

Research Article

Remaining Useful Life Prediction of Rolling Bearings Using PSR, JADE, and Extreme Learning Machine

Yongbin Liu,^{1,2} Bing He,¹ Fang Liu,^{1,2} Siliang Lu,² Yilei Zhao,¹ and Jiwen Zhao²

¹Department of Mechanical Engineering, Anhui University, Hefei 230601, China

²National Engineering Laboratory of Energy-Saving Motor & Control Technology, Anhui University, Hefei 230601, China

Correspondence should be addressed to Fang Liu; ufun@ahu.edu.cn

Received 15 January 2016; Revised 31 March 2016; Accepted 3 April 2016

Academic Editor: Giuseppe Vairo

Copyright © 2016 Yongbin Liu 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.

Rolling bearings play a pivotal role in rotating machinery. The degradation assessment and remaining useful life (RUL) prediction of bearings are critical to condition-based maintenance. However, sensitive feature extraction still remains a formidable challenge. In this paper, a novel feature extraction method is introduced to obtain the sensitive features through phase space reconstitution (PSR) and joint with approximate diagonalization of Eigen-matrices (JADE). Firstly, the original features are extracted from bearing vibration signals in time and frequency domain. Secondly, the PSR is applied to embed the original features into high dimensional phase space. The between-class and within-class scatter (SS) are calculated to evaluate the feature sensitivity through the phase point distribution of different degradation stages and then different weights are assigned to the corresponding features based on the calculated SS. Thirdly, the JADE is employed to fuse the weighted features to obtain the advanced features which can better reflect the bearing degradation process. Finally, the advanced features are input into the extreme learning machine (ELM) to train the RUL prediction model. A set of experimental case studies are carried out to verify the effectiveness of the proposed method. The results show that the extracted advanced features can better reflect the degradation process compared to traditional features and could effectively predict the RUL of bearing.

1. Introduction

Currently, Prognostics and Health Management (PHM) is an important research area in industry, which aims at increasing availability and decreasing the downtime by predicting the residual life of machine. Rolling bearing as a fundamental component is widely used in the rotary machine. Bearing failure is one of the principal reasons for the machine damage [1]. In brief, condition monitoring of bearing is becoming more and more significant, especially in the RUL prediction, which can reduce the costs of maintenance. There are three main approaches, namely, model based, data driven, and hybrid [2–4]. Among them, the data driven methods are easier to deploy than others, which are based on the relation between degradation state and measurement data [5].

One of the challenges for RUL prediction based on data driven is sensitive feature extraction from vibration

signals that are always mixed by noise [6, 7]. Many modern mathematical methods are employed to overcome this challenge. These methods offer systematic, scientific, and effective tactics for extracting the features that can reflect the bearing degradation. There are some representative feature extraction methods via statistical calculation, such as time domain analysis, frequency domain analysis, and time-frequency domain analysis [8–10]. However, these original features are often redundant and with high dimensionality. Therefore, dimensionality reduction methods such as the principal component analysis (PCA) and the independent component analysis (ICA) are employed to reduce the redundancy for machine fault diagnosis [11–13]. On the other hand, the nonlinear dimensionality reduction methods are developed in recent years, such as local embedding structure (LLE) [14], local tangent space information (LTSA) [15], and local

adjacency relations (LE) [16]. However, they are often with complex algorithm and need a large amount of calculation [17].

In this paper, a new dimensionality reduction technique is introduced based on approximate diagonalization of Eigen-matrices (JADE) algorithm, which is originally used in blind identification. Generally, the mixed model without environmental noise can be effectively identified by JADE. However, it has a poor performance when the signals are mixed with background noise [18]. To overcome this problem, a feature weighted fusion method combining the phase space reconstitution (PSR) [19] with the between-class and within-class scatters evaluation [20] is proposed in this paper. As different kinds of signals demonstrate different structure in phase trajectory, the PSR can better grasp the nature and regularity of the time series.

After extracting feature from vibration signals, machine learning algorithms can be used to establish the bearing RUL prediction model, such as the support vector machine (SVM) [21–23] and the ANN [24]. In recent years, a new machine learning algorithm called extreme learning machine (ELM) has been introduced by Huang et al. It is a novel algorithm of single hidden layer feed-forward neural networks (SLFNs), which has an extremely fast learning speed and advantageous generalization capability through randomly choosing hidden nodes and analytically determining output weights [25]. Moreover, ELM has a commendable performance in both classification and regression with a simple structure. Compared with back-propagation (BP) feed-forward network learning algorithm, the training speed of ELM is much faster while obtaining better generalization [26, 27]. In view of these advantages, the ELM is selected in this paper to establish bearing RUL prediction model.

The rest of the paper is organized as follows. Section 2 gives a brief introduction of PSR, JADE, and ELM. The proposed RUL prediction procedure for bearings is illustrated in Section 3. Experimental validations and comparison with other linear dimensionality reduction methods are presented in Section 4. Finally, the conclusions are drawn in Section 5.

2. Theory Background

2.1. Phase Space Reconstitution (PSR). Chaotic time series analysis is a method for nonlinear signal processing. By extending the time series to high dimensional phase space, this method can grasp the nature of time series much better. The research of PSR on chaotic time series was started from Packard et al. [28]. Suppose that the time series is $x = (x_1, x_2, \dots, x_n)$, the embedding dimension is chosen as d , and the time delay is selected as τ . And then a phase point is constructed as

$$X_i(d) = [x_i, x_{i+\tau}, \dots, x_{i+(d-2)\tau}, x_{i+(d-1)\tau}], \quad (1)$$

where n is the number of time series and i is the index of the phase point. However, the acquired signals are blended with noise, so the parameters should be determined under a reasonable rule [29]. In this paper, the autocorrelation

function is employed to identify the time delay, which is defined as

$$C_i(\tau) = \frac{(1/n) \sum_{i=1}^n [x(i+\tau) - \bar{x}][x(i) - \bar{x}]}{(1/n) \sum_{i=1}^n [x(i) - \bar{x}]^2}, \quad (2)$$

where $\bar{x} = (1/n) \sum_{i=1}^n x(i)$. According to Rosenstein's theories [30], the minimum τ that makes $C_i(\tau)$ lower than $1 - 1/e$ can be considered as the most suitable time delay. And Cao's algorithm [31] is employed to identify the embedding dimension. Suppose that the nearest phase point to $X_i(d)$ is $X_{\text{near}(i,d)}(d)$. In accordance with the embedding theorem, if d is the suitable embedding dimension, then any two points that are close to the d dimensional reconstructed phase space will continue to be close to the $d + 1$ dimensional space

$$a(i, d) = \frac{\|X_i(d+1) - X_{\text{near}(i,d+1)}(d+1)\|}{\|X_i(d) - X_{\text{near}(i,d)}(d)\|},$$

$$E(d) = \frac{1}{n-d\tau} \sum_{i=1}^{n-d\tau} a(i, d), \quad (3)$$

$$E_1(d) = \frac{E(d+1)}{E(d)}.$$

With the increase of d , $E_1(d)$ will be gradually stabilized; therefore, the minimum d that stabilizes $E_1(d)$ is set as the embedding dimension. However, it is difficult to determine whether the curve of $E_1(d)$ is stable due to the limited length of samples. Therefore, supplementary criteria are proposed as shown in

$$E^*(d) = \frac{1}{n-d\tau} \sum_{i=1}^{n-d\tau} |x_{i+d\tau} - x_{\text{near}(i,d+d\tau)}|, \quad (4)$$

$$E_2(d) = \frac{E^*(d+1)}{E^*(d)},$$

where $E_2(d)$ will be changed with the increasing of d if the sample is deterministic. Besides, more details of the description are presented in [32].

