

Research Article

Bearing Fault Diagnosis Based on Improved Locality-Constrained Linear Coding and Adaptive PSO-Optimized SVM

Haodong Yuan, Jin Chen, and Guangming Dong

State Key Laboratory of Mechanical System and Vibration, Shanghai Jiao Tong University, Shanghai 200240, China

Correspondence should be addressed to Guangming Dong; gmdong@sjtu.edu.cn

Received 28 April 2017; Accepted 6 July 2017; Published 16 August 2017

Academic Editor: Damijan Markovic

Copyright © 2017 Haodong Yuan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A novel bearing fault diagnosis method based on improved locality-constrained linear coding (LLC) and adaptive PSO-optimized support vector machine (SVM) is proposed. In traditional LLC, each feature is encoded by using a fixed number of bases without considering the distribution of the features and the weight of the bases. To address these problems, an improved LLC algorithm based on adaptive and weighted bases is proposed. Firstly, preliminary features are obtained by wavelet packet node energy. Then, dictionary learning with class-wise K-SVD algorithm is implemented. Subsequently, based on the learned dictionary the LLC codes can be solved using the improved LLC algorithm. Finally, SVM optimized by adaptive particle swarm optimization (PSO) is utilized to classify the discriminative LLC codes and thus bearing fault diagnosis is realized. In the dictionary learning stage, other methods such as selecting the samples themselves as dictionary and *K*-means are also conducted for comparison. The experiment results show that the LLC codes can effectively extract the bearing fault characteristics and the improved LLC outperforms traditional LLC. The dictionary learned by class-wise K-SVD achieves the best performance. Additionally, adaptive PSO-optimized SVM can greatly enhance the classification accuracy comparing with SVM using default parameters and linear SVM.

1. Introduction

Rolling bearings have been widely used in rotating machinery. However, rolling bearings easily suffer faults due to their structural characteristics and serious mechanical failure maybe occurs. Hence, it is very essential to conduct fault diagnosis of rolling bearings. Vibration signals are usually used to monitor the bearing operating conditions. Feature extraction and pattern recognition are two crucial steps for machinery fault diagnosis using vibration signals. As for feature extraction of mechanical vibration signals, many methods based on the time domain, frequency domain, or time-frequency analysis have been proposed, such as envelope analysis [1], short time Fourier transform [2], wavelet transform [3, 4], Wigner-Ville distribution (WVD) [5], empirical mode decomposition [6], and Hilbert-Huang transform [7].

Sparse representation is more adaptive to represent a signal based on redundant dictionary comparing with the aforementioned signal processing methods, which has been

widely applied in the field of image processing, for example, image denoising [8], image compression [9], and face recognition [10]. Recently, many researchers pay attention to the methods of machinery fault diagnosis based on sparse representation. Feng and Chu implemented matching pursuit for gear damage detection [11]. Liu et al. proposed a feature extraction scheme with sparse coding and realized bearing fault diagnosis [12]. Tang et al. used shift-invariant dictionary learning method to decompose the vibration signal into a series of latent components and then detected the bearing and gear faults [13]. Chen et al. proposed a new method, namely, Sparse Extraction of Impulse by Adaptive Dictionary (SpaEIAD), to extract impulse components of vibration signals with heavy noise and detected the gearbox fault using the proposed method [14].

As shown in [15], locality is more essential than sparsity. Based on [15], locality-constrained linear coding (LLC) algorithm is proposed to reduce computation cost, which employed the locality constraints to map feature in its local

coordinate system and ensure that similar features have similar codes [16]. LLC has been widely used for feature extraction of images such as WCE image [17], lepidopteran images [18], CT Images [19], and facial image [20]. In the area of machinery fault diagnosis, LLC has also attracted extensive attention. Wang and Liu employed LLC for extracting the characteristics of time–frequency images of bearing vibration signals [21]. Li et al. utilized locality-constrained sparse coding method for bearing fault diagnosis and proved that locality-constrained sparse coding is effective for the feature extraction of bearing fault [22]. However, in traditional LLC each feature is encoded with a fixed number of bases, which may produce weak responses in the low-density area and poor representation in the high-density area. To tackle this problem, adaptive number of bases can be found based on the distribution of the features [17]. Moreover, in traditional LLC the weight information of the bases is ignored. Weight of the bases can be used to boost the discriminative power of LLC codes [23]. To address the two problems, an improved LLC algorithm based on adaptive and weighted bases is proposed in this paper.

Proper dictionary is very important for sparse representation, which should reflect the latent characteristic patterns of signals so as to obtain better representation. The dictionary can be predefined; however it is always difficult to know the prior knowledge of the signal in practical application. The dictionary can also be constructed by selecting the training samples directly. Nevertheless, when dealing with large-scale dataset the performance of this method is usually not good so that dictionary learning method that learns a small-sized dictionary is very essential. Some dictionary learning methods have been proposed, such as K -means [24], K -means singular value decomposition (K-SVD) [25], method of optimal directions (MOD) [26], and shift-invariant sparse coding [27]. K-SVD algorithm has proved to be very effective in image compression and denoising [8, 28]. Feature extraction methods based on K-SVD algorithm have been employed for machinery fault diagnosis recently. Feng and Liang exploited shift-invariant K -means singular value decomposition (SI-K-SVD) to extract the components of planetary gearbox signals; subsequently background noise is suppressed and the vibration patterns are obtained [29]. In this study, K-SVD algorithm is applied to learn the dictionary for solving LLC codes.

With regard to the intelligent diagnosis of mechanical fault, it is essentially a pattern recognition problem. Many pattern recognition methods have been employed for mechanical fault diagnosis, such as artificial neural network (ANN) [30], support vector machine (SVM) [31–33], Bayesian classifiers [34], and hidden Markov model (HMM) [35]. SVM is a popular machine learning method based on statistical learning theory [36], which employs structural risk minimization principle, and the problems of overfitting, local minimum, and slow convergence speed can be overcome compared with neural network. Moreover, when dealing with small training samples and high-dimensional data SVM is more suitable. If the data cannot be linearly separated, a kernel function can be used in SVM to map the data into high-dimensional feature space. Linear kernel, polynomial kernel, and radial basis function (RBF) kernel are commonly used kernel functions.

The penalization coefficient and the parameter of kernel function have a great influence on the performance of SVM. The optimization algorithm particle swarm optimization (PSO) [37] can be employed to optimize the parameters of SVM. SVM with PSO-optimized parameters has been widely applied for machinery fault diagnosis. Liu et al. proposed a hybrid intelligent diagnosis method of bearing fault based on empirical model decomposition (EMD), distance evaluation technique, and wavelet support vector machine (WSVM) with PSO [32]. Standard PSO algorithm easily falls into the local optimum and the convergence is slow. To this end, adaptive PSO algorithm with adaptive weight and acceleration coefficients is adopted for the optimization of the parameters of SVM in this paper [38].

In our work, a novel method based on improved LLC algorithm and adaptive PSO-optimized SVM is proposed for bearing fault diagnosis. Firstly, preliminary feature is extracted from the bearing vibration signals under different operating conditions. Then, class-wise K-SVD algorithm is employed to learn the dictionary for solving LLC codes. Afterwards, based on the learned dictionary LLC codes can be solved using the improved LLC algorithm and employed as the input feature vectors of SVM. Finally, adaptive PSO-optimized SVM is utilized to distinguish different operating conditions of bearings and hence the bearing fault diagnosis is realized.

The paper is organized as follows: in Section 2 the feature extraction method based on improved LLC is presented. In Section 3 the classification method based on adaptive PSO-optimized SVM is described. In Section 4 the summary of the proposed bearing fault diagnosis method is demonstrated. An experiment of rolling bearings with artificial fault is conducted to validate the proposed method in Section 5. Finally, the conclusions are drawn in Section 6.

2. Feature Extraction Based on Improved Locality-Constrained Linear Coding

Firstly, feature extraction based on traditional methods is employed to acquire the preliminary feature. Then two stages including dictionary learning and LLC codes solving are conducted for feature extraction based on sparse LLC codes which can be regarded as high level sparse features.

2.1. Preliminary Feature Based on Wavelet Packet Node Energy. Wavelet packet node energy (WPNE) is adopted for the preliminary feature extraction [34, 39]. Wavelet transform is very efficient for local time–frequency analysis of nonstationary signals. Based on discrete wavelet transform (DWT), discrete wavelet packet transform (DWPT) continues to decompose the detailed coefficients of DWT to finer frequency bands.

Denote a signal $\mathbf{x} \in R^N$, wavelet packet transform is conducted, and we can get the wavelet packet coefficients $W_{j,n}$ corresponding to wavelet node (j, n) , where j is the decomposition level and $n = 0, \dots, 2^{j-1}$ is the node number with the same level. At time t the coefficient of node (j, n) is denoted as $W_{j,n,t}$, where $t = 0, 1, \dots, N/2^j - 1$. The reconstructed signal obtained by node (j, n) is $X_{j,n,k}$, where

$k = 0, 1, \dots, N - 1$. The energy of the signal can be expressed by the square of 2-norm of signal. The energy of wavelet packet node (j, n) and corresponding reconstructed signal $\mathbf{X}_{j,n}$ is equal, which is denoted as $E_{j,n}$:

$$E_{j,n} = \|\mathbf{W}_{j,n}\|^2 = \|\mathbf{X}_{j,n}\|^2. \quad (1)$$

And the total energy E of wavelet packet nodes corresponding to depth j is equal to the energy of signal \mathbf{x} ; that is,

$$E = \|\mathbf{x}\|^2 = \sum_{n=0}^{2^j-1} \|\mathbf{W}_{j,n}\|^2. \quad (2)$$

The wavelet packet node (j, n) corresponds to the frequency range $[f_s \cdot n/2^{j+1}, f_s \cdot (n + 1)/2^{j+1}]$, where f_s is sampling frequency, so we can obtain the energy distribution $p_{j,n}$ corresponding to the frequency domain, which is the relative wavelet packet node energy:

$$p_{j,n} = \frac{E_{j,n}}{E}. \quad (3)$$

In this paper the vibration signal is decomposed with 3-level DWPT and eight wavelet packet nodes corresponding to different frequency bands are produced. The mother wavelet Daubechies 2 (db2) is used since it is effective for bearing fault diagnosis [4]. Feature vectors can be formulated as follows:

$$\mathbf{T} = [p_0, p_1, p_2, p_3, p_4, p_5, p_6, p_7]. \quad (4)$$

In this way an 8-dimensional feature vector is constructed for each signal, which is used for further feature extraction as shown below.

