Journal of Sound and Vibration 360 (2016) 277–299

Contents lists available at ScienceDirect

Journal of Sound and Vibration

journal homepage: www.elsevier.com/locate/jsvi

A fault diagnosis scheme for rolling bearing based on local mean decomposition and improved multiscale fuzzy entropy

Yongbo Li, Minqiang Xu*, Rixin Wang, Wenhu Huang

Department of Astronautical Science and Mechanics, Harbin Institute of Technology (HIT), No. 92 West Dazhi Street, Harbin 150001, People's Republic of China

ARTICLE INFO

Article history:
Received 10 December 2014
Received in revised form
19 August 2015
Accepted 9 September 2015
Handling Editor: I. Lopez Arteaga

ABSTRACT

This paper presents a new rolling bearing fault diagnosis method based on local mean decomposition (LMD), improved multiscale fuzzy entropy (IMFE), Laplacian score (LS) and improved support vector machine based binary tree (ISVM-BT). When the fault occurs in rolling bearings, the measured vibration signal is a multi-component amplitude-modulated and frequency-modulated (AM–FM) signal. LMD, a new self-adaptive time-frequency analysis method can decompose any complicated signal into a series of product functions (PFs), each of which is exactly a mono-component AM–FM signal. Hence, LMD is introduced to preprocess the vibration signal. Furthermore, IMFE that is designed to avoid the inaccurate estimation of fuzzy entropy can be utilized to quantify the complexity and self-similarity of time series for a range of scales based on fuzzy entropy. Besides, the LS approach is introduced to refine the fault features by sorting the scale factors. Subsequently, the obtained features are fed into the multi-fault classifier ISVM-BT to automatically fulfill the fault pattern identifications. The experimental results validate the effectiveness of the methodology and demonstrate that proposed algorithm can be applied to recognize the different categories and severities of rolling bearings.

1. Introduction

The rolling bearings are widely used and key components of rotating machinery and their condition monitoring techniques are always a central topic for the maintenance of rotating machinery [1]. Due to the direct relationship between the vibration and the structure of the rotating machine, numerous researches have been focused on the vibration analysis method in recent years [2,3]. In general, vibration analysis method can be summarized into three steps: data acquisition, fault feature extraction and fault pattern classification. Due to the nonlinear and non-stationary characteristics of the vibration signal, it is hard to extract the fault information and many signal processing techniques have been developed to extract fault features[4]. For example, Yang et al. applied empirical mode decomposition method (EMD) to decompose the bearing vibration signal and the energy entropy of intrinsic mode functions (IMFs) was used as the fault feature to fulfill the bearing fault diagnosis [5]. Lin et al. [6] presented a novel approach based on multiscale entropy and wavelet transform to extract the fault features of misaligned motors. LMD and multiscale entropy were combined to classify the rolling bearing fault condition by Liu et al. [7]. Zheng et al. [8] put forward a new approach called local characteristic-scale decomposition

* Corresponding author. Tel.: +86 451 86414320.
E-mail addresses: liyongbo0532@126.com (Y. Li), xumqhit@126.com (M. Xu), wangrx@hit.edu.cn (R. Wang), leobo28@foxmail.com (W. Huang).

http://dx.doi.org/10.1016/j.jsv.2015.09.016
0022-460X/

(LCD) to preprocess the rolling bearing vibration signal and took the fuzzy entropy of intrinsic scale components (ISCs) as the inputs of the adaptive neuro-fuzzy inference systems (ANFIS) classifier.

The current time-frequency analysis techniques are mainly composed of two classes. The first one is that some parameters needed to be set before analyzing the vibration signal. The popular and well-known example is the wavelet transform (WT) [9], which can decompose a multi scales into several scale time-frequency components, WT is widely applied to diagnostic the rolling bearing fault. However, the wavelet basis function is needed to predefined, and different choices of the wavelet basis function will have great influence on final results. Therefore, it doesn't have the nature of self-adaptive feature. As an improved approach, another one is self-adaptive time-frequency decomposition techniques, the typical example is EMD [10], which can self-adaptively decompose any complicated signal into sum of IMFs. Unlike the wavelet transform, EMD performs the decomposition according to natural oscillations embedded in the vibration signal. Unfortunately, EMD has the end effect and mode mixing problems, which limit its applications [11]. A new self-adaptive time-frequency analyzing method, Local mean decomposition (LMD) was proposed by Smith in 2005 [12]. LMD can automatically decompose a multi-component AM–FM signal into a number of PFs and each PF is a mono-component AM–FM signal. Also, the comparisons of LMD and EMD have been done and the merits of the LMD have been verified [13]. When the rolling bearing fault occurs, the collected vibration signal often presents AM–FM features. Consequently, LMD is suitable for processing the rolling bearing fault vibration signal.

Naturally, after LMD decomposition, a major focus is how to extract feature and identify the fault patterns by using the obtained PFs. Many studies have been done to investigate the feature extraction methods. A statistical measure method, called approximate entropy (ApEn) was put forward by Pincus, which was successfully applied to physiological time series analysis [14]. However, ApEn is heavily dependent on the data length and its estimated value is uniformly lower than that expected for short records as well [15,16]. To overcome the weakness of ApEn, Richman and Moorman proposed Sample entropy (SampEn), which had attracted a lot of attentions [16]. Although SampEn can improve performance, it results in an unacceptable result when applied to actual data analysis. In regard to this disadvantage, Costa [17] put forward a multiscale entropy procedure to estimate the complexity of the original time series over a range of scales. Multiscale entropy (MSE) was firstly applied to rolling bearing fault diagnosis by Zhang et al. [15] and the effectiveness of MSE was validated by analyzing the complex time series. However, the short-time analysis of MSE is far more optimal. Focus on the drawbacks of MSE, Wu et al. proposed modified Multiscale entropy (MMSE) to improve the performance of MSE, which can obtain excellent results for the short-term time series short-time time series [18].

As an alternative approach of SamEn, the fuzzy entropy (FuzzyEn) was proposed by Chen et al., recently [19]. Compared with SamEn calculation procedure, FuzzyEn replaces the Heaviside function with fuzzy membership function with a better continuity, which was successfully applied to rolling bearing fault pattern recognition by Zheng et al. [4]. Also the concept of the coarse-graining procedure combined with FuzzyEn (MFE) was developed to evaluate the self-similarity of original data [20]. Nevertheless, MFE method also produces uncertain and unsatisfactory analysis for short-term data. There remains a need for a reliable method that can overcome the weakness of MFE. In this paper, a novel approach called improved multiscale fuzzy entropy (IMFE) is proposed, and the effectiveness of IMFE is verified by the simulation signal and actual experiment data.

Hence, IMFE method is utilized to extract the fault features from the rolling bearings, then the obtained fault features are fed into a multi-classifier to fulfill the fault diagnosis. However, the feature vectors extracted using IMFE method are high dimension with information redundancy, it is time-consuming and often results in unacceptable classification accuracy when the entire obtained features are adopted as inputs of SVM-BT. In this paper, we introduce an effective approach called Laplacian Score (LS) to choose the first several important scale factors to construct the new fault feature vectors [21]. By the virtue of the LS approach, the fault feature vectors can be automatically ranked according to their importance and correlations with the main fault information [22], then we select the first four important scale factors as the new fault feature vectors. By using LS approach, it can not only reduce the data dimension but also enhance the identification accuracy greatly.

Naturally, after extracting feature vectors, a multi-fault classifier is employed to automatically identify the fault categories and severities of rolling bearings. Various artificial intelligence pattern recognition techniques in recent years are used for fault detection in rotating machines, such as artificial neural networks (ANN) [23], adaptive neuro-fuzzy inference system (ANFIS) [24] and Self-organizing maps (SOM) [25]. However, some limitations that restrict their applications. For example, ANN, which adopts of empirical risk minimization (ERM) for classification is likely lead to failure because of the insufficient training sample and unreasonable structure design [26]. As an alternative approach to other classifiers, support vector machine based on binary tree (SVM-BT) [27] has superior recognition rates in comparison to other classification methods, which is not only effective in making a reliable decision for a smaller number of datasets but also has good generalization capability [27].

However, the SVM-BT often suffers from the hierarchical structures decision problems. In this paper, we introduce a novel hierarchical structures design for SVM-BT, called improved SVM-BT (ISVM-BT), which is based on the combination of inter-class Euclidean distance (ED) [28] and intra-class sample distribution [29,30]. ISVM-BT can not only reflect the class separability more comprehensively but also improve the classification accuracy obviously. Hence, a novel rolling bearing fault diagnosis approach based on LMD, IMFE, LS and ISVM-BT is proposed in this paper.

This paper is organized into six sections. In Section 2 a brief review of preprocessing method based on LMD is provided. Section 3 describes the basis of IMFE and validates the superiority of IMFE using the simulation signal. Section 4 presents the

LS for scale factor selection briefly. In Section 5 the ISVM-BT method is introduced, meanwhile, the calculation procedures of ISVM-BT are illustrated. Section 6 discusses experimental results for the proposed rolling bearing fault diagnosis method. Finally, conclusions are drawn in Section 7.

2. Local mean decomposition and principal PF component selection

2.1. Review of LMD method

LMD method can decompose a complicated signal into a set of product functions (PFs), each of which is the product of an envelope signal and a purely frequency modulated signal. Given any signal $x(t)$, it can be decomposed by LMD method in the following way:

(1) Identify all local extrema n_i of the original signal $x(t)$, and then calculate the mean value m_i of two successive extrema n_i and n_{i+1},and envelope estimate a_i according to Eqs. (1) and (2).

$$m_i = (n_{i+1} + n_i)/2 \tag{1}$$

$$a_i = |n_{i+1} - n_i|/2 \tag{2}$$

(2) All mean values m_i and envelope estimate a_i are connected by straight lines. The local mean function $m_{11}(t)$ and the amplitude function $a_{11}(t)$ are obtained by applying the moving averaging method to smooth the local means and envelope estimates, respectively.

(3) The local mean function $m_{11}(t)$ is subtracted from the original data $x(t)$, and the residual signal is denoted as $h_{11}(t)$.

$$h_{11}(t) = x(t) - m_{11}(t) \tag{3}$$

$h_{11}(t)$ is then divided by the amplitude function $a_{11}(t)$.

$$s_{11}(t) = h_{11}(t)/a_{11}(t) \tag{4}$$

The envelope $a_{12}(t)$ of $s_{11}(t)$ can then be calculated by steps (1) and (2). If the envelope function $a_{12}(t) = 1$, stop the procedure and take $s_{12}(t)$ as the first purely frequency modulated signal. If the envelope function $a_{12}(t) \neq 1$, regard $s_{11}(t)$ as the original signal and repeat the steps(1)–(4) continues n times until a purely frequency modulated signal $s_{1n}(t)$ is obtained.

$$\begin{cases} h_{11}(t) = x(t) - m_{11}(t) \\ h_{12}(t) = s_{11}(t) - m_{12}(t) \\ \vdots \\ h_{1n}(t) = s_{1(n-1)}(t) - m_{1n}(t) \end{cases} \tag{5}$$

in which

$$\begin{cases} s_{11}(t) = h_{11}(t)/a_{11}(t) \\ s_{12}(t) = h_{12}(t)/a_{12}(t) \\ \vdots \\ s_{1n}(t) = h_{1n}(t)/a_{1n}(t) \end{cases} \tag{6}$$

The objective of the iterations is to obtain a signal whose envelope function should satisfy $a_{1n}(t) = 1$.