2.2. Joint Approximate Diagonalization of Eigen-Matrices (JADE). The JADE algorithm is widely used in blind source separation (BSS) [33]; it is based on a standard linear data model that can be expressed as follows:

$$x(t) = As(t) + n(t), \quad (5)$$

where $x(t)$ is the source signal that mixed with noise, $n(t)$ is additive noise, and $s(t)$ is the source signal, which can be obtained by transfer matrix A . The JADE can be summarized as follows:

(a) Signal whitening is as follows:

$$z(t) = Wx(t),$$

$$W = [(\mu_1 - \sigma)^{-1/2} h_1, \dots, (\mu_n - \sigma)^{-1/2} h_n]^H, \quad (6)$$

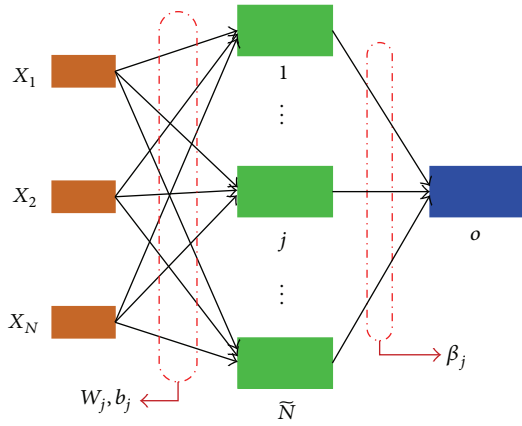


FIGURE 1: The structure of ELM.

where $z(t)$ is whitened signal, $\mu_1, \mu_2, \dots, \mu_n$ is the eigenvalue of sample, h_1, h_2, \dots, h_n is the corresponding eigenvector, and σ is the noise variance.

- (b) Calculate the fourth-order cumulants [34]

$$Q_z = \{ \text{Cum}(z_i, z_j, z_k, z_l) \mid 1 \leq i, j, k, l \leq d \}, \quad (7)$$

where d is the dimensionality of the vector z .

- (c) Calculate the unitary matrix

$$\widehat{U} = \arg_{\min} \sum_i \text{off}(U^\# Q_z(M_i) U), \quad 1 \leq i \leq n, \quad (8)$$

where $\#$ denotes the pseudo-inverse, “off” is the square of the nondiagonal elements, “ \arg_{\min} ” is the argument of a complex number, and U is a rotation matrix.

- (d) The single model can be separated as

$$s(t) = \widehat{A}x(t), \quad (9)$$

where $\widehat{A} = \widehat{U}W^\#$.

In this paper, JADE is firstly applied in dimensionality reduction and feature extraction. The single model can be obtained through Eigen-decomposition of the fourth-order cumulative matrix of the multivariate data. This not only is simpler than PCA and fast-ICA algorithm but also can improve the robustness of the results. However, the original features have different units and dimension, and it is difficult to distinguish between the contributions of each feature to the overall feature.

2.3. Extreme Learning Machine (ELM). ELM is an efficient learning algorithm for classification and prediction in SLFNs, the structure of ELM as shown in Figure 1. It can randomly generate the bias and input weights of hidden nodes. Moreover, it can immediately get the result without the puzzle of a local minimum [35].

Suppose there are N samples (X_i, t_i) , where X_i denotes the feature and t_i denotes the target value. Then the model of SLFNs with \widetilde{N} hidden nodes can be described as below:

$$\sum_{i=1}^{\widetilde{N}} \beta_i g_i(X_j) = \sum_{i=1}^{\widetilde{N}} \beta_i g(W_i \cdot X_j + b_i) = o_j, \quad (10)$$

$$j = 1, \dots, N,$$

where $W_i = [w_{i,1}, w_{i,2}, \dots, w_{i,\widetilde{N}}]^T$ is the weight vector that connects the i th hidden and the input nodes and b_i is the i th bias of hidden layer nodes, and they are randomly assigned. $\beta_i = [\beta_{i,1}, \beta_{i,2}, \dots, \beta_{i,\widetilde{N}}]^T$ is the weight vector connecting the i th hidden and the output nodes and o is the output value. To minimize the error between output o_j and target t_j , which is represented as $\sum_{j=1}^N \|o_j - t_j\| = 0$, there is an existing optimal value of β_i such that

$$\sum_{i=1}^{\widetilde{N}} \beta_i g(W_i \cdot X_j + b_i) = t_j, \quad j = 1, \dots, N. \quad (11)$$

Then above N equation can be written compactly as

$$H\beta = T, \quad (12)$$

where

$$H(W, b, X) = \begin{bmatrix} g(W_1 \cdot X_1 + b_1) & \cdots & g(W_{\widetilde{N}} \cdot X_1 + b_{\widetilde{N}}) \\ \vdots & \cdots & \vdots \\ g(W_1 \cdot X_N + b_1) & \cdots & g(W_{\widetilde{N}} \cdot X_N + b_{\widetilde{N}}) \end{bmatrix}_{N \times \widetilde{N}}. \quad (13)$$

Thus the output weight β can be written as

$$\beta = H^+ T, \quad (14)$$

where $H^+ = (H^T H)^{-1} H^T$.

Although the ELM is faster than other machine learning algorithms, its accuracy is still dependent on feature extraction.

3. The Proposed Method for Bearing RUL Prediction

In the abovementioned methods, the PSR has superior merits for characterizing nonlinear time series. The JADE focuses on feature fusion or dimensionality reduction. The ELM is an effective learning algorithm with fast training speed. A bearing RUL prediction scheme combined with the three methods is introduced in this paper.

TABLE 1: List of the original extracted features.

Time domain		Frequency domain	
$F1 = \frac{1}{N} \sum_{i=1}^N x_i$	$F2 = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}$	$F9 = \frac{1}{N} \sum_{i=1}^N s_i$	$F10 = \frac{1}{N} \sum_{j=1}^N \left(s_j - \frac{1}{N} \sum_{i=1}^N s_i \right)^2$
$F3 = \left[\frac{1}{N} \sum_{i=1}^N \sqrt{ x_i } \right]^2$	$F4 = \frac{1}{N} \sum_{i=1}^N x_i $	$F11 = \frac{(1/N) \sum_{j=1}^N (s_j - (1/N) \sum_{i=1}^N s_i)^3}{(\sqrt{F10})^3}$	$F12 = \frac{\sum_{i=1}^N f_i s_i}{\sum_{j=1}^N s_j}$
$F5 = \frac{1}{N} \sum_{i=1}^N x_i^3$	$F6 = \frac{\sqrt{(1/N) \sum_{i=1}^N x_i^2}}{F4}$	$F13 = \sqrt{\frac{1}{N} \sum_{i=1}^N s_i (f_i - F12)^2}$	$F14 = \sqrt{\frac{\sum_{i=1}^N f_i^2 s_i}{\sum_{j=1}^N s_j}}$
$F7 = \frac{\max(x)}{(1/N) \sum_{i=1}^N x_i }$	$F8 = \frac{(1/N) \sum_{i=1}^N x_i^4}{\left(\sqrt{(1/N) \sum_{j=1}^N x_j^2} \right)^4}$	$F15 = \sqrt{\frac{\sum_{i=1}^N f_i^4 s_i}{\sum_{j=1}^N f_j^2 s_j}}$	$F16 = \frac{\sum_{i=1}^N f_i^2 s_i}{\sqrt{\sum_{j=1}^N s_j \sum_{k=1}^N f_k^4 s_k}}$

Where S_i is a spectrum for $i = 1, 2, \dots, N$, N is the number of spectrum lines, and f_i is the frequency value of the i th spectrum line.

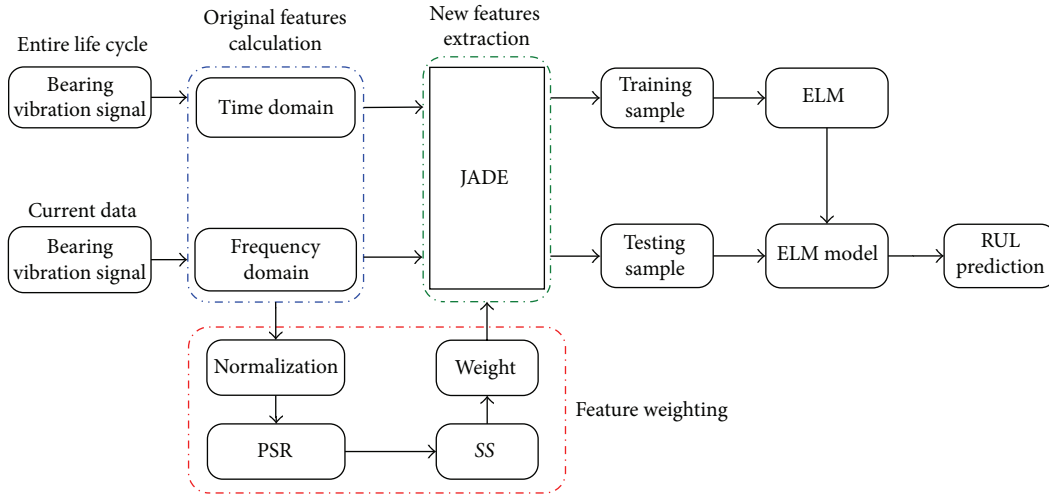


FIGURE 2: Principle of the proposed method based on PSR and JADE.

