Nai-quan Su ✉
Zhi-Jie Zhou
Qing-hua Zhang
Shao-lin Hu
Xiao-xiao Chang

https://doi.org/10.21278/TOF.464036521
ISSN 1333-1124
eISSN 1849-1391

FAULT FEATURE EXTRACTION OF BEARINGS FOR THE PETROCHEMICAL INDUSTRY AND DIAGNOSIS BASED ON HIGH-VALUE DIMENSIONLESS FEATURES

Summary

The time and frequency domain features of a petrochemical unit have a variety of effects on the fault type of bearings, and the signal exhibits nonlinearity, unpredictability, and ergodicity. The detection system's important data are disrupted by noise, resulting in a huge number of invalid and partial records. To reduce the influence of these factors on feature extraction, this work presents a method for the fault feature extraction of bearings for the petrochemical industry and for diagnosis based on high-value dimensionless features. Effective data are extracted from the obtained data using a complex data preprocessing approach, and the dimensionless index is expressed. Then, based on the distribution rule of the dimensionless index, the high-value dimensionless features are retrieved. Finally, to ensure sample completeness, a high-value dimensionless feature augmented model is developed. This approach is applied to the bearing fault experiment platform of a petrochemical unit to effectively classify the bearing fault features, which benefits theoretical guidance for the feature extraction of bearings for a petrochemical unit.

Key words: *petrochemical unit, bearing, high-value dimensionless, feature extraction*

1. Introduction

Petrochemical unit bearings usually operate for a long period in extreme temperatures and severe environments, at high speeds and with a large load. As a result, accurate classification and identification of defect types is challenging. The concept of fault diagnosis is the extraction of high-value defect features. Currently, research on signal analysis, fault feature extraction, and enhanced features mostly focuses on the following three aspects.

(1) The Fourier series extends the time-frequency domain method. The frequency domain signal is converted from the time domain signal. The inverse Fourier transform method is used to convert the frequency domain feature to the time-domain feature. The time-domain property can indicate signal properties such as the amplitude of the signal changing over time, whereas the frequency-domain property can reflect the signal's periodicity and bandwidth variation. The

authors of [1] intended to promote demand for high-quality products. The usage of monitoring systems has become critical, and monitoring systems based on time-frequency signals, images, and convolutional neural networks have been developed. The authors of [2] sought to enhance methods for measuring, processing, and analysing infrasound noise generated by operational wind turbines, as well as to investigate the applicability of the time-frequency ridge transformation to the analysis of wind turbine infrasound data. Meanwhile, the authors of [3] created a straightforward sensor-based system for detecting and identifying train issues. They suggested a linear and quadratic time-frequency analysis of vibration for the purpose of identifying and locating faults. Given the difficulty of extracting and recognizing internal combustion engine time-frequency features, researchers have coupled feature optimization and support vector machine (SVM) classification recognition to provide a new approach for complicated structure failure signal recognition [4,5]. The authors of [6] used a time-frequency feature and adaptive algorithms to perform failure signal recognition when maintaining and identifying wind turbine electronic control system operating data.

(2) The dimensionless parameter describes the relation between two-dimensional parameters. The acquired result is used to define a new parameter, which is insensitive to amplitude and frequency in the normal state and is less affected by machine operating conditions [7,8]. Through dimensionless indices, researchers have attempted to increase diagnostic accuracy. A new intelligent fault diagnosis approach based on composite multi-scale dimensionless indicators and affinity propagation clustering to identify the working conditions of mechanical components is proposed in [9]. Considering that traditional dimensionless indices usually suffer low diagnostic accuracy for mechanical components, a machinery fault diagnosis scheme using redefined dimensionless indicators and minimum redundancy maximum relevance (mRMR) feature selection is proposed in [10]. Meanwhile, the authors of [11] suggested that data fusion in the process of the traditional fault diagnosis method is not sufficiently accurate and they proposed a data fusion method based on mutual dimensionless features. According to the difficulty in early feature extraction of the unit operation system, scholars have analysed the dimensionless index feature distribution trend and fault sensitivity to solve the difficult problem of finding dimensionless index sensitive factors [12].

(3) The representation of sample features is a challenge in the context of fault diagnostics and artificial intelligence. The feature must be enhanced in a data-driven way to account for the lack of representation and completeness. Sample feature augmentation is the selection of secondary data using an SVM. The SVM affects the classification interface and reflects the category edge information [13,14]. In [15], the authors proposed data augmentation optimized for a generative adversarial network (GAN) that effectively leveraged the augmented data to improve the learning of the discriminator and generator. Further, a multi-subject functional magnetic resonance imaging (fMRI) data augmentation method to address the two above-mentioned challenges is proposed in [16], which can improve the decoding accuracy of the target subject. The authors of [17] investigated the small sample problem restriction in the pattern recognition field and proposed an extracting and increasing capacity algorithm by normal feature distribution boundary conditions. In [18], an overlapping of the bearing fault dimensionless index ranges and poor fault diagnosis as well as a proposed fault diagnosis of the rotating machinery based on a dimensionless index and a two-sample distribution test are demonstrated.

A thorough research review highlighted the primary challenges associated with the fault feature extraction of bearings for a petrochemical unit. The first major issue was the residue, noise, and conflicting nature of the original data, which introduced uncertainty into the bearing fault data. There was no unambiguous quantitative distinction between the characteristics. Finally, a wide range of overlapping features contributed to the difficulties in extracting petrochemical unit bearing fault features.

The main contributions and innovation of this work are as follows:

- According to the complex system preprocessing method, effective data were extracted from the collected data, and a dimensionless index was expressed.
- The high-value dimensionless features were extracted based on the dimensionless index distribution law.
- A high-value dimensionless feature augmented model was established to ensure sample completeness.

