Mechanism and Machine Theory 98 (2016) 114–132

Contents lists available at ScienceDirect

Mechanism and Machine Theory

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

Hierarchical fuzzy entropy and improved support vector machine based binary tree approach for rolling bearing fault diagnosis

Yongbo Li, Minqiang Xu *, Haiyang Zhao, 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 22 August 2014
Received in revised form 17 November 2015
Accepted 19 November 2015
Available online 5 January 2016

Keywords:
Hierarchical fuzzy entropy (HFE)
Laplacian score (LS)
Improved support vector machine based binary tree (ISVM-BT)
Fault feature extraction

ABSTRACT

A novel rolling bearing fault diagnosis method based on hierarchical fuzzy entropy (HFE), Laplacian score (LS) and improved support vector machine based binary tree (ISVM-BT) is proposed in this paper. Focus on the difficulty of extracting fault feature from the non-linear and non-stationary vibration signal under complex operating conditions, HFE method is utilized for fault feature extraction. Compared with multi-scale fuzzy entropy (MFE) method, HFE method considers both the low and high frequency components of the vibration signals, which can provide a much more accurate estimation of entropy. Besides, Laplacian score (LS) method is introduced to refine the fault feature 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 demonstrate that the proposed method is effective in recognizing the different categories and severities of rolling bearings faults.

1. Introduction

Rolling bearings are important and fragile parts in the industry filed, which are frequently damaged under harsh environment, resulting in huge economic loss if not detected in time [1]. Much research in recent years has focused on the fault diagnosis of rolling bearings. Among them, the vibration analysis method has been widely applied for diagnosing the rolling bearing fault due to its intrinsic merits of revealing bearing failure [2–4]. It is generally accepted that vibration analysis method consists of two aspects: fault feature extraction and fault pattern classification [5]. However, due to the influence of kinds of factors such as friction, clearances, overloading and so on, the measured vibration signal often represents non-linear and non-stationary characteristics, which is a challenge to extract the underlying fault features in complex vibration signals. Therefore, many researchers have suggested ways of fault feature extraction. For example, correlation dimension was proposed by Logan to identify the rolling bearing operating conditions [6]. However, correlation dimension method has been largely unsuccessful to short data. To avoid the drawbacks of correlation dimension, Pincus [7] introduced approximate entropy (ApEn), which was successfully applied to monitor the rolling bearing health conditions by Yan [2]. Unfortunately, the various lengths of data will affect the ApEn calculation significantly, especially for processing short time series. The estimated value is uniformly lower than that expected one as well [8]. In order to overcome the disadvantages of ApEn, Richman and Moorman proposed Sample entropy (SampEn), which had been widely used in a range of applications both in physiological and machinery signals [9]. Although SampEn can improve performance in contrast to ApEn, it measures the complexity of data in a single scale, which may lead to inaccurate results. Costa [10] proposed a multi-scale entropy procedure to calculate the SampEn over a range of scales. Multi-scale entropy (MSE) has been widely applied quantify the time series' complexity. For instance, Wu applied MSE to

* Corresponding author. Tel.: +86 451 86414320.
E-mail addresses: liyongbo0532@126.com (Y. Li), xumqhit@126.com (M. Xu), weiyu1219@126.com (H. Zhao), leobo28@foxmail.com (W. Huang).

http://dx.doi.org/10.1016/j.mechmachtheory.2015.11.010
0094-114X/

diagnose the rolling bearing fault [11], and Liu conducted the bearing fault pattern identifications by the combination of LMD and MSE [5]. However, MSE only takes into account the low frequency components due to the progressive smoothing operations. Besides, MSE is not well adapted to the practical measured bearing fault vibration signals, which contain rich frequencies and the fault information may be embedded in both lower and higher frequency components. Hence, it is not adequate to extract the fault feature by using MSE method. To overcome the disadvantages of MSE, Jiang et al. proposed hierarchical entropy (HE) method to estimate the complexity of data recently [12], which is successfully applied to analyze the cardiac interbeat interval time series. HE was firstly applied to identify different rolling bearing fault patterns by Zhu et al. [13], and the superiority of HE was validated by comparing with MSE method.

Based on the SampEn, a novel approach called fuzzy entropy (FuzzyEn) was proposed by Chen et al. [14], which replaced the Heaviside function with fuzzy membership function with a better continuity. FuzzyEn was applied to rolling bearing fault pattern recognition by Zheng et al. [15]. Recently, the multi-scale fuzzy entropy (MFE) was developed to enhance the physical meanings and statistical sense of FuzzyEn [15]. However, the coarse-grained procedure used in MFE essentially represents a linear smoothing, which only captures the low frequency components, ignoring the high frequency components [16]. Furthermore, the calculation of MFE is time-consuming when analyzing the long time series. Hence, a novel approach called hierarchical fuzzy entropy (HFE) is proposed in this paper to overcome the weakness of MFE. Compared with MFE, HFE has some advantages as follows: firstly, for each scale of time series, it considers a lower frequency component produced by averaging the components in the previous scale, as well as a higher frequency component produced by taking the difference of two consecutive scales; secondly, the hierarchical decomposition has higher computational efficiency than the coarse-graining procedure, so HFE needs fewer time than MFE for the long time series analysis. Therefore, HFE is utilized to extract the fault feature from rolling bearing vibration signals in this paper.

Therefore, HFE is taken as a feature extractor to extract the fault information from the vibration signals in this paper. However, the feature vectors extracted from vibration signals using HFE are high dimension with information redundancy, which will make the diagnosis accuracy decreasing and time consuming. In this paper, we introduce an effective approach called Laplacian Score (LS) to select the first several important scale factors to construct the new fault feature vectors [17]. By virtue of the LS method, the fault feature vectors can be automatically ranked according to their importance and correlations with the main fault information [4], and then we select the first four important scale factors as the new fault feature vectors. By using LS, it can not only reduce the data dimension but also enhance the identification accuracy greatly.

Normally, after obtaining the fault features using HFE, another focus is to achieve the fault pattern recognition by using a multi-fault classifier. Support vector machine (SVM) based on statistical learning theory was put forward by Vapnik [18], which had been demonstrated to be effective in making a reliable decision for a smaller number of datasets. Since SVM has high accuracy and good generalization capabilities, it has been widely applied in fault diagnosis and classification field [19].

SVM was originally designed as a binary classifier, nevertheless, the practical identification problem is mostly the multi-class problem. So far, varieties of techniques for solving the multi-class problem based on SVM have been developed, such as: one against one (OAO) [20] one against all (OAA) [21] decision directed acyclic graph SVM(DDAGSVM) [22] and support vector machines based on binary tree (SVM-BT) [19]. Compared with other classifiers, SVM-BT has the superiorities as follows: fewer sub-classifiers, none of unclassifiable region and good classification performance, which is suitable for practical application of multi-class fault recognitions [19].

A major concern of SVM-BT is the adoption of hierarchical structures for the training of a multi-class SVM, which is vitally important for classification performance. Furthermore, the hierarchical structures should be designed before training each sub-classifiers of SVM, and the much research in recent years has focused on hierarchical structures design for SVM-BT. Among many measures, inter-class Euclidean distance (ED) has been one of the most effective methods to quantify the similarity of two classes, inter-class ED is used to construct the binary tree hierarchy [23], in which the class with bigger distance from other classes will be separated earlier and in the higher node of the binary tree architecture. In addition, intra-class sample distribution was introduced to design the hierarchical structures of binary tree [24,25]. Namely, the class with wider distribution range will be classified earlier. The inter-class ED and intra-class sample distribution are both the approaches to describe the distinguishability from different aspects. Considering the merits of the two approaches, in this paper, we introduce a novel measure to construct the hierarchical structure of binary tree, called improved SVM-BT (ISVM-BT),which takes advantages of the above two approaches and reflects the class separability more comprehensively.