2.2. Dictionary Learning. Proper dictionary selection is important for sparse features, which should fully reflect the characteristics of signals in order to better extract sparse features. However, predefined dictionaries such as Dirac dictionary, Fourier dictionary, wavelet dictionary, and Gabor dictionary usually cannot match the characteristics of signals well. When the dataset is getting large, the methods of using all the training samples directly or randomly selecting some training samples as dictionary may not perform well so learning a dictionary with small size which can better capture the signal characteristics is essential. Assume that there are a set of signals $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in R^{n \times N}$; dictionary learning will learn an optimal dictionary \mathbf{D} with specified K dictionary atoms which is adaptive to the signals. More generally, we should optimize both dictionary \mathbf{D} and sparse coefficients \mathbf{S} using sparsity constraint of l_0 norm as follows:

$$\begin{aligned} (\mathbf{D}, \mathbf{S}) = \arg \min_{\mathbf{D}, \mathbf{S}} \quad & \|\mathbf{X} - \mathbf{D}\mathbf{S}\|_F^2 \\ \text{s.t.} \quad & \|\mathbf{s}_i\|_0 \leq T, \quad \forall i, \end{aligned} \quad (5)$$

where \mathbf{s}_i is sparse coefficient set corresponding to the signal \mathbf{x}_i ; \mathbf{S} are sparse coefficients corresponding to all signals. T is the sparsity prior which means the nonzero number in \mathbf{s}

should not be more than T . The notation F stands for the Frobenius norm.

We can also use sparsity constraint of l_1 norm as follows:

$$(\mathbf{D}, \mathbf{S}) = \arg \min_{\mathbf{D}, \mathbf{S}} \quad \|\mathbf{X} - \mathbf{D}\mathbf{S}\|_F^2 + \beta \sum_{i=1}^N \|\mathbf{s}_i\|_1, \quad (6)$$

where β is a scalar controlling the contribution of the sparsity term.

In (5) and (6) there are two issues.

(1) *Coefficients Solving.* To solve the sparse coefficients with a specified dictionary, there are many algorithms for approximately doing this, including l_1 norm based basis pursuit (BP) [40], l_0 norm based matching pursuit (MP) [41], and orthogonal matching pursuit (OMP) [42]. In this paper, we employ the OMP method for solving the sparse coefficients because OMP performs better and runs faster.

(2) *Dictionary Learning.* In order to find a dictionary \mathbf{D} which best matches the structures in the signals, some algorithms have been developed for dictionary learning, for example, K-means [24], K-SVD [25], MOD [26], and shift-invariant sparse coding [27].

Through dictionary learning, the dictionary atoms can capture the characteristic patterns of the signals. Subsequently, the signals can be sparsely represented based on the learned dictionary. The nonzero coefficient represents activation of a characteristic pattern; thus the characteristic pattern of the signal is more explicit than preliminary characteristic. Hence, sparse codes can be taken as high level representations of the signals.

A redundant dictionary can be learned based on K-SVD algorithm. K-SVD algorithm that solves (5) can learn a dictionary with small size based on the training samples. A two-step iterative algorithm using singular value decomposition is utilized for the optimization of object function demonstrated in the above equation. Finally a small-sized dictionary which best suits the training samples can be learned with K-SVD algorithm [25].

Using K-SVD algorithm, we can learn dictionary using the samples that contain all classes or learn a subdictionary separately based on the samples of each class, which is class-wise K-SVD. For class-wise K-SVD, assuming that there are L classes of signals, the learned dictionary corresponding to each class named subdictionary $\{\mathbf{D}_i \mid i = 1, 2, \dots, L\}$ (where i indicates the class label) can be obtained with K-SVD algorithm. Each subdictionary contains k atoms. Then a whole redundant dictionary \mathbf{D} with K atoms can be formulated by concatenating the subdictionaries:

$$\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_L]. \quad (7)$$

2.3. Locality-Constrained Linear Coding. Assuming that the signal set $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in R^{n \times N}$, after the dictionary $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_K] \in R^{n \times K}$ has been learned, sparse codes of the training samples and test samples can be solved by (5) or (6). However, since locality certainly leads to sparsity while sparsity cannot necessarily lead to locality, locality is

more essential than sparsity [15]. Therefore, in LLC locality constraint is substituted for sparsity constraint as follows [16]:

$$\begin{aligned} \min_{\mathbf{s}} \quad & \sum_{i=1}^N \|\mathbf{x}_i - \mathbf{D}\mathbf{s}_i\|^2 + \lambda \|\mathbf{I}_i \odot \mathbf{s}_i\|^2 \\ \text{s.t.} \quad & \mathbf{1}^T \mathbf{s}_i = 1, \quad \forall i, \end{aligned} \quad (8)$$

where \odot represents element-wise multiplication. \mathbf{I}_i denotes the locality adaptor, which allows different freedom for each base proportional to its similarity to the signal \mathbf{x}_i :

$$\mathbf{I}_i = \exp\left(\frac{\text{dist}(\mathbf{x}_i, \mathbf{D})}{\sigma}\right), \quad (9)$$

where $\text{dist}(\mathbf{x}_i, \mathbf{D}) = [\text{dist}(\mathbf{x}_i, \mathbf{d}_1), \dots, \text{dist}(\mathbf{x}_i, \mathbf{d}_K)]^T$ and $\text{dist}(\mathbf{x}_i, \mathbf{d}_j)$ is the Euclidean distance between \mathbf{x}_i and \mathbf{d}_j . The locality adaptor \mathbf{I}_i is further normalized to $(0, 1]$ by subtracting $\max(\text{dist}(\mathbf{x}_i, \mathbf{D}))$ from $\text{dist}(\mathbf{x}_i, \mathbf{D})$. The parameter σ is utilized to adjust the weight decay speed of the locality adaptor. The constraint $\mathbf{1}^T \mathbf{s}_i = 1$ means the requirements of shift invariance of the LLC codes. The parameter λ denotes the weight of locality constraint term.

Comparing with sparse codes solved by (5) or (6), LLC codes own the property of local smooth sparsity so that similar codes can be produced for similar signals. In addition, (8) has analytical solution for each signal \mathbf{x}_i :

$$\begin{aligned} \tilde{\mathbf{s}}_i &= (\mathbf{C}_i + \lambda \text{diag}(\mathbf{I}_i)) \setminus \mathbf{1} \\ \mathbf{s}_i &= \frac{\tilde{\mathbf{s}}_i}{\mathbf{1}^T \tilde{\mathbf{s}}_i}, \end{aligned} \quad (10)$$

where $\mathbf{C}_i = (\mathbf{D} - \mathbf{1}\mathbf{x}_i^T)^T (\mathbf{D} - \mathbf{1}\mathbf{x}_i^T)$ is the data covariance matrix. Hence, LLC codes can be solved very fast using the analytical solution while sparse code solved by (5) or (6) always needs iterative optimization process and thus has high computation complexity.

In order to improve the encoding speed, fast approximation LLC can be achieved by selecting only k ($k \ll K$) nearest neighbors of \mathbf{x}_i as the local bases \mathbf{D}_i . Thus (8) is converted to

$$\begin{aligned} \min_{\tilde{\mathbf{s}}} \quad & \sum_{i=1}^N \|\mathbf{x}_i - \mathbf{D}_i \tilde{\mathbf{s}}_i\|^2 \\ \text{s.t.} \quad & \mathbf{1}^T \tilde{\mathbf{s}}_i = 1, \quad \forall i, \end{aligned} \quad (11)$$

Using fast approximation LLC, the computation complexity can be reduced from $O(K^2)$ to $O(K+k^2)$. In this paper, fast approximation LLC is employed.

2.4. Improved LLC Based on Adaptive and Weighted Bases

2.4.1. Adaptive LLC. In traditional LLC, each signal is encoded with a fixed number of bases. However, considering the distribution of the feature vectors, LLC with a fixed number of bases may produce weak responses in the low-density area and poor representation in the high-density area [43]. In order to solve this problem, adaptive bases can be

selected based on the feature density. The traditional LLC is reformulated as

$$\begin{aligned} \min_{\mathbf{s}} \quad & \sum_{i=1}^N \|\mathbf{x}_i - \mathbf{D}\mathbf{s}_i\|^2 + \lambda \|\mathbf{a}\mathbf{I}_i \odot \mathbf{s}_i\|^2 \\ \text{s.t.} \quad & \mathbf{1}^T \mathbf{s}_i = 1 \quad \forall i, \end{aligned} \quad (12)$$

where λ represents the weight of adaptive locality constraint term. $\mathbf{a}\mathbf{I}_i$ denotes the adaptive locality adaptor:

$$\begin{aligned} \mathbf{a}\mathbf{I}_i &= \exp\left(\frac{\text{dist}(\mathbf{x}_i, \mathbf{D})}{\sigma}\right) \\ \mathbf{d}_j &\in N_{q_i}(\mathbf{x}_i), \end{aligned} \quad (13)$$

where $\text{dist}(\mathbf{x}_i, \mathbf{D}) = [\text{dist}(\mathbf{x}_i, \mathbf{d}_1), \dots, \text{dist}(\mathbf{x}_i, \mathbf{d}_j)]^T$. $N_{q_i}(\mathbf{x}_i)$ represents the set of q_i nearest neighbors of \mathbf{x}_i .

For each signal \mathbf{x}_i , the variance V_i of the distances to all bases can be computed:

$$V_i = \text{Var}\{\text{dist}(\mathbf{x}_i, \mathbf{d}_1), \dots, \text{dist}(\mathbf{x}_i, \mathbf{d}_K)\}. \quad (14)$$

As for the signal \mathbf{x}_i , if the distance variance V_i is large which means the signal locates in the low-density area with relatively few neighbors, fewer nearest neighbors can represent the signal enough. On the contrary, if the distance variance V_i is small which means the signal locates in the high-density area with more neighbors, then larger number of nearest bases should be selected to better represent the signal. Accordingly, the number of bases q_i corresponding to the signal \mathbf{x}_i can be adaptively adjusted depending on the distance variance V_i as follows:

$$q_i = \text{int}\left[Q \exp\left(-\frac{V_i - \min(V)}{\max(V) - \min(V)}\right)\right], \quad (15)$$

where $V = [V_1, V_2, \dots, V_N]$ denotes the distance variance of all the signals. The symbol int represents the round operation. The adaptive coding parameter Q is a specified scalar. After the adaptive number of nearest bases q_i is obtained, the LLC codes can be solved using fast approximation LLC as the above section.

2.4.2. Weighted LLC. A dictionary contains both sample information and weight information [23]. The sample information is the base vector of the dictionary. The weight information is the weight of a base depending on the proportion of training samples assigned to the base. However, the traditional LLC lacks the weight information and assumes that every base owns the same weight. In order to boost the discriminative ability of the LLC codes, the weight information of the dictionary is considered.