2. The process of high-value dimensionless feature extraction

The definition of the high-value dimensionless feature is based on five dimensionless characteristics. The many dimensionless properties exert little or no influence on one another and can be easily differentiated. The feature extraction technique, which uses the same dimensionless index as the abscissa and frequency as the ordinate and is based on the dimensionless index normal distribution, results in the extraction of features that do not overlap. This section discusses the process of extracting high-value dimensionless characteristics.

2.1 Construction of the wavelet denoising model and data augmentation

Wavelet denoising is the mathematical processing of wavelet coefficients with varying scales. This is due to the fact that the effective and external noise signals have distinct wavelet coefficient properties, allowing for the implementation of wavelet coefficient processing on the noisy signal. The creation of a wavelet denoising model and data augmentation are detailed in this section. First, we decomposed the number of wavelet decomposition layers and established a threshold to obtain effective signals. Second, we examined the wavelet denoising impact and signal reconstruction. Third, after denoising, the effective data were lost, and the cubic spline data interpolation function was used for data augmentation.

(1) Wavelet transform analysis

Wavelet transform uses two or more wavelet bases to approximate the original function. Assuming that the signal $x(t)$ is a square-integrable function, we have

$$W_f(a,b) = \int_{-\infty}^{+\infty} |a|^{-\frac{1}{2}} x(t) \cdot \psi^* \left(\frac{t-b}{a} \right) \mathrm{d}t \,, \tag{1}$$

where a denotes the scale factor, b is the translation factor, $\psi^*(t)$ is the conjugate function of the wavelet function $\psi(t)$, and $\left|W_f\right|_i$ is the i optimal scale size of the wavelet coefficient.

The petrochemical unit vibration signal is often disturbed by noise to form a large amount of invalid data. In response to this problem, the wavelet transform method was used to decompose the noisy signal, compress the energy to relatively small and large wavelet coefficients, and select an appropriate threshold to filter the noise signal and retain the useful sample signal. The estimated value is

$$S = THR(f(t), \lambda) \,, \tag{2}$$

where *THR* denotes the threshold processing function, and λ represents the set threshold. The generally set threshold parameter is λ [19]

$$\lambda = 1.4826\sqrt{2\log n} \cdot MAD\left(\left|W_f\right|_i\right), \tag{3}$$

where *MAD* denotes the average absolute deviation, n is the original signal $x(t)$ which is decomposed by n-layers of wavelet packets.

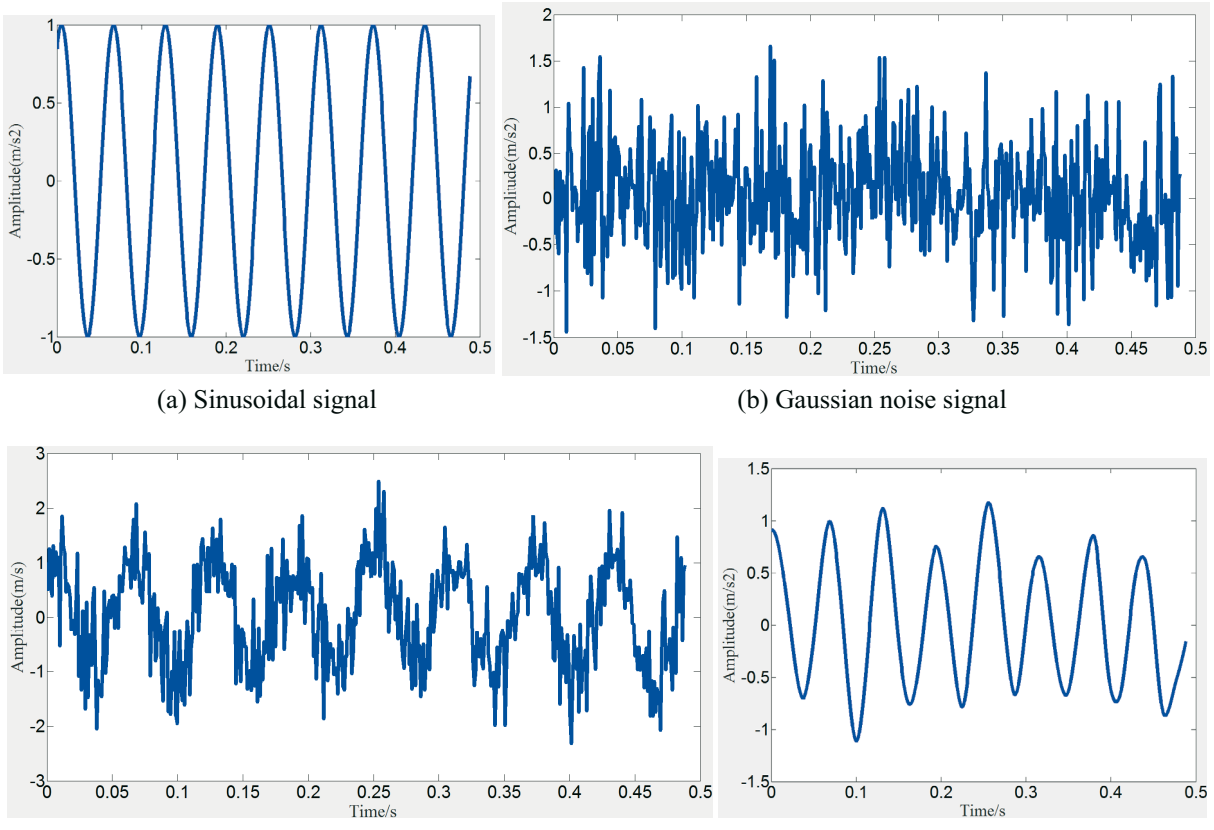
2) Denoising model construction

A denoising model was constructed to grasp the effect of wavelet denoising and signal reconstruction. Assuming that the original signal was sinusoidal, the noise signal was a Gaussian white noise. The Gaussian white noise probability density function obeys the normal distribution, and its density function is

$$P(x)=\frac{1}{\sqrt{2\pi\sigma}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \tag{4}$$

where a sinusoidal signal was added with a Gaussian white noise to simulate a noise environment. μ and σ^2 are the mean and variance of the Gaussian distribution, respectively. When $\mu=0,\ \sigma^2=1$, the distribution is referred to as a standard normal distribution. *A* simulation was performed using a sinusoidal signal and a Gaussian white noise with a noise power of −5 dB in the same dimension.