(4) The corresponding envelope is obtained by Eq. (7).

$$a_1(t) = a_{11}(t)a_{12}(t)\cdots a_{1n}(t) = \prod_{q=1}^{n} a_{1q}(t) \tag{7}$$

The envelope function $a_1(t)$ is then multipied by the purely frequency modulated signal $s_{1n}(t)$ and then the first product function $\mathrm{PF}_1(t)$ can be written as

$$\mathrm{PF}_1(t) = a_1(t)s_{1n}(t) \tag{8}$$

(5) This derived $PF_1(t)$ is then subtracted from the original time series $x(t)$,resulting in a new function $u_1(t)$, which represents a smoothed version of the original data since the highest frequency oscillations have been removed from it. Regard

$u_1(t)$ as a new data and repeat the above procedure k times until $u_k(t)$ is a constant or contains no more oscillations.

$$\begin{cases} \mathrm{u}_1(t) = x(t) - \mathrm{PF}_1(t) \\ \vdots \\ \mathrm{u}_k(t) = u_{k-1}(t) - \mathrm{PF}_k(t) \end{cases} \tag{9}$$

Thus, the original signal can be reconstructed k product functions and a residual $u_k(t)$ according to

$$x(t) = \sum_{p=1}^{k} \mathrm{PF}_k(t) + \mathrm{u}_k(t) \tag{10}$$

where $u_k(t)$ is the residue and k is the number of PF components.

2.2. The optimum PF component selection

As mentioned above, a number of PFs can be obtained by using LMD method, and the selection of PF component used to conduct the fault feature extraction is a key step for fault diagnosis. Since kurtosis is effective to characterize the impulsive feature, the PF with the highest kurtosis value indicate that it contains most impulsive characteristics [31]. Kurtosis is then

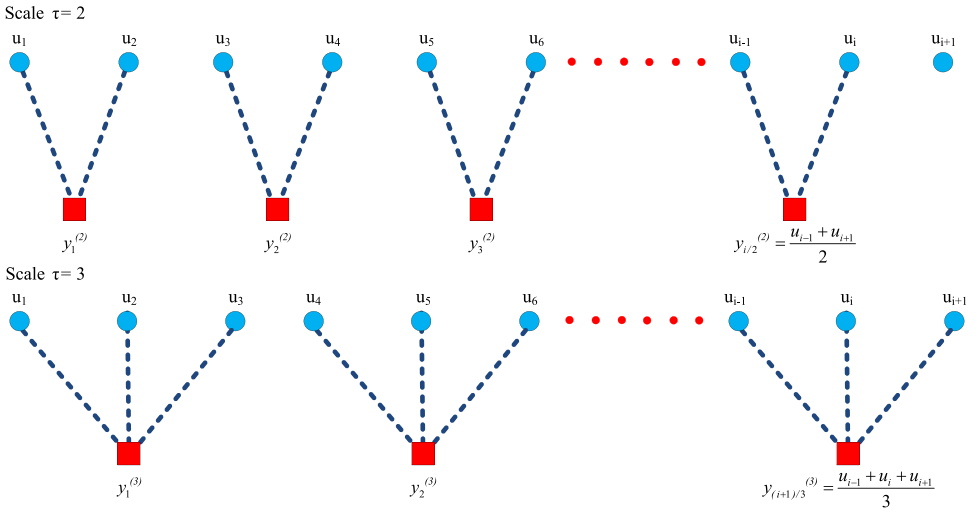

Fig. 1. The schematic illustration of the coarse-grained procedure for scale factor τ=2 and τ=3.

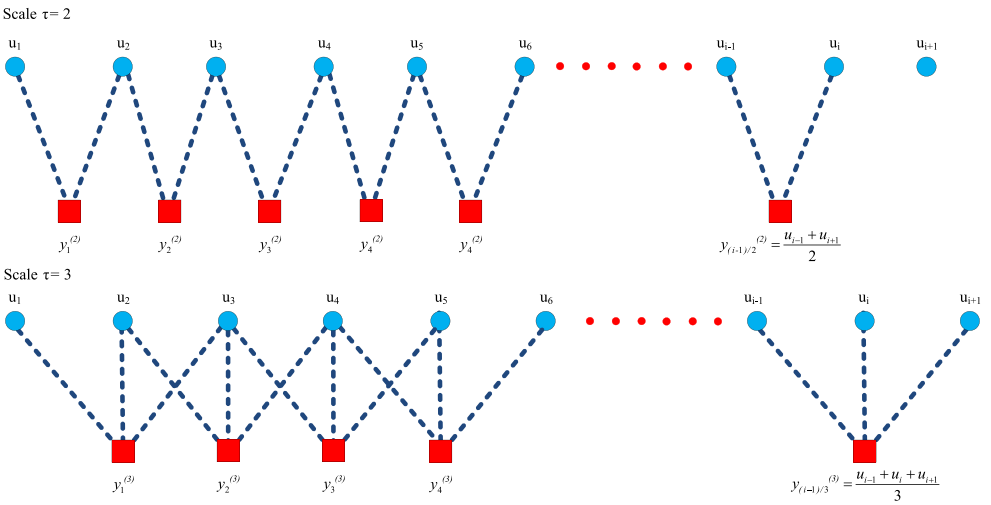

Fig. 2. The schematic illustration of the improved coarse-grained procedure for scale factor τ=2 and τ=3.

taken as the criterion to select the optimum PF (OPF) component that contains the most fault information from the LMD decomposition results. The main procedures of OPF selection with kurtosis can be described as follows:

(1) LMD method is firstly used to preprocess the vibration signals under different health conditions and then a number of PF components are obtained.
(2) Kurtosis is then applied to calculate each PF component according to Eq. (11). Also the normalized kurtosis value can be got by Eq. (12).

$$K_i = \frac{1}{n}\sum_{k=1}^{n} \mathrm{PF}_{ik}^4 \tag{11}$$

$$U_i = \frac{K_i}{\sum_{i=1}^{m} K_i} \tag{12}$$

where K_i represents the kurtosis value of the ith PF component, n is the length of the data. U_i is the normalized kurtosis value of the ith PF component, m is number of PF components.
(3) The U_i values are then used to select the OPF component, which contains the most fault information.

3. Improved multiscale fuzzy entropy

3.1. Fuzzy entropy

The definition and calculation steps of ApEn and SampEn are described by [32]. As the similarity definition of the two vectors is mainly according to the Heaviside function, which is jumping. However, in the actual times series, the boundaries of the two classes are mostly ambiguous, the Heaviside function is unsuitable to measure the similarity of two vectors. As an improvement of ApEn, SamEn, FuzzyEn replaces the Heaviside function with a Gaussian function. Due to the continuity of the exponential function, the FuzzyEn can avoid the drawbacks of ApEn and SamEn effectively [19]. The detailed steps of FuzzyEn are listed as follows:

(1) Given a time series with the length $N\{u(i), i=1,2,\cdots,\mathrm{N}\}$, then the m dimensional vector at time i can be constructed as

$$X_i^m = \{u(i), u(i+1), \cdots, u(i+m+1)\} - u_0(i), i = 1,2,\cdots N-m+1 \tag{13}$$

where X_i^m is a new time series, and the $u_0(i)$ represents the mean value of the m consecutive $u(i)$values.

$$u_0(i) = \frac{1}{m}\sum_{k=0}^{m-1} u(i+k) \tag{14}$$

(2) Define the maximum distance between X_i^m and X_j^m as d_{ij}^m

$$d_{ij}^m = d[X_i^m, X_j^m] = \max_{k\in(0,m-1)}\{|[u(i+k)-u_0(i)] - [u(j+k)-u_0(j)]|\}$$
$$i,j = 1,2,\ldots N-m, i\neq j. \tag{15}$$

(3) We can obtain the similarity degree D_{ij}^m by using the exponential function (namely, fuzzy function) $\mu(d_{ij}^m, n, r)$

$$D_{ij}^m = \mu(d_{ij}^m, n, r) = e^{-\ln 2(d_{ij}^m/r)^n} \tag{16}$$

where n and r are the gradient and the width of the border, respectively.
(4) The $\varphi^m(n,r)$ is then defined as follows:

$$\varphi^m(n,r) = \frac{1}{N-m}\sum_{i=1}^{N-m}\left(\frac{1}{N-m-1}\sum_{\substack{j=1\\ j\neq i}}^{N-m} D_{ij}^m\right) \tag{17}$$

(5) Repeat Eqs. (15)–(17) for obtaining $m+1$ dimensional, and $\varphi^{m+1}(n,r)$ can be described as

$$\varphi^{m+1}(n,r)=\frac{1}{N-m}\sum_{i=1}^{N-m}\left(\frac{1}{N-m-1}\sum_{\substack{j=1\\ j\neq i}}^{N-m}D_{ij}^{m+1}\right) \tag{18}$$

(6) Then the Fuzzy entropy of the time series $\{x(i), i=1,2,\cdots,N\}$ can be defied as

$$\mathrm{FuzzyEn}(m,n,r)=\lim_{N\to\infty}\left[\ln\varphi^{m}(n,r)-\ln\varphi^{m+1}(n,r)\right] \tag{19}$$

If N is finite, FuzzyEn(m, n, r) can be expressed as

$$\mathrm{FuzzyEn}(m,n,r,N)=\ln\varphi^{m}(n,r)-\ln\varphi^{m+1}(n,r) \tag{20}$$

3.2. The basis of multiscale fuzzy entropy

The multiscale analysis algorithm was developed by Costa [17] to quantify the complexity of time series in the real world. Because the application of single scale entropy algorithm to actual measurement time series may produce the unreliable results, Costa introduced a coarse-grained procedure to yield a series of scale time series. Based on the concept of multiscale analysis, MFE method was firstly proposed by Zheng et al. [20] and applied to rolling bearing fault diagnosis. MFE algorithm contains two steps. Firstly, apply the coarse-grained procedure to get multiple scale time series from the original time series.

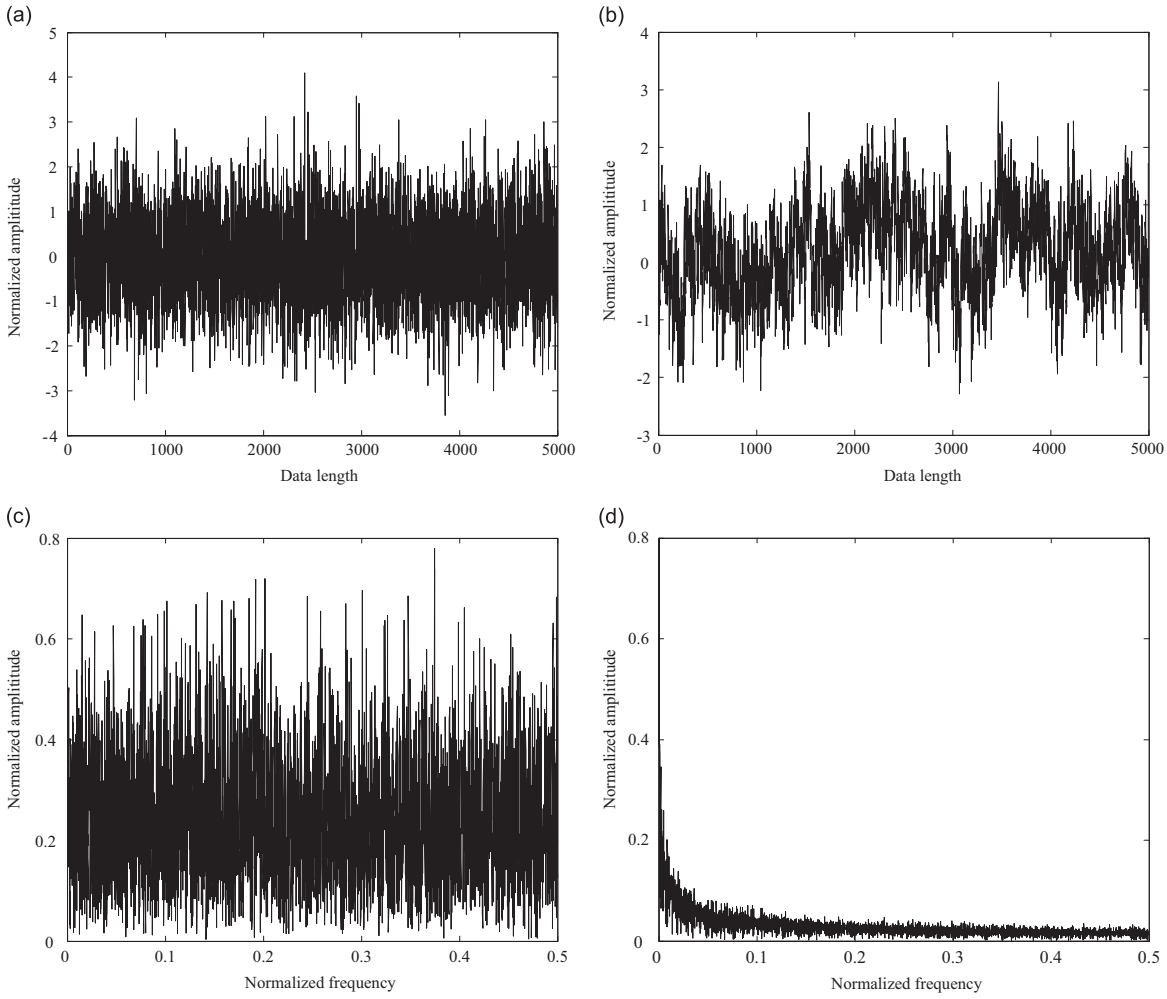

Fig. 3. (a) Waveform of white noise, (b) waveform of 1/*f* noise, (c) FT spectrum of white noise and (d) FT spectrum of 1/*f*.