In the dictionary training stage, the dictionary is learned by the training samples. After the dictionary is obtained, for each training sample the Euclidean distance between the sample and each base can be calculated and the nearest base is treated as the base that the sample is assigned to. Then for each base \mathbf{d}_i , the accumulated number of the training samples

N_i assigned to the base can be acquired and the weight of the base is

$$w_i = \frac{N_i}{N_t}, \quad (16)$$

where N_t is the number of the total training samples.

Assuming that k nearest neighbors are selected using fast approximation LLC, the corresponding weight vector is $w = [w_1, w_2, w_j, \dots, w_k]$, where w_j denotes the weight of the j th selected base. The weight vector should be normalized as follows:

$$\bar{w}_j = \frac{w_j}{\sum_{j=1}^k w_j}. \quad (17)$$

Suppose that the solved LLC code of signal \mathbf{x} corresponding to selected base using fast approximation LLC is $\mathbf{s} = [s_1, s_2, \dots, s_k]$; the weighted code can be acquired by

$$\bar{\mathbf{s}} = [\bar{w}_1 s_1, \bar{w}_2 s_2, \dots, \bar{w}_k s_k]. \quad (18)$$

2.4.3. Improved LLC Based on Adaptive and Weighted Bases. Considering both adaptive number of bases and the weight of the selected bases, an improved LLC algorithm based on adaptive and weighted bases is proposed. Assume that for a signal \mathbf{x} using improved LLC based on adaptive bases, q nearest bases are selected, the LLC code of the signal corresponding to selected q bases is $\mathbf{s} = [s_1, s_2, \dots, s_q]$, and the corresponding weight vector is $w = [w_1, w_2, \dots, w_j, \dots, w_q]$. Then the weight vector is normalized by

$$\bar{w}_j = \frac{w_j}{\sum_{j=1}^q w_j}. \quad (19)$$

Finally, the weighted code can be solved as follows:

$$\bar{\mathbf{s}} = [\bar{w}_1 s_1, \bar{w}_2 s_2, \dots, \bar{w}_q s_q]. \quad (20)$$

With regard to the LLC codes corresponding to the unselected bases, they are set to zeros directly. The improved LLC codes (ILLC codes) of the training samples and test samples which are solved by the improved LLC algorithm based on adaptive and weighted bases can be used as feature vectors directly. Thus K -dimensional feature vector can be obtained for each sample.

3. Classification Based on APSO-Optimized SVM

3.1. Support Vector Machine. Support vector machine (SVM) is a popular machine learning method for the problems of classification, regression, and outlier detection, which is based on statistical learning theory. The basic SVM is applied to two-class classification problems, mapping the data into the high-dimensional feature space and finding the optimal hyperplane by maximizing the margin between the two classes. In practical application there are always multiple classes to discriminate, that is, multiclass classification. Many

methods have been proposed to solve the multiclass problems by combining the two-class classification, such as one-against-one, one-against-all, and binary tree architecture. In this paper, LIBSVM is used for multiclass classification with one-against-one method [44]. Suppose there are L classes; $C_L^2 = L(L-1)/2$ two-class classifiers are constructed.

The effect of kernel function is mapping the data into the high-dimensional feature space and hence makes the data linearly separable for classification in SVM. Common SVM kernel functions include linear, polynomial, sigmoid, and radial basis function (RBF) kernels. RBF kernel is widely used for nonlinear classification and thus RBF kernel function $K(x_i, x_j) = e^{-g\|x_i - x_j\|^2}$ is utilized, where g can stand for the width of RBF kernel function. In addition, the other important parameter is the penalization parameter c . Therefore, the parameters (c, g) need be optimized to obtain the best performance of SVM. Firstly, the data should be linearly normalized to the range of $[0, 1]$ and cross validation based on the parameters (c, g) is conducted to get the highest cross validation accuracy. Then the corresponding best parameters are employed to train on the training set and subsequently SVM model can be acquired. Finally, the test set is predicted using the SVM model. Adaptive particle swarm optimization (APSO) is implemented for the optimization of the parameters of SVM (c, g) as shown below.

3.2. Adaptive Particle Swarm Optimization. Particle swarm optimization (PSO) is an evolutionary algorithm based on swarm intelligence [37]. Comparing with genetic algorithm, PSO does not have the cross and mutation operations and the particles are updated only by the internal velocity so the algorithm is simple and easier to implement.

Assume that the particle swarm owns N particles and the i th particle is denoted as $\mathbf{p}_i = (p_{i1}, p_{i2}, \dots, p_{iK})$ (where K denotes the parameter numbers, in this paper $K = 2$). The corresponding velocity of the i th particle is denoted as $\mathbf{v}_i = (v_{i1}, v_{i2}, \dots, v_{iK})$. During the iteration, the best value of the i th particle and the best particle in all the particles are denoted as $pbest_i$ and $gbest$, respectively, among which $pbest_i$ represents the local best value of the i th particle while $gbest$ represents the global best value. At first the particles of the particle swarm are randomly initialized within the specified range of parameter value. Then during the k th iteration the particles and the corresponding velocities are updated as follows:

$$\begin{aligned} v_{ij}(k+1) &= w \cdot v_{ij}(k) + c_1 r_1 (pbest_{ij}(k) - p_{ij}(k)) \\ &\quad + c_2 r_2 (gbest_j(k) - p_{ij}(k)) \end{aligned} \quad (21)$$

$$p_{ij}(k+1) = p_{ij}(k) + v_{ij}(k+1),$$

where w named weight is elastic coefficient of velocity for velocity update, which indicates the effect of the previous velocity on the current velocity. c_1 and c_2 are positive scalars named acceleration coefficients, among which c_1 reflects the local search ability while c_2 reflects the global search ability. r_1 and r_2 are random variables with uniform distribution limited in $[0, 1]$.

Each iteration can be regarded as one generation. When the maximum number of iterations, namely, maximum generations, has been reached, the iteration is terminated. Finally, the global best value can be found corresponding to the highest cross validation accuracy. The fitness function of PSO is also the cross validation accuracy.

In PSO algorithm, it is easy to fall into the local optimum and the convergence is slow. The weight w is very important to balance exploration and exploitation and thus w has great influence on the accuracy and efficiency of the search. In many papers, the weight is linearly decreased with the number of iterations. However, the iteration number cannot well indicate the convergence state of the particle swarm. The average velocity of the particle swarm is large in the early iteration while it becomes small in the later iteration. Therefore, the average velocity of the particle swarm can better reflect the convergence state as follows:

$$\bar{V}(k) = \frac{1}{NK} \sum_{i=1}^N \sum_{j=1}^K |v_{ij}(k)|. \quad (22)$$

Based on the average velocity of the particle swarm, adaptive weight w can be acquired by

$$w(k) = 1 - \exp\left(-\frac{\bar{V}(k)}{V_{\max}}\right) \quad (23)$$

$$V_{\max} = \frac{1}{K} \sum_{j=1}^K |v_{j\max}|,$$

where $v_{j\max}$ is the specified maximum corresponding to the j th parameter. It can be found that the weight w is adaptively updated according to the average velocity of the particle swarm.

Moreover, the acceleration coefficients c_1 and c_2 are also adaptively updated by

$$c_1(k) = c_0 \left(1 - \exp\left(-\frac{\bar{V}(k)}{V_{\max}}\right)\right) \quad (24)$$

$$c_2(k) = c_0 - c_1(k),$$

where c_0 is a specified constant. It indicates that the acceleration coefficient c_1 for local search is decreased while c_2 for global search is increased with the decline of the average velocity.

Therefore, in the improved PSO algorithm with adaptive weight and acceleration coefficients, namely, adaptive PSO (APSO), the particles and the corresponding velocities can be updated as follows:

$$\begin{aligned} v_{ij}(k+1) &= w(k) \cdot v_{ij}(k) \\ &+ c_1(k) r_1 (pbest_{ij}(k) - p_{ij}(k)) \\ &+ c_2(k) r_2 (gbest_j(k) - p_{ij}(k)) \end{aligned} \quad (25)$$

$$p_{ij}(k+1) = p_{ij}(k) + v_{ij}(k+1).$$

The adaptive PSO algorithm can not only improve the global search ability but also accelerate the convergence compared with standard PSO algorithm [38].

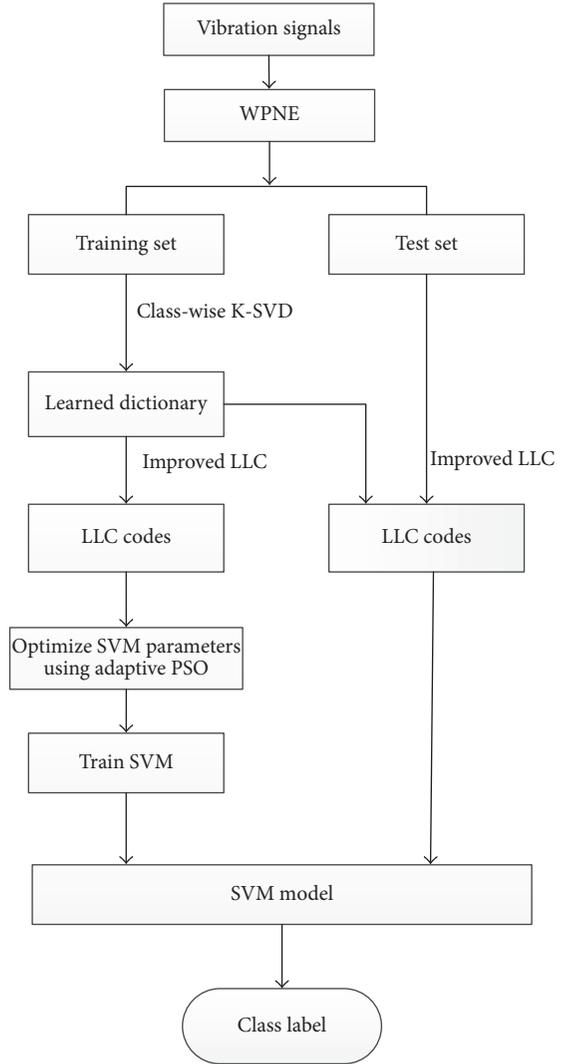


FIGURE 1: Flow chart of the proposed diagnosis scheme.

4. Bearing Fault Diagnosis Model Based on ILIC Codes and APSO-Optimized SVM

In this study, a bearing fault diagnosis model based on ILIC codes and APSO-optimized SVM is proposed. The proposed fault diagnosis method includes six stages: preliminary feature extraction, dictionary learning, sparse feature extraction base on ILIC codes, optimization of the parameters of SVM, SVM model training, and fault diagnosis. The flow chart of the proposed scheme is demonstrated in Figure 1. The detailed description is as follows.