Figure 1 presents the signal change process. The original signal exhibited clear data and high visibility, as the white Gaussian noise signal exhibited randomness. The signal demonstrated randomness when the original signal was doped with noise. After using wavelet denoising, the original signal was significantly restored, and the noise interference was reduced.



(a) Sinusoidal signal

(b) Gaussian noise signal

(c) Noisy signal

(d) Signal after denoising

Fig. 1 Signal changing process

A comparison of Figure 1(a) and Figure 1 (d) revealed an effective data loss phenomenon mainly caused by the wavelet function selection, the decomposition layer number, and the threshold setting, which led to data loss in the signal reconstruction process. Although the noisy signal can obtain a large amount of effective data after denoising, effective data loss of the complex systems will affect the feature representation.

(3) Data augmentation after denoising

The signal data must be augmented to solve the problem of effective data loss and feature representation completeness. For the main augmentation process, the cubic spline data interpolation function was developed and its change rule was deduced from the existing data to achieve data completeness.

2.2 Dimensionless index

The dimensional parameters often change due to the influence of load, speed, and external noise. The dimensionless index must be introduced to reduce the different factors that influence the dimensional parameters and fault signal excessive overlapping problem.

The dimensionless index is obtained by comparing two-dimensional parameters, which are not sensitive to disturbance in the rotating machinery and can better express the fault characteristics [20]. The dimensionless indices are the waveform index (i.e., the degree of deviation and distortion of the waveform compared with the sine wave), the impulse index (indicating the impact nature of the wave), the margin index (indicating the waveform size), the peak index (i.e., whether or not the waveform has an impact; it is an index of the peak height), and the kurtosis index (i.e., waveform sharpness, whether or not there is an impact), which are defined as

$$\zeta_x = \frac{\left[\int_{-\infty}^{+\infty} |x|^l p(x)\mathrm{d}x\right]^{\frac{1}{l}}}{\left[\int_{-\infty}^{+\infty} |x|^m p(x)\mathrm{d}x\right]^{\frac{1}{m}}} = \frac{\sqrt[l]{E(|x|^l)}}{\sqrt[m]{E(|x|^m)}} \tag{5}$$

where x denotes the amplitude of the vibrational random time domain signal, $p(x)$ is the signal x probability density function, and l and m are the numerator and denominator coefficients, respectively. The dimensionless index obtained by Eq. (5) is presented in Table 1 [21].

The following equation verifies that the dimensionless feature in the normal state is less affected by the working conditions and is only sensitive to changes in the probability density function $p(x)$

$$I_f = \frac{\lim\limits_{l\to\infty}\left[\int_{-\infty}^{+\infty} |x|^l p(x)\mathrm{d}x\right]^{\frac{1}{l}}}{\left[\int_{-\infty}^{+\infty} |x| p(x)\mathrm{d}x\right]} = \frac{X_{\max}}{|X|}. \tag{6}$$

Where $\frac{X_{\max}}{|X|} = \frac{X_{\max}}{\frac{x_1 + x_2 + \cdots + x_n}{N}}$, the maximum value is $X_{\max} = \max\{|x_i|\}$, and when the working conditions change, the collected sample data will also change, and we obtain

$$\begin{aligned} \frac{X'_{\max}}{|X|'} &= \frac{X_{\max} + (\pm a_i)}{\frac{(x_1 \pm a_1) + (x_2 \pm a_2) + \cdots + (x_n \pm a_n)}{N}} \\ &= \frac{X_{\max}}{\frac{x_1 + x_2 + \cdots + x_n}{N}} + \frac{(\pm a_i)}{\frac{(\pm a_1) + (\pm a_2) + \cdots + (\pm a_n)}{N}} \\ &= I_f + \frac{(\pm a_i)}{\frac{(\pm a_1) + (\pm a_2) + \cdots + (\pm a_n)}{N}} \end{aligned} \tag{7}$$

Table 1 Dimensionless index formula

Dimensionless index	Coefficient of numerator *l* and denominator *m*	Calculation formula
Waveform Index	$l=2;\ m=1$	$S_f=\frac{\left[\int_{-\infty}^{+\infty}\lvert x\rvert^2\ p(x)\mathrm{d}x\right]^{\frac{1}{2}}}{\left[\int_{-\infty}^{+\infty}\lvert x\rvert\ p(x)\mathrm{d}x\right]}=\frac{X_{rms}}{\lvert X\rvert}$
Pulse Index	$l\to\infty;\ m=1$	$I_f=\frac{\lim\limits_{l\to\infty}\left[\int_{-\infty}^{+\infty}\lvert x\rvert^l\ p(x)\mathrm{d}x\right]^{\frac{1}{l}}}{\left[\int_{-\infty}^{+\infty}\lvert x\rvert\ p(x)\mathrm{d}x\right]}=\frac{X_{\max}}{\lvert X\rvert}$
Margin Index	$l\to\infty;\ m=\frac{1}{2}$	$CL_f=\frac{\lim\limits_{l\to\infty}\left[\int_{-\infty}^{+\infty}\lvert x\rvert^l\ p(x)\mathrm{d}x\right]^{\frac{1}{l}}}{\left[\int_{-\infty}^{+\infty}\lvert x\rvert^{\frac{1}{2}}\ p(x)\mathrm{d}x\right]^2}=\frac{X_{\max}}{X_r}$
Peak index	$l\to\infty;\ m=2$	$C_f=\frac{\lim\limits_{l\to\infty}\left[\int_{-\infty}^{+\infty}\lvert x\rvert^l\ p(x)\mathrm{d}x\right]^{\frac{1}{l}}}{\left[\int_{-\infty}^{+\infty}\lvert x\rvert^2\ p(x)\mathrm{d}x\right]^{\frac{1}{2}}}=\frac{X_{\max}}{X_{rms}}$
Kurtosis index	*N/A*	$K_v=\frac{\int_{-\infty}^{\infty}x^4p(x)\mathrm{d}x}{\left(\int_{-\infty}^{\infty}x^2p(x)\mathrm{d}x\right)^2}$