Then, calculate the FuzzyEn at each coarse-grained time series. Two procedures of MFE algorithm are briefly described as follows.

(1) To obtain the coarse-grained time series at a scale factor of τ, the original time series is divided into disjointed windows of length τ, and the data points are averaged inside each window. Namely, the coarse-grained time series at a scale factor of τ (τ is a positive integer), y_j^τ can be constructed according to Eq. (21) and an example of the coarse-grained procedure is illustrated in Fig. 1.

$$y_j^\tau = \frac{1}{\tau}\sum_{i=(j-1)\tau+1}^{j\tau} u_i \qquad 1 \le j \le \frac{N}{\tau} \tag{21}$$

In MFE analysis, the FuzzyEn of each coarse-grained time series is calculated based on Eqs. (13)–(19) and then plotted as the function of the scale factor τ, which can be expressed as

$$\mathrm{MFE}(x,\tau,m,r) = \mathrm{FuzzyEn}(y_j^{\tau,m,r)}. \tag{22}$$

Note that the r in the calculation for different scales is same, which is obtained by the $r = \lambda * SD$ and SD is the standard deviation of the original time series.

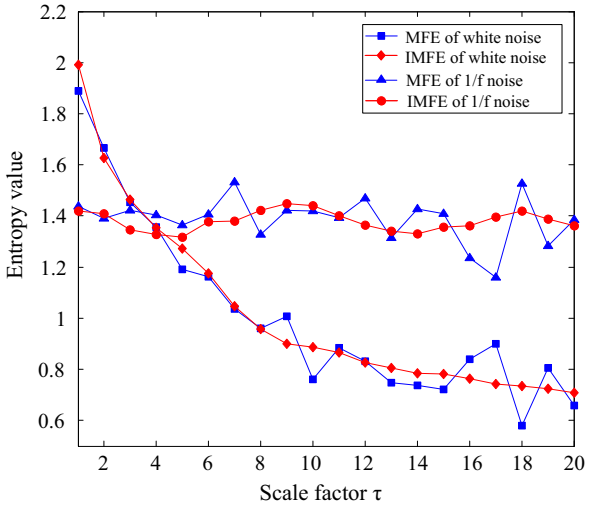

Fig. 4. MFE and IMFE analysis of white and 1/*f* noises.

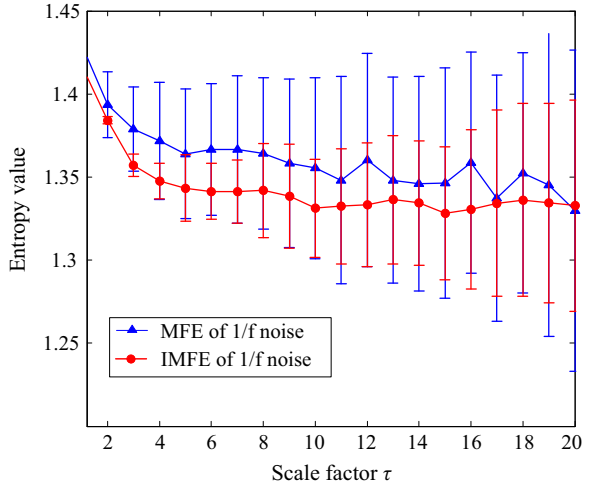

Fig. 5. MFE and IMFE curves of 1/*f* noise.

3.3. Improved multiscale fuzzy entropy

From Fig. 1, we know that the coarse-graining procedure in MFE can be simply considered as the procedure of averaging the original time series within a window of length τ and then downsampling by a factor of τ. However, the imprecise and unreliable results may occur in the process of the downsampling at a certain time scale. To overcome the disadvantage of the coarse-graining procedure in MFE, IMFE method is introduced in this paper, which is composed of two procedures: (1) The time series on different time scales is obtained by performing a moving-averaging procedure; (2) The complexities of the new moving-averaged data are calculated by FuzzyEn with a time delay τ. The detailed descriptions can be summarized as follows:

(1) Based on the improved coarse-graining procedure, a new vector time series vector called H^{τ} at a scale factor of τ is constructed as

$$\mathbf{H}_j^{\tau}=\frac{1}{\tau}\sum_{i=j}^{j+\tau-1}u_i \qquad 1\le j\le N-\tau+1 \tag{23}$$

where N and τ are the length of original time series and scale factor, respectively. By Eq. (23) the original time series$u(i)$ is divided into τ coarse grained vector series $\mathbf{H}^{\tau}$ with length of $N-\tau+1$. The schematic illustration of the improved coarse-grained procedure is shown in Fig. 2.

(2) The derived new time series $\mathbf{H}^{\tau}$ is calculated by a modified FuzzyEn, the modified FuzzyEn is defined as IMFE value with a time delay δ, it can be expressed as follows:

$$\mathrm{IMFE}(u,m,\tau,r)=\mathrm{FuzzyEn}(\mathbf{H}^{\tau},m,\delta=\tau,r) \tag{24}$$

Compared with traditional FuzzyEn, the modified FuzzyEn adds a time delay parameterδ ($\delta=\tau$) in the constraction of the m dimensional vector at time$i$. Given a time series, $N\{u(i),i=1,2,\cdots,N\}$, the new m dimensional vector $X_i^m(\delta)$ can be formed as

$$X_i^m(\delta)=\{u(i),u(i+\delta),\cdots,u(i+(m-1)\delta)\}-u_0(i),i=1,2,\cdots,N-m\delta \tag{25}$$

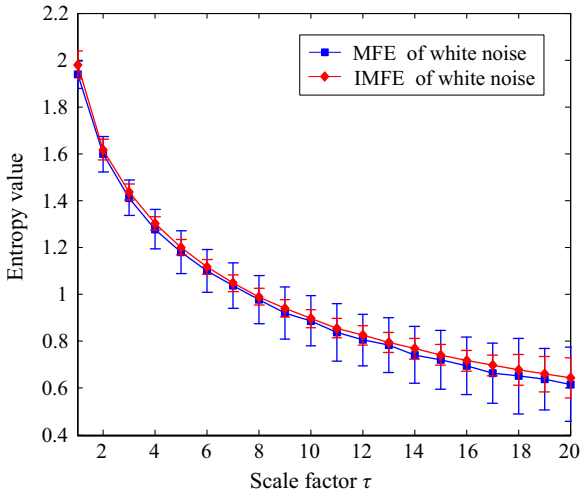

Fig. 6. MFE and IMFE curves of white noise.

Table 1
The FuzzyEn values obtained by MFE and IMFE with different data lengths (τ=20).

	Methods	Data length							
		500	1000	1500	2000	3000	5000	8000	10,000
1/*f* noise	MFE	1.5531	1.4450	1.4115	1.4061	1.3891	1.3668	1.3651	1.3505
	IMFE	1.4466	1.3854	1.3790	1.3666	1.3622	1.3576	1.3568	1.3508
White noise	MFE	0.8222	0.7556	0.7393	0.5743	0.6362	0.5836	0.6098	0.6253
	IMFE	0.7348	0.6554	0.6459	0.6434	0.6204	0.6168	0.6198	0.6199

Table 2
Statistic information of data sets.

Database	Number of training data	Number of testing data	Number of classes	Number of features
Wine	89	89	3	13
Segment	2100	210	7	19

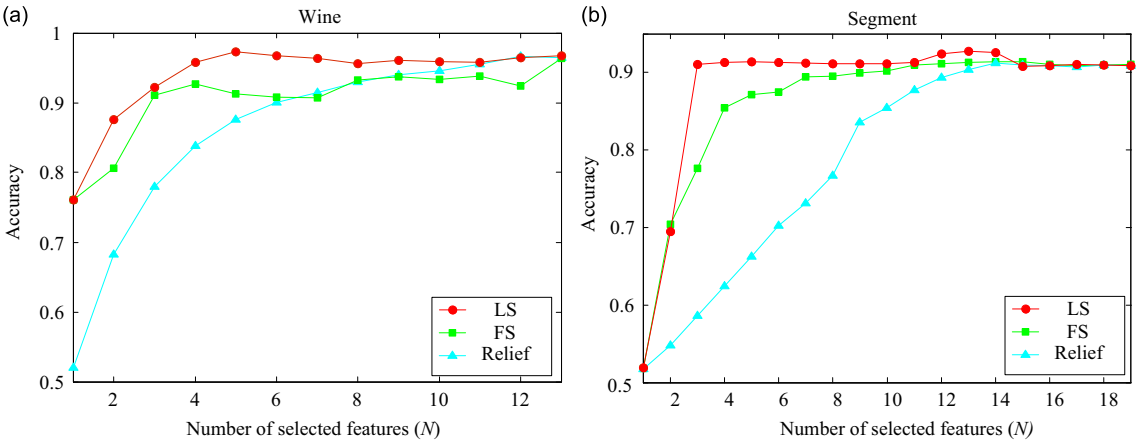

Fig. 7. The classification results of different feature selection methods using the UCI database: (a) Wine and (b) Segment.

where $X_i^m(\delta)$ is a new time series, and the $u_0(i)$ represents the mean value of the m $u(i)$ values.

$$u_0(i) = \frac{1}{m}\sum_{k=0}^{m-1} u(i+k\delta) \tag{26}$$

The following procedures are the same as the traditional FuzzyEn calculation. For the sake of convenient application, the Matlab code of the IMFE method is also listed in Appendix A.

From the above analysis, we know that the length of the moving-averaged time series H^τat a scalar of τ is $(N-\tau+1)$ and the total number of vectors used to compute FuzzyEn with time delay δaccording to Eqs. (25) and (26) is $(N-(m+1)\tau+1)$. Whereas, in the conventional MFE method, the length of vectors used is $(\frac{N}{\tau}-m+1)$, which is much smaller than those in IMFE method. Give an exmple to illustrate the advantage of IMFE alogrithm. Set $N=1500$, $m=2$ and $\tau=15$. In IMFE method, there are 1454 vectors used to calculate FuzzyEn, while there are only 99 vectors used in the original MFE method. Therefore, we can get the conclusion that the proposed IMFE method can provide a more precise and accurate estimation of entropy in comparison with traditional MFE method, which is especial effective for the short-term time series analysis.

3.4. The parameter selection of IMFE

We need to set four parameters before using IMFE, including embedding dimension m, boundary width and gradient of the exponential function r and n, and the scale factor τ. Since unsuitable m will result in loss of information, generally, is set to 2; r is set by 0.1–0.25 multiplied by the standard deviation (SD), here$r=0.15*SD$; n determines the boundary gradient of the exponential function, it is convenient to fix n to 2; since a too large τ will affect the computation efficiency, while a too small τ cannot extract enough information. In this paper, the scale factor τ is selected as 20.