The rest of this paper is organized as follows: in Section 2, we firstly recall SampEn, FuzzyEn and MFE methods. In Section 3, the proposed HFE method is described detailed. The ISVM-BT method is introduced, meanwhile, the superiority of ISVM-BT is validated in Section 4. The proposed method based on HFE and ISVM-BT is introduced briefly and experiment data analysis is presented in Section 5. Finally, conclusions are given in Section 6.

2. Review of SampEn, FuzzyEn and multi-scale fuzzy entropy

2.1. Sample entropy

To overcome the drawbacks of approximate entropy (ApEn), the sample entropy (SampEn) is proposed by Richman [9], which can provide more precise complexity estimation. For a given time series $\{x(i), i = 1, 2, \cdots, N\}$, the main calculation procedures of SampEn can be written as follows.

(1) Construct an m dimension vector according to the Eq. (1):

$$X_i^m = \{x(i), x(i+1), \ldots, x(i+m+1)\}, 1 \le i \le N-m+1 \tag{1}$$

(2) Define the distance d_{ij}^m for each pair of vectors (X_i^m, X_j^m) as

$$d_{ij}^m = \left\| X_i^m - X_j^m \right\|, 1 \le i, j \le N-m+1 \tag{2}$$

(3) Give a tolerance threshold r, calculate the total number of d_{ij}^m, when $d_{ij}^m < r$, and the rate of total number is denoted as B_i, for $1 \le N - m$, it can be written as

$$B_i^m(r) = \frac{1}{N-M-1} B_i \tag{3}$$

(4) The average value of $B_i^m(r)$ is then divided by N-m, which can be expressed as

$$B^m(r) = \frac{1}{N-M-1} \sum_{i=1}^{N-m} B_i^m(r). \tag{4}$$

(5) Calculate the $B^{m+1}(r)$ for dimension $m + 1$ by repeating the steps (1)–(4).

$$B^{m+1}(r) = \frac{1}{N-M} \sum_{i=1}^{N-m} B_i^{m+1}(r) \tag{5}$$

(6) The SampEn of the time series is then obtained by

$$\mathrm{Samp\,En} = \lim_{N\to\infty} \left\{ -\ln \left[\frac{B^{m+1}(r)}{B^m(r)} \right] \right\} \tag{6}$$

when N is finite, it can be written as

$$\mathrm{Samp\,En}(m, r, N) = -\ln \left[\frac{B^{m+1}(r)}{B^m(r)} \right]. \tag{7}$$

Since the similarity definition of the two vectors is dependent on the Heaviside function, it has the jumping character, which is not consistent with the ambiguous behavior of the actual times series' boundaries. That's why the SampEn may yield imprecise estimation. In order to overcome the drawbacks of SampEn, an alternative method FuzzyEn is developed to measure the complexity of a time series.

2.2. Fuzzy entropy

As an improvement of SampEn, FuzzyEn replaces the Heaviside function with a Gaussian function. Due to the continuity of the potential function, the FuzzyEn can avoid the drawbacks of the ApEn and SampEn effectively [14]. The main steps of FuzzyEn are described as follows:

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

$$X_i^m = \{u(i), u(i+1), \ldots, u(i+m+1)\} - u_0(i), i = 1, 2, \ldots, N-m+1 \tag{8}$$

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.

$$u_0(i) = \frac{1}{m} \sum_{k=0}^{m-1} u(i+k) \tag{9}$$

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

$$\begin{aligned} &d_{ij}^m = d\left[X_i^m, X_j^m\right] = \max_{k\in(0,m-1)} \{|[u(i+k) - u_0(i)] - [u(j+k) - u_0(j)]|\} \\ &i, j = 1, 2, \ldots N-m, i \ne j. \end{aligned} \tag{10}$$

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

$$D_{ij}^m = \mu\left(d_{ij}^m, n, r\right) = e^{-\ln 2 \left(d_{ij}^m / r\right)^n} \tag{11}$$

where n and r are the gradient and the width of the border, respectively.

(4) The $\varphi^m(n, r)$ and is then defined as follows

$$\varphi^m(n,r) = \frac{1}{N-m}\sum_{i=1}^{N-m}\left(\frac{1}{N-m-1}\sum_{\substack{j=1\\ j\neq i}}^{N-m} D_{ij}^m\right) \tag{12}$$

(5) Repeat the above steps (9)–(12) for obtaining $m + 1$ dimensional, and the $\varphi^{m+1}(n, r)$ can be described as

$$\varphi^{m+1}(n,r) = \frac{1}{N-m}\sum_{i=1}^{N-m}\left(\frac{1}{N-m-1}\sum_{\substack{j=1\\ j\neq i}}^{N-m} D_{ij}^{m+1}\right). \tag{13}$$

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

$$FuzzyEn(m,n,r) = \lim_{N\to\infty}\left[\ln\varphi^m(n,r) - \ln\varphi^{m+1}(n,r)\right]. \tag{14}$$

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

$$FuzzyEn(m,n,r,N) = \ln\varphi^m(n,r) - \ln\varphi^{m+1}(n,r). \tag{15}$$

2.3. The basis of multi-scale fuzzy entropy

In order to measure the complexity of time series more accurate and reliable, the multi-scale analysis algorithm was proposed by Costa [10], in which a coarse-grained procedure is used to preprocess the time series and resulting in a series of low frequency components of original data. Multi-scale fuzzy entropy (MFE) is developed by Zheng et al. [15], based on the concept of multi-scale analysis. Two procedures of the 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

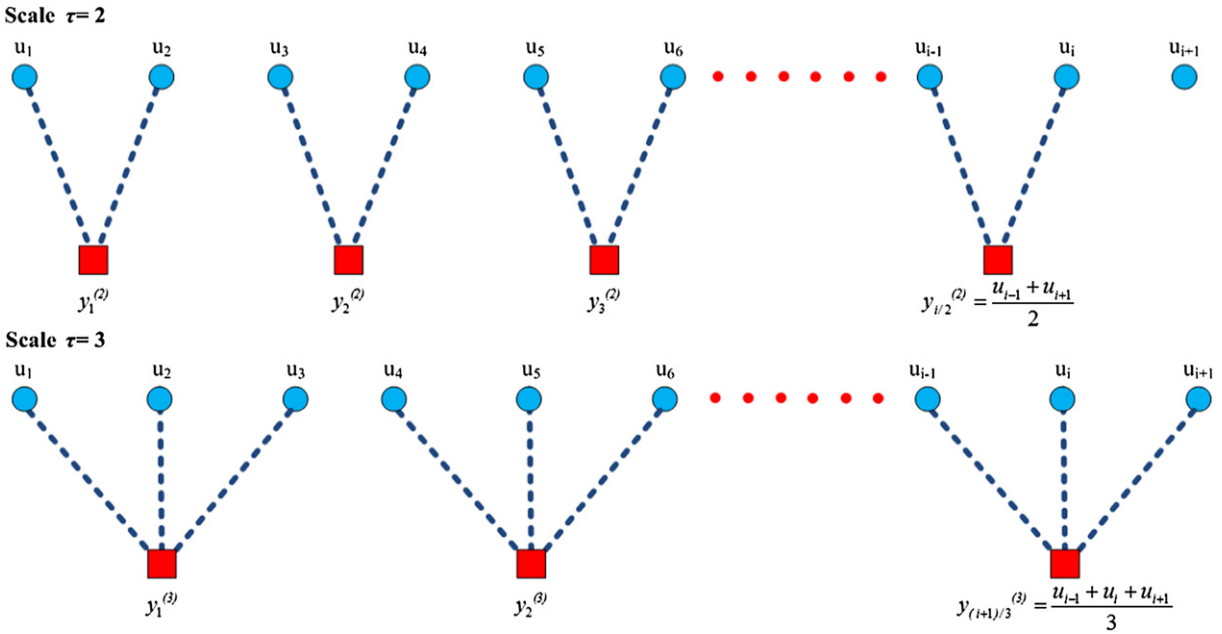

Fig. 1. The schematic illustration of the coarse-grained procedure for scale factor $\tau = 2$ and $\tau = 3$.