(1) *Preliminary Feature Extraction.* The bearing vibration signal is decomposed with 3-level DWPT and 8-dimensional wavelet packet node energy is obtained for each signal.

(2) *Dictionary Learning.* The dictionary for solving ILIC codes is learned by class-wise K-SVD algorithm based on the training set.



FIGURE 2: The rolling bearing test rig.

(3) *Sparse Feature Extraction Base on ILLC Codes.* Based on the learned dictionary, the sparse LLC codes of the training set and test set can be solved with improved LLC algorithm. Since the sparse LLC codes are discriminative, they can be used for the input feature vector of the classifiers.

(4) *Optimization of the Parameters of SVM.* Adaptive particle swarm optimization (APSO) is carried out for the optimization of the parameters of SVM c and g in order to obtain the best parameters c and g corresponding to the highest cross validation accuracy based on the training set.

(5) *SVM Model Training.* Under the best parameters c and g , SVM model is trained using the training set.

(6) *Fault Diagnosis.* The class label of the test sample can be acquired based on the trained SVM model and thus the bearing fault diagnosis is realized.

5. Experiment and Analysis

5.1. Description of Data Set. An artificial fault experiment on rolling bearings is carried out to verify the proposed method based on ILLC codes and adaptive PSO-optimized SVM. The bearing test rig is demonstrated in Figure 2. The machine shaft is driven by an AC motor using rubber belts and shaft coupling. The rolling bearing is employed to support the shaft, whose outer race is fixed while the inner race rotates with the shaft. A data acquisition system (NI PXI-1042) is installed to acquire the vibration signals from the rolling bearing. In the data acquisition system, a bracket is stuck on the rolling bearing and an accelerometer (Kistler 8791A250) is mounted on the bracket. A series of GB203 rolling bearings are applied to the experiment. Single point faults generated by electrodischarge machining are conducted on the surface of the inner race, outer race, and rolling element of the bearings and inner race fault, outer race fault, and ball fault are introduced, respectively. Therefore, the rolling bearings have four states including normal, inner race fault (IRF), outer race fault (ORF), and ball fault (REF). The rotating speed is 720 r/min. The vibration signals are collected with the sampling frequency 25.6 kHz.

In the experiment, there are 20,480 points in each example and 120 examples are collected in total under four operating states with 30 examples for each state. Time series with 2048 points are obtained through truncation on each example and thus 1200 data samples can be acquired under four states with 300 data samples for each state. The raw vibration signals of rolling bearings under four different states are shown in Figure 3.

With regard to each state, 150 data samples are selected randomly as the training set and the rest are taken as the test set. Therefore, 600 training samples and 600 test samples are formed. Totally there are four classes to distinguish different running status.

5.2. Feature Extraction Based on Improved LLC. Firstly, wavelet packet transformation using db2 wavelet basis was carried out on both training samples and test samples and the preliminary features based on wavelet packet node energy were obtained subsequently. With respect to each signal, 8-dimensional preliminary feature was acquired. The preliminary features based on WPNE under four different states are demonstrated in Figure 4. It can be found that although the difference of the features corresponding to four classes is relatively obvious, it is still hard to distinguish so further feature extraction based on high level sparse LLC codes is implemented to enhance the discriminative ability.

Then dictionary learning was conducted using class-wise K-SVD algorithm based on the training samples. The dictionary atom number K corresponding to each class is specified as 10 and the sparsity prior T is set to 8. The learned dictionary with totally 40 atoms is shown in Figure 5. As shown in the figure, the learned dictionary atoms corresponding to different classes exhibit large difference while the dictionary atoms corresponding to the same class are more similar on the whole.

The method randomly selecting some training samples as a subdictionary for each class and then concatenating the subdictionaries to form a redundant dictionary was also conducted. For fair comparison, the selected number per class is the same as the proposed method. Moreover, we also employed dictionary learning methods based on

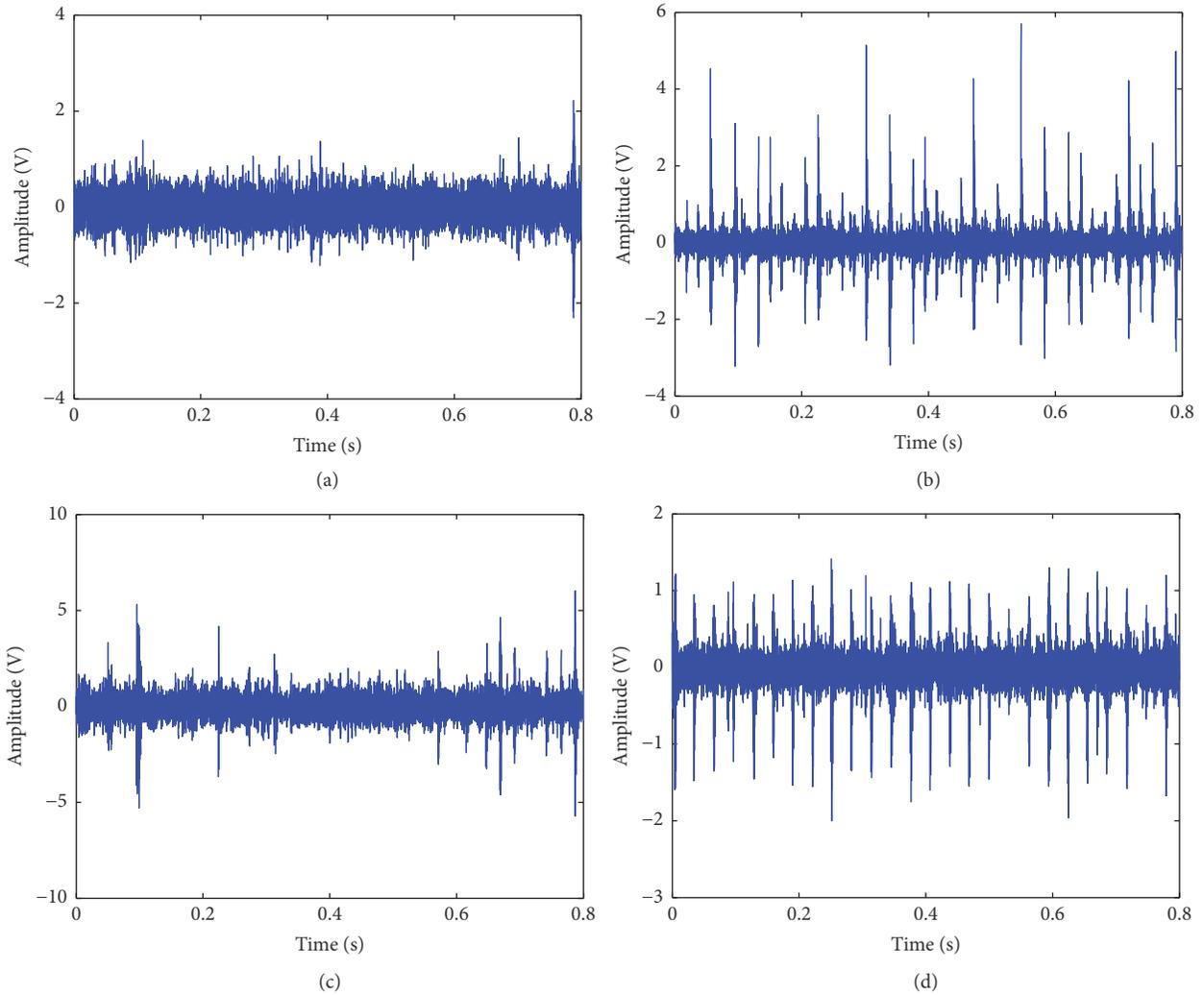


FIGURE 3: Vibration signals under four states: (a) normal; (b) inner race fault; (c) ball fault; (d) outer race fault.

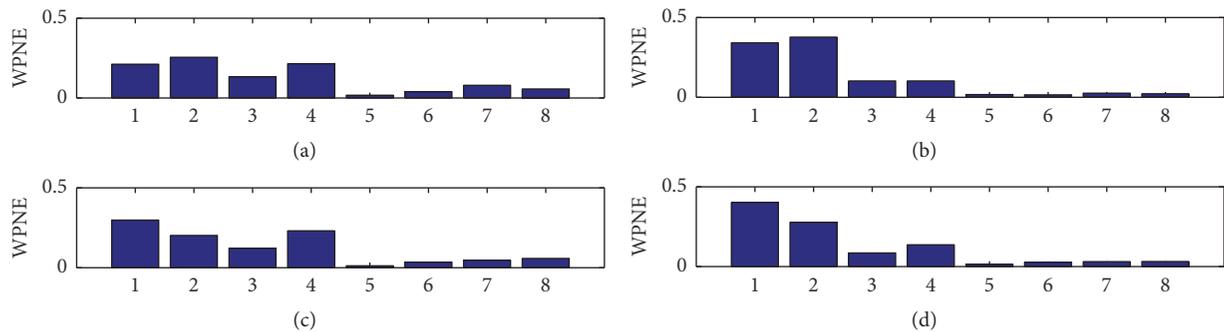


FIGURE 4: WPNE for four classes: (a) normal; (b) inner race fault; (c) ball fault; (d) outer race fault.

class-wise K -means and K -SVD for comparison. Class-wise K -means learn a subdictionary for each class using K -means algorithm and then a whole redundant dictionary can be formed by concatenating the subdictionaries while K -SVD directly learns the whole dictionary based on all the samples

containing four classes. The dictionary size is also kept the same as the proposed method. During the dictionary learning stage of K -SVD, the sparsity priors are also set to 8.