3.5. Comparison between IMFE and MFE

In order to verify the effectiveness of IMFE method, a synthetic noise signal: white noise and $1/f$ noise are utilized to conduct the comparisons between IMFE and MFE methods. The numerical results of the white noise and $1/f$ noise are plotted in Fig. 3 (a) and (b), respectively. Also we conduct the Fourier transform to the white noise and $1/f$ noise, which are illustrated in Fig. 3 (c) and (d), respectively. As can be seen from their spectrum, the $1/f$ noise is more complex than white noise.

To begin with, the IMFE and MFE are employed to analyze the white and $1/f$ noises across 20 scales and the obtained results are shown in Fig. 4, from which the following conclusions can be drawn. Firstly, FuzzyEn values of $1/f$ noise are larger than that of white noise over the whole scales except the beginning scale factors ($\tau=1$, 2 and 3), which agrees with the intuitive results drawn in the spectrum analysis results. Secondly, the FuzzyEn curve of white noise obtained by IMFE method decreases monotonically with τ increasing, whereas the FuzzyEn curve obtained by traditional MFE method fluctuates significantly [17]. Tirdly, the FuzzyEn curve of $1/f$ noise obtained by IMFE method is more smooth and steady than

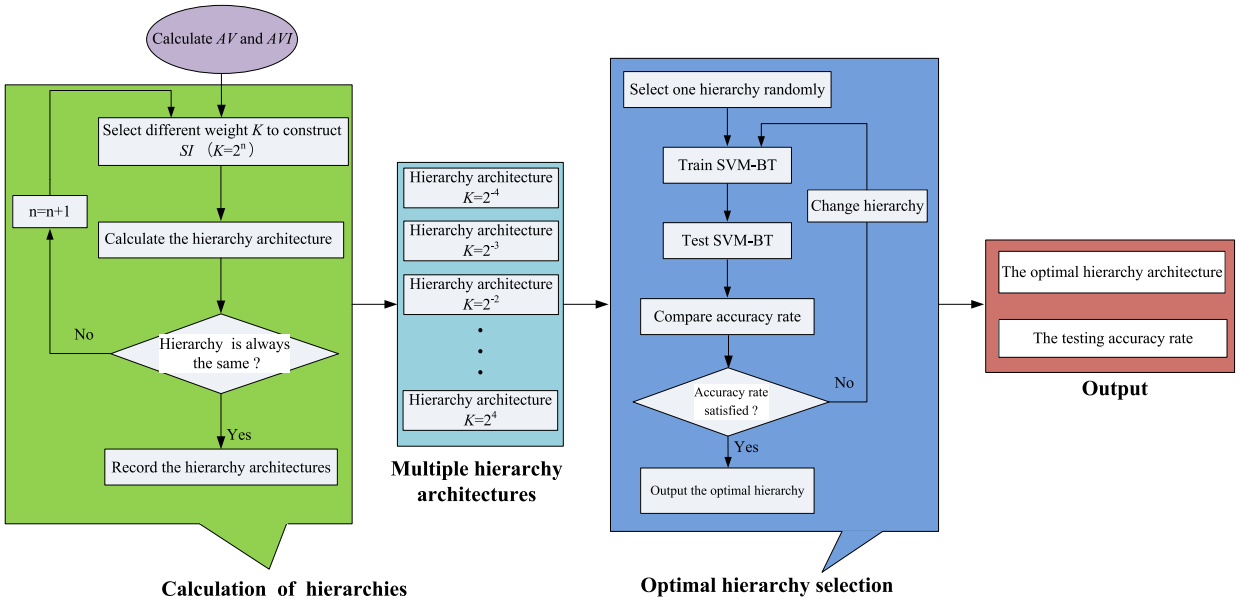

Fig. 8. Framework of ISVM-BT calculation procedures.

that using MFE method. The conclusions indicate that the IMFE method has a more stable performance than traditional MFE method, which can provide a more accurate FuzzyEn estimation of nonlinear and non-stationary signals.

Then to further investigate the estimate performance of the IMFE and MFE methods, we apply IMFE and MFE to analyse 100 independent white and 1/*f* noises, each of which contains 1000 data points. The error bar calculated from 100 independent noise signals at each scale are shown in Figs. 5 and 6, respectively. It is well known that the error bar at each scale indicates the standard deviation (SD) of an FuzzyEn value. As seen the error bar of 1/*f* noise in Fig. 5, the fluctuation of the mean curve obtained through MFE is larger than that of IMFE and the SD of MFE is more than that of IMFE. For white noise in Fig. 6, although the mean curve obtained using MFE and IMSE are nearly equal, the SD of IMFE is less than that of MFE. The comparison results indicate that IMFE has a better performance in estimating the complexity, which can provide a more precise estimation of entropy.

Furthermore, the effect of data length on the MFE and IMFE is also investigated, FuzzyEn is tested on simulated white and 1/*f* with various date lengths ($N=500$, 1000, 1500, 2000, 3000, 5000, 8000, 10,000) at the scale 20 ($\tau=20$). 100 groups of independent noises are employed in each case to calculate the means. The results are listed in Table 1.

From Table 1, it can be concluded that MFE is susceptible to be affected by the data length, especially when the data length is smaller than 2000. Compared with MFE, IMFE method has some superiority in avoiding the undefined FuzzyEn and providing a more reliable and accurate estimation. Hence, the IMFE is utilized to extract fault features from the nonlinear and non-stationary rolling bearing vibration signal under various operating conditions.

4. Feature selection using laplacian score algorithm

The fault features extracted by IMFE in 20 scales are high dimension, if they are all fed into multi-classifier to complete the fault patterns identification, it will not only enhance computation time but also result in information inefficient for rolling bearing fault diagnosis. Therefore, it is still necessary for us to refine the obtained feature vectors.

To avoid the drawbacks of high dimension of feature vector, the Laplacian Score (LS) algorithm is introduced to automatically choose the optimum feature vectors according to their importance and distinguishability. LS method is fundamentally founded on Laplacian Eigenmapsand Locality Preserving Projection. The basic idea of LS is to evaluate the importance of a feature by its power of locality preserving.

Given m data and each data has n features. Suppose that L_r represents the Laplacian Score of the rth feature, $r=1,\ldots, n$. Let f_{ri} represents the ith sample of the rth feature, $i=1, 2, \ldots, m$. The main calculation procedures of LS algorithm can be written as follows [20,21]:

(1) Construct a nearest neighbor graph G with m nodes, where the ith node corresponds to x_i. Then and edge is put between nodes i and j, if x_i and x_j are "close" (for example x_i is among k nearest neighbors of x_j, or x_j is among k nearest neighbors of x_i). When the label information is available, one can put an edge between two nodes sharing the same label.

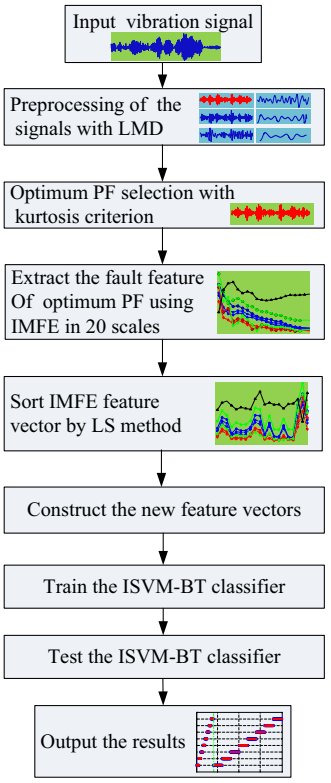

Fig. 9. Flowchart of the proposed algorithm.

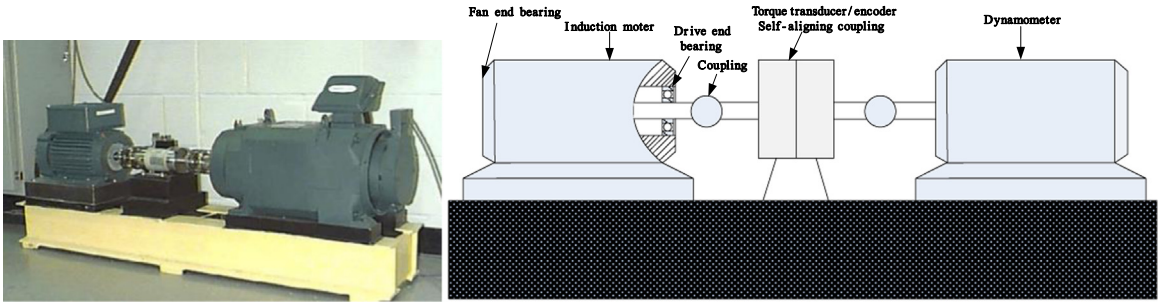

Fig. 10. The rolling bearing experiment system and its sketch.

(2) The weight matrix $\boldsymbol{S}_{ij}$ of the models can be defined as

$$\mathbf{S}_{ij} = \begin{cases} e^{\frac{\|x_i - x_j\|}{t}} & \text{if nodes } i \text{ and } j \text{ are connected} \\ 0 & \text{otherwise} \end{cases} \tag{27}$$

where t is a suitable constant.

(3) For the rth feature, $\mathbf{f}_r$ can be expressed as

$$\mathbf{f}_r = [f_{r1}, f_{r2}, \cdots, f_{rm}]^T, \mathbf{D} = \mathrm{diag}(\mathbf{SI}), \mathbf{I} = [1, \cdots, 1]^T, \mathbf{L} = \mathbf{D} - \mathbf{S} \tag{28}$$

where the matrix $\mathbf{L}$ is called graph Laplacian. Let

$$\tilde{\mathbf{f}_r} = \mathbf{f}_r - \frac{\mathbf{f}_r^T \mathbf{DI}}{\mathbf{I}^T \mathbf{DI}} \mathbf{I} \tag{29}$$

Table 3
The detailed description of the experimental data sets.

Fault class	Fault diameter(mm)	Fault severity	Number of training data	Number of test data	Class label
IRF	0.1778	Slight	10	40	1
	0.3556	Medium	10	40	2
	0.5334	Severe	10	40	3
ORF	0.1778	Slight	10	40	4
	0.3556	Medium	10	40	5
BF	0.1778	Slight	10	40	6
	0.7112	Very severe	10	40	7
Normal	0		10	40	8

(4) The Laplacian Score of the *r*th feature can be written as follows:

$$\mathbf{L}_r = \frac{\sum_{ij}(f_{ri}-f_{rj})^2 \mathbf{S}_{ij}}{Var(\mathbf{f}_r)} = \frac{\tilde{\mathbf{f}}_r^{\,T}\mathbf{L}\tilde{\mathbf{f}}_r}{\tilde{\mathbf{f}}_r^{\,T}\mathbf{D}\mathbf{f}_r} \tag{30}$$

where $Var(\mathbf{f}_r)$ is the estimated variance of the *rth* feature value.

It should be noted that the bigger $\mathbf{S}_{ij}$ value indicates the smaller the Laplacian Score value, which is the characteristic of good feature. Therefore, the scales with lower LS values are taken as the important feature vectors.

In order to illustrate the advantage of LS approach for feature selection, we conduct the comparisons among LS, Fisher score (FS) [33] and Relief [34]. Meanwhile, the UCI database [35]: wine and segment are utilized to estimate the clustering performance of each feature ranking technique, whose information is summarized in Table 2.

The ISVM-BT is used as the classifier here, and the classification results of different feature selection methods varied with the number of features are presented in Fig. 7. It can be clearly observed from Fig. 7 that the feature ranking techniques can significantly affect the classification accuracies. Compared with FS and Relief, LS method has superiority in ranking the features over the most number of selected features (such as N: 3–8). In contrast, FS and Relief methods result in lower accuracy and such performance might be of limited use in roller bearing pattern recognition. This is due to the basic idea of LS is to evaluate the importance of a feature by its power of locality preserving, which can choose the optimum feature vectors according to their importance and distinguishability.