(τ is a positive integer), y_j^τ can be constructed according to the Eq. (16), and an example of the coarse-grained procedure is illustrated in Fig. 1.

$$y_j^\tau = \frac{1}{\tau}\sum_{i=(j-1)\tau+1}^{j\tau} u_i \quad 1 \le j \le \frac{N}{\tau} \tag{16}$$

(2) In the MFE analysis, the FuzzyEn of each coarse-grained time series is calculated according to the Eqs. (8)–(15) and then plotted as the function of the scale factor τ, which can be expressed as

$$MFE(x, \tau, m, r) = FuzzyEn\left(y_j^\tau, m, r\right). \tag{17}$$

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.

From the above procedures of MFE, the advantage of using MFE method is that it can estimate the complexity of the original time series over different scales. In fact, MSE method only considers the lower frequency components of the original time series, eliminating high-frequency components, which is far more optimal for analyzing the signals with extensive frequencies information from low to high frequencies.

3. Hierarchical fuzzy entropy (HFE)

3.1. The introduction of HFE

Focus on the drawbacks of MFE method for analyzing the higher frequency components, an alternative approach called hierarchical fuzzy entropy (HFE) is proposed in this paper, which is composed of hierarchical procedure and fuzzy entropy calculation. The HFE method is briefly summarized as follows.

To begin with, given a time series with the length N, $\{u(i), i = 1, 2, \cdots, N\}$, then an averaging operator Q_0 defined for the time series can be expressed as

$$Q_0(u) = \frac{u(2j) + u(2j+1)}{2} \quad j = 0, 1, 2, \cdots, 2^{n-1}. \tag{18}$$

Note that $N = 2^n$, n is a positive integer. The time series $Q_0(u)$ is the low frequency of the original series $u(i)$ with length of 2^{n-1} at scale 2. Simultaneously, another operator Q_1 is described as follows

$$Q_1(u) = \frac{u(2j) - u(2j+1)}{2} \quad j = 0, 1, 2, \cdots, 2^{n-1}. \tag{19}$$

Compared with $Q_0(u)$, $Q_1(u)$ describes the original series $u(i)$ from a high frequency. Also the length of $Q_1(u)$ is 2^{n-1} at scale 2. It can be got that the original time series can be reconstructed by using $Q_0(u)$ and $Q_1(u)$ according to

$$u = \left\{\left(Q_0(u)_j + Q_1(u)_j\right), \left(Q_0(u)_j - Q_1(u)_j\right)\right\}, j = 0, 1, 2, \cdots, 2^{n-1}. \tag{20}$$

Namely, the time series $Q_0(u)$ and $Q_1(u)$ constitute a two-scale analysis for the time series $u(i)$. When $j = 0$ or 1, the operator Q_j has a matrix representation as

$$Q_j(u) = \begin{bmatrix} \frac{1}{2} & \left(\frac{-1}{2}\right)^j & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \left(\frac{-1}{2}\right)^j & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & \frac{1}{2} & \left(\frac{-1}{2}\right)^j \end{bmatrix}_{2^{n-1} \times 2^n}. \tag{21}$$

Note that the size of the matrix Q_j can be adaptively changing depending on the length of the original time series $N = 2^n$.

Secondly, to realize the multi-scale analysis for the time series $u(i)$, the operators need to be utilized repeatedly. For an $k \in N$,and let $[\gamma_1, \gamma_2, \cdots, \gamma_n] \in \{0, 1\}$, the integer e can be written as

$$e = \sum_{j=1}^{k} \lambda_j 2^{k-j}. \tag{22}$$

Note that, e is non-negative integer. Once the e is fixed, a unique vector $[\gamma_1, \gamma_2, \cdots, \gamma_n]$ can be obtained by Eq. (22). Namely, the calculated vector is the binary representation of e. Based on the vector$[\gamma_1, \gamma_2, \cdots, \gamma_n]$, the hierarchical components of original time series $u(i)$ can be denoted as

$$u_{k,e} = Q_{\gamma_n} \cdot Q_{\gamma_{n-1}} \cdot \cdots \cdot Q_{\gamma_1}(u) \tag{23}$$

where k represents the k^{th} layer of the hierarchical analysis. $u_{k,0}$, $u_{k,1}$ are the low frequency and high frequency component of the time series $u(i)$ at scale $k + 1$, respectively. Then it can be deduced that: for different k and e, the signals $u_{k,e}$ consist of the hierarchical decomposition of signal $u(i)$ in multi-scale. Actually, the component of $u_{k,0}$ $(k = 1, 2 \cdots, N)$ is exactly the same as the MFE analysis of the time series $u(i)$ at scale 2^k. Fig. 2 gives the hierarchical decomposition of $u(i)$ in 4 scales by using the node tree diagram.

Finally, the HFE is obtained by calculating FuzzyEn of the obtained hierarchical component. Two procedures of the HFE algorithm are briefly described as follows.

(1) Perform the procedure of hierarchical decomposition to obtain the low frequency and high frequency component of the time series $u(i)$, each hierarchical component can be expressed as

$$u_{k,e} = Q_{\gamma_n} \cdot Q_{\gamma_{n-1}} \cdot \cdots \cdot Q_{\gamma_1}(u) \tag{24}$$

(2) In the HFE analysis, the FuzzyEn of each hierarchical component is calculated based on Eqs. (9)–(16) and then plotted as the function of the scale factor τ, which can be expressed as Eq. (25).

$$\mathrm{HFE}(u, k, e, m, r) = \mathrm{FuzzyEn}\left(u_{k,e}, m, r\right) \tag{25}$$

For the sake of convenient application, the Matlab code of the HFE method is also listed in Appendix A.

3.2. The parameter selection of HFE

We need to set four parameters before using HFE, including embedding dimension m, boundary width and gradient of the exponential function r and n, and the scale of hierarchical decomposition k. A detailed description of each parameter of FuzzyEn has been done in literature [15]. 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 k will affect the computation efficiency and lead to few points of each hierarchical component, while a too small k can't obtain enough hierarchical components from high to low frequencies. In this paper, the k is selected as 3.