After the dictionary has been learned, the sparse LLC codes of the training samples and test samples can be

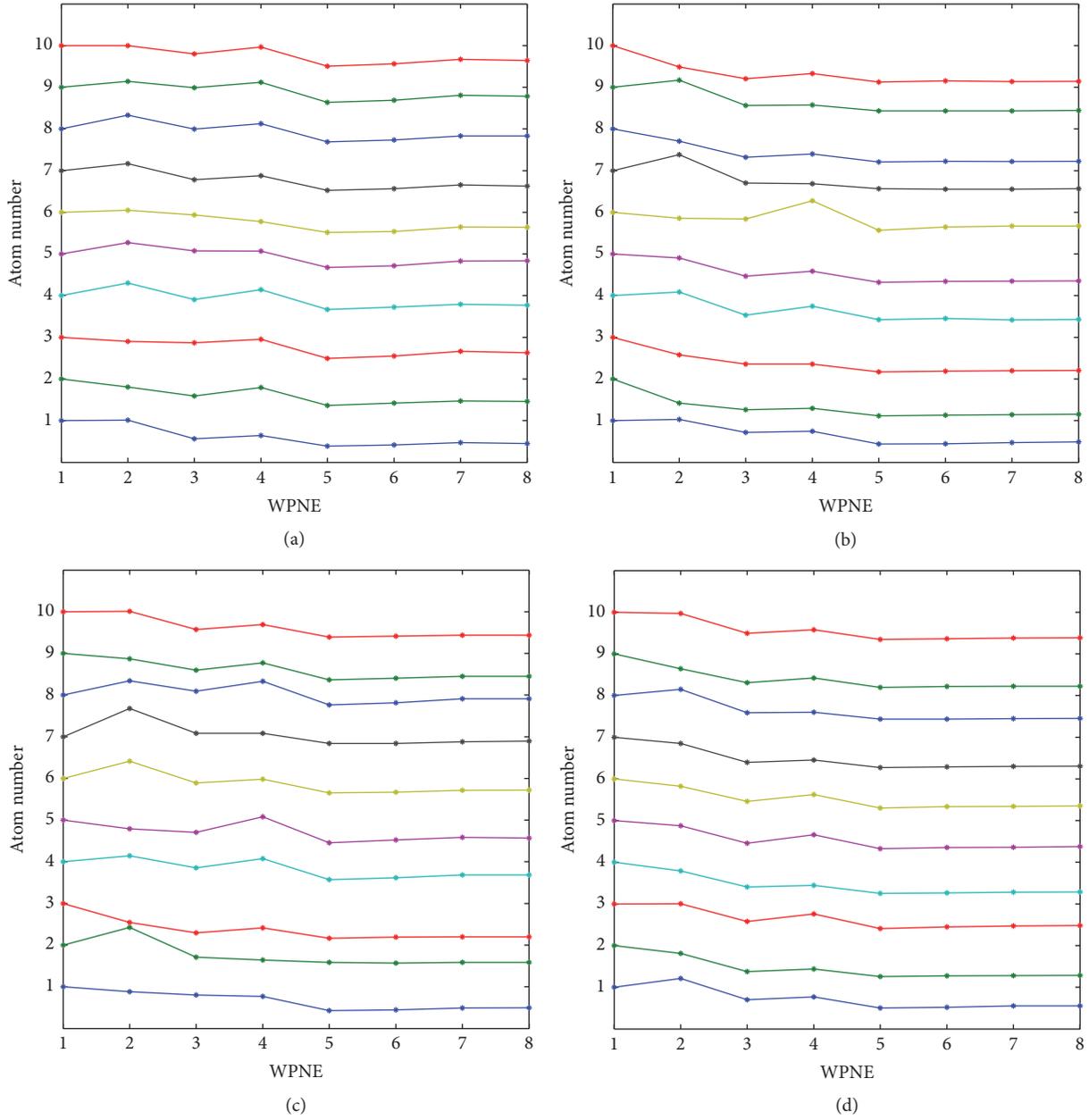


FIGURE 5: The learned dictionary atoms corresponding to four classes: (a) normal; (b) inner race fault; (c) ball fault; (d) outer race fault.

solved using improved LLC algorithm based on adaptive and weighted bases. The parameter Q is set to 10. Hence, 40-dimensional feature vector is obtained for each signal, which can be used as the input of the classifier. Using improved LLC algorithm, the sparse LLC codes of four test samples corresponding to four classes and the sum of absolute sparse LLC codes of different test samples corresponding to the same class are shown in Figures 6 and 7, respectively, where the dictionary atom numbers 1~10, 11~20, 21~30, and 31~40, respectively, represent the 1st class (norm state), the 2nd class (inner race fault), the 3rd class (ball fault), and the 4th class (outer race fault). The sum of absolute sparse LLC codes of different test samples using traditional LLC is also shown in Figure 8.

As can be seen from Figure 6, for a test sample with class label i , the corresponding LLC code has significant values in the dictionary atoms corresponding to the i th class. As shown in Figure 7, for the test samples belonging to the same class, the peak values are mostly located at the dictionary atoms belonging to the corresponding class, which indicates that the improved LLC algorithm can ensure that similar signals belonging to the same class will have similar sparse codes and thus the sparse LLC codes are discriminative for classification. Comparing Figure 7 with Figure 8, it can be seen that the improved LLC codes have larger values in the dictionary atoms belonging to the same classes and smaller values in the dictionary atoms belonging to the other classes

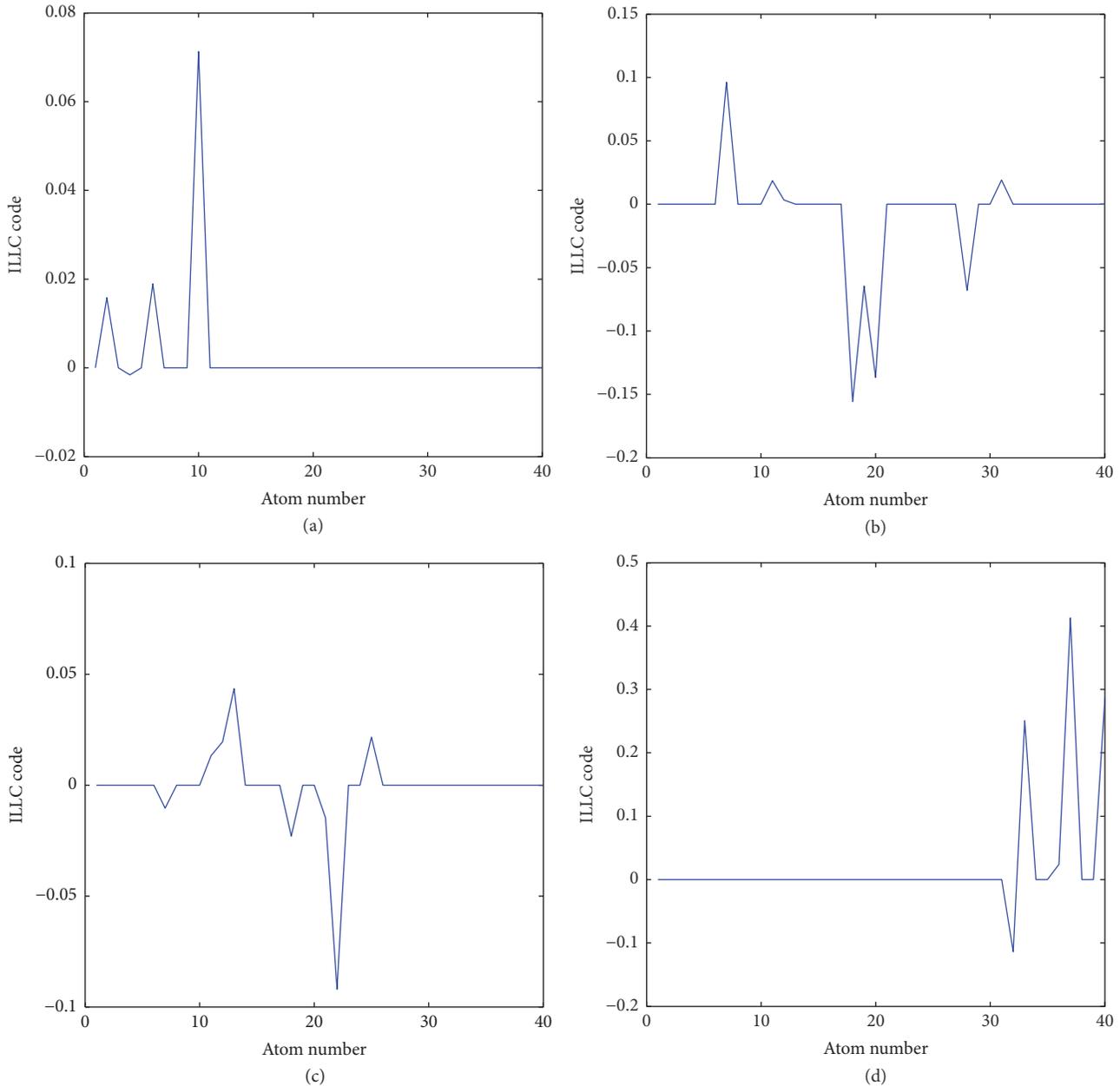


FIGURE 6: Improved LLC codes corresponding to four classes: (a) normal; (b) inner race fault; (c) ball fault; (d) outer race fault.

and thus are more discriminative than traditional LLC codes. Therefore, the improved LLC codes can effectively reflect the fault characteristics of rolling bearings and well distinguish different bearing states.

5.3. Fault Diagnosis with ILLC Codes. After feature extraction the classification methods based on SVM were employed. Firstly, we used linear SVM. Then the effect of preliminary feature using different methods and the influence of dictionary learning and LLC parameter sets was discussed. Finally, the bearing fault diagnosis based on ILLC codes and APSO-optimized SVM was conducted.

5.3.1. Diagnosis Results Based on Linear SVM. Combining different methods for generating dictionary with improved LLC algorithm, four feature extraction methods including randomly selecting some training samples as dictionary, class-wise K -means, K -SVD, and class-wise K -SVD as mentioned before were implemented, which are denoted as original, CK-means, K -SVD, and CK-SVD, respectively. Combining class-wise K -SVD with traditional LLC algorithm was also conducted and denoted as CKSVD-LLC. The classification results are shown in Table 1. The detailed results of the classification accuracy for each class are shown in Figure 9.