LS method not only avoids the high dimension but also provides important information about the feature vector. In this manner, LS method is applied to select the important feature vector, leading to obvious enhancement of the identification efficiency.

5. The improved SVM-BT

Support Vector Machines based Binary Tree (SVM-BT) takes the advantage of both the efficient computation of the tree architecture and the high reorganization accuracy of SVMs. For a *k*-classes problem, only $(k-1)$ SVM classifiers are needed by applying the binary tree architecture. Nevertheless the super performance of SVM-BT on the pattern recognition is heavily dependent on the sub-classifiers' assignment in hierarchical structures of BT. There are two common approaches to design the hierarchical structures of BT. The first one is applying the average distances between classes[26]. Namely the bigger average distances value indicates the better separability and then it should be put at the top of the hierarchical structures of BT. The second one is applying distribution information inside one class[27], in which the class with wider sample distribution will be put at the top of the hierarchical structures of BT. Based on the merits of the two approaches, a new approach called improved SVM-BT(ISVM-BT) is proposed to design the hierarchical structures of SVM-BT, which is based on the combination of average distance between classes and sample distribution inside one class. The major advantage of the ISVM-BT lies in the combination of the two separability measures in a single SVM-BT, which can provide a more comprehensive reflection of sample separability and enhance the accuracy of pattern recognition significantly.

In this paper, the Euclidean distance(ED) is selected as the two separability measure standards. The proposed approach to design the hierarchical structures of SVM-BT consists of two algorithms, namely the average intra-class ED and the average inter-class ED. In fact, the average intra-class ED and the average inter-class ED represent the average distance between classes and sample distribution inside one class, respectively.

Algorithm 1. The average intra-class ED

Give a data set inside one class $\{a_i, i=1, 2, \ldots k_a\}$, the definition of the average intra-class ED is described as follows:

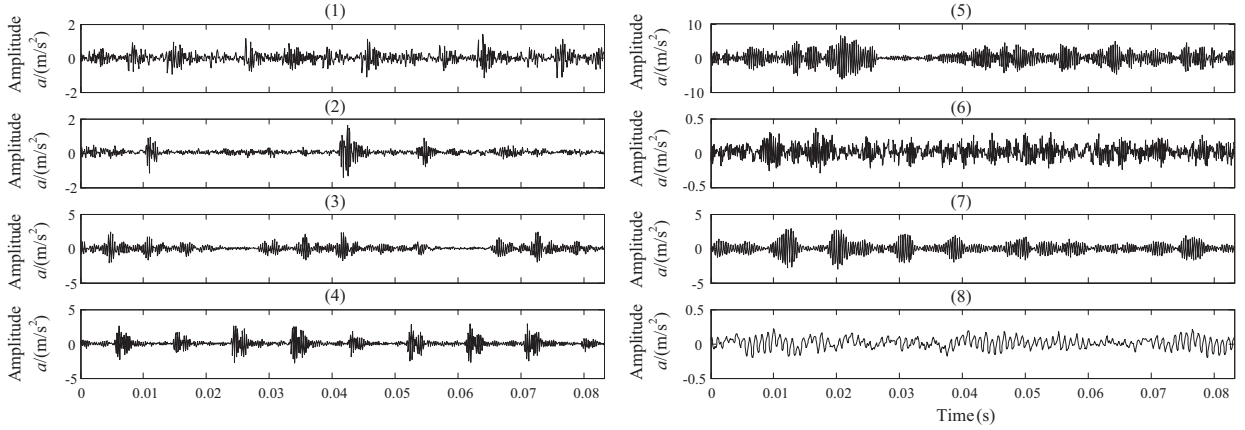

Fig. 11. The vibration acceleration signal of each rolling bearing condition.

Firstly, calculate the intra-class ED of one sample:

$$\mathrm{d}_{\mathrm{ij}}^{a} = \mathrm{d}(a_{\mathrm{i}}, a_{\mathrm{j}}) \tag{31}$$

Secondly, calculate the intra-class ED among different samples inside one class.

$$D_i^a = \frac{1}{k_a - 1} \sum_{j=1}^{k_a} d_{ij}^a, \quad i \neq j \tag{32}$$

Thirdly, calculate the average intra-class ED of one class.

$$\mathrm{AV}^a = \frac{1}{k_a} \sum_{j=1}^{k_a} D_i^a \tag{33}$$

Note that the smaller average intra-class Euclidean distance value suggests more concentrated distribution.

Algorithm 2. The average inter-class ED

Give two classes sample sets: $\{a_i, i=1, 2, \ldots k_a\}$ and $\{b_i, i=1, 2, \ldots, k_b\}$, (Note that $a_{\mathrm{i}} \in$ class A, $b_{\mathrm{j}} \in$ class B). The average inter-class ED is defined as following.

Firstly, calculate the inter-class ED among different categories:

$$\mathrm{d}_{\mathrm{ij}}^{ab} = \mathrm{d}(a_{\mathrm{i}}, b_{\mathrm{j}}) \tag{34}$$

Secondly, calculate the average ED from sample a_{i} to all samples of class B:

$$\mathrm{D}_i^{ab} = \frac{1}{k_b} \sum_{j=1}^{k_b} \mathrm{d}_{ij}^{ab} \tag{35}$$

Thirdly, calculate the average inter-class ED between class A and class B:

$$AVI^{ab} = \frac{1}{k_a} \sum_{i=1}^{k_a} \mathrm{D}_i^{ab} \tag{36}$$

Note that the bigger average inter-class ED value indicates higher divisibility between classes.

The aim of this study is to develop a separation criterion, which enables separate the class with bigger distance from other classes and wider sample distribution within itself at first. Hence, the separability measure $\mathrm{I}_{A,B}$ can be defined by the combination of AV^{ab} and AVI^{ab} using the weight K, it can be written as

$$I_{A,B} = AVI^{ab} + K(AV^a + AV^b) \tag{37}$$

where AVI^{ab} represents the average inter-class ED between class A and class B, AV^a and AV^b represent the average intra-class ED inside class A and class B, K is the weight coefficient.

Based on the $\mathrm{I}_{A,B}$, the detailed procedures of the ISVM-BT can be described as follows:

(1) According to the training data, the average intra-class ED AV and inter-class ED AVI are calculated, and then give the range of the weight K. (In this paper, $K_n = 2^n$, $-4 \le n \le 4$, n is integer value);

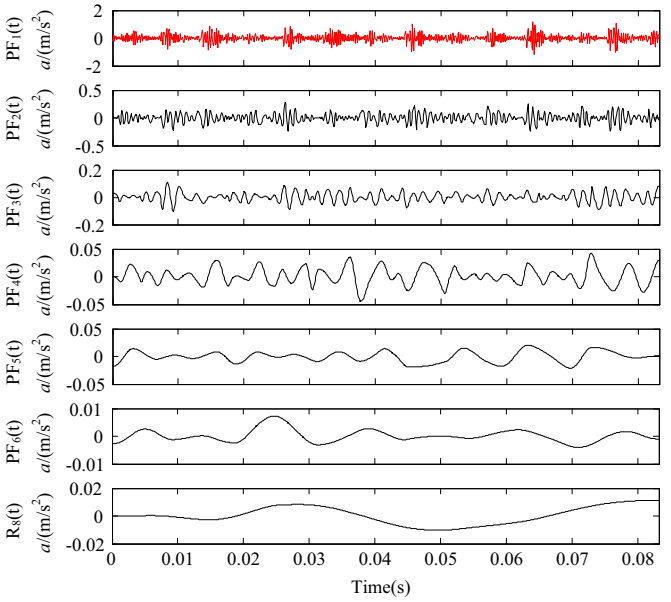

Fig. 12. LMD decomposition results of the vibration acceleration of bearing with inner race fault condition. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

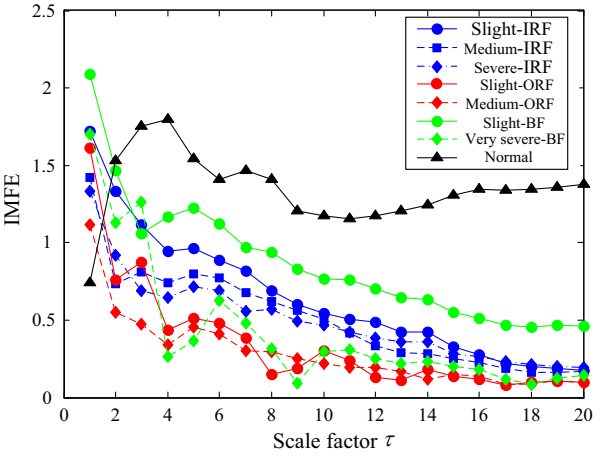

Fig. 13. IMFE over 20 scales of PF1 components derived from LMD method with the average of fifty trails.

(2) Compute the separability measure $\mathbf{SI}=\mathrm{I}_{i,j}$, $i,j=1,2,\ldots,N$, $i\neq j$ for a given K and then construct the symmetric matrix as follows:

$$\mathbf{SI}=\begin{bmatrix} 0 & \mathrm{I}_{1,2} & \mathrm{L} & \mathrm{I}_{1,N-1} & \mathrm{I}_{1,N} \\ \mathrm{I}_{2,1} & 0 & \mathrm{L} & \mathrm{I}_{2,N-1} & \mathrm{I}_{2,N} \\ \mathrm{M} & \mathrm{M} & 0 & \mathrm{M} & \mathrm{M} \\ \mathrm{I}_{N-1,1} & \mathrm{I}_{N-1,2} & \mathrm{L} & 0 & \mathrm{I}_{N-1,N} \\ \mathrm{I}_{N,1} & \mathrm{I}_{N,2} & \mathrm{L} & \mathrm{I}_{N,1\mathrm{N}-1} & 0 \end{bmatrix} \tag{38}$$

It should be noted that: the weight K selection is based on the index of 2, and the initial value of weight K is 2^{-4}.

(3) Ensure hierarchical structures of BT by sorting the summation value of each row of the matrix **SI**;

(4) Change the weight K_n, and then repeat the steps (2) and (3) to generate a series of hierarchical structures of BT;

(5) Select one hierarchical structures of BT with weight K_n, and establish sub-classifiers of SVM according to the hierarchical structures. For k-class task, it will generate k-1 sub-classifiers.

(6) For a certain hierarchical structures of BT with weight K_n, adopt the testing data to test the SVM-BT, generating a testing accuracy rate.

(7) Let $n=n+1$ and repeat steps (5) and (6) until $n=4$, then stop.

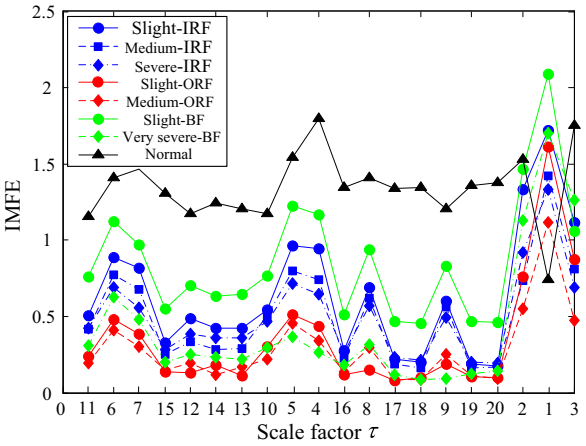

Fig. 14. IMFE over 20 scales of PF1 components reordered by LS.