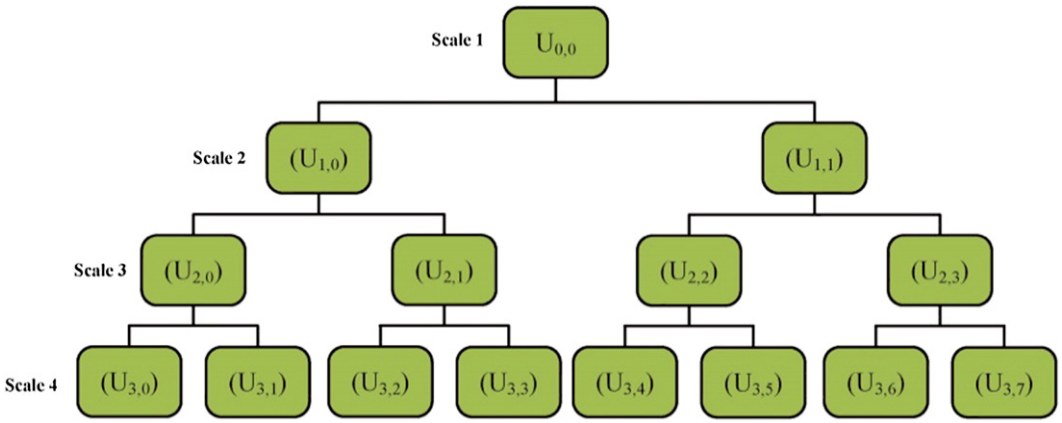

Fig. 2. Hierarchical tree diagram of time series $u(i)$ with four scales.

3.3. Comparison of HFE and MFE

In order to verify the superiority of IMFE method, the synthetic noise signals: white noise and 1/*f* noise are applied to conduct the comparisons between HFE and MFE method, which are referred to the work [4,12].

To begin with, the numerical results of the two synthetic noise signals with data lengths N = 5000 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 white noise is more complex than 1/*f* noise.

Secondly, to investigate the estimate performance of the HFE and MFE methods, the HFE and MFE are employed to analyze the white and 1/*f* noises across 8 scales. Besides, 100 independent noise signals are used to calculate the error bar at each scale, the obtained figure is shown in Fig. 4. As shown in Fig. 4, it can be found that the FuzzyEn of 1/*f* noise obtained by HFE decreases monotonically, while FuzzyEn of white noise is smooth and steady. The conclusions are consistent well with the results published in reference [12]. Also it can be seen that the FuzzyEn of white noise is higher than that of 1/*f* noise over the whole scales except beginning scale ($\tau = 1, 2$), which agrees with the intuitive results drawn in the Fig. 3**(c) and (d)**. However, the confusing conclusions are drawn using the MFE method, which implies that MFE can not depict the complexity of the white noise and 1/*f* noise accurately. Since two noise signals contain extensive frequencies from low to high frequencies, only consideration of lower frequency components in MFE is not unsuitable.

Lately, the SD (standard deviation) of HFE method should be discussed. Since the error bar at each scale indicates the SD of an FuzzyEn value, as seen in Fig. 4, the SD of HFE method is slight, which indicates that HFE has a stable performance of estimating the complexity. Hence, the HFE is utilized to extract fault features from the non-linear and non-stationary rolling bearing vibration signal under complex operating conditions.

4. Feature selection using Laplacian score algorithm

After extracting features from the vibration signal by using HFE, it is still necessary for us to refine the obtained feature vectors. In this paper, the LS algorithm is applied to automatically choose the optimum feature vectors. Laplacian score (LS) is fundamentally

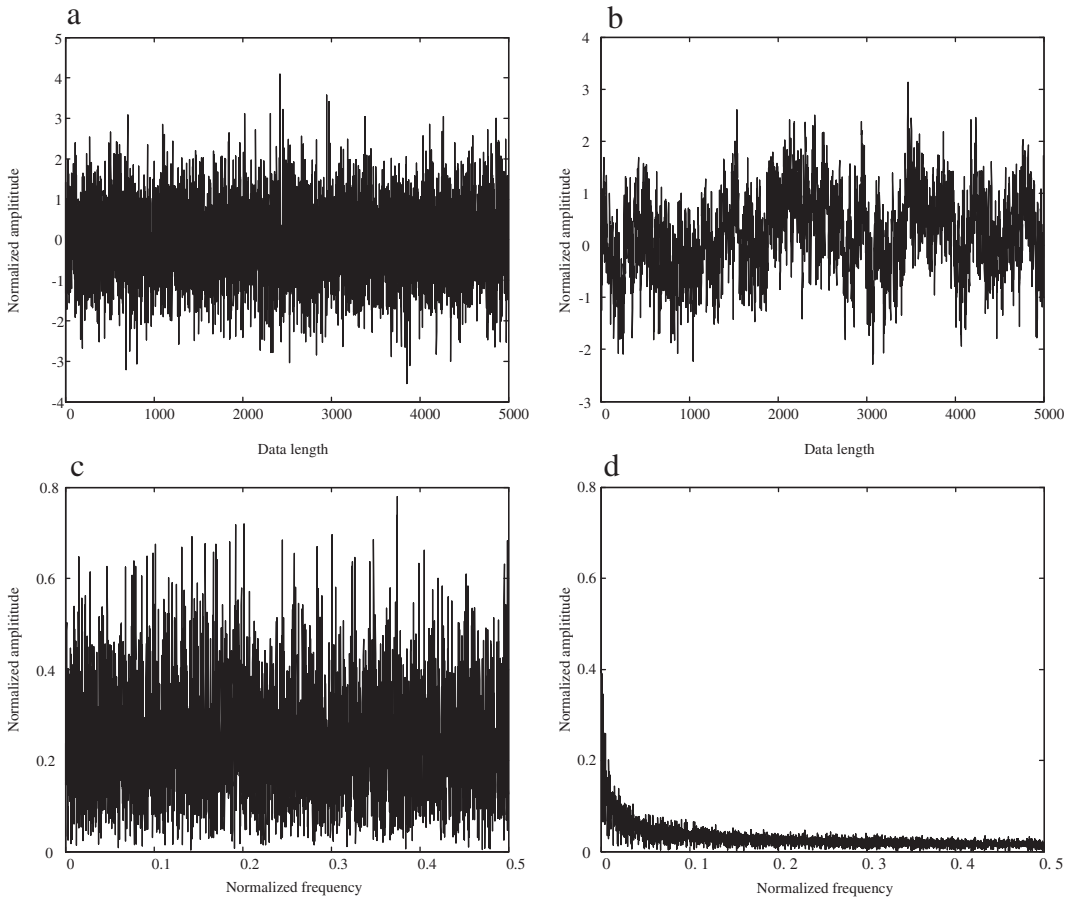

Fig. 3. (a) Waveform of white noise, (b) waveform of 1/f noise, (c) FT spectrum of white noise and (d) FT spectrum of 1/f noise.

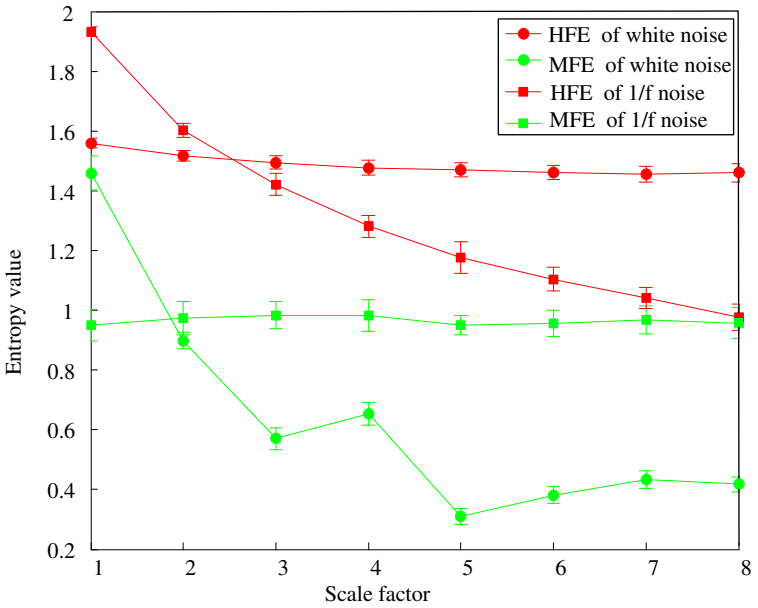

Fig. 4. HFE and MFE curves of white noise and 1/*f* noise.

founded on Laplacian Eigenmaps and 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 *r*-th feature, $r = 1,\ldots, n$. Let f_{ri} represent the *i*-th sample of the *r*-th feature, $i = 1, 2, \ldots, m$. The main calculation procedures of LS algorithm can be written as follows [17].