The scheme is exhibited in Figure 2 and can be described as follows.

Step 1 (original feature extraction). A total of 16 statistical features are shown in Table 1 including 8 time domain features and 8 frequency domain features. $F1$ to $F8$ are the mean, root mean square, RMS amplitude, absolute average, skewness, waveform index, impulsion index, and kurtosis index, respectively. $F9$ – $F16$ are the frequency domain feature, which indicates the degree of dispersion or concentration of the spectrum, and the change of the dominant frequency band.

Step 2 (PSR for each feature). The autocorrelation function is employed to select the time delay and Cao's algorithm is used to determine the embedding dimension. All the features are reconstructed by PSR with the same parameter.

Step 3 (between-class and within-class scatter (SS) for feature weighting). Based on the feature phase space distribution, SS is used to describe the classification ability of features quantitatively for different degradation states. Suppose that

there are C classes that are established by degradation states and the i th class vector is $x^i = (x_1^i, x_2^i, \dots, x_{n_i}^i)$, where n_i is the number of samples in this class. Then SS is defined as follows:

$$\begin{aligned}
 SS &= \text{trace} \left(\frac{S_b}{S_w} \right), \\
 S_b &= \sum_{i=1}^C p_i \|m_i - m\|^2, \\
 S_w &= \sum_{i=1}^C p_i \frac{1}{n_i} \sum_{k=1}^{n_i} \|x_k^i - m_i\|^2, \\
 p_i &= \frac{n_i}{\sum_{j=1}^C n_j}, \\
 m_i &= \frac{1}{n_i} \sum_{k=1}^{n_i} x_k^i, \\
 m &= \sum_{i=1}^C p_i m_i,
 \end{aligned} \tag{15}$$

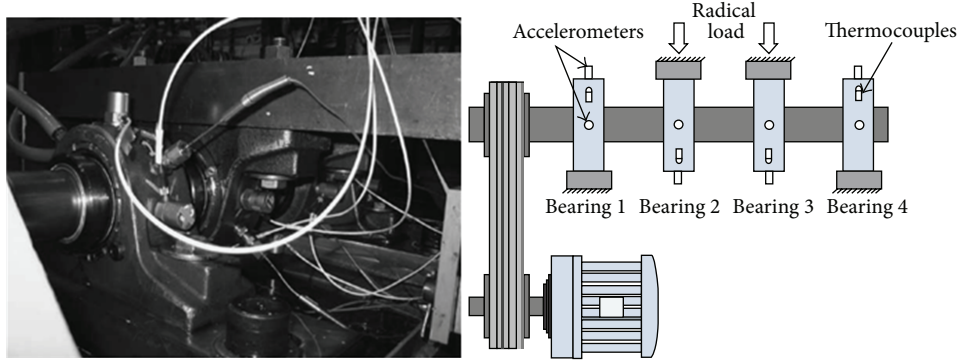


FIGURE 3: Bearing test platform.

where m_i is the mean vector for samples in the i th class and m is the mean vector for all classes. The between-class scatter S_b denotes the scattered degree in different classes while the within-class scatter S_w indicates the concentration among the same class. Therefore, the classification ability can be evaluated through the SS. The weight of each original feature is calculated as follows:

$$w_i = \frac{SS_i}{\sum_{j=1}^{16} SS_j}. \quad (16)$$

Step 4 (JADE for feature fusion). In this step, every feature that is extracted in Step 1 can be seen as a “source signal,” and the JADE is introduced to separate the single model from the 16 source signals that have been normalized and weighted. However, the separated unidimensional vector is mixed by noise. Therefore, the exponentially weighted moving average (EWMA) is used for denoising, which is defined as follows:

$$x_t = (1 - \alpha) x_{t-1} + \alpha x_t, \quad (17)$$

where $\alpha \in [0, 1]$ is the smoothing factor, and it generally takes between 0.05 and 0.25 [36].

Step 5 (prediction model training). Based on Steps 1–4, the fused feature vector with real RUL (X_j, t_j) is input into ELM, where j is the index of the samples. The bias and input weights of hidden nodes in ELM are randomly generated. Then the model with \tilde{N} hidden nodes can be constructed as $\sum_{i=1}^{\tilde{N}} \beta_i g_i(X_j) = \sum_{i=1}^{\tilde{N}} \beta_i g(W_i \cdot X_j + b_i) = o_j$, where w_i is the weight connecting the i th hidden and the input nodes, b_i is the bias of hidden layer nodes, which are randomly assigned, and o is the output value. β_i is the weight connecting the i th hidden and output nodes, where the best value can minimize the error between output o_j and real RUL t_j . And then the fused feature vector of testing samples is input into the trained prediction model for running bearing RUL prediction.

4. Experimental Results and Analysis

4.1. Experimental Data Set Description. In order to evaluate the effectiveness of the proposed feature fusion scheme for

RUL prediction, the entire life cycle bearing data originated from the Center for Intelligent Maintenance Systems [37] is analyzed. The experimental data sets are collected from run-to-failure bearing test on the designed test platform (Figure 3). Four bearings are installed on a shaft. The rotation speed is kept at 2000 r/min by using an AC motor coupled to the shaft via rub belts. A 6000 lb radial load is applied onto the bearing and shaft by using a spring mechanism. A PCB 353B33 High Sensitivity Quartz ICP Accelerometer was installed on bearing housing. A vibration data sample of 20480 points was collected every 10 minutes by a National Instruments DAQCard-6062E data acquisition card. The sampling rate is set as 20 kHz, the vibration data are recorded on the test bench, and more details can be found in [38].

4.2. Original Feature Extraction. In this paper, 100 groups of training samples are selected from the run-to-failure bearing data, and the data samples adjacent to the training data samples are regarded as testing samples. The length of every group's data is set to 1680 points. Take outer-race-fault case as an example; the acquired run-to-failure bearing data is $(X_1, X_2, \dots, X_{984})$, where X_i is the i th acquired data sample and i ranges from 1 to 984. The training samples can be selected as $(X_{90}, X_{99}, \dots, X_{981})$ and the testing samples can be selected as $(X_{91}, X_{100}, \dots, X_{982})$, where $X_j = (x_1, x_2, \dots, x_{1680})$ and j is the index of samples and ranges from 1 to 100.

The 16 original features (as shown in Table 1) are extracted from the training samples and illustrated in Figures 4–6. However, these features are not robust in characterizing the degradation process. Take $F12$ as an example; it performs well in inner-race-fault case and roller-fault case but performs badly in outer-race-fault case. Therefore, it is very important to extract a robust and favorable feature from the original features, which can better express the bearing degradation. In addition, from these extracted features, it can probably be found that the time ratio of each operation stage (including normal, slight degradation, and serious degradation) in the outer-race-fault case, inner-race-fault case, and roller-fault case is about 0.5 : 0.4 : 0.1, 0.6 : 0.3 : 0.1, and 0.5 : 0.2 : 0.3, respectively.