In the formula, $X_{\max}$ is the maximum value, X_r is the root square amplitude, X_{rms} is the root mean square value, and $|X|$ is the average amplitude [22].

The random signal probability density function is

$$p(x)=\lim_{\Delta x\to 0}\frac{P_r\left[x<x(t)<x+\Delta x\right]}{\Delta x}=\lim_{\Delta x\to 0}\frac{1}{\Delta x}\left[\lim_{T\to\infty}\frac{T_x}{T}\right]. \tag{8}$$

In the formula, P_r represents a continuous random variable with any real number $x<x(t)<x+\Delta x$, T_x is the time sum of the signal amplitude falling between x and $x+\Delta x$, and T is the sample length.

Although the working condition changes, the data $\Delta x\to 0$ ($\Delta x=\frac{(\pm a_i)}{\frac{(\pm a_1)+(\pm a_2)+\cdots+(\pm a_n)}{N}}\to 0$) are collected in a unit time in such a way that the dimensionless feature will be less affected by the working condition, and it is only sensitive to the change in the probability density function, which is

$$\frac{X'_{\max}}{|X|'}=\frac{X_{\max}}{|X|}=I_f. \tag{9}$$

The dimensionless index is less affected by the changes in the working conditions. For further research, the dimensionless feature of the mathematical relationship calculation is still less affected by the changes in the working conditions. Assuming that the dimensionless features F_1 and F_2 are examples for verification, we have $F_1 = \frac{a_1(x)}{a_2(x)}$ and $F_2 = \frac{b_1(x)}{b_2(x)}$, where $a_1(x)$, $a_2(x)$, $b_1(x)$, and $b_2(x)$ are all dimension parameters [25].

Assuming F_1+F_2, there is

$$\begin{aligned} F_1 + F_2 &= \frac{a_1(x)}{a_2(x)} + \frac{b_1(x)}{b_2(x)} \\ &= \frac{a_1(x)b_2(x) + a_2(x)b_1(x)}{a_2(x)b_2(x)} = \frac{F'_{12}(x)}{F''_{22}(x)}. \end{aligned} \tag{10}$$

Assuming F_1-F_2, there is

$$\begin{aligned} F_1 - F_2 &= \frac{a_1(x)}{a_2(x)} - \frac{b_1(x)}{b_2(x)} \\ &= \frac{a_1(x)b_2(x) - a_2(x)b_1(x)}{a_2(x)b_2(x)} = \frac{F'_{12}(x)}{F''_{22}(x)}. \end{aligned} \tag{11}$$

Assuming $F_1 \cdot F_2$, there is

$$F_1 \cdot F_2 = \frac{a_1(x)}{a_2(x)} \cdot \frac{b_1(x)}{b_2(x)} = \frac{a_1(x)b_1(x)}{a_2(x)b_2(x)} = \frac{F''_{11}(x)}{F''_{22}(x)}. \tag{12}$$

Assuming $F = \frac{F_1}{F_2}$, there is

$$F = F_1 / F_2 = \frac{a_1(x)}{a_2(x)} \Big/ \frac{b_1(x)}{b_2(x)} = \frac{a_1(x)b_2(x)}{a_2(x)b_1(x)} = \frac{F''_{12}(x)}{F''_{21}(x)}. \tag{13}$$

The dimensionless index added, subtracted, multiplied, and divided using mathematical operations is still dimensionless, further demonstrating that the dimensionless index is less affected by the changes in the working conditions.

2.3 The extraction method of the high-value dimensionless feature

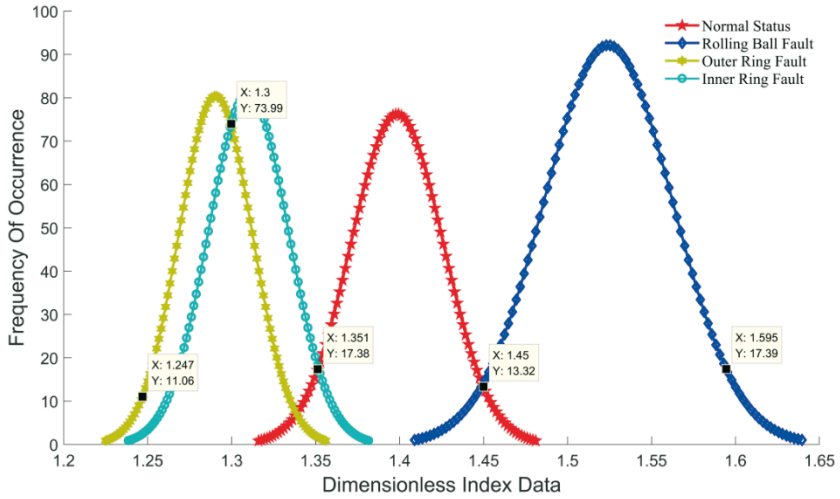
In this section, the method of normal distribution is used to observe dimensionless feature distribution, and then, according to the dimensionless feature distribution, high-value dimensionless features are extracted. Figure 2 shows the normal distribution of the waveform index and kurtosis index. It can be seen from the figure that the waveform index and kurtosis index of the bearing outer and inner ring faults excessively overlap, and the normal state and ball fault show small overlapping. The interval distribution range of the non-overlapping part of the waveform index and kurtosis index is as follows:

The distribution of the waveform index from the bearing outer ring fault interval was (1.246 1.29). The bearing inner ring fault interval was (1.29 1.351). The normal-state interval was (1.351 1.449). The bearing inner ring fault interval was (1.449 1.594), indicating a non-overlapping part.