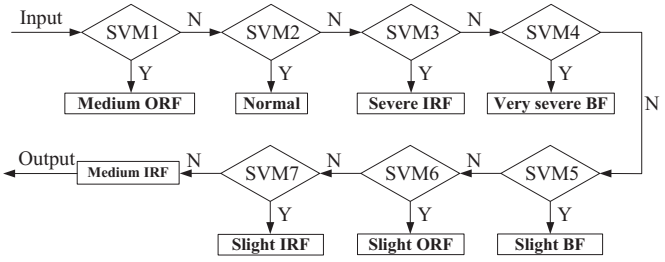

Fig. 15. The order of multi-fault classifier under optimum hierarchy structure.

(8) All the testing accuracy rates for weight K_n can be obtained, and determine the optimum hierarchical structures of BT according to the highest testing accuracy rate.

Based on the above steps, the framework of the ISVM-BT calculation procedures are illustrated in Fig. 8.

The kernel function plays a crucial role in SVM, which can not only reduce the computational load but also solve the high-dimensional transformation effectively. There are different kernel functions used in SVM, the most common kernel functions used in SVM are listed as follows [36,37]:

(i) Polynomial kernel

$$K(x,x_i)=(<x\cdot x_i>+c)^d \tag{39}$$

(ii) Linear kernel

$$K(x,x_i)=<x\cdot x_i> \tag{40}$$

(iii) Radial basis function (RBF) kernel

$$K(x,x_i)=\exp\{-r\|x-x_i\|^2\} \tag{41}$$

where $\gamma>0$, and d and γ are the kernel parameters.

In these kernel functions, RBF kernel is employed in this paper due to its universal application and good performance [38,39].

In the experiments, the parameters: penalty parameter C and the kernel parameter γ of SVM are optimized by Genetic Algorithm (GA) algorithm for each multi-class SVM method. The parameters are set in the following intervals: C (0.1, 1000) and γ (0.001, 10) according to the literature [40]. In order to achieve better estimates of the error rates of the classifiers, 10-fold cross-validation (CV) method is employed. For the 10-fold CV scheme, the training dataset is randomly partitioned into ten subsets. Each sub-set is validated on the classifier that was trained using the other nine subsets. The process was repeated 10 times, the error rate of the classifier is then given by the average of the error rates taken in each test fold. Hence,

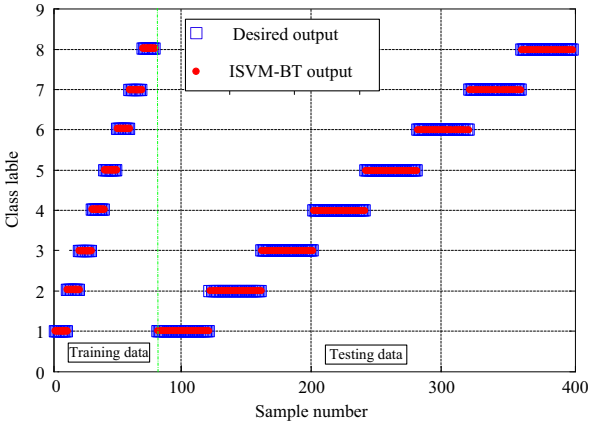

Fig. 16. Classification results of the proposed method.

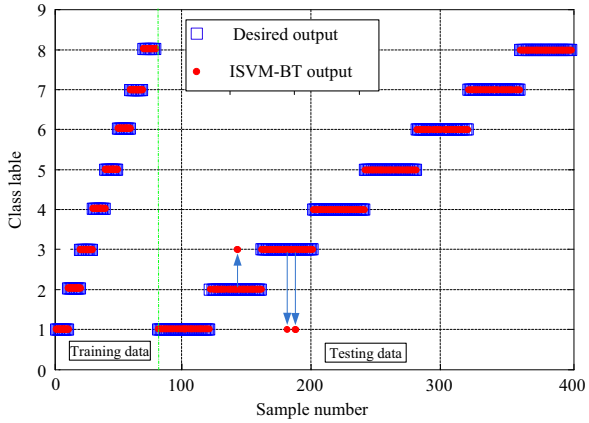

Fig. 17. Classification results of IMFE without preprocessing by LMD method.

GA results are implemented 10 times for each training–validation–test set partition of a dataset with different initial populations. Then select the best parameter $[\gamma, C]$ according to the highest cross-validation accuracy on the validation set (average of the 10 classification accuracy rates).

6. The proposed fault diagnosis method and experimental validation

6.1. The fault feature extraction based on LMD and IMFE

Based on the superiorities of LMD, IMFE, LS and SVM-BT, a novel rolling bearing fault diagnosis approach is presented in this paper, it can be summarized as follows:

(1) The measured vibration signals are firstly preprocessed by LMD method and a set of PF components are obtained.
(2) Calculate and sum the normalized kurtosis value of each PF under different conditions and then choose the OPF with highest normalized kurtosis value;
(3) IMFE method is employed to calculate the selected OPF components under different scales. In the whole paper, we define the scale factor τ from 1 to 20(τ=1–20) and the FuzzyEn values of each coarse grained time series acquired by Eq. (26) is computed with the dimension m=2 and tolerance r=0.15*SD.
(4) LS approach is utilized to sort the 20 feature vectors from low to high values according to their importance and divisibility.
(5) Choose the first 5 scale factors with least LS values to construct the new fault feature vector;
(6) The obtained new fault features are fed into fault classifier ISVM-BT to identify the different health conditions.

A flow chart of the proposed algorithm is presented in Fig. 9.

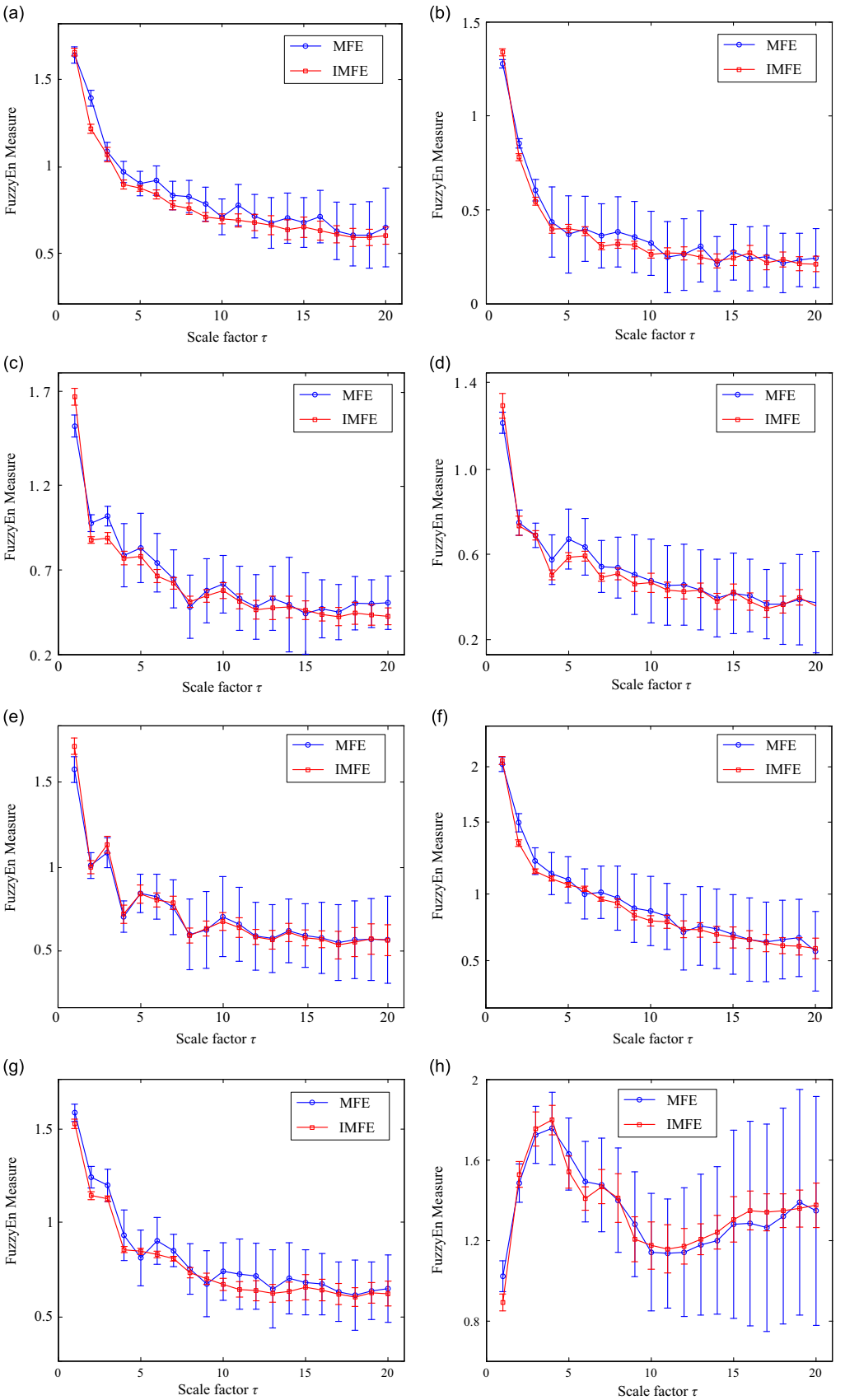

Fig. 18. MFE and IMFE results of rolling bearing vibration signals: (a) slight IRF; (b) medium IRF; (c) severe IRF; (d) slight ORF; (e) medium ORF; (f) slight BF; (g) very severe BF; and (h) normal.

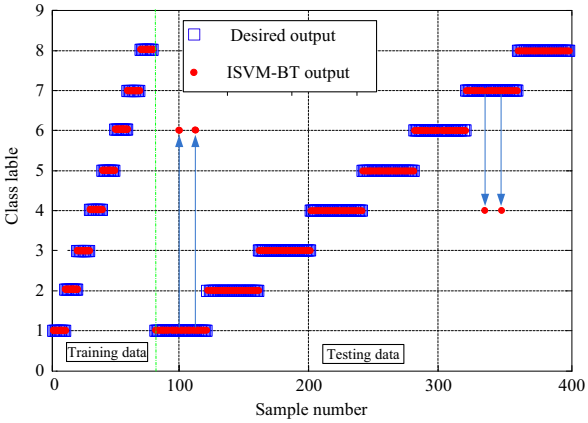

Fig. 19. Classification results of ISVM-BT using MFE method.

Table 4
Classification accuracy of each algorithm using IMFE and MFE as feature extractor.

Methods	OAO *R* (%)	OAA *R* (%)	Intra-BT *R* (%)	Inter-BT *R* (%)	ISVM-BT *R* (%)
IMFE	99.38	98.75	99.06	99.38	**100**
MFE	98.43	**96.88**	98.12	98.43	**98.75**

Table 5
Confusion matrix for OAA and MFE method.

Slight IRF	Medium IRF	Severe IRF	Slight ORF	Medium ORF	Slight BF	Very severe BF	Normal	Classified as
35	0	0	0	0	5	0	0	Slight-IRF
2	38	0	0	0	0	0	0	Medium-IRF
0	0	40	0	0	0	0	0	Severe-IRF
0	0	0	40	0	0	0	0	Slight-ORF
0	0	0	0	40	0	0	0	Medium-ORF
0	0	0	0	0	40	0	0	Slight-BF
0	0	0	3	0	0	37	0	Very severe BF
0	0	0	0	0	0	0	40	Normal

Table 6
Recognition accuracy of each class obtained using OAA method.

Actual class	Recognition accuracy (%)							
	Slight-IRF	Medium-IRF	Severe-IRF	Slight-ORF	Medium-ORF	Slight-BF	Very Severe BF	Normal
Slight-IRF	87.5	0	0	0	0	12.5	0	0
Medium-IRF	5	95	0	0	0	0	0	0
Severe-IRF	0	0	100	0	0	0	0	0
Slight-ORF	0	0	0	100	0	0	0	0
Medium-ORF	0	0	0	0	100	0	0	0
Slight-BF	0	0	0	0	0	100	0	0
Very Severe BF	0	0	0	7.5	0	0	92.5	0
Normal	0	0	0	0	0	0	0	100

6.2. Experimental validation

The purpose of this section is to examine the utility of the proposed algorithm for analyzing the real rolling bearing vibration data. The test bearing data are obtained from the Bearing Data Center of Case Western Reserve University Bearing

Table 7
The accuracy rates of the SVM outputs with different number of inputs.