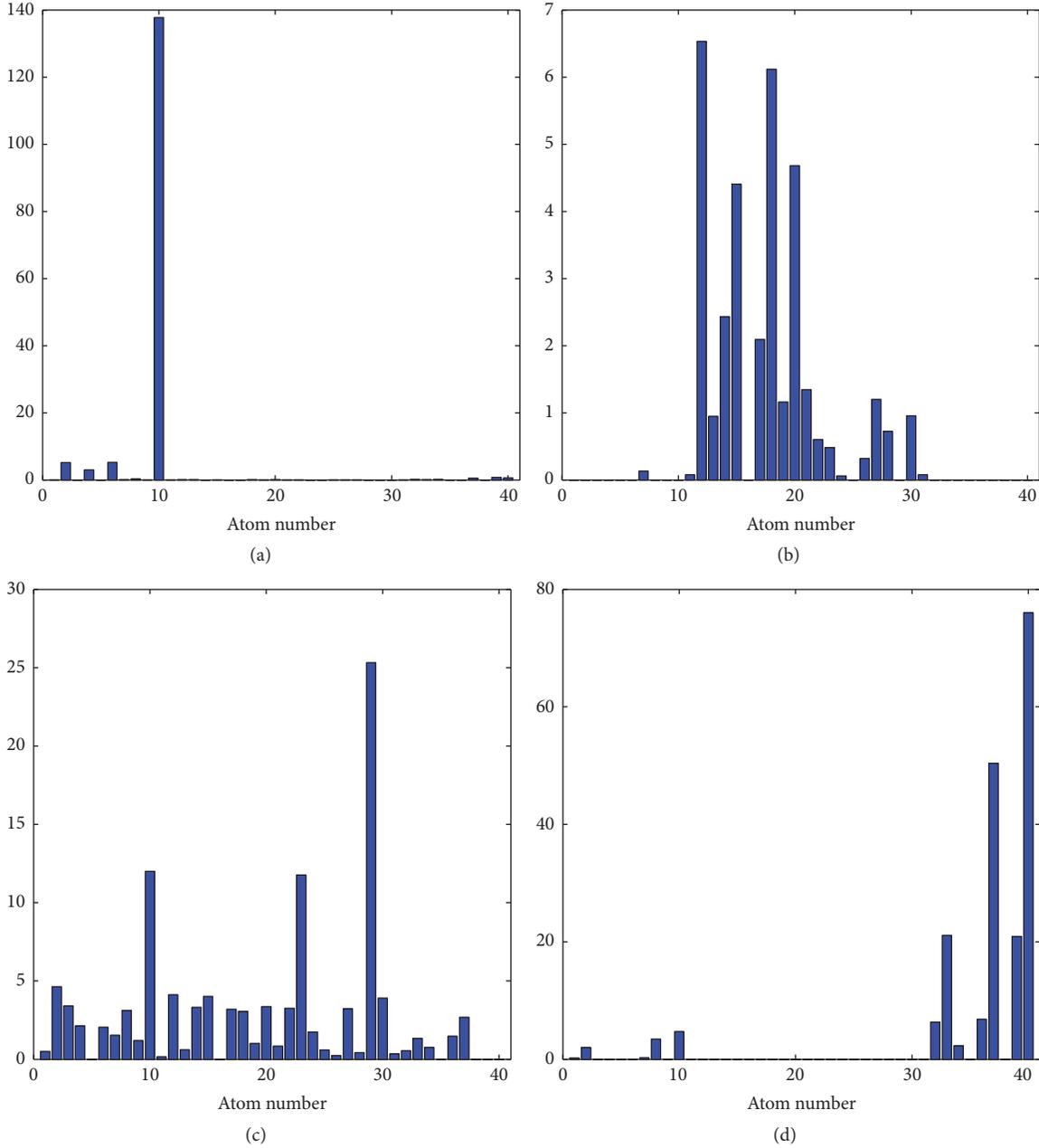


FIGURE 7: Sum of absolute ILCC codes corresponding to four classes using improved LLC: (a) normal; (b) inner race fault; (c) ball fault; (d) outer race fault.

TABLE 1: The classification accuracy of different methods (%).

Original	CK-means	K-SVD	CK-SVD	CKSVD-LLC
80.8	85.0	85.2	89.0	84.7

From Table 1, it can be found that the dictionary has a significant influence on the classification accuracy. The feature extraction method based on improved LLC codes by combining class-wise K-SVD with improved LLC algorithm acquires the highest accuracy, which validates the superiority of the improved LLC codes. The accuracy of class-wise

K-SVD is higher than class-wise K -means, which indicates that K-SVD is more effective for dictionary learning. We can also find that class-wise K-SVD outperforms K-SVD. Based on the same dictionary learned by class-wise K-SVD, the improved LLC is superior to traditional LLC, which shows the effectiveness of the improved LLC algorithm. As shown in Figure 9, the classification accuracy of ball fault (REF) is the lowest for all the methods, which indicates that the ball fault is more complex to distinguish. The proposed method based on class-wise K-SVD and improved LLC has the highest accuracy for ball fault (REF), which shows that the proposed method can improve the classification accuracy of ball fault.

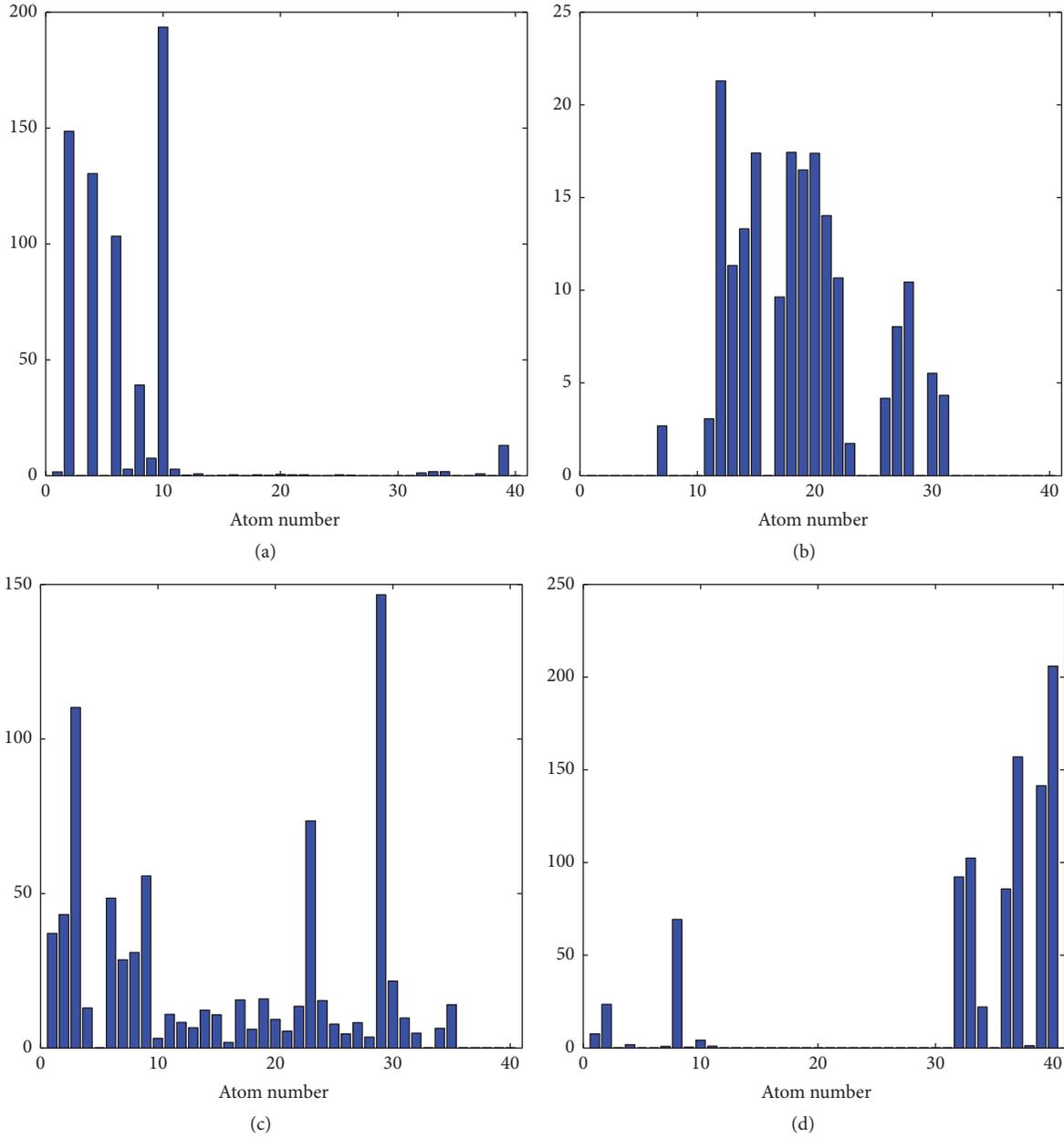


FIGURE 8: Sum of absolute LLC codes corresponding to four classes using traditional LLC: (a) normal; (b) inner race fault; (c) ball fault; (d) outer race fault.

5.3.2. Comparison with Different Preliminary Features. Different preliminary features based on direct use of the raw signal or the statistical features in time domain [45] were produced for comparison. The nine time-domain features include standard deviation, root mean square (RMS), kurtosis, skewness, crest factor, peak-peak, impulse factor, clearance factor, and shape factor. Then the sparse features based on class-wise K-SVD and improved LLC were extracted and linear SVM was employed as the classifier. The parameters of dictionary learning and LLC are the same as Section 5.2. Moreover, directly using preliminary features as the input of SVM was also conducted. The classification results are shown in Figure 10, where No FE (no feature extraction) represents

the method that directly uses the raw signal. For each bar, the different color on the top means the improvement of the high level feature based on improved LLC. As shown in Figure 10, for both directly using preliminary feature and LLC codes, WPNE owns the best accuracy. With respect to each preliminary feature, the improved LLC performs better than direct use of preliminary feature especially direct use of the raw signal, which shows the superiority of the sparse feature based on improved LLC.

5.3.3. Influence of Dictionary Learning and LLC Parameter Sets. The parameter influence on the method that combines class-wise K-SVD, improved LLC, and linear SVM was

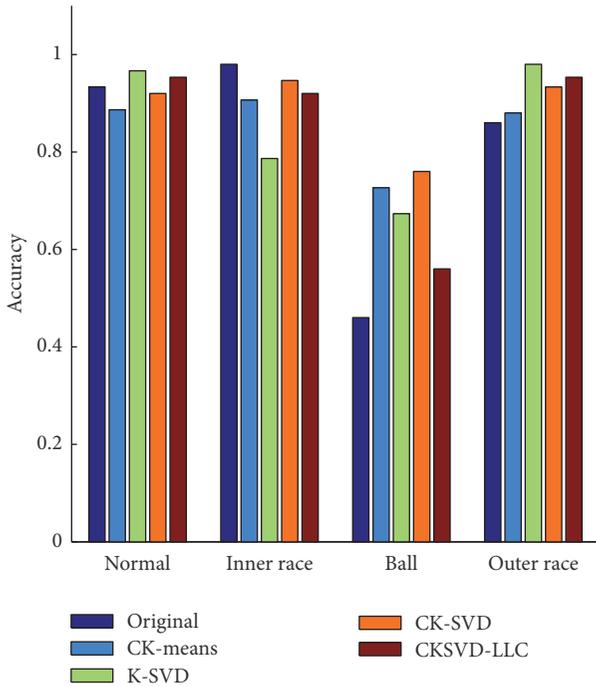


FIGURE 9: Detailed classification results.