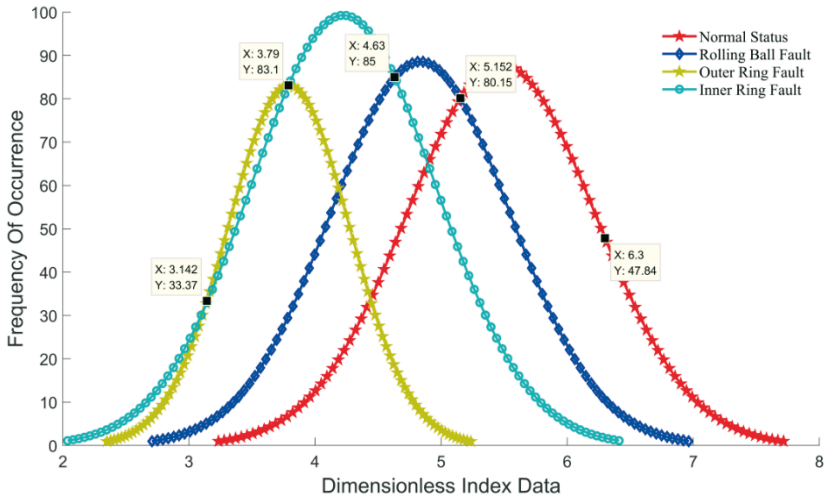
The kurtosis index distribution demonstrated many overlapping parts of different fault features. The bearing outer ring fault interval was (3.142 3.79), the bearing inner ring fault

interval was (3.79 4.63), and the bearing ball fault interval was (4.63 5.152). The normal-state interval was (5.152 6.233), indicating a non-overlapping part.

The dimensionless features of the non-overlapping regions extracted are called high-value dimensionless features.



(a) waveform index



(b) kurtosis index

Fig. 2 Distribution of the waveform index and kurtosis index in different bearing faults

3. Bearing fault high-value dimensionless feature extraction

Experimental research was conducted with the help of the fault diagnosis platform, data resources, and petrochemical monitoring equipment test conditions (Figure 3). The petrochemical monitoring and diagnosis platform can simulate different typical fault states, such as bearing outer and inner ring fault, bearing ball fault, and gear fault. Moreover, this platform can achieve a composite fault design through faulty component combinations. The experimental platform consisted of four parts, namely, an electric motor (the motor speed was 1000 min^{-1}), load controller, gearbox, and air compressor.

The fault types in this experimental study are the bearing inner and outer ring fault, and the bearing ball fault. The rolling bearing parameters are as follows: bearing outer diameter 110 mm; bearing inner diameter 50 mm; and thickness 27 mm. The rolling diameter and rolling number are 14 mm and 13, respectively. Figure 4 presents normal and faulty bearings.

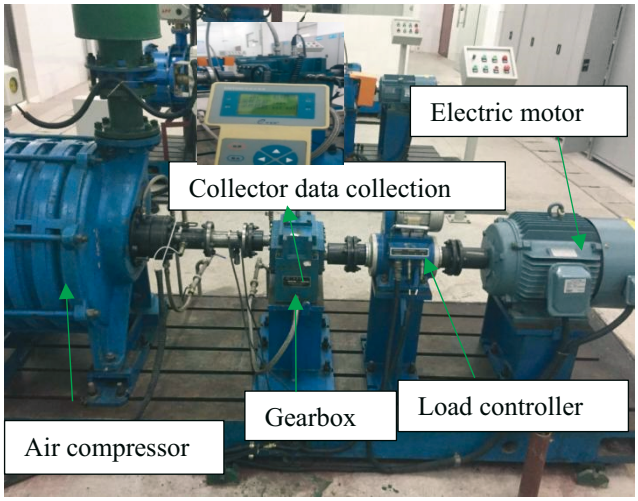


Fig. 3 Petrochemical monitoring equipment fault diagnosis platform

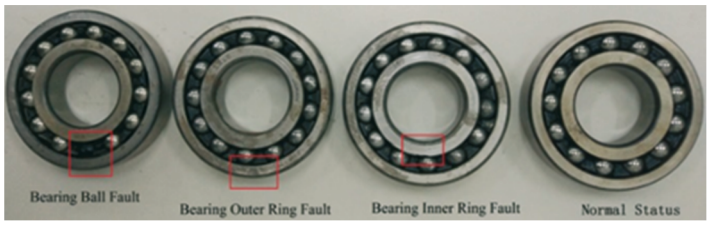


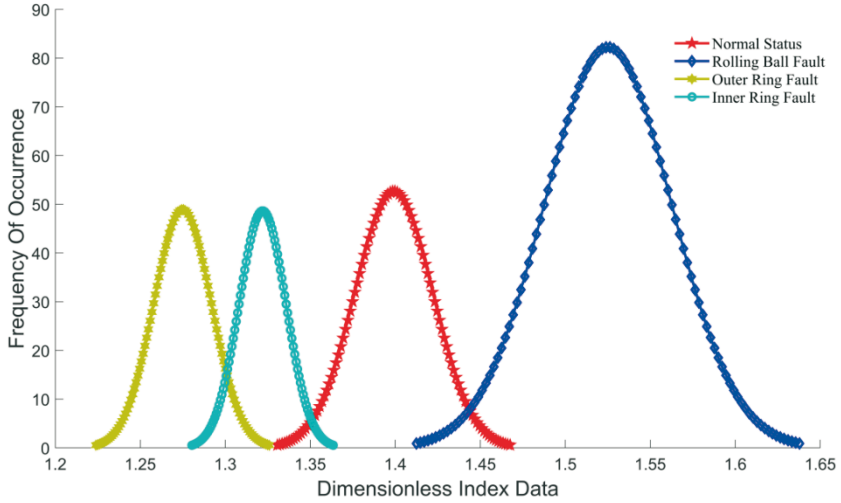
Fig.4 Presents normal and fault bearings