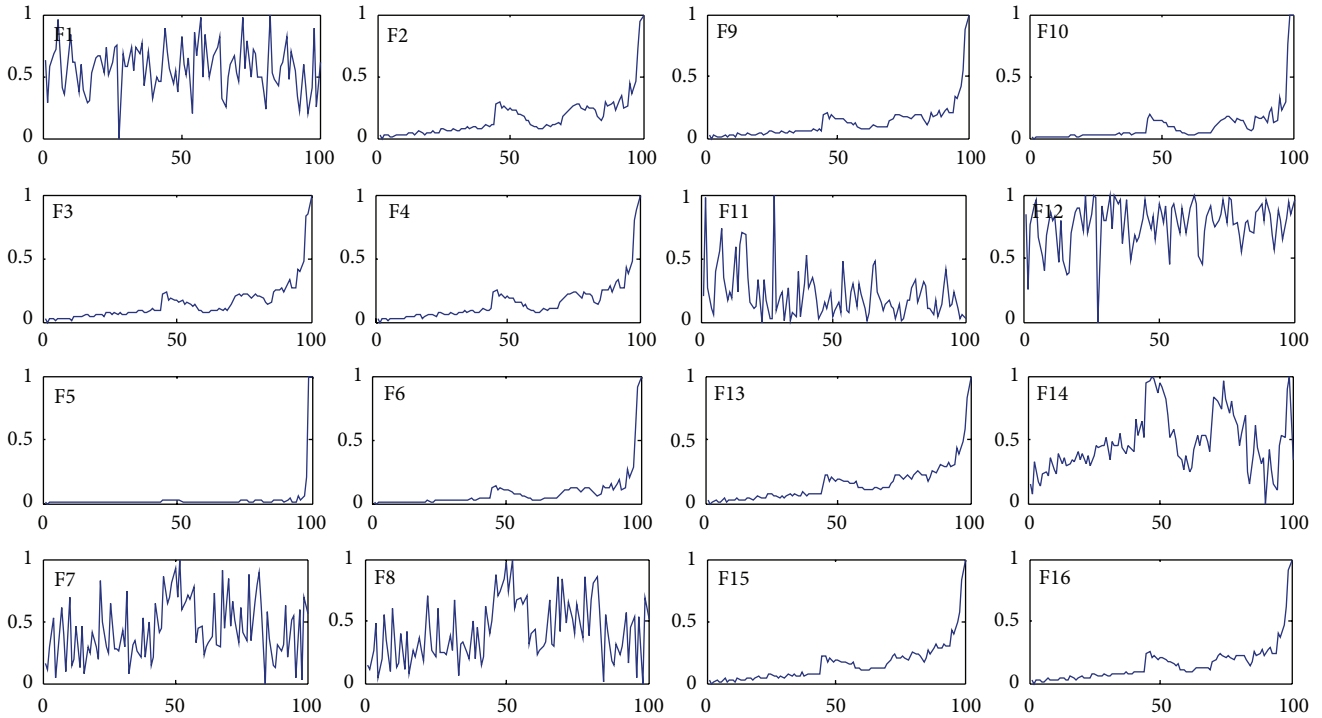


FIGURE 4: The extracted 16 features of outer-race-fault case.

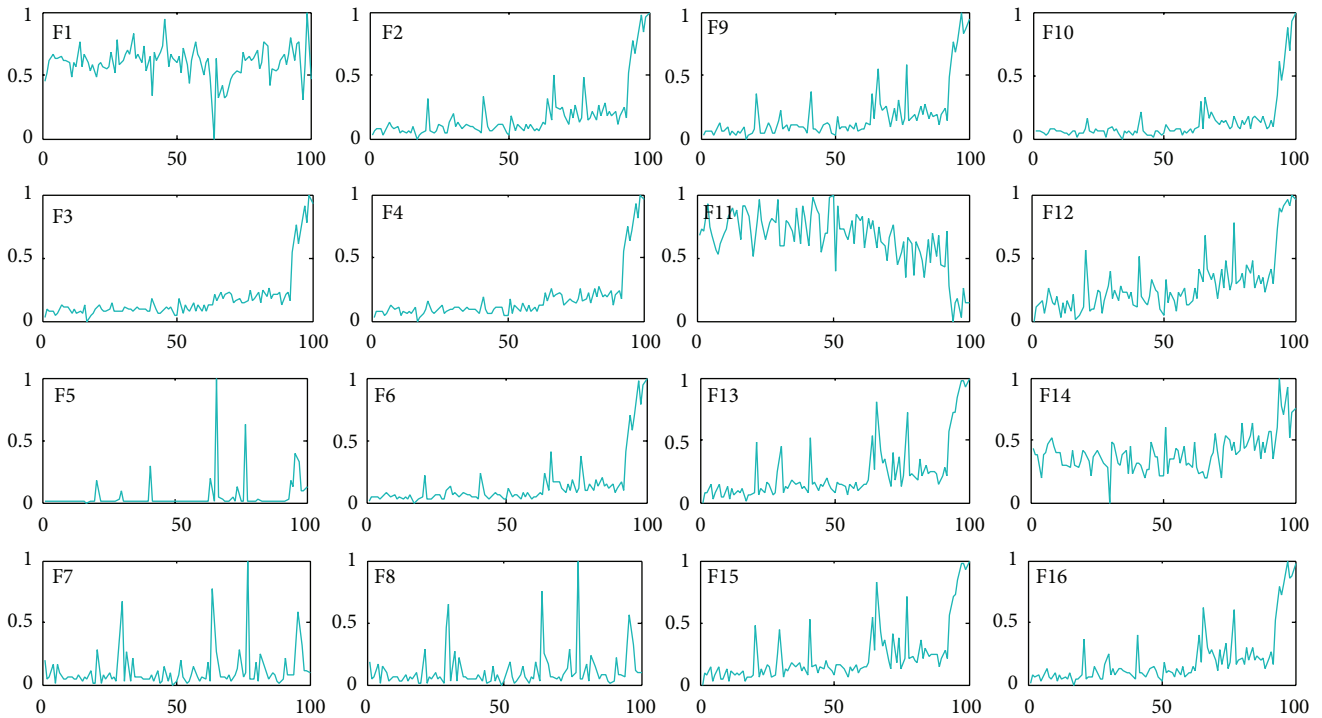


FIGURE 5: The extracted 16 features of inner-race-fault case.

4.3. Phase Space Reconstitution. As the high dimensional phase space can grasp the nature of time series much better, the PSR is then applied to reconstruct the original feature vector to the phase space before the following feature contribution evaluation using SS.

In this paper, the autocorrelation function and Cao's algorithm are employed to identify the time delay and embedding dimension. Take $F1$ of outer-race-fault case as an example; the results are shown in Figure 7. Figure 7(a) shows that $C_i(\tau)$ changes with the increase of delay time, and the

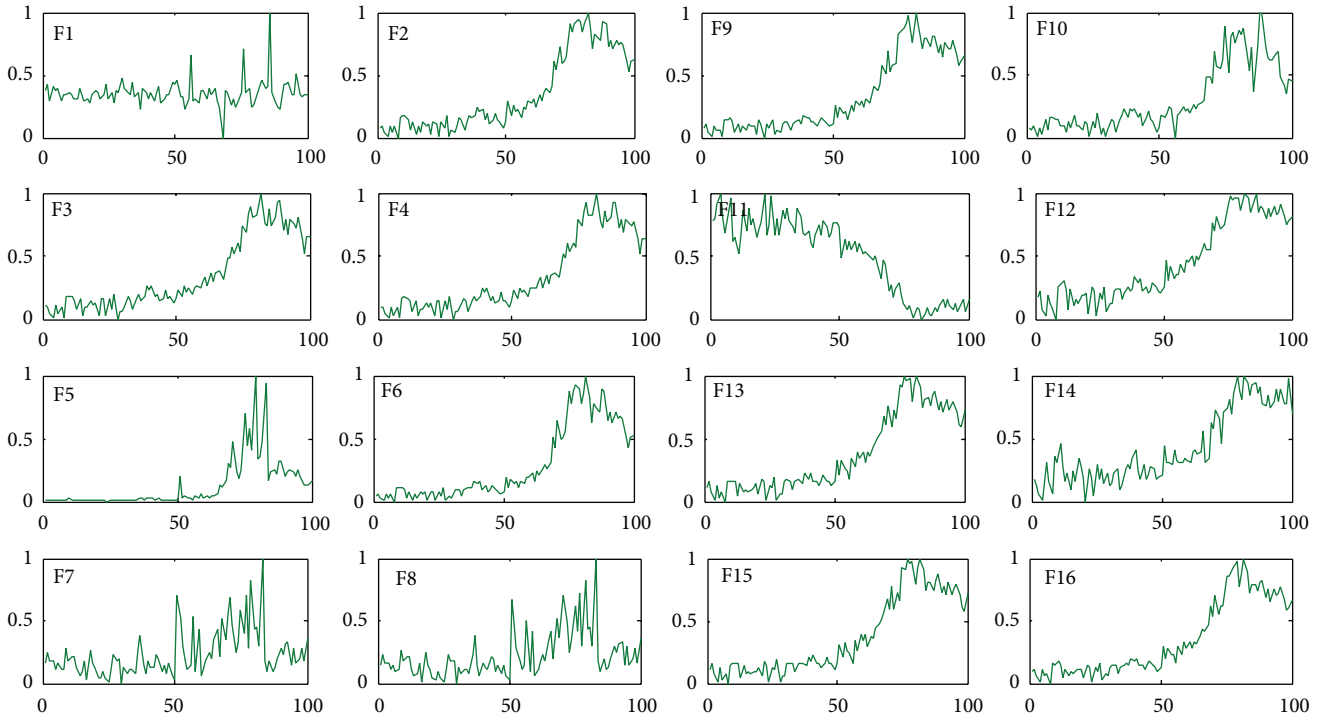


FIGURE 6: The extracted 16 features of roller-fault case.