The number of inputs	2	3	4	5	6
Accuracy rate (%)	100	96.88	100	100	98.12

Table 8
Comparisons between the current work and some published work.

References	Machine element	Fault types	Fault severity levels	Feature extraction method and Classifier used	Classified states	Maximum classification efficiency	Denoising. technique	Feature selection method
Wu et al. [18]	Rolling element bearings	ORF, IRF and BF	Single	MPE and SVM	4	100%	NA	NA
Vakharia et al. [44]	Rolling element bearings	ORF, IRF and BF	Single	Different attribute filters and SVM, ANN	4	97.5%	Wavelet denoising	NA
Liu et al. [7]	Rolling element bearings	ORF, IRF and BF	Single	MSE and SVM	4	100%	LMD denoising	NA
Li et al. [43]	Rolling element bearings	ORF, IRF and BF	Single	MPE and SVM	4	100%	LMDdenoising	LS
Shama et al. [42]	Rolling element bearings	ORF, IRF and BF	Multiple	Different attribute filters and SVM, ANN	7	100%	NA	Three attribute filters
Li et al. (Present work)	Rolling element bearings	ORF, IRF and BF	Multiple	IMFE and improved SVM-BT	8	100%	LMD denoising	LS

Note: MPE is multiscale permutation entropy, ANN is artificial neural network and LMD is local mean decomposition.

Data Center [41]. Fig. 10 gives the experiment system and its sketch. The 6205-2RS JEM SKF deep groove ball bearing is used in this test. The vibration signals of bearing were collected under four conditions including inner race fault condition, outer race fault condition, ball fault condition and the normal condition. In each bearing fault condition, the bearing was seeded with signal point using the electro-discharge machining with fault diameters of 0.1778 mm, 0.3556 mm, 0.5334 mm and 0.7112 mm. An accelerometer was mounted on the front section end to collect the vibration signal. Besides, the sampling frequency is 12,000 Hz and the shaft rotating speed of the motor was 1797 rpm without motor load.

The experimental vibration signals are composed of four fault categories and each fault category contains different levels of severity. Based on the different fault categories and various fault degrees, actually, the experimental analysis is an eight-class recognition problem. The vibration signals in the experiment are split into several non-overlapping segments with the length of 1000. Hence, 50 samples are obtained for each bearing condition, and there are total 400 samples, out of which 80 samples will be randomly selected to train the ISVM-BT classifier and the residual 320 samples are used for testing. The detailed numbers of samples description for each health condition are shown in Table 3. The time domain waveforms of bearing vibration signals with different fault categories and severities as well as normal condition are illustrated in Fig. 11.

Since the measured vibration signal of mechanical system with fault often represents the nonlinear and non-stationary characteristics, it is hard to distinguish the fault categories from each other only according to the time domain waveforms in Fig. 11. Hence, it's necessary to preprocess the original vibration signals by applying LMD method. The vibration signals under different health conditions can be decomposed into a sum of PF components by using LMD method. To save space, only the decomposition results of inner race bearing fault with slight severity are shown in Fig. 12 as a representative.

Naturally, according to the flowchart of the proposed fault diagnosis model, the kurtosis criterion is then used to choose the optimum PF (OPF) component with highest normalized kurtosis value. Take the rolling bearing with inner race fault for example, the normalized kurtosis value of each PF component is 0.3298, 0.2457, 0.1553, 0.1055, 0.0918, 0.0757, respectively. It is easily observed that the PF1 component has the highest normalized kurtosis value, thus the PF1 component (denoted in red color) for different conditions are selected as the OPF component for further analysis.

IMFE method is then utilized to extract the fault features from the selected OPF component under 20 scales for each bearing condition, which means the dimension of obtained feature space is 20 in the beginning analysis. The IMFEs over 20 scales of the bearing data under 8 health conditions are presented in Fig. 13. It can be seen from Fig. 13 that the rolling bearing with normal condition has the largest FuzzyEn values among the whole scales except the first scale, which indicates the normal bearing condition is more complex than the defective bearing conditions. Moreover, the FuzzyEn values of rolling bearing with ball fault is larger than that of rolling bearing with inner race fault and outer race fault, and the rolling bearing with outer race fault has the relative smallest FuzzyEn values (the order of FuzzyEn from high to low is ball fault > inner race fault > outer race fault).

It can be explained in the following way [20]. When the rolling bearing works in a healthy condition, the vibration signals over most scales are random and irregular. Therefore it has lower self-similarity and higher probability to generate new modes, leading to larger FuzzyEn values. Whereas, the vibration signals of rolling bearing with fault conditions are no longer random, which have lower probability to generate new modes, leading to lower FuzzyEn values. For further explanation of the arrangement of the rolling bearing with defective conditions, it is well known that the rolling elements and inner race are both rotating with the roller of the motor, moreover, the rolling element has its own rotation, whereas, the outer race is fixed. Therefore, the vibration signals of rolling bearing with ball fault are more random and complex than inner race fault, leading to larger FuzzyEn values. Also the FuzzyEn values of rolling bearing with inner race fault are larger than that of outer race fault.

Based on the discussion above, although IMFE method can be employed to characterize the complexity and discriminate the normal bearing condition and three fault categories with slight severity, it is still hard to distinguish the same fault category with different degrees (such as the inner race fault with medium and serve degrees) and different fault categories with various fault degrees (such as outer race fault with medium degree and ball fault with very serve degree) only from the IMFE curves in Fig. 13. Therefore, a multi-fault classifier is applied to automatically recognize the various fault categories and severities of rolling bearings. In this paper, a multi-fault classifier based on ISVM-BT is used to fulfill eight classes' classification, which has a higher classification accuracy compared with other classifiers.

However, if IMFE over 20 scales are all taken as the feature vector, it will be time-consuming as well as the classification accuracy rate decreasing. In this paper, the LS algorithm is utilized to rank feature according to their importance. According to the procedures described in Section 4, the new order of IMFE can be listed as follows:

$$LS_{11} < LS_6 < LS_7 < LS_{15} < LS_{12} < LS_{14} < LS_{13} < LS_{10} < LS_5 < LS_4 < LS_{16} < LS_8 < LS_{17} < LS_{18} < LS_9 < LS_{19} < LS_{20} < LS_2 < LS_1 < LS_3.$$

Note that the subscript of LS is the scale factors and the IMFE values are ranked and replotted in Fig. 14. Then the first front five features ($\tau = 11, 6, 7, 15, 12$) with most important information are chosen as new feature vectors. Naturally, after calculating the feature vectors using IMFE and LS methods, the ISVM-BT classifier is needed to automatically complete the fault conditions identifications. Besides, 80 samples are selected randomly from the whole data set as the training data and the residual 320 samples as the testing data, then the training samples are used to train the ISVM-BT to get the optimum BT architectures. Based on the procedures described in Section 5, the average intra-class ED AV and inter-class ED AVI are firstly calculate to ensure the hierarchical structure of BT by using weight $K(K=2^{-4}, 2^{-3},\ldots, 2^{3}, 2^{4})$. The optimum hierarchical structure of BT is shown in Fig. 15.

Furthermore, the 10-fold CV has been used in this paper to give a more accurate estimate of the performance for our method. The 10-fold CV is a standard way of predicting the error rate of a learning technique [42], it can give statistically unbiased result and avoid overfitting problem while dividing data into training and testing set. The classification results of the proposed method using 10-fold CV method are shown in Fig. 16, which includes the ISVM-BT outputs and the desired outputs about the training and testing samples. As can be seen, there are no training and testing samples misclassified and the average recognition accuracy reaches to 100 percent, it is demonstrated that the new proposed approach performs a well classification result, which is exactly suitable and effective in rolling bearing fault diagnosis.

To validate the necessity of preprocessing the rolling bearing vibration signals, IMFE method is directly used to calculate the FuzzyEns of the original signals. Through the same process as the above-mentioned in the proposed method (such as the same number of training and testing data and ISVM-BT classification process), the desired outputs and actual outputs of ISVM are shown in Fig. 17, it can be observed that 3 testing samples with inner race fault are misclassified into the wrong fault degrees, with a classification accuracy of 99.06 percent, which is lower than the proposed method in this paper. Therefore, the comparison results demonstrate the neccesanility of preprocessing of original vibration signals using LMD method. It is because that the interference noise can be restrained and the fault information can be highlighted by LMD decomposition, it is essential to decompose the vibration signals before extracting fault characteristics.

In order to verify the superiority of IMFE, a detailed comparison is made between IMFE and MFE by analyzing 50 independent PF1 components under 8 health bearing conditions. Fig. 18 represents the means and SDs of the corresponding features obtained using the IMFE and MFE methods. The following conclusions can be drawn from Fig. 18. Firstly, for each rolling bearing health condition, the mean curves of the FuzzyEn values derived from IMFE are really close to those derived from MFE. Secondly, compared with MFE method, smaller SDs of the features can be obtained by using IMFE method, especially when the scale factor is over 10. The above conclusions are consistent well with the analysis results of simulations using white and $1/f$ noises.

To further illustrate the fault features extracted using IMFE have higher distinguishability than that of MFE, the fault features obtained by MFE method are also fed into a classifier to distinguish the various health bearing conditions. The number of training and testing data and other conditions remains the same as the mentioned above and the classification accuracy based on MFE is 98.75 percent, with 4 testing samples misclassified. The ISVM-BT outputs for training and testing samples are shown in Fig. 19. It can be easily observed from Fig. 19 that two testing samples with slight inner race fault are misclassified into slight ball fault and two testing samples with very serve ball fault are misclassified into slight outer race fault.

The comparison results provide compelling evidence that IMFE can provide much more accurate estimation of entropy values with higher distinguishability than MFE. The above analysis results can be explained by the fact that when MFE method is used to analyze the short time series, the calculation points decrease exponentially with scale factor increasing, it can not only give rise to questionable and uncertain estimation of entropy values but also increase the SDs of features. The

proposed IMFE method can avoid the above drawbacks of MFE effectively and obtain more precise estimation of entropy, generating higher classification accuracy than that of MFE.

Eventually, for comparison purpose, the other multi-fault classifiers based on SVM such as: OAO, OAA, Intra-BT, Inter-BT and ISVM-BT are all used to solve the eight-class recognition problem. Besides, the training and testing data are the same in each algorithm. The classification accuracy of each multi-fault classifier using IMFE and MFE as feature extractor are summarized in Table 4.

It can clearly observed in Table 4 that the highest classification accuracy of 100 percent is obtained using the proposed method (ISVM-BT and IMFE) and the classification results of the proposed method are shown in Fig. 16. While the classification accuracy of 98.75 percent is obtained using ISVM-BT and MFE with 4 testing samples misclassified, which is shown in Fig. 19. Meanwhile, the least classification accuracy of 96.88 percent is obtained using the OAA and MFE, with 10 testing samples misclassified. To illustrate the reasons for the above results, the confusion matrix for OAA and MFE method is presented in Table 5. Meanwhile, Table 6 shows the detailed recognition accuracy of each class obtained using OAA method.

Seen from Table 5, it is clearly observed that five testing samples with slight inner race fault are misclassified into slight ball fault, two testing samples with medium inner race fault are misclassified into slight degree and three testing samples with very serve ball fault are misclassified into slight outer race fault.

It can be explained in the following way. Firstly, compared with IMFE methods, MFE method could produce uncertain and unsatisfactory analysis for short-term data. When MFE is applied to analyze the short time series, the calculation points decrease exponentially with scale factor increasing, it can not only give rise to questionable and uncertain estimation of entropy values but also increase the SDs of features, which would result in lower distinguishability. On the other hand, the comparison between the classifiers OAA and ISVEM-BT indicate that OAA has lower classification performance in distinguishing different categories and severities of rolling bearings with the same input features. This would further reduce the identification accuracy.

