feature learning with joint distribution adaptation," in *Proc. IEEE IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2200–2207.

[26] W. Qian, S. Li, P. Yi, and K. Zhang, "A novel transfer learning method for robust fault diagnosis of rotating machines under variable working conditions," *Measurement*, vol. 138, pp. 514–525, May 2019.

[27] Z.-H. Liu, B.-L. Lu, H.-L. Wei, X.-H. Li, and L. Chen, "Fault diagnosis for electromechanical drivetrains using a joint distribution optimal deep domain adaptation approach," *IEEE Sensors J.*, vol. 19, no. 24, pp. 12261–12270, Dec. 2019.
[28] T. Han, C. Liu, W. Yang, and D. Jiang, "Deep transfer network with joint distribution adaptation: A new intelligent fault diagnosis framework for industry application," *ISA Trans.*, vol. 97, pp. 269–281, Feb. 2020.
[29] Z. Wu, H. Jiang, K. Zhao, and X. Li, "An adaptive deep transfer learning method for bearing fault diagnosis," *Measurement*, vol. 151, Feb. 2020, Art. no. 107227.
[30] M. Zhang, A. McDaid, A. J. Veale, Y. Peng, and S. Q. Xie, "Adaptive trajectory tracking control of a parallel ankle rehabilitation robot with joint-space force distribution," *IEEE Access*, vol. 7, pp. 85812–85820, Jul. 2019.
[31] S. Zhang, S. Zhang, B. Wang, and T. G. Habetler, "Deep learning algorithms for bearing fault diagnostics—A comprehensive review," *IEEE Access*, vol. 8, pp. 29857–29881, Feb. 2020.
[32] M. Kuncan, K. Kaplan, M. R. Minaz, Y. Kaya, and H. M. Ertunç, "A novel feature extraction method for bearing fault classification with one dimensional ternary patterns," *ISA Trans.*, vol. 100, pp. 346–357, May 2020.
[33] J. Tian, L. Liu, F. Zhang, Y. Ai, R. Wang, and C. Fei, "Multi-domain entropy-random forest method for the fusion diagnosis of inter-shaft bearing faults with acoustic emission signals," *Entropy*, vol. 22, no. 1, p. 57, Jan. 2019.
[34] J. Zhang, Z. Yin, P. Chen, and S. Nichele, "Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review," *Inf. Fusion*, vol. 59, pp. 103–126, Jul. 2020.
[35] X. Song, J. Zhao, J. Song, F. Dong, L. Xu, and J. Zhao, "Local demagnetization fault recognition of permanent magnet synchronous linear motor based on S-transform and PSO–LSSVM," *IEEE Trans. Power Electron.*, vol. 35, no. 8, pp. 7816–7825, Aug. 2020.
[36] W. Lu, Q. Peng, Y. Cui, Z. Huang, J. Teng, and W. Hu, "Structural response estimation method based on particle swarm optimisation/support vector machine and response correlation characteristics," *Measurement*, vol. 160, Aug. 2020, Art. no. 107810.
[37] A. Bemani, Q. Xiong, A. Baghban, S. Habibzadeh, A. H. Mohammadi, and M. H. Doranehgard, "Modeling of cetane number of biodiesel from fatty acid methyl ester (FAME) information using GA-, PSO-, and HGAPSO-LSSVM models," *Renew. Energy*, vol. 150, pp. 924–934, May 2020.
[38] B. Sun, X. Jiang, K.-C. Yung, J. Fan, and M. G. Pecht, "A review of prognostic techniques for high-power white LEDs," *IEEE Trans. Power Electron.*, vol. 32, no. 8, pp. 6338–6362, Aug. 2017.
[39] J. Cui, R. Yu, D. Zhao, J. Yang, W. Ge, and X. Zhou, "Intelligent load pattern modeling and denoising using improved variational mode decomposition for various calendar periods," *Appl. Energy*, vol. 247, pp. 480–491, Aug. 2019.
[40] B. Sun, Y. Li, Z. Wang, Z. Li, Q. Xia, Y. Ren, Q. Feng, D. Yang, and C. Qian, "Physics-of-failure and computer-aided simulation fusion approach with a software system for electronics reliability analysis," *Eksploatacja Niezawodnosc-Maintenance Rel.*, vol. 22, no. 2, pp. 340–351, Feb. 2020.

ZHANG XU was born in 1994. She received the B.S. degree in communication engineering from Chongqing Jiaotong University, Chongqing, China, in 2016, where she is currently pursuing the master's degree with the College of Information Science and Engineering. Her research interests include fault diagnosis and fault-tolerant control of dynamic systems, analysis and design of complex systems, and reliability engineering.

HUANG DARONG (Member, IEEE) was born in 1978. He received the B.S. degree in applied mathematics from the Hubei National Institute, Hubei, China, in 2000, the M.S. degree in applied mathematics from Liaoning University, Liaoning, China, in 2003, and the Ph.D. degree in control theory and control engineering from Chongqing University, Chongqing, China, in 2006.

Since 2011, he has been a Professor with the College of Information Science and Engineering, Chongqing Jiaotong University, Chongqing. His research interests include fault diagnosis and fault-tolerant control of dynamic systems, analysis and design of complex systems, big data analysis of transport systems, reliability engineering, and so on.

GUOXI SUN received the B.Sc. degree from the Department of Precision Instrument and Mechanology, Tianjin University, China, in 1992, and the M.Sc. and Ph.D. degrees from the South China University of Technology, in 1999 and 2006, respectively.

Since 2015, he has been a Professor with the Department of Computer and Electronic Information, Guangdong University of Petrochemical Technology. His main research interests include control algorithms, artificial intelligence, and fault diagnosis.

WANG YONGCHAO received the B.S. degree in electrical engineering and automation from Beijing Information Science and Technology University, Beijing, China, in 2011, and the M.S. degree in control theory and control engineering from the University of Science and Technology Beijing, Beijing, in 2015.

Since 2015, he has been with the Xinjiang Electric Power Research Institute, State Grid Xinjiang Electric Power Company Ltd. His main research interests include control theory and engineering and electric energy metering technical work.