(1) Construct a nearest neighbor graph *G* with *m* nodes, where the *i*-th node corresponds to x_i. Then an 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.

(2) The weight matrix S_{ij} of the models can be defined as

$$S_{ij} = \begin{cases} e^{\left\|\frac{x_i - x_j}{t}\right\|} & \text{if nodes i and j are connected} \\ 0 & \text{otherwise} \end{cases} \tag{26}$$

where *t* is a suitable constant.

(3) For the *r*-th feature, f_r can be expressed as

$$f_r = [f_{r1}, f_{r2}, \cdots, f_{rm}]^T, \quad D = diag(SI), I = [1, \cdots, 1]^T, L = D - S. \tag{27}$$

where the matrix *L* is called graph Laplacian. Let

$$\tilde{f}_r = f_r - \frac{f_r^T DI}{I^T DI} I. \tag{28}$$

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

$$L_r = \frac{\sum_{ij}\left(f_{ri} - f_{rj}\right)^2 S_{ij}}{Var(f_r)} = \frac{\tilde{f}_r{}^T L \tilde{f}_r}{\tilde{f}_r{}^T D \tilde{f}_r} \tag{29}$$

where $\mathrm{Var}(f_r)$ is the estimated variance of the *r*-th feature value.

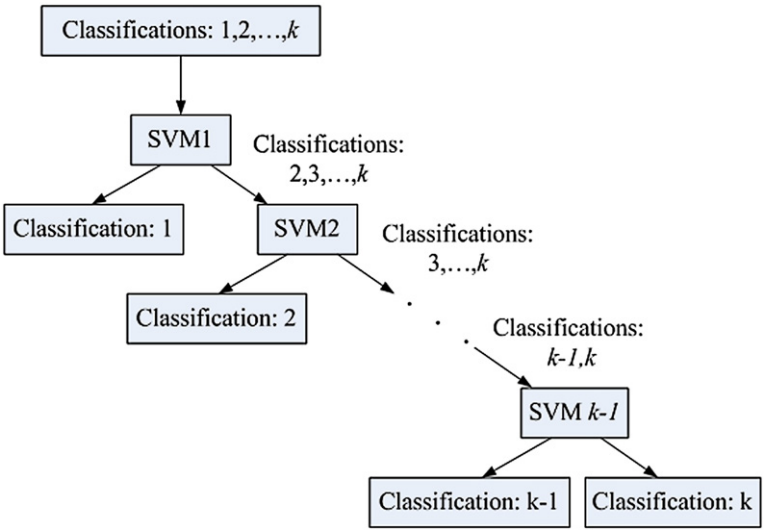

Fig. 5. Illustration of the model of SVM binary tree architecture.

It should be noted that the bigger S_{ij} value indicates the smaller the Laplacian Score value, which is the characteristic of a good feature. Therefore, the scales with lower LS values are taken as the important feature vectors. In this paper, the LS method is introduced to avoid the high dimension and enhance the identification efficiency.

5. The improved SVM-BT a comparisons

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 [18]. By virtue of the binary tree architecture, only $(k-1)$ SVM classifiers are needed to classify the k categories, and an example of SVM-BT that achieves pattern recognition problem using binary tree is shown in Fig. 5. Although the SVM-BT approach improves the performance of pattern recognition, it is heavily dependent on the sub-classifiers' assignment of hierarchical structures of BT. If sub-classifiers of SVM are arrayed randomly in hierarchical structures of BT, the recognition performance is far from optimal [26]. Much research in recent years has focused on the hierarchical structures of BT construction method.

It is generally accepted that average distances between classes can be utilized to measure the separability between classes, and the bigger average distance value means better separability. The average distance between classes is widely used to design the hierarchical structures of BT [23]. Namely, the class which has bigger distance from other classes is firstly separated. Recently, F.M. Tang proposed the distribution information inside one class to estimate the separability, in which the class with wider sample distribution will be firstly separated [24].

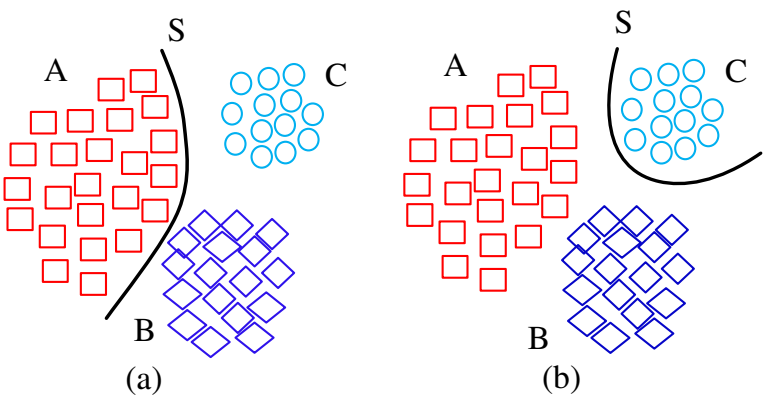

Fig. 6. Different order of classification for SVM-BT.

Take three classes for example to illustrate the two algorithms, which are shown in Fig. 6. As seen from Fig. 6, class *A* has the widest sample distribution, class *B* takes the second place and class *C* has the most concentrated sample distribution with the biggest distance between classes. If the average intra-class ED is applied to design the hierarchical structure of SVM-BT, class *A* is classified firstly and the separating hyperplane *S* is constructed shown in Fig. 6**(a)**. While the average inter-class ED is adopted, class *C* is classified firstly and the separating hyperplane *S* is constructed shown in Fig. 6**(b)**. Both the two algorithms have steady performance of distinguishability and good extend capability.

Based on the merits of the two separability measures, a new approach 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 improved SVM-BT (ISVM-BT) is designed to separate the class which has both bigger distance from other classes and wider sample distribution within itself firstly. The greatest advantage over existing SVM-BTs, however, 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 order to realize the proposed approach, it is necessary to select suitable parameters to satisfy the two separability measure standards. Since Euclidean distance(ED) has the advantages of efficient calculation and easy high dimensional space conversion, it has been widely used to compute the distance of sample set. Therefore, in this paper ED is taken as separability measure parameter.