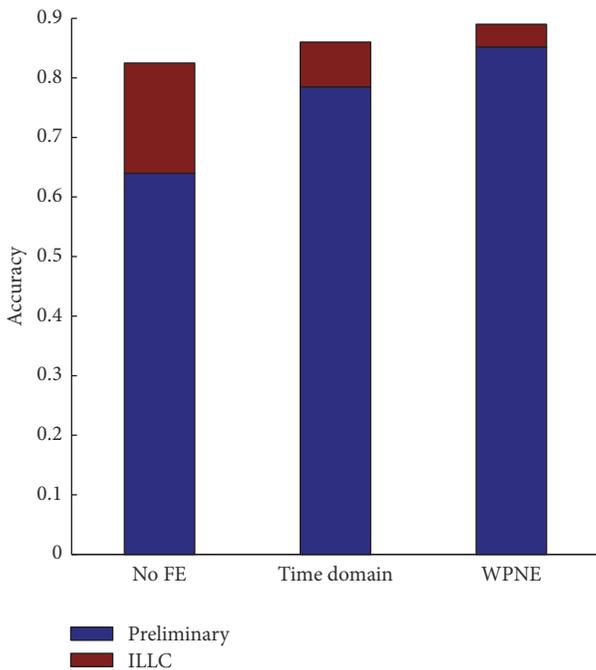


FIGURE 10: The classification results using different preliminary features.

investigated. The major parameters including the dictionary atom number k corresponding to each class, namely, the whole dictionary size K and adaptive coding parameter Q in improved LLC, were discussed.

Firstly, k was set to 10 and the adaptive coding parameter Q varied between 3 and 15. Then different sizes of

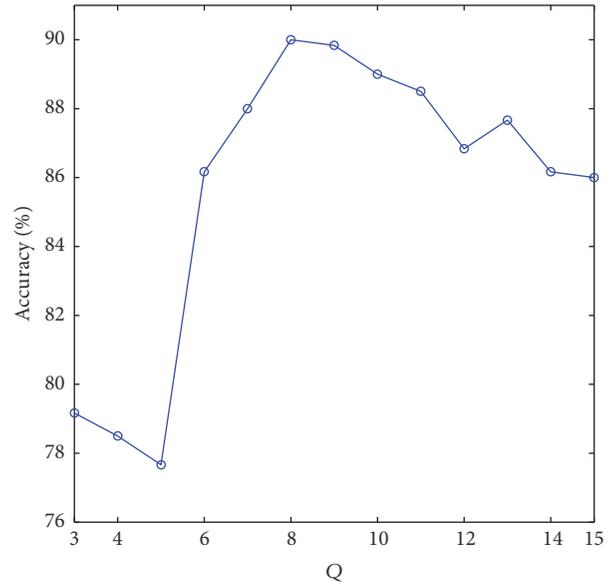


FIGURE 11: The classification accuracy of improved LLC under different Q .

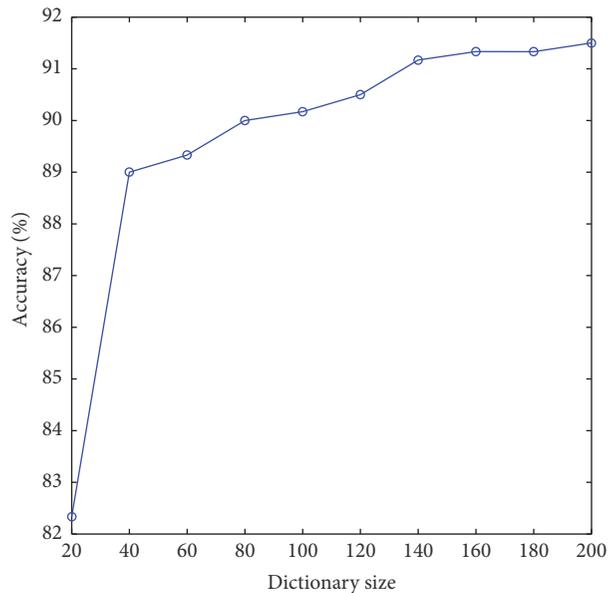


FIGURE 12: The classification accuracy of improved LLC under different dictionary size K .

dictionaries with atom number k corresponding to each class varied from $\{5, 10, 15, 20, 25, 30, 35, 40, 45, 50\}$ and thus the whole dictionary size K varied from $\{20, 40, 60, 80, 100, 120, 140, 160, 180, 200\}$ were employed. The classification results are shown in Figures 11 and 12, respectively, and the dictionary training time with varying dictionary size is demonstrated in Figure 13.

It can be found in Figure 11 that the parameter Q in improved LLC has large influence on the classification result. The best classification accuracy is acquired at $Q = 8$. From Figure 12 we can see that the classification accuracy rises

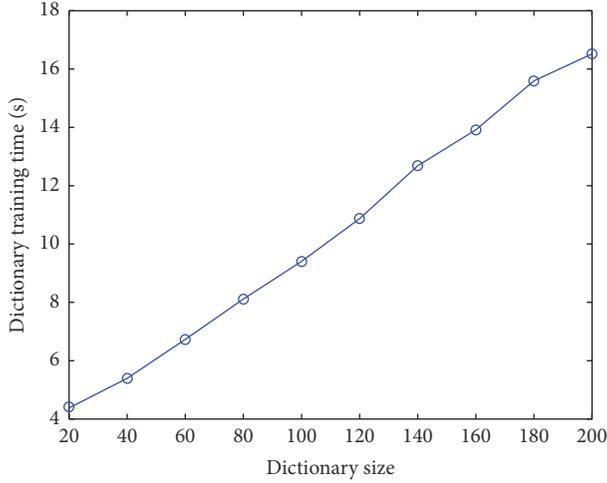


FIGURE 13: Dictionary training time with varying dictionary size K .

rapidly when the dictionary size is small while the increase slows down when the dictionary size becomes large enough. Overall, the increase of dictionary size K can improve the classification accuracy. However, it is shown in Figure 13 that the training time grows rapidly with the increase of K so the parameter K should be properly specified considering the computational complexity and classification accuracy simultaneously.

5.3.4. Diagnosis Results Based on APSO-Optimized SVM. Based on class-wise K-SVD and improved LLC codes, adaptive PSO was implemented for the optimization of the parameters c and g of SVM with RBF kernel function. The parameters of dictionary learning and LLC remain the same as Section 5.2. c and g are limited in $[2^{-10}, 2^{10}]$ and 5-fold cross validation is conducted. The population number is 20, the maximum iteration number is 100, and the fitness function is the cross validation accuracy. The parameter c_0 is set to 3. Standard PSO with the same parameters was also carried out for comparison. The parameters w , c_1 , and c_2 for standard PSO are, respectively, set to 1, 1.5, and 1.5. The fitness curves of standard PSO and adaptive PSO are shown in Figures 14 and 15, respectively. From Figures 14 and 15, it can be found that the convergence of adaptive PSO is faster than standard PSO. Moreover, the final cross validation accuracy of adaptive PSO is a little higher than standard PSO.

After the best parameters c and g are obtained, based on the training set optimized SVM model can be acquired using the optimized parameters. Finally, the test set can be predicted by the SVM model. The default parameters c and g of SVM are $c = 1$ and $g = 1/n$ (n is the data dimension) in libsvm [44]. SVM with default parameters and linear SVM were also performed for comparison. The classification results and computation time are shown in Table 2. From the results we can find that the classification accuracy of the proposed method based on improved LLC and adaptive PSO-optimized SVM is the highest, which indicates that the proposed method can effectively distinguish different status of rolling bearings and hence the bearing fault diagnosis is

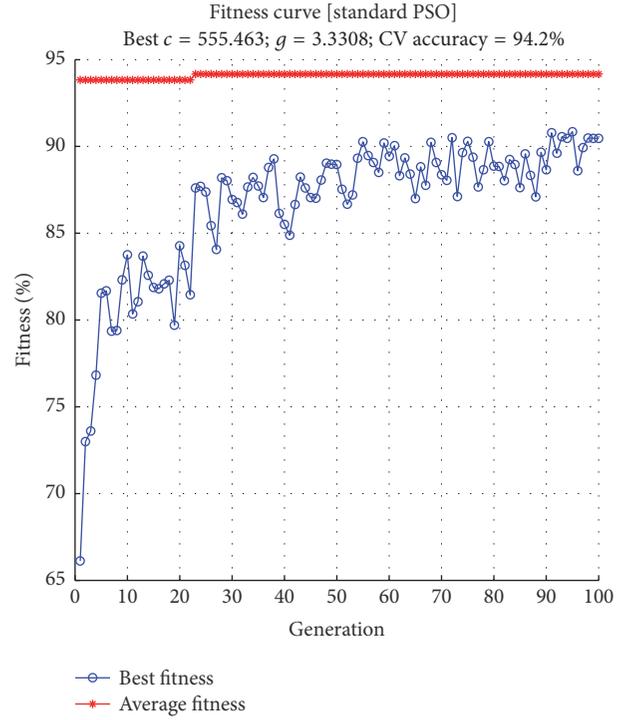


FIGURE 14: Fitness curve of standard PSO.

TABLE 2: The classification accuracy using linear and RBF SVM (%).

	Linear	Default	PSO	APSO
Normal	92.0	94.7	97.3	96.7
IRF	94.7	95.3	94.7	94.0
REF	76.0	76.7	90.0	93.3
ORF	93.3	96.0	98.7	99.3
Average	89.0	90.7	95.2	95.8
Time/s	0.2420	0.5157	147.0129	67.9978

realized. The accuracy is relatively low when using the default parameters, which shows that the parameters c and g have great effect on the performance of SVM with RBF kernel function and thus parameter optimization is essential for SVM with RBF kernel function. Comparing with linear SVM, although the computation time of adaptive PSO-optimized SVM is much longer, the accuracy is obviously higher, which proves the effectiveness of RBF kernel. Comparing with standard PSO, the adaptive PSO has higher accuracy and shorter computation time, which validates the superiority of adaptive PSO.