Through comparing the classification results, the conclusions can be got as follows. To begin with, the classification accuracy rate using IMFE is higher than that of MFE in each multi-fault classifier, which reinforces superiority of IMFE over MFE. Secondly, of all the multi-fault classifiers, ISVM-BT has the highest classification accuracy rate, which verifies the advantage of ISVM-BT in classification performance. Lastly, the results rule out the possibility that the above advantages of the proposed approach are a result of occasionality. Therefore, the comparison results demonstrate that the proposed approach is effective in detecting rolling bearing faults.

Also, a basic problem about the number of the selected new features as inputs of ISVM-BT needs to be addressed here. The training accuracy rates with different number of inputs are listed in Table 7. As we can see, 4 or 5 is the suitable number to conduct the fault identification with a higher accuracy rate, while smaller or larger number can both result in a lower accuracy rate, which can be partly explained by the fewer features with fewer fault information, while more features with information redundancy.

In order to illustrate the potential application of proposed methodology in bearing fault diagnosis, a comparative study between the present work and published literature is presented in Table 8 [18,42,43,44]. The comparing items include the machine elements used, fault type, fault severity levels, feature extraction method and classifier used, classified states, maximum classification efficiencies, denoising technique and feature selection method.

7. Conclusions

A novel rolling bearing fault diagnosis algorithm based on LMD, IMFE, LS and ISVM-BT is presented in this paper. In the proposed method, LMD is employed to preprocess the vibration signal, resulting in a sum of PF components. Then, the optimum PF (OPF) component is selected from the PF components according to their kurtosis features. IMFE is taken as the feature extractor to calculate the multiscale FuzzyEn of the OPF component. Furthermore, to solve the selection problem of the obtained features, the LS approach is introduced to automatically select the best scale factor according to their importance and distinguishability. In addition, the ISVM-BT classifier is adopted to fulfill the fault classification. The preliminary simulation demonstrates that the IMFE has a more stable performance and higher divisibility for short time series analysis. Finally, the experimental rolling bearing fault diagnosis confirms that the proposed approach has superior performance in identifying different categories and severities of rolling bearings. Moreover, the proposed method is promising, which is not limited to rolling bearing fault diagnosis but could be applied in fault diagnosis of other mechanical equipment.

Acknowledgments

The research is supported by the National Natural Science Foundation of China (No. 11172078) and the Important National Basic Research Program of China (973 Program – 2012CB720003), and the authors are grateful to all the reviewers and the editor for their valuable comments.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.jsv.2015.09.016.

References

[1] H. Ocak, K.A. Loparo, F.M. Discenzo, Online tracking of bearing wear using wavelet packet decomposition and probabilistic modeling: a method for bearing prognostics, *Journal of Sound and Vibration* 302 (2007) 951–961.
[2] Z. Feng, M.J. Zuo, Vibration signal models for fault diagnosis of planetary gearboxes, *Journal of Sound and Vibration* 331 (2012) 4919–4939.
[3] S. Janjarasjitt, H. Ocak, K.A. Loparo, Bearing condition diagnosis and prognosis using applied nonlinear dynamical analysis of machine vibration signal, *Journal of Sound and Vibration* 317 (2008) 112–126.
[4] C.S. Park, Y.C. Choi, Y.H. Kim, Early fault detection in automotive ball bearings using the minimum variance cepstrum, *Mechanical Systems and Signal Processing* 38 (2013) 534–548.
[5] Y. Yang, D.J. Yu, J.S. Cheng, A roller bearing fault diagnosis method based on EMD energy entropy and ANN, *Journal of Sound and Vibration* 294 (2006) 269–277.
[6] J.L. Lin, J.Y. Liu, C.W. Li, L.F. Tsai, H.Y. Chung, Motor shaft misalignment detection using multiscale entropy with wavelet denoising, *Expert Systems with Applications* 37 (2010) 7200–7204.
[7] H. Liu, M. Han, A fault diagnosis method based on local mean decomposition and multi-scale entropy for roller bearings, *Mechanism and Machine Theory* 75 (2014) 67–78.
[8] J.D. Zheng, J.S. Cheng, Y. Yang, A rolling bearing fault diagnosis approach based on LCD and fuzzy entropy, *Mechanism and Machine Theory* 70 (2013) 441–453.
[9] R. Rubini, U. Meneghetti, Application of the envelope and wavelet transform analyses for the diagnosis of incipient faults in ball bearings, *Mechanical Systems and Signal Processing* 15 (2001) 287–302.
[10] N.E. Huang, Z. Shen, R. Long, C. Wu, H. Shih, Q. Zheng, C. Yen, C. Tung, H. Liu, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proceedings of the Royal Society of London Series A—Mathematical Physical and Engineer Sciences* 454 (1998) 903–995.
[11] Z.K. Peng, P.W. Tse, F.L. Chu, An improved Hilbert–Huang transform and its application in vibration signal analysis, *Journal of Sound and Vibration* 286 (2005) 187–205.
[12] J.S. Smith, The local mean decomposition and its application to EEG perception data, *Journal of the Royal Society Interface* 2 (2005) 443–445.
[13] J.S. Cheng, Y. Yang, A rotating machinery fault diagnosis method based on local mean decomposition, *Digital Signal Processing* 22 (2012) 356–366.
[14] S.M. Pincus, Approximate entropy as a measure of system complexity, *Proceedings of the National Academy of Sciences* 88 (1991) 2297–2301.
[15] L. Zhang, G. Xiong, H. Liu, Bearing fault diagnosis using multi-scale entropy and adaptive neuro-fuzzy inference, *Expert Systems with Applications* 37 (2010) 6077–6085.
[16] J.S. Richman, J.R. Moorman, Physiological time-series analysis using approximate entropy and sample entropy, *American Journal of Physiology—Heart and Circulatory Physiology* 278 (2000) H2039–H2049.
[17] M. Costa, A.L. Goldberger, C.K. Peng, Multiscale entropy analysis of complex physiologic time series, *Physical Review Letters* 89 (2002) 068102.
[18] S.D. Wu, P.H. Wu, C.W. Wu, J.J. Ding, C.C. Wang, Bearing fault diagnosis based on multiscale permutation entropy and support vector machine, *Entropy* 14 (2012) 1343–1356.
[19] W. Chen, J. Zhuang, W. Yu, Z. Wang, Measuring complexity using FuzzyEn, ApEn, and SampEn, *Medical Engineering & Physics* 31 (2009) 61–68.
[20] J.D. Zheng, J.S. Cheng, Y. Yang, S.R. Luo, A rolling bearing fault diagnosis method based on multi-scale fuzzy entropy and variable predictive model-based class discrimination, *Mechanism and Machine Theory* 78 (2014) 187–200.
[21] X. He, D. Cai, P. Niyogi, Laplacian score for feature selection, *Advances in Neural Information Processing System* (2005) 507–551.
[22] Stefan Knerr, L.éon Personnaz, G.érard Dreyfus, Single-layer learning revisited: a stepwise procedure for building and training a neural network, *Neurocomputing* (1990) 41–50.
[23] J.D. Wu, Y.J. Tsai, Speaker identification system using empirical mode decomposition and an artificial neural network, *Expert Systems with Applications* 38 (2011) 6112–6117.
[24] K. Polat, S. Güneş, An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease, *Digital Signal Processing* 17 (2007) 702–710.
[25] S. Haykin, *Neural Networks. A Comprehensive Foundation*, Pearson Prentice Hall Publications, Ontario, Canada, 2005.
[26] P.K. Kankar, C. Satish, S.P. Sharma, Fault diagnosis of ball bearings using machine learning methods, *Expert Systems with Applications* 38 (2011) 1876–1886.
[27] S. Cheong, S. Hoon, S. Lee, Support vector machines with binary tree architecture for multi-class classification, *Neural Information Processing—Letters and Reviews* 2 (2004) 47–51.
[28] G. Madzarov, D. Gjorgjevikj, Evaluation of distance measures for multi-class classification in binary SVM decision tree, *Artificial Intelligence and Soft Computing* (2010) 437–444.
[29] F.M. Tang, Z.D. Wang, J.Y. Chen, On multiclass classification methods of support vector machines, *Control and Decision* 20 (2005) 746–749.
[30] H.Y. Zhao, M.Q. Xu, J.D. Wang, An improved binary tree SVM and application for fault diagnosis, *Journal of Vibration Engineering* 26 (2013) 764–770.
[31] C. Li, M. Liang, Continuous-scale mathematical morphology-based optimal scale band demodulation of impulsive feature for bearing defect diagnosis, *Journal of Sound and Vibration* 331 (2012) 5864–5879.
[32] J.M. Yentes, N. Hunt, K.K. Schmid., The appropriate use of approximate entropy and sample entropy with short data sets, *Annals of Biomedical Engineering* 41 (2013) 349–365.
[33] Y.W. Chen, C.J. Lin, Combining SVMs with various feature selection strategies, *Feature Extraction* (2006) 315–324.
[34] M. Robnik-Šikonja, I. Kononenko, Theoretical and empirical analysis of ReliefF and RReliefF, *Machine Learning* 53 (2003) 23–69.
[35] B. Catherine, J.M. Christopher, *UCI Repository of Machine Learning Databases*, Department of Information and Computer Science, University of California, 1998.
[36] B. Samanta, K.R. Al-Balushi, S.A. Al-Araimi, Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection, *Engineering Applications of Artificial Intelligence* 16 (2003) 657–665.
[37] S. Sathiya Keerthi, Asymptotic behaviors of support vector machines with Gaussian kernel, *Neural Computation* 15 (2003) 1667–1689.
[38] B. Scholkopf, K.K. Sung, F. Girosi, Comparing support vector machines with Gaussian kernels to radial basis function classifiers, *IEEE Transactions on Signal Processing* 45 (1997) 2758–2765.
[39] J. Kamruzzaman, R.K. Begg, Support vector machines and other pattern recognition approaches to the diagnosis of cerebral palsy gait, *IEEE Transactions on Biomedical Engineering* 53 (2006).
[40] A. Subasi, A decision support system for diagnosis of neuromuscular disorders using DWT and evolutionary support vector machines, *Signal Image and Video Processing* 9 (2015) 399–408.
[41] Bearing Data Center, Case Western Reserve University, ⟨http://csegroups.case.edu/bearing datacenter/pages/download-data-file⟩.

[42] Aditya Sharma, M. Amarnath, P.K. Kankar, Feature extraction and fault severity classification in ball bearings, *Journal of Vibration and Control* (2014) 1–17.
[43] Y.B. Li, M.Q. Xu, Y. Wei, W.H. Huang, A new rolling bearing fault diagnosis method based on multiscale permutation entropy and improved support vector machine based binary tree, *Measurement* (2015).
[44] V. Vakharia, V.K. Gupta, P.K. Kankar, A multiscale entropy based approach to select wavelet for fault diagnosis of ball bearings, *Journal of Vibration and Control* (2014) 1–9.

Yongbo Li received the Master's degree in Harbin Engineering University (HRBEU), Harbin, China, in 2012. Now he is a Ph.D. student in Department of Astronautical Science and Mechanics, Harbin Institute of Technology (HIT), Harbin, China. His research interests include signal processing, fault diagnosis, fault feature extraction and pattern identification. He is an author/co-author of over 20 peer-reviewed papers on journals and conference proceedings.

Minqiang Xu graduated in Electronics, The Peking University, Beijing, China, in 1983, his Master's degree in Nuclear Physics from Northeast Normal University, China, in 1989, and his Ph.D. degrees in general mechanics from the Harbin Institute of Technology (HIT), Harbin, China, in 1999. He is a Professor in Department of Astronautical Science and Mechanics, Harbin Institute of Technology. His research interests include dynamics control, signal processing, fault diagnosis and spacecraft fault diagnosis.