The proposed way to design the hierarchical structures of SVM-BT consists of two algorithms, namely the average intra-class ED and the average inter-class ED, which are outlined later. Essentially, 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

The average intra-class ED is used to estimate the distribution information inside one class. The smaller average intra-class Euclidean distance value suggests more concentrated distribution. 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:

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

$$d_{ij}{}^{a} = d\left(a_i, a_j\right) \tag{30}$$

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

$$D_i^a = \frac{1}{k_a - 1}\sum_{j=1}^{k_a} d_{ij}^a, \quad i \neq j \tag{31}$$

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

$$AV^a = \frac{1}{k_a}\sum_{j=1}^{k_a} D_i^a \tag{32}$$

Algorithm 2. The average inter-class ED

The average inter-class ED is utilized to measure the distribution information among classes. The bigger average inter-class ED value indicates higher divisibility between classes. Give two classes sample sets: $\{a_i, i = 1,2,\ldots k_a\}$ and $\{b_i, i = 1,2,\ldots k_b\}$(note that $a_i \in$class *A*, $b_j \in$class *B*). The average inter-class ED is defined as following.

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

$$d_{ij}{}^{ab} = d\left(a_i, b_j\right) \tag{33}$$

Secondly, calculate the average ED from sample a_i to all samples of class *B*.

$$D_i^{ab} = \frac{1}{k_b}\sum_{j=1}^{k_b} d_{ij}^{ab} \tag{34}$$

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

$$AVI^{ab} = \frac{1}{k_a}\sum_{i=1}^{k_a} D_i^{ab} \tag{35}$$

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

$$I_{A,B} = AVI^{ab} + K\left(AV^{a} + AV^{b}\right) \tag{36}$$

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, and K is the weight coefficient.

Based on the $I_{A,B}$, 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 the range of the weight K is given. (In this paper, $K_n = 2^n, -4 \le n \le 4$, n is integer value).
(2) Compute the separability measure $SI = I_{i,j}$, $i,j = 1,2,\ldots,N$, $i \ne j$ for a given K and then construct the symmetric matrix as follows

$$SI = \begin{bmatrix} 0 & I_{1,2} & \cdots & I_{1,N-1} & I_{1,N} \\ I_{2,1} & 0 & \cdots & I_{2,N-1} & I_{2,N} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ I_{N-1,1} & I_{N-1,2} & \cdots & 0 & I_{N-1,N} \\ I_{N,1} & I_{N,2} & \cdots & I_{N,1N-1} & 0 \end{bmatrix}. \tag{37}$$

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 (a detailed illustration is shown in Fig. 5).
(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.
(8) All the testing accuracy rates for weight K_n can be obtained, and the optimum hierarchical structures of BT according to the highest testing accuracy rate is determined.

According to the steps above, the flowchart of the optimum hierarchical structures of BT selection is illustrated in Fig. 7. Hence, the ISVM-BT has better classification performance, which can be used to complete the fault type recognition automatically.

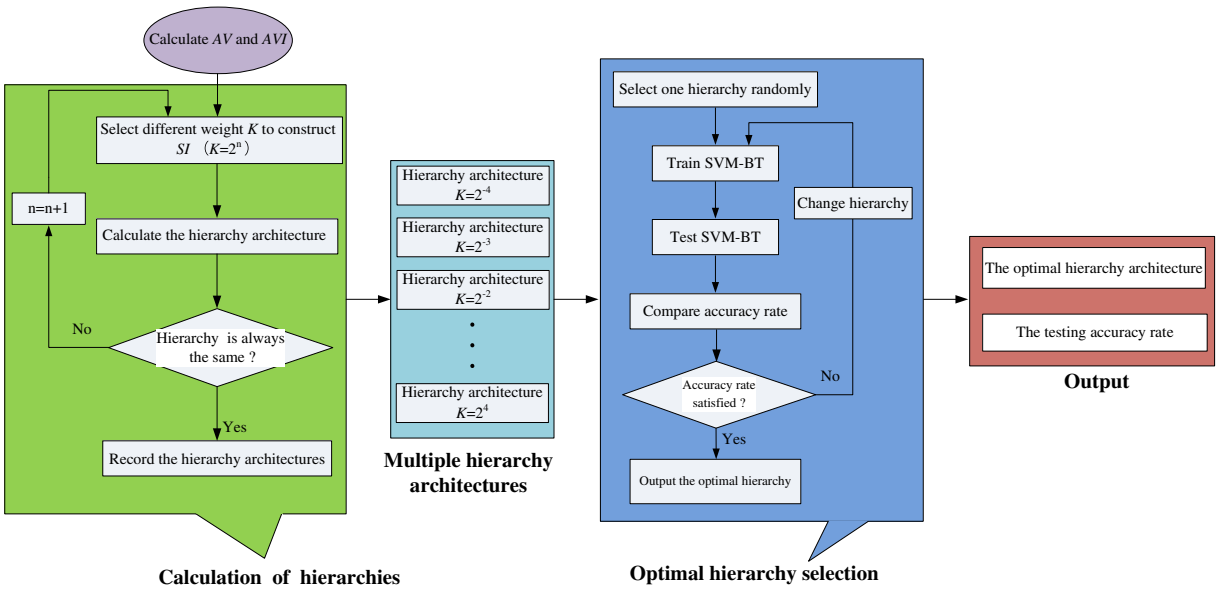

Fig. 7. Framework of ISVM-BT calculation procedure.

6. The proposed fault diagnosis method and applications

6.1. The proposed fault diagnosis method

A new rolling bearing fault diagnosis method is proposed based on the advantages of HFE, LS and ISVM-BT, the exact procedures of the proposed method can be summarized as follows:

(1) HFE method is utilized to calculate the rolling bearing vibration signal under different conditions. In this paper, the hierarchical decomposition of bearing vibration signal is defined in 4 scales (resulting in 8 child nodes) and the FuzzyEn values of each component derived from hierarchical decomposition are computed with the dimension $m = 2$ and tolerance $r = 0.15$*SD, and then 8 features are obtained to represent the fault information of rolling bearing under different conditions.
(2) After calculation of HFE, LS is employed to rank the 8 features according to their importance from low to high score. Then the first four important features with least scores were chosen to construct the new fault feature vector.
(3) The obtained new feature vectors are divided into training data set and testing data set, and the training data are used to train the improved SVM-BT to determine the optimum hierarchical structures of BT.
(4) The testing data are fed into multi-fault classifier of ISVM-BT to achieve the automatic fault patterns recognition according to the optimum hierarchical structures of BT determined in step (3).

It should be noted that since the penalty parameter C and the kernel parameter γ obtained by GA are not fixed, we calculate 5 times and select the best [γ,C] according to the highest accuracy rate. Also the highest accuracy rate is taken as the final training and testing accuracy rate.

6.2. Experimental analysis

In this paper, the rolling bearing experimental data analyzed are kindly provided by Bearing Data Center of Case Western Reserve University [27]. And the experimental data collection apparatus and description are given in detail in literature [15]. The proposed method for rolling bearing fault diagnosis is applied to the experimental data to validate its effectiveness. The experiment system and its sketch are given in Fig. 8, which consists of driving motor, a torque sensor, and a dynamometer used to control the torque load. The 6205-2RS JEM SKF deep groove ball bearing is utilized in this experiment, which is at the drive end. The fault bearings are using the electro-discharge machining with fault diameters of 0.1778 mm, 0.3556 mm, 0.5334 mm, and 0.7112 mm. The vibration signals of bearing were collected under four conditions including the inner race fault condition, the outer race fault condition, and the ball fault condition and normal condition. An accelerator was mounted on the front section end to collect the vibration signal with a sampling frequency of 12,000 Hz. Also the motor speed is set to be 1730 rpm with motor load about 2206.50 W and the sample time is 1 s.