An analysis of the waveform and kurtosis indices revealed that although the distributions of the different faults of the same dimensionless index overlapped, a sensitive interval existed between each fault. Therefore, we performed high-value dimensionless feature extraction based on the fault feature sensitivity interval to solve the problems of feature overlapping, coupling, and poor sensitivity. The same dimensionless index was taken as the abscissa, and the frequency was the ordinate. According to the dimensionless index distribution, two adjacent sample curve intersections were chosen as the high-value dimensionless feature. Table 2 presents the high-value dimensionless feature interval.

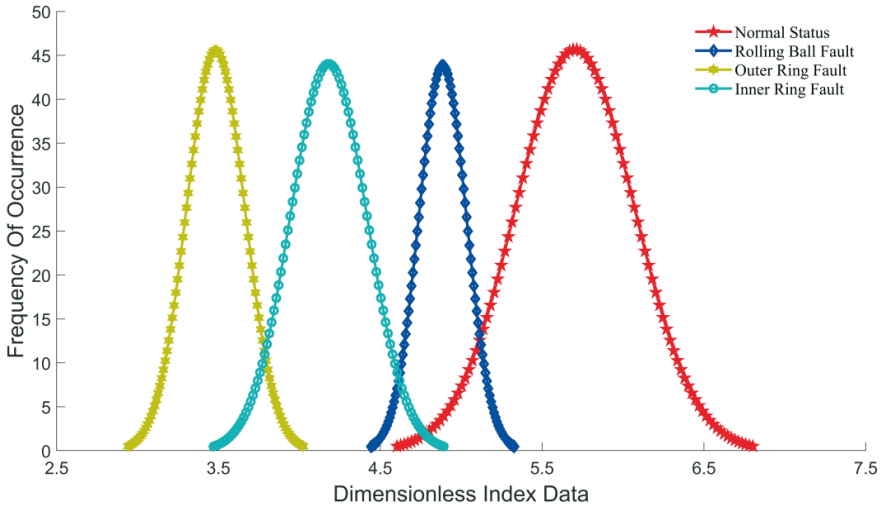
Table 2 Interval of high-value dimensionless feature

Index \ Status	Normal Status	Bearing Ball Fault	Bearing Outer Ring Fault	Bearing Inner Ring Fault
Waveform Index	(1.351 1.450)	(1.450 1.595)	(1.247 1.300)	(1.300 1.351)
Peak Index	(1.001 1.861)	(0.267 0.993)	(3.000 3.780)	(3.780 5.000)
Margin Index	(1.739 2.991)	(0.532 1.735)	(4.1 5.949)	(5.949 8.000)
Pulse Index	(1.524 2.504)	(0.388 1.496)	(3.8 4.936)	(4.936 7.000)
Kurtosis Index	(5.154 6.300)	(4.633 5.152)	(3.142 3.79)	(3.79 4.633)

According to the high-value dimensionless features extracted in Figure 2, Figure 5 is augmented to ensure a balance of the extracted dimensionless features. This section establishes random sampling generation based on the Markov chain sampling model to achieve an augmented sample and ensure extracted high-value dimensionless feature completeness.



(a) High-value waveform index



(b) High-value kurtosis index

Fig. 5 Distribution of different faults of bearing high-value waveform index and kurtosis index

The high-value waveform index distribution in Figure 5(a) revealed that the high-value dimensionless features were augmented by the Markov chain random sampling model. We observed fewer intersections and high distinguishability. Meanwhile, the high-value kurtosis index distribution in Figure 5 (b) exhibited a significantly different fault distinction.

4. Evaluation of the petrochemical unit bearing fault high-value dimensionless feature

This section evaluates the high-value dimensionless features using the SVM + PSO (Particle swarm optimization) intelligent pattern recognition method for evaluation. The model output result was utilized as the evaluation index.

The data processing process: (1) Using the collector to randomly extract 1000 sets of fault sample features on the rotating machinery fault diagnosis platform, 70% of the sample features are used as training samples, and 30% of the sample features are used as test samples. (2) The training sample is used as the label fault sample feature of SVM + PSO; the test sample is used as the unknown unlabelled fault sample feature of SVM + PSO. (3) Without considering the unknown unlabelled fault sample features, SVM + PSO finds the optimal hyperplane to separate the two types of labelled samples.

We selected the high-value dimensionless features as the training and test samples and set three conditions of use: (1) the same training sample and different test samples; (2) the same test sample and different training samples; and (3) random training and test samples. These conditions will help reduce the fault sample evaluation process contingency and ensure accurate evaluation. Tables 3 to 5 present the dimensionless feature evaluation results. Tables 6 to 8 show the high-value dimensionless feature evaluation results.

The evaluation results in Tables 3 to 5 indicate the normal-state dimensionless feature evaluation results: 35%, 23.67%, 53.5%, 100%, 100%, and 100%. The bearing outer ring fault dimensionless feature evaluation results were 100%, 0%, 0%, 0%, 89.85%, and 84.77%. Meanwhile, the bearing inner ring fault dimensionless feature evaluation results were 100%, 100%, 100%, 100%, 69.65%, and 77.56%, and the bearing ball fault dimensionless feature evaluation result was 100%.