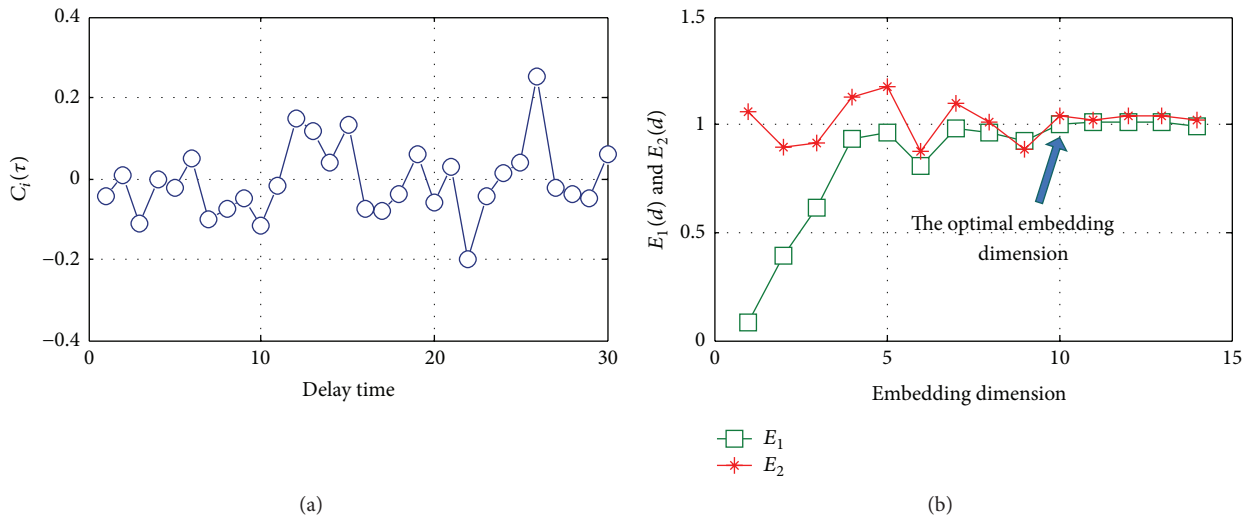


FIGURE 7: (a) Curve of $C_i(\tau)$ and (b) curve of $E_1(d)$ and $E_2(d)$.

minimum τ that makes $C_i(\tau)$ lower than $1 - 1/e$ could be considered as the most optimal time delay. From Figure 7(b), it can be seen that $E_1(d)$ keeps stable when d is larger than 10, and $E_2(d)$ randomly fluctuates with d . Consequently, the time delay is set to 1 and embedding dimension is set to 10. Then other features are processed as the abovementioned methods, and the phase points distribution of different operational states is shown in Figures 8–10 (two dimensions are chosen

to reflect the phase point distribution for better observation experience).

4.4. Feature Fusion Based on JADE and SS. SS is then employed to evaluate the contributions of the feature. As shown in Figure 11, the abscissa is the feature number which has been described in Table 1. The ordinate is the value of SS which is the evaluation index of feature. The greater the

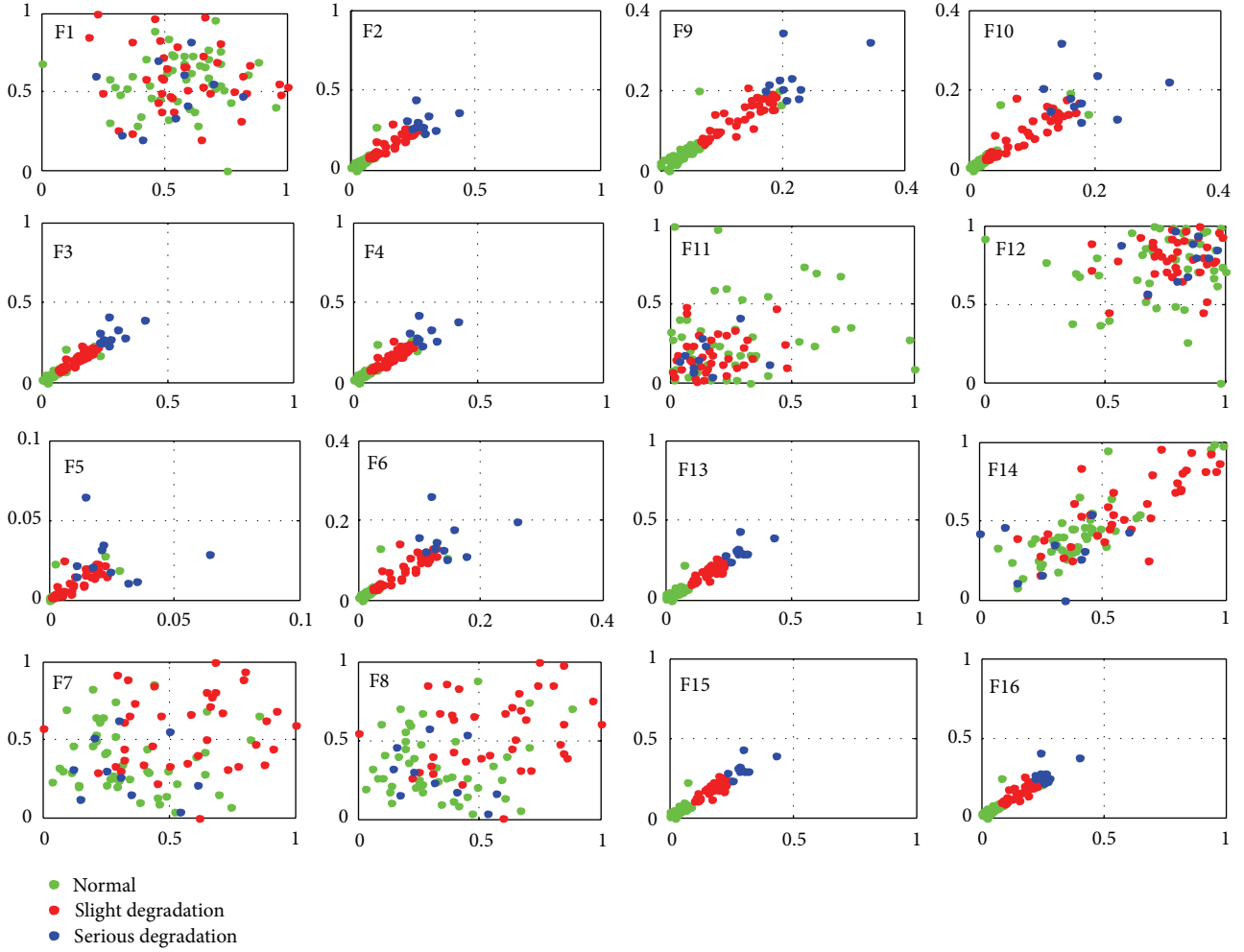


FIGURE 8: The phase points distribution of outer-race-fault case in two dimensions.

SS, the better the identifiable capacity of bearing degradation states. Take inner-race-fault case as an example; it can be seen that $F1$, $F5$, $F7$, and $F8$ cannot effectively characterize or distinguish different degradation stages, and $F3$ provides excellent preference for distinguishing of different degradation states. The weights of every feature can be calculated as $w_i = SS_i / \sum_{j=1}^{16} SS_j$. Besides, $F12$ performs well in roller-fault case but performs badly in outer-race-fault case. Therefore, it is very important to extract a robust feature that could be adapted to different cases.