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 sizes, actually, the experimental analysis is a ten-class recognition problem. The collected data sets are divided into several non-overlapping segments with the length $N = 2048$. There are 50 samples for each bearing condition, and there are total 500 samples, in which 100 samples will be randomly selected as the training data, and the residual 400 samples will be testing data. The detailed numbers of samples description for each bearing condition are shown in Table 1. The time domain waveforms and frequency spectrum of vibration signals under fault bearings with different types and severities as well as normal condition are illustrated in Figs. 9 and 10 As can be seen, it is difficult to identify the different fault types and severities accurately just by the time domain and frequency domain analysis, since the measured vibration signal often represents non-linear and non-stationary characteristics. To extract more fault information, HFE is applied to analyze the rolling bearing vibration signals.

Naturally, FuzzyEn is firstly performed to calculate the entropy values of 8 child nodes of hierarchical decomposition in four scales for each bearing condition, which means the dimension of obtained feather space is 8 in the following analysis. The HFE

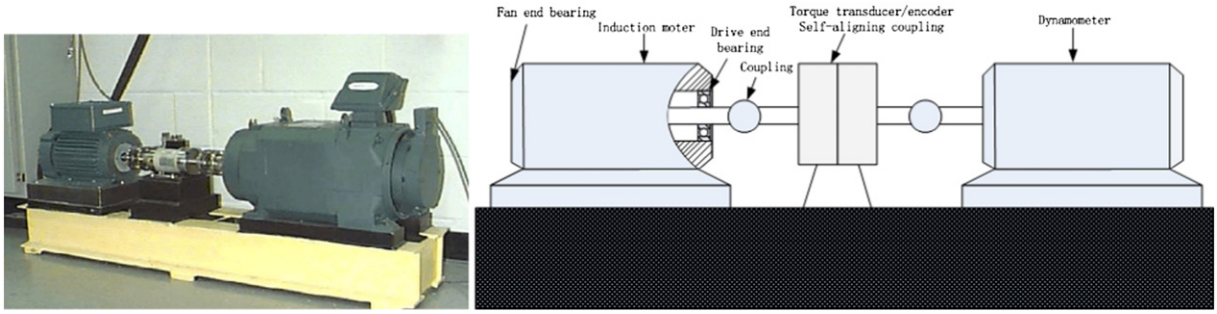

Fig. 8. The rolling bearing experiment system and its sketch.

Table 1
The detailed description of numbers of the experimental data sets.

Fault class	Fault size (mm)	Fault severity	Class label	Number of training data	Number of testing data
IRF	0.1778	Slight	1	10	40
	0.3556	Medium	2	10	40
	0.5334	Severe	3	10	40
	0.7112	Very severe	4	10	40
ORF	0.1778	Slight	5	10	40
	0.3556	Medium	6	10	40
	0.5334	Severe	7	10	40
BF	0.1778	Slight	8	10	40
	0.7112	Very severe	9	10	40
Normal	0		10	10	40

of bearing data under 10 health conditions is presented in Fig. 11. As can be seen, HFE of normal bearing condition has the largest value in the first hierarchical decomposition node, and then exhibits monotonically decreased trend with the increment of child nodes. Since the components derived from hierarchical decomposition are listed from low to high frequencies, HFE curve of normal condition demonstrates that the information is mainly embedded in low frequency component, which can be explained by the fact that there is no high frequency impulse induced by fault when no fault occurs [13]. Simultaneously, the bearing signals with various fault conditions have high HFE values both in the low frequency components (such as: node 3) and high frequency components (such as: node 7), it implies that the fault information is stored in both low and high frequency components, which is consistent with the characteristics of the measured multi-component, amplitude-modulated (AM) and frequency-modulated (FM) signals (the high frequency is modulated by the shock impulse generated by the local defect) when the bearing works under fault conditions.

From Fig. 11, it can be observed that applying HFE method to characterize the complexity of bearing vibration signals exhibits good stability and high distinguishability, even the trend of HFE values with different health conditions is similar. Therefore, HFE can be utilized to extract fault feather from the bearing vibration signals, and the obtained feather vectors are desired to have better separability to recognize the different fault categories.

However, if the HFE values of the 8 hierarchical decomposition nodes are all fed into the classifier to accomplish the patterns identifications, it will be time-consuming and result in the classification accuracy rate decreasing. In this paper, the LS algorithm is utilized to rank feature according to their importance, and the new orders of MPE can be shown as follows:

$$LS_2 < LS_4 < LS_1 < LS_6 < LS_8 < LS_5 < LS_3 < LS_7$$

where the subscript of *LS* represents the scale factors. The PEs are ranked and replotted in Fig. 12 and then the HFEs with nodes ($e = 2, 4, 1$ and 6) are selected to form the new feature vector.

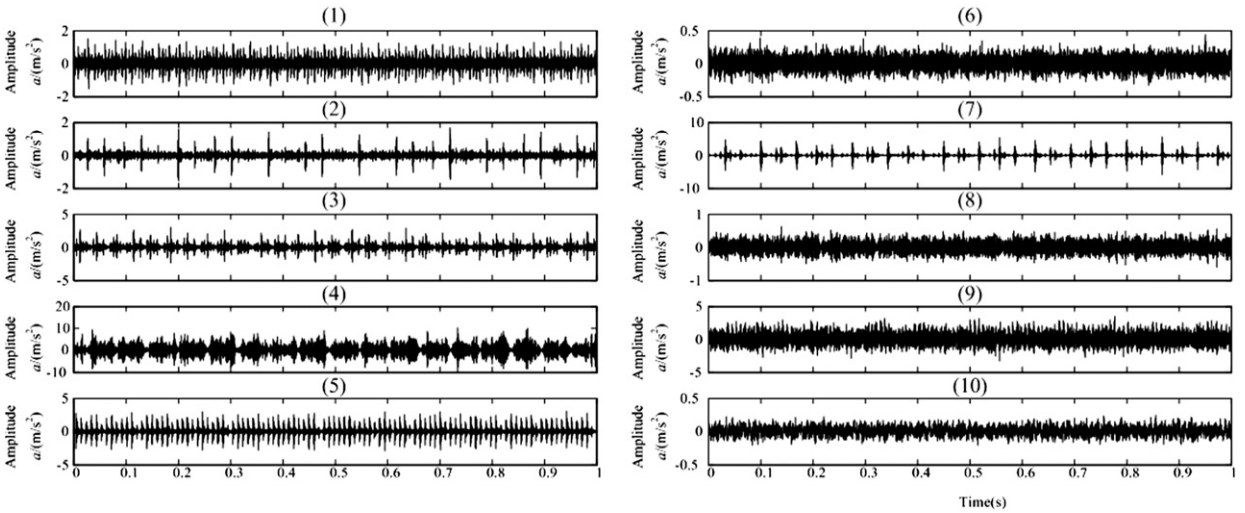

Fig. 9. The waveforms of rolling bearing vibration signal under ten different conditions.