Table 3 The dimensionless feature evaluation results of the same training sample and different test samples

Failure category	Training samples	Test samples	Correct number	Number of errors	Correct rate
Normal Status	600	200	70	130	35%
		300	71	229	23.67%
Bearing Ball Fault		200	200	0	100%
		300	300	0	100%
Bearing Outer Ring Fault		200	0	200	0%
		300	0	200	0%
Bearing Inner Ring Fault		200	200	0	100%
		300	300	0	100%

Table 4 The dimensionless feature evaluation results of the same test sample and different training samples

Failure category	Training samples	Test samples	Correct number	Number of errors	Correct rate
Normal Status	700	200	107	93	53.5%
	800		200	0	100%
Bearing Ball Fault	700		200	0	100%
	800		200	0	100%
Bearing Outer Ring Fault	700		0	200	0%
	800		0	200	0%
Bearing Inner Ring Fault	700		200	0	100%
	800		200	0	100%

Table 5 The dimensionless feature evaluation results of random training and test samples

Failure category	Training samples	Test samples	Correct number	Number of errors	Correct rate
Normal Status	703	197	197	0	100%
	792	208	208	0	100%
Bearing Ball Fault	695	205	205	0	100%
	810	190	190	0	100%
Bearing Outer Ring Fault	703	197	177	20	89.85%
	803	197	167	30	84.77%
Bearing Inner Ring Fault	699	201	140	61	69.65%
	795	205	159	46	77.56%

Tables 3 to 5 present the recognition effect of the normal state and bearing outer and inner ring fault, in which the changes in the training and test samples were more obvious. The main

reason for this is the dimensionless sample feature data coupling and overlapping, which results in poor robustness and evaluation results. Tables 6 to 8 present the evaluation results of the high-value dimensionless features.

Table 6 High-value dimensionless feature evaluation results of the same training sample and different test samples

Failure category	Training samples	Test samples	Correct number	Number of errors	Correct rate
Normal Status	600	200	200	0	100%
		300	300	0	100%
Bearing Ball Fault		200	200	0	100%
		300	300	0	100%
Bearing Outer Ring Fault		200	200	0	100%
		300	300	0	100%
Bearing Inner Ring Fault		200	200	0	100%
		300	300	0	100%

Table 7 High-value dimensionless feature evaluation results of the same test sample and different training samples

Failure category	Training samples	Test samples	Correct number	Number of errors	Correct rate
Normal Status	700	200	200	0	100%
	800		200	0	100%
Bearing Ball Fault	700		200	0	100%
	800		200	0	100%
Bearing Outer Ring Fault	700		200	0	100%
	800		200	0	100%
Bearing Inner Ring Fault	700		200	0	100%
	800		200	0	100%

Table 8 High-value dimensionless feature evaluation results of random training and test samples

Failure category	Training samples	Test samples	Correct number	Number of errors	Correct rate
Normal Status	691	209	209	0	100%
	808	192	192	0	100%
Bearing Ball Fault	689	211	211	0	100%
	781	219	219	0	100%
Bearing Outer Ring Fault	712	188	188	0	100%
	811	189	189	0	100%
Bearing Inner Ring Fault	708	192	192	0	100%
	800	200	200	0	100%

The evaluation results in Tables 6 to 8 indicate the high-value dimensionless sample feature evaluation result of 100%. The evaluation results reveal that the high-value dimensionless features extracted are beneficial in eliminating overlap and coupling between each fault and in improving the classification and recognition effect of the faults. At the same time, the evaluation results of the high-value dimensionless features were less affected by the training and test samples.

5. Conclusion

Verification was performed using the high-value dimensionless features of the normal state, bearing ball fault, and bearing inner and outer ring faults of the petrochemical unit bearing. The experimental results demonstrate that the extracted high-value dimensionless

features assist in reducing the influence of nonlinearity, unpredictability, and ergodicity and so improve the effectiveness of fault classification and recognition. Additionally, the results of the evaluation of the high-value dimensionless features are less affected by the training and test samples and demonstrate a high degree of robustness.

Acknowledgements

The research was partially supported by the Key Project of the Natural Science Foundation of China (Grant No. 61933013) and the National Nature Science Foundation of China (Grant Nos. 61673127, 61973094). The work described in this paper was also partially supported by the Natural Science Foundation of Guangdong (2022A1515010599) and Guangdong University key platform Youth Innovative Talents Project (2019KQNCX085), the Maoming Science and Technology Project (No. 201805, 2020522, 2020S004), and the PhD project of Guangdong University of Petrochemical Technology (No. 2020bs006).

REFERENCES

[1] Hubner, H.B.; Duarte, M.; Silva, R. Automatic Grinding Burn Recognition Based on Time- Frequency Analysis and Convolutional Neural Networks, International Journal of Advanced Manufacturing Technology 2020, 110(7-8),1833- 1849. https://doi.org/10.1007/s00170-020-05902-w

[2] Boczar, T.; Zmarzy, D.; Kozio, M.; Wotzka, D. The Application of Time-Frequency Ridge Transformation for the Analysis of Infrasound Signals Generated by Wind Turbines, Applied Acoustics 2021, 177. https://doi.org/10.1016/j.apacoust.2021.107961

[3] Barman, J.; Hazarika, D. Linear and Quadratic Time-Frequency Analysis of Vibration for Fault Detection and Identification of NFR Trains, IEEE Transactions on Instrumentation and Measurement 2020, 69(11), 8902 - 8909. https://doi.org/10.1109/TIM.2020.2998888

[4] Durak, L.; Arikan, O. Short-Time Fourier Transform: Two Fundamental Properties and an Optimal Implementation, IEEE Transactions on Signal Processing, 2003, 51(5), 1231-1242. https://doi.org/10.1109/TSP.2003.810293

[5] Cai, Y.P.; Li, A.H.; He, Y.P.; Wang, T.; Feng, G.Y. Fault Diagnosis Method Based on Vibration Spectrum Time-Frequency Image Feature & SVM Parameters Synchronization Optimization Recognition, Transactions of CSICE 2012, 30(4), 377-383.