6. Conclusion

A novel method based on improved LLC algorithm and adaptive PSO-optimized SVM is proposed for fault diagnosis of rolling bearing. The experimental results show that the improved LLC algorithm is superior to traditional LLC. Moreover, different dictionary has large influence on the LLC codes. The dictionary learned by class-wise K-SVD algorithm can produce more discriminative LLC codes than

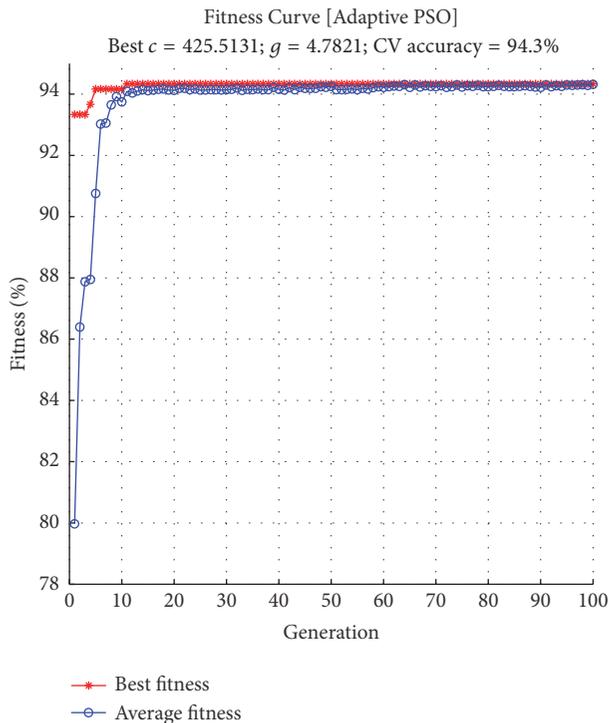


FIGURE 15: Fitness curve of adaptive PSO.

the dictionary generated by selecting the samples themselves or other dictionary learning methods such as K -means and K -SVD. Some other preliminary features are also employed for comparison, which shows that the high level sparse feature based on improved LLC outperforms the preliminary feature, especially when directly using the raw signal. The optimization results of SVM with RBF kernel function show that parameters selection is very important and adaptive PSO can greatly enhance the performance of SVM. Possible future work includes using other sparse coding algorithms to extract the sparse features of bearing vibration signals more effectively.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

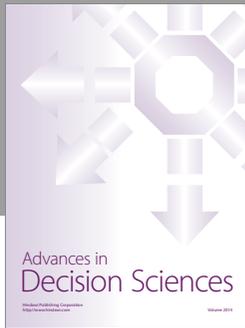
Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant nos. 51575339 and 51475286).

References

- [1] D. Ho and R. B. Randall, "Optimisation of bearing diagnostic techniques using simulated and actual bearing fault signals," *Mechanical Systems and Signal Processing*, vol. 14, no. 5, pp. 763–788, 2000.
- [2] H. Liu, L. Li, and J. Ma, "Rolling bearing fault diagnosis based on STFT-deep learning and sound signals," *Shock and Vibration*, vol. 2016, Article ID 6127479, 12 pages, 2016.
- [3] G. Tang, X. Wang, and Y. He, "A novel method of fault diagnosis for rolling bearing based on dual tree complex wavelet packet transform and improved multiscale permutation entropy," *Mathematical Problems in Engineering*, vol. 2016, Article ID 5432648, 13 pages, 2016.
- [4] X. S. Lou and K. A. Loparo, "Bearing fault diagnosis based on wavelet transform and fuzzy inference," *Mechanical Systems and Signal Processing*, vol. 18, no. 5, pp. 1077–1095, 2004.
- [5] N. Baydar and A. Ball, "A comparative study of acoustic and vibration signals in detection of gear failures using Wigner-Ville distribution," *Mechanical Systems and Signal Processing*, vol. 15, no. 6, pp. 1091–1107, 2001.
- [6] D. Yu, J. Cheng, and Y. Yang, "Application of EMD method and Hilbert spectrum to the fault diagnosis of roller bearings," *Mechanical Systems and Signal Processing*, vol. 19, no. 2, pp. 259–270, 2005.
- [7] H. Li, Y. Zhang, and H. Zheng, "Hilbert-Huang transform and marginal spectrum for detection and diagnosis of localized defects in roller bearings," *Journal of Mechanical Science and Technology*, vol. 23, no. 2, pp. 291–301, 2009.
- [8] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3736–3745, 2006.
- [9] L. Peotta, L. Granai, and P. Vandergheynst, "Image compression using an edge adapted redundant dictionary and wavelets," *Signal Processing*, vol. 86, no. 3, pp. 444–456, 2006.
- [10] M. Yang, L. Zhang, J. Yang, and D. Zhang, "Robust sparse coding for face recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*, pp. 625–632, Providence, RI, USA, June 2011.
- [11] Z. Feng and F. Chu, "Application of atomic decomposition to gear damage detection," *Journal of Sound and Vibration*, vol. 302, no. 1-2, pp. 138–151, 2007.
- [12] H. Liu, C. Liu, and Y. Huang, "Adaptive feature extraction using sparse coding for machinery fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 25, no. 2, pp. 558–574, 2011.
- [13] H. Tang, J. Chen, and G. Dong, "Sparse representation based latent components analysis for machinery weak fault detection," *Mechanical Systems and Signal Processing*, vol. 46, no. 2, pp. 373–388, 2014.
- [14] X. F. Chen, Z. Du, J. Li, X. Li, and H. Zhang, "Compressed sensing based on dictionary learning for extracting impulse components," *Signal Processing*, vol. 96, pp. 94–109, 2014.
- [15] K. Yu, T. Zhang, and Y. Gong, "Nonlinear learning using local coordinate coding," in *Proceedings of the 22nd International Conference on Neural Information Processing Systems*, pp. 2223–2231, Vancouver, British Columbia, Canada, December 2009.
- [16] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained linear coding for image classification," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '10)*, pp. 3360–3367, IEEE, June 2010.
- [17] Y. Yuan, B. Li, and M. Q.-H. Meng, "WCE abnormality detection based on saliency and adaptive locality-constrained linear coding," *IEEE Transactions on Automation Science and Engineering*, vol. 14, no. 1, pp. 149–159, 2017.
- [18] L. Zhu, Z. Zhang, and D. Zhang, "Recognising lepidopteran images based on locality-constrained linear coding and SVM," *Oriental Insects*, vol. 49, no. 1-2, pp. 36–46, 2015.

- [19] Y. Yuan, A. Hoogi, C. F. Beaulieu, M. Q.-H. Meng, and D. L. Rubin, "Weighted locality-constrained linear coding for lesion classification in CT images," in *Proceedings of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC '15)*, pp. 6362–6365, Milan, Italy, August 2015.
- [20] Y. Liang, M. Song, J. Bu, and C. Chen, "Colorization for gray scale facial image by locality-constrained linear coding," *Journal of Signal Processing Systems*, vol. 74, no. 1, pp. 59–67, 2014.
- [21] W.-G. Wang and Z.-S. Liu, "Classification of time-frequency images based on locality-constrained linear coding optimization model for rotating machinery fault diagnosis," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 229, no. 18, pp. 3350–3360, 2015.
- [22] Y. Li, S. Bu, Z. Liu, and C. Zhang, "Mechanical fault diagnosis of rolling bearing based on locality-constrained sparse coding," in *Proceedings of the IEEE Conference on Prognostics and Health Management (PHM '15)*, pp. 1–7, Austin, TX, USA, June 2015.
- [23] J. Yang and Z. Ma, "Action recognition using weighted locality-constrained linear coding," *IEICE Transactions on Information and Systems*, vol. E98D, no. 2, pp. 462–466, 2015.
- [24] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, Springer, Boston, MA, USA, 1992.
- [25] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.
- [26] K. Engan, S. O. Aase, and J. H. Husøy, "Multi-frame compression: Theory and design," *Signal Processing*, vol. 80, no. 10, pp. 2121–2140, 2000.
- [27] R. Grosse, R. Raina, H. Kwong, and A. Y. Ng, "Shift-invariant sparse coding for audio classification," in *Proceedings of the Twenty-Third Conference on Uncertainty in AI*, pp. 149–158, Vancouver, BC, Canada, July 2007.
- [28] O. Bryt and M. Elad, "Compression of facial images using the K-SVD algorithm," *Journal of Visual Communication and Image Representation*, vol. 19, no. 4, pp. 270–282, 2008.
- [29] Z. Feng and M. Liang, "Complex signal analysis for planetary gearbox fault diagnosis via shift invariant dictionary learning," *Measurement*, vol. 90, pp. 382–395, 2016.
- [30] B. Samanta and K. R. Al-Balushi, "Artificial neural network based fault diagnostics of rolling element bearings using time-domain features," *Mechanical Systems and Signal Processing*, vol. 17, no. 2, pp. 317–328, 2003.
- [31] Q. Hu, Z. He, Z. Zhang, and Y. Zi, "Fault diagnosis of rotating machinery based on improved wavelet package transform and SVMs ensemble," *Mechanical Systems and Signal Processing*, vol. 21, no. 2, pp. 688–705, 2007.
- [32] Z. Liu, H. Cao, X. Chen, Z. He, and Z. Shen, "Multi-fault classification based on wavelet SVM with PSO algorithm to analyze vibration signals from rolling element bearings," *Neurocomputing*, vol. 99, pp. 399–410, 2013.
- [33] J. Cui, M. Shan, R. Yan, and Y. Wu, "Aero-engine fault diagnosis using improved local discriminant bases and support vector machine," *Mathematical Problems in Engineering*, vol. 2014, Article ID 283728, 9 pages, 2014.
- [34] Y. Feng and F. S. Schlindwein, "Normalized wavelet packets quantifiers for condition monitoring," *Mechanical Systems and Signal Processing*, vol. 23, no. 3, pp. 712–723, 2009.
- [35] H. T. Zhou, J. Chen, G. M. Dong, H. C. Wang, and H. D. Yuan, "Bearing fault recognition method based on neighbourhood component analysis and coupled hidden Markov model," *Mechanical Systems and Signal Processing*, vol. 66–67, pp. 568–581, 2016.
- [36] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, Berlin, Germany, 1995.
- [37] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of the IEEE International Conference on Neural Networks (ICNN '95)*, vol. 4, pp. 1942–1948, Perth, WA, Australia, November 1995.
- [38] S. Zhai and T. Jiang, "A novel sense-through-foilage target recognition system based on sparse representation and improved particle swarm optimization-based support vector machine," *Measurement*, vol. 46, no. 10, pp. 3994–4004, 2013.
- [39] Z. Xia, S. Xia, L. Wan, and S. Cai, "Spectral regression based fault feature extraction for bearing accelerometer sensor signals," *Sensors*, vol. 12, no. 10, pp. 13694–13719, 2012.
- [40] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.
- [41] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.
- [42] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *Institute of Electrical and Electronics Engineers. Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.
- [43] Q. Fang, J. Sang, and C. Xu, "Saliency aware locality-preserving coding for image classification," in *Proceedings of the 2012 13th IEEE International Conference on Multimedia and Expo (ICME '12)*, pp. 260–265, Melbourne, VIC, Australia, July 2012.
- [44] C. Chang and C. Lin, "LIBSVM: a Library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1–27, 2011.
- [45] Y. Zhang, H. Zuo, and F. Bai, "Classification of fault location and performance degradation of a roller bearing," *Measurement*, vol. 46, no. 3, pp. 1178–1189, 2013.



Hindawi

Submit your manuscripts at
<https://www.hindawi.com>