The JADE is then applied to fuse the weighted features. In order to improve the sensitivity and reliability of the fused feature in the bearing slight degradation, the EWMA is employed for denoising according to (17), where the smoothing factor is selected as 0.2. After all of the steps, the new feature vector can be extracted and shown in Figure 12.

4.5. Prediction Performance. In this step, the fused feature vector and the corresponding real RUL are input into the ELM to construct the RUL prediction model. The data samples adjacent to the training data samples are used to test the model. Results are shown in Figure 13. It can be seen

that the predicted RUL is coinciding fairly well with the real RUL, which indicates that the feature fusion method and constructed ELM model can achieve an ideal result. However, the RUL prediction accuracy is relatively high in the first stage as the vibration is relatively stable. The slight fault stage is always complex and sharp fluctuation also happens at this stage. The vibration at the serious fault stage is more regular, which can lead to higher prediction accuracy.

The PCA and ICA as the representative linear dimensionality reduction methods are applied on the same data to compare with the proposed method. The root mean square error (RMSE) is selected to evaluate the RUL prediction accuracy which can be defined as

$$RMSE = \sqrt{\sum_{i=1}^n (t_i - y_i)^2}, \quad i = 1, 2, \dots, n, \quad (18)$$

where t_i is the real RUL and y_i is the predicted RUL.

Considering the problem of repeatability, the ELM is applied on the same training data 10 times, and the average of the RMSE is selected to evaluate the prediction accuracy. The results of comparison are shown in Tables 2–4. It can

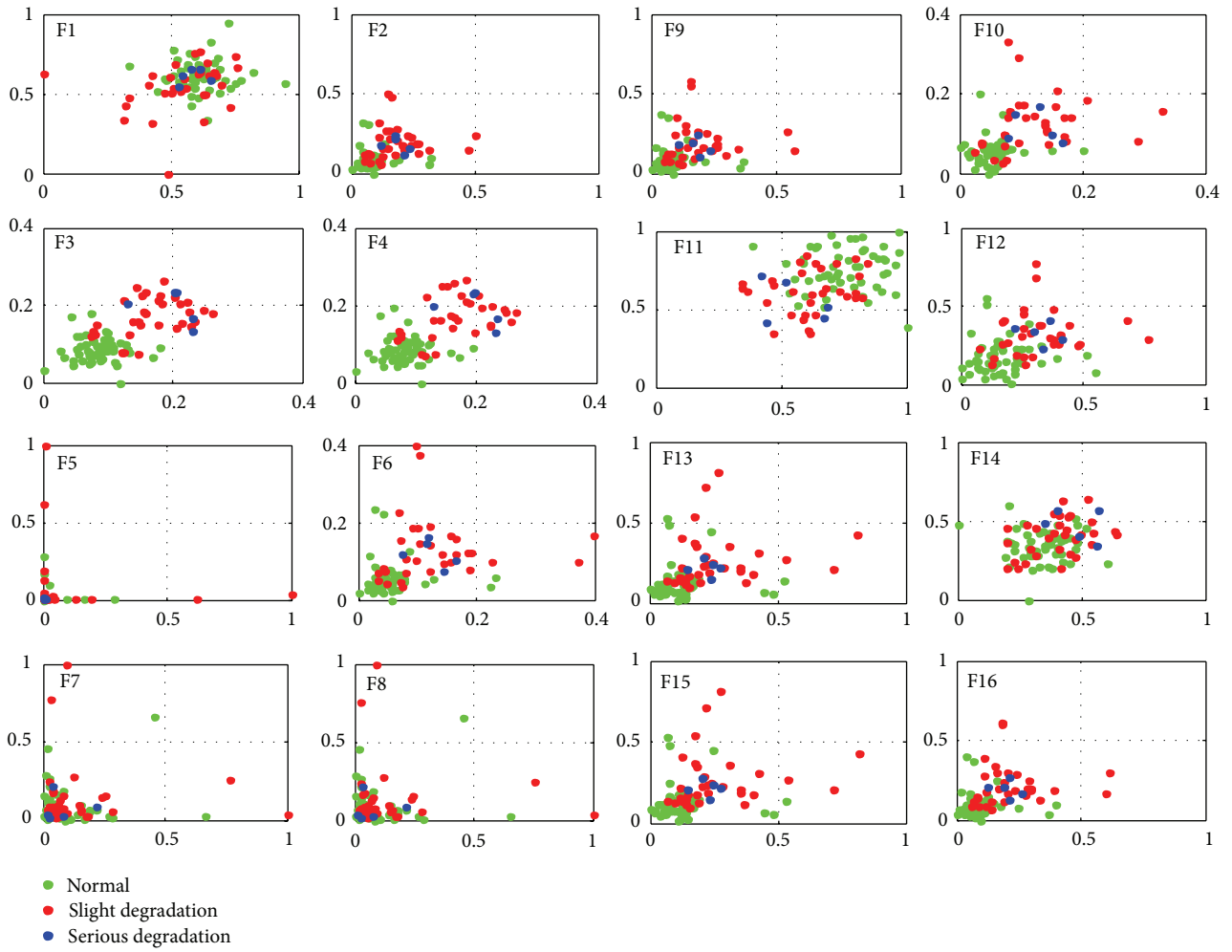


FIGURE 9: The phase points distribution of inner-race-fault case in two dimensions.

TABLE 2: Predicted accuracy of outer-race-fault case.

	Training		Testing	
	Number of training data	RMSE	Number of testing data	RMSE
PCA	100	12.549	100	13.526
ICA	100	12.605	100	14.108
JADE	100	8.872	100	10.018

TABLE 3: Predicted accuracy of inner-race-fault case.

	Training		Testing	
	Number of training data	RMSE	Number of testing data	RMSE
PCA	100	16.073	100	17.548
ICA	100	16.165	100	17.103
JADE	100	14.552	100	14.846

be seen that the proposed feature extraction is better than the traditional two methods, which proves the merits of the proposed method for fusing original features.

5. Conclusions

A novel feature extraction method that integrates the PSR, SS, and JADE is introduced in this paper. The contribution of the original features is considered according to SS that derived from phase point distribution after PSR. The JADE

TABLE 4: Predicted accuracy of roller-fault case.

	Training		Testing	
	Number of training data	RMSE	Number of testing data	RMSE
PCA	100	9.668	100	9.124
ICA	100	9.650	100	9.077
JADE	100	9.105	100	8.227