[6] Xiao, C.; Wang, Y.L. Fault Diagnosis Research of Wind Turbines Electronic Control System Based on the Advanced Control Principles, Journal of North China Institute of Aerospace Engineering 2017, 27(2), 8-11.

[7] Wang, Z.W.; Zhang, Q.H.; Xiong, J.B.; Ming, X.; Sun, G.X.; He, J. Fault Diagnosis of a Rolling Bearing Using Wavelet Packet Denoising and Random Forests, IEEE Sensors Journal 2018, 17(17), 5581- 5588. https://doi.org/10.1109/JSEN.2017.2726011

[8] Hoang, N.B.; Kang,H.J. A Model-based Fault Diagnosis Scheme for Wheeled Mobile Robots, International Journal of Control Automation and Systems 2014, 12(3), 637-651. https://doi.org/10.1007/s12555-013-0012-1

[9] Hu, Q.; Zhang, Q.H.; Si, X.S.; Sun, G.X.; Qin, A.S. Intelligent Fault Diagnosis Approach Based on Composite Multi-scale Dimensionless Indicators and Affinity Propagation Clustering, IEEE Sensors Journal 2020, 20(19), 11439 - 11453. https://doi.org/10.1109/JSEN.2020.2995817

[10] Hu, Q.; Si, X.S.; Qin, A.S.; Lv, Y.R.; Zhang, Q.H. Machinery Fault Diagnosis Scheme Using Redefined Dimensionless Indicators and mRMR Feature Selection, IEEE Access 2020, 8, 40313- 40326. https://doi.org/10.1109/ACCESS.2020.2976832

[11] Xiong, J.B.; Zhang, Q.H.; Wan, J.F.; Liang, L.; Cheng, P.; Liang, Q. Data Fusion Method Based on Mutual Dimensionless, IEEE/ ASME Transactions on Mechatronics 2018, 23(2), 506-517. https://doi.org/10.1109/TMECH.2017.2759791

[12] Gu, Y.J.; Jia, Z. W.; Yin, C.T.; Ren, Y. Non-dimensional Parameters Trend Analysis Method in the Early Warning and Diagnosis of Wind Turbine Gearbox Failures, Journal of Vibration and Shock 2017, 36(19), 213- 220.

[13] Chandra, M.A.; Bedi, S.S. Survey on SVM and Their Application in Image Classification, International Journal of Information Technology 2018, 2,1-11. https://doi.org/10.1007/s41870-017-0080-1

[14] Qiao, N.S.; Zhang, F. Method of Defect Image Edge Information Extraction of Printed Circuit Board, Computer Engineering and Applications 2015, 51(20), 11-15+27.

[15] Tran, N.T.; Tran, V.H.; Nguyen, N.B.; Nguyen, T.K.; Cheung, N.M. On Data Augmentation for GAN Training, IEEE Transactions on Image Processing 2021, 30, 1882-1897. https://doi.org/10.1109/TIP.2021.3049346

[16] Li, D.; Du, C.D.; Wang, S.P.; Wang, H.; He, H. Multi-subject Data Augmentation for Target Subject Semantic Decoding with Deep Multi-view Adversarial Learning, Information Sciences 2021, 547(2), 1025-1044. https://doi.org/10.1016/j.ins.2020.09.012

[17] Huang, R.; He, M.Y.; Yang, S.J. A Margin-based Feature Extraction Algorithm for the Small Sample Size Problem, Chinese Journal of Computers 2007, 30(7), 1173-1178.

[18] Su, N.Q.; Li, X.; Zhang, Q.H.; Huang, C.L. Fault Diagnosis of Rotating Machinery Based on the Dimensionless Index and Two-sample Distribution Test, Journal of Computers 2020, 31(3), 1-10.

[19] Chang, S.-G.; Yu, B. Adaptive Wavelet Thresholding for Image Denoising and Compression, IEEE Transactions on Image Processing 2000, 9(9), 1532- 1546. https://doi.org/10.1109/83.862633

[20] Xiong, J.B.; Zhang, Q.H.; Sun,G.X.; Peng, Z.P.; Liang,Q. Fusion of the Dimensionless Parameters and Filtering Methods in Rotating Machinery Fault Diagnosis, Journal of Networks 2014, 9(5), 1201-1207.

[21] Su, N.Q.; Li, X.; Zhang, Q.H. Fault Diagnosis of Rotating Machinery Based on Wavelet Domain Denoising and Metric Distance, IEEE Access 2019, 7, 73262-73270. https://doi.org/10.1109/ACCESS.2019.2920939

[22] Zhang, Q.H. Research on Fault Diagnosis Technology of Unit Based on Artificial Immune System[D], South China University of Technology 2004.

Submitted: 17.11.2021

Accepted: 14.6.2022

Nai-quan Su*
Guangdong Provincial Key Lab of Petrochemical Equipment and Fault Diagnosis, Guangdong University of Petrochemical Technology, Maoming 525000, China.
High-Tech Institute of Xi'an, Xi'an 710025, China

Zhi-Jie Zhou
High-Tech Institute of Xi'an, Xi'an 710025, China

Qing-hua Zhang
Shao-lin Hu
Xiao-xiao Chang
Guangdong Provincial Key Lab of Petrochemical Equipment and Fault Diagnosis, Guangdong University of Petrochemical Technology, Maoming 525000, China

*Corresponding author:
534306286@qq.com