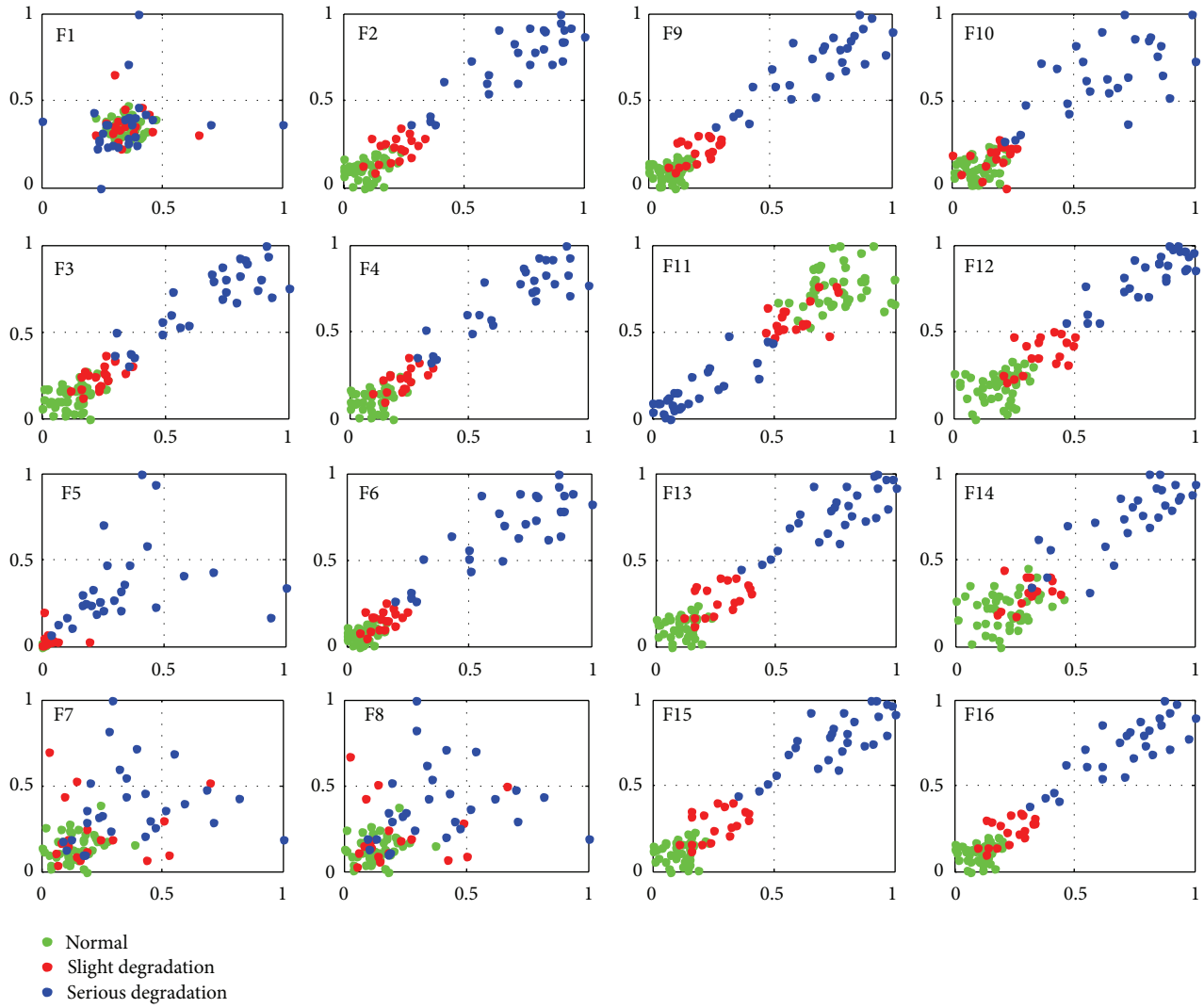


FIGURE 10: The phase points distribution of roller-fault case in two dimensions.

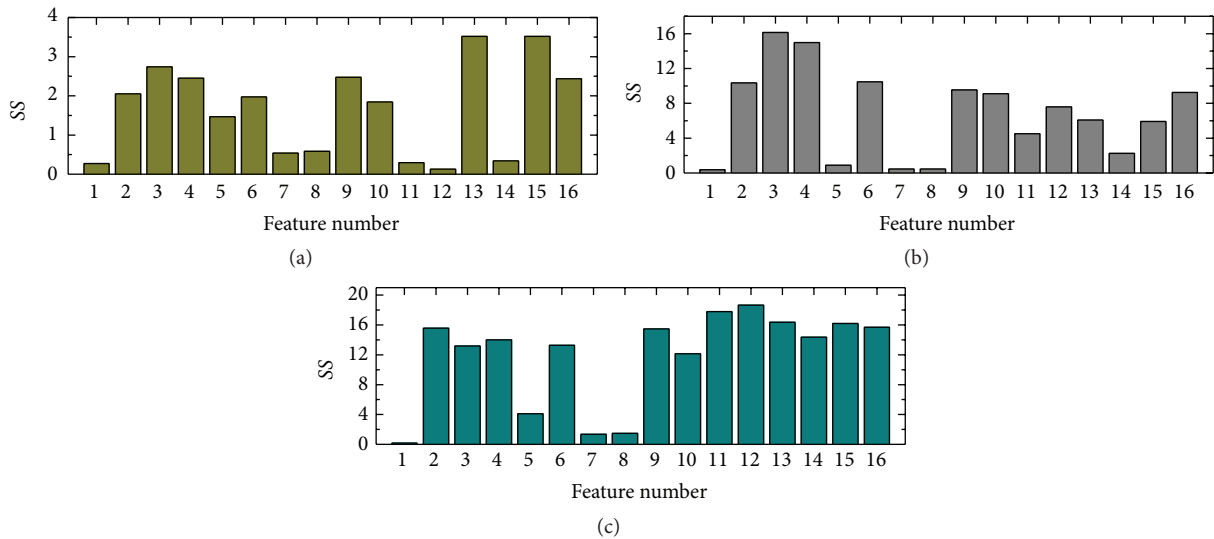


FIGURE 11: The SS of 16 features in (a) outer-race-fault case, (b) inner-race-fault case, and (c) roller-fault case.

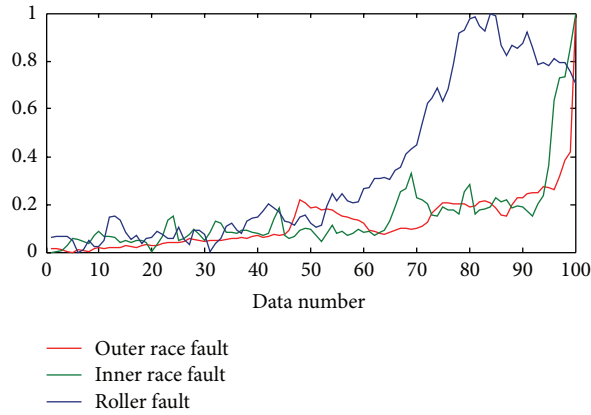


FIGURE 12: The extracted features of bearing degradation in three cases.

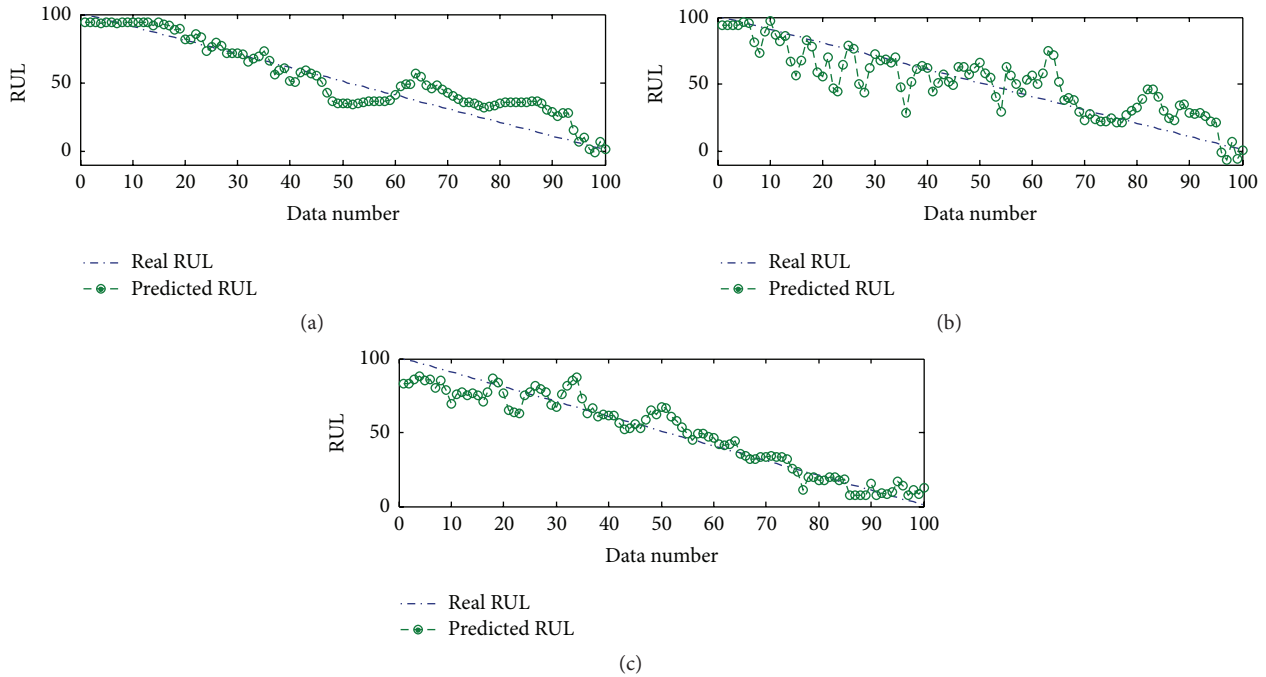


FIGURE 13: Predicted RUL: (a) outer-race-fault case, (b) inner-race-fault case, and (c) roller-fault case.

is employed to fuse the features. Based on the extracted advanced features, the ELM is employed to construct the RUL prediction model.

A set of experimental case studies are presented to verify the effectiveness of the proposed method. The results indicate that the proposed method is appropriate for RUL prediction and even performs better than the traditional methods. The introduced JADE joint PSR and SS feature extraction method is a multifeature fusion method that can inhibit the nonsensitive features while enhancing the superior features. It is also suitable for other machine RUL predictions, such as cutting tools and gears.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

The authors would like to thank the Center for Intelligent Maintenance Systems (IMS), University of Cincinnati, for providing free download of the rolling element bearing fault data sets. This work is supported by the National Natural Science Foundation of China (51505001), Education Department Key Projects of Anhui Province (KJ2013A010), and Natural Science Foundation of Anhui Province (1508085SQE212 and 1408085ME81).

References

- [1] S. Nandi, H. A. Toliyat, and X. Li, "Condition monitoring and fault diagnosis of electrical motors—a review," *IEEE Transactions on Energy Conversion*, vol. 20, no. 4, pp. 719–729, 2005.
- [2] A. Heng, S. Zhang, A. C. C. Tan, and J. Mathew, "Rotating machinery prognostics: state of the art, challenges and opportunities," *Mechanical Systems and Signal Processing*, vol. 23, no. 3, pp. 724–739, 2009.
- [3] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical Systems and Signal Processing*, vol. 20, no. 7, pp. 1483–1510, 2006.
- [4] J. Z. Sikorska, M. Hodkiewicz, and L. Ma, "Prognostic modelling options for remaining useful life estimation by industry," *Mechanical Systems and Signal Processing*, vol. 25, no. 5, pp. 1803–1836, 2011.
- [5] K. Javed, R. Gouriveau, N. Zerhouni, and P. Nectoux, "Enabling health monitoring approach based on vibration data for accurate prognostics," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 1, pp. 647–656, 2015.
- [6] Q. He, X. Ding, and Y. Pan, "Machine fault classification based on local discriminant bases and locality preserving projections," *Mathematical Problems in Engineering*, vol. 2014, Article ID 923424, 12 pages, 2014.
- [7] S. Zhang and W. Li, "Bearing condition recognition and degradation assessment under varying running conditions using NPE and SOM," *Mathematical Problems in Engineering*, vol. 2014, Article ID 781583, 10 pages, 2014.
- [8] X. Wang, Y. Zheng, Z. Zhao, and J. Wang, "Bearing fault diagnosis based on statistical locally linear embedding," *Sensors*, vol. 15, no. 7, pp. 16225–16247, 2015.
- [9] 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.
- [10] 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.
- [11] A. Malhi and R. X. Gao, "PCA-based feature selection scheme for machine defect classification," *IEEE Transactions on Instrumentation and Measurement*, vol. 53, no. 6, pp. 1517–1525, 2004.
- [12] K. Schweizer, P. C. Cattin, R. Brunner, B. Müller, C. Huber, and J. Romkes, "Automatic selection of a representative trial from multiple measurements using Principle Component Analysis," *Journal of Biomechanics*, vol. 45, no. 13, pp. 2306–2309, 2012.
- [13] C. Zang, M. I. Friswell, and M. Imregun, "Structural damage detection using independent component analysis," *Structural Health Monitoring*, vol. 3, no. 1, pp. 69–83, 2004.
- [14] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation*, vol. 15, no. 6, pp. 1373–1396, 2003.
- [15] Z. Zhang and H. Zha, "Principal manifolds and nonlinear dimensionality reduction via tangent space alignment," *SIAM Journal on Scientific Computing*, vol. 26, no. 1, pp. 313–338, 2004.
- [16] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [17] Q. Jiang, M. Jia, J. Hu, and F. Xu, "Machinery fault diagnosis using supervised manifold learning," *Mechanical Systems and Signal Processing*, vol. 23, no. 7, pp. 2301–2311, 2009.
- [18] X. Song, D. Ta, and W. Wang, "Analysis of superimposed ultrasonic guided waves in long bones by the joint approximate diagonalization of eigen-matrices algorithm," *Ultrasound in Medicine and Biology*, vol. 37, no. 10, pp. 1704–1713, 2011.
- [19] G. F. Wang, Y. B. Li, and Z. G. Luo, "Fault classification of rolling bearing based on reconstructed phase space and Gaussian mixture model," *Journal of Sound and Vibration*, vol. 323, no. 3–5, pp. 1077–1089, 2009.
- [20] X. Ding, Q. He, and N. Luo, "A fusion feature and its improvement based on locality preserving projections for rolling element bearing fault classification," *Journal of Sound and Vibration*, vol. 335, pp. 367–383, 2015.
- [21] S. Dong and T. Luo, "Bearing degradation process prediction based on the PCA and optimized LS-SVM model," *Measurement*, vol. 46, no. 9, pp. 3143–3152, 2013.
- [22] W. Caesarendra, A. Widodo, and B.-S. Yang, "Combination of probability approach and support vector machine towards machine health prognostics," *Probabilistic Engineering Mechanics*, vol. 26, no. 2, pp. 165–173, 2011.
- [23] C. Sun, Z. Zhang, and Z. He, "Research on bearing life prediction based on support vector machine and its application," *Journal of Physics: Conference Series*, vol. 305, no. 1, Article ID 012028, 2011.
- [24] G. S. Vijay, H. S. Kumar, P. P. Srinivasa, N. S. Sriram, and R. B. K. N. Rao, "Evaluation of effectiveness of wavelet based denoising schemes using ANN and SVM for bearing condition classification," *Computational Intelligence and Neuroscience*, vol. 2012, Article ID 582453, 12 pages, 2012.
- [25] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, 2006.
- [26] F. Liu, Y. Liu, F. Chen, and B. He, "Residual life prediction for ball bearings based on joint approximate diagonalization of eigen matrices and extreme learning machine," *Proceedings of the Institution of Mechanical Engineers—Part C*, 2015.
- [27] A. Grigorievskiy, Y. Miche, A.-M. Ventelä, E. Séverin, and A. Lendasse, "Long-term time series prediction using OP-ELM," *Neural Networks*, vol. 51, pp. 50–56, 2014.
- [28] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, "Geometry from a time series," *Physical Review Letters*, vol. 45, no. 9, pp. 712–716, 1980.
- [29] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence, Warwick 1980*, vol. 898 of *Lecture Notes in Mathematics*, pp. 366–381, Springer, Berlin, Germany, 1981.
- [30] M. T. Rosenstein, J. J. Collins, and C. J. De Luca, "A practical method for calculating largest Lyapunov exponents from small data sets," *Physica D. Nonlinear Phenomena*, vol. 65, no. 1–2, pp. 117–134, 1993.
- [31] L. Cao, "Practical method for determining the minimum embedding dimension of a scalar time series," *Physica D: Nonlinear Phenomena*, vol. 110, no. 1–2, pp. 43–50, 1997.
- [32] Q. Zhang, P. W.-T. Tse, X. Wan, and G. Xu, "Remaining useful life estimation for mechanical systems based on similarity of phase space trajectory," *Expert Systems with Applications*, vol. 42, no. 5, pp. 2353–2360, 2015.
- [33] J. F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *IEE Proceedings, Part F: Radar and Signal Processing*, vol. 140, no. 6, pp. 362–370, 1993.
- [34] J.-F. Cardoso, "High-order contrasts for independent component analysis," *Neural Computation*, vol. 11, no. 1, pp. 157–192, 1999.

- [35] H.-G. Zhang, S. Zhang, and Y.-X. Yin, "A novel improved ELM algorithm for a realindustrial application," *Mathematical Problems in Engineering*, vol. 2014, Article ID 824765, 7 pages, 2014.
- [36] J. M. Lucas and M. S. Saccucci, "Exponentially weighted moving average control schemes: properties and enhancements," *Technometrics*, vol. 32, no. 1, pp. 1–29, 1990.
- [37] "Bearing Data of Center for Intelligent Maintenance Systems (IMS)," University of Cincinnati, <http://ti.arc.nasa.gov/tech/dash/pcoe/prognostic-data-repository/#bearing>.
- [38] H. Qiu, J. Lee, J. Lin, and G. Yu, "Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics," *Journal of Sound and Vibration*, vol. 289, no. 4-5, pp. 1066–1090, 2006.



Hindawi

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