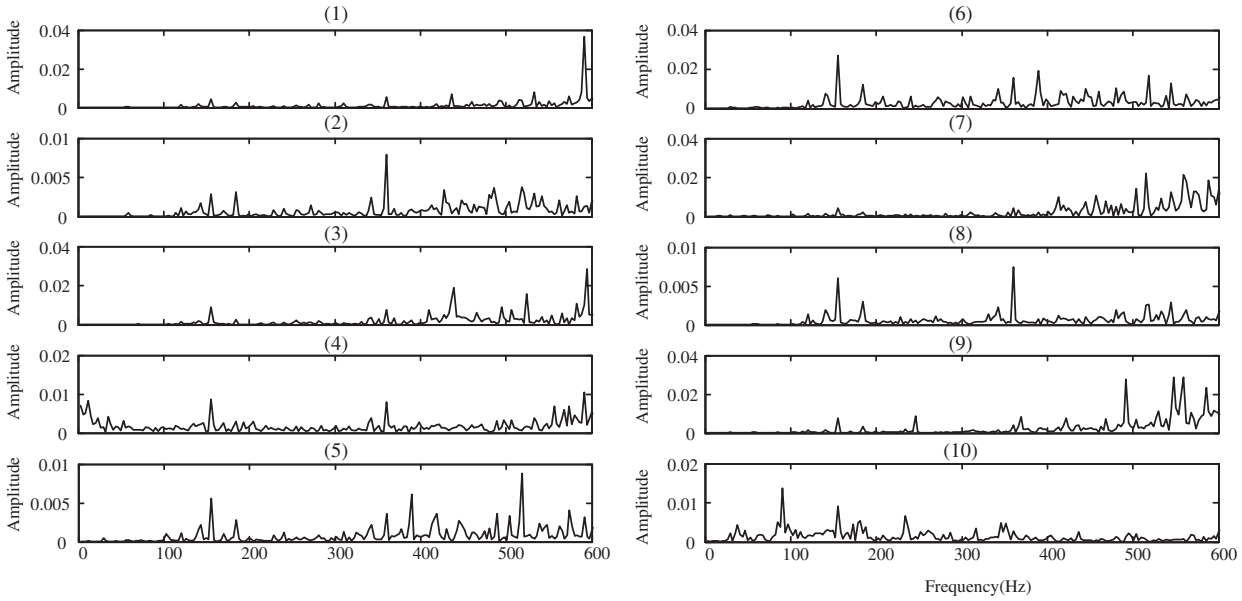

Fig. 10. The frequency spectrums of rolling bearing vibration signal under ten different conditions.

Naturally, after obtaining the new feature vectors using LS method, a multi-class support vector machine (SVM) is needed to automatically accomplish the fault conditions identifications. According to step 2 described in Section 6.1, 100 samples are selected randomly from the whole data set as the training data and the residual 400 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 5.1, the average intra-class ED *AV* and inter-class ED *AVI* are firstly calculated and then the hierarchical structure of BT was ensured by using weight $K(K = 2^{-4}, 2^{-3},..., 2^{3}, 2^{4})$. The classification results of testing data obtained by different weight K are listed in Table 2. It can be observed that the highest classification accuracy 100% can be obtained when the weight $K = 2^{1}$, and the

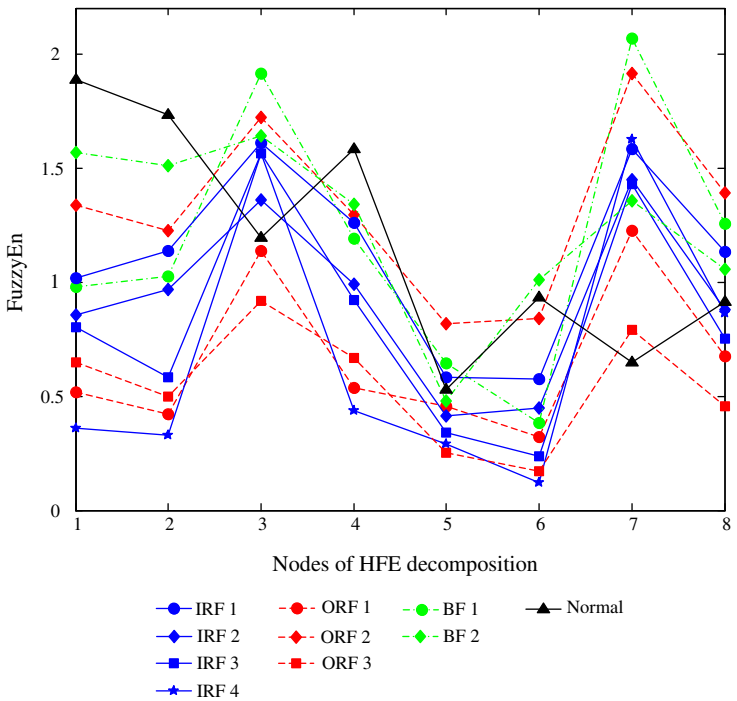

Fig. 11. HFE values of the 8 hierarchical decomposition nodes for analyzing 10 health bearing conditions.

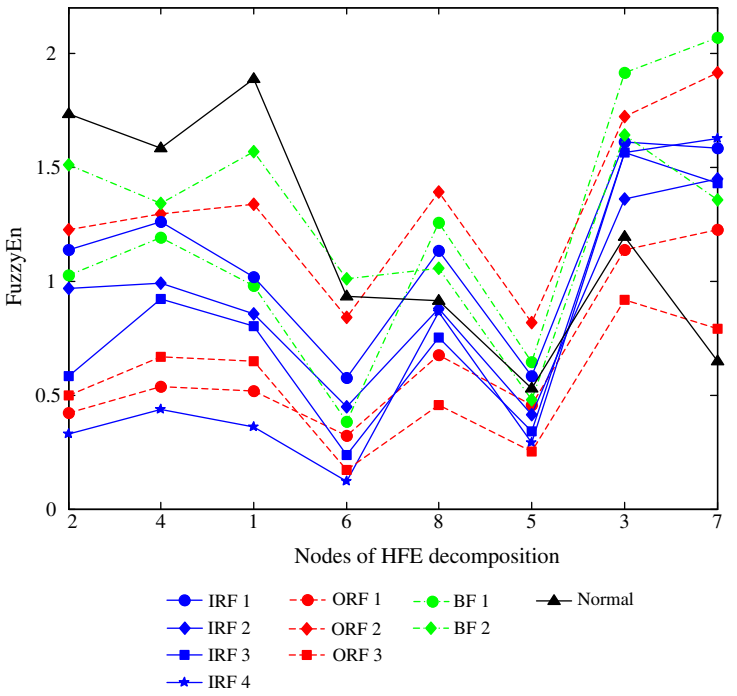

Fig. 12. HFE values of the 8 hierarchical decomposition nodes reordered by LS.

optimum hierarchical structure of BT is as follows: O2 > N > O3 > I4 > B2 > B1 > I3 > O1 > I1 > I2. Note that: *N* represents normal condition; I1, I2, I3, I4 represent the inner race fault with slight, medium, severe and very severe conditions respectively; O1, O2, O3 represent outer race fault with slight, medium and severe conditions; B1, B2 represent ball fault with slight and very severe conditions.

In order to verify the superiority of ISVM-BT, the single Intra-BT and Inter-BT are also employed to conduct the multi-class identifications. For comparison purpose, we take same training and testing data to train and test Intra-BT algorithm and Inter-BT algorithm. The classification results of the proposed method are shown in Fig. 13. As can be seen, there are no testing samples misclassified and the average recognition accuracy reaches 100%. Also the classification results of single Intra-BT and Inter-BT are shown in Figs. 14 and 15. It can be found that four and five testing samples are misclassified, respectively. The inner race fault, outer race fault and ball fault may be misclassified to wrong fault degrees and class labels. The average recognition accuracies of single Intra-BT and Inter-BT are 99.0% and 98.75%, which is lower than the proposed method in this paper. The comparison results validate the superiority of ISVM-BT's advantage in recognition accuracy. Besides, the preliminary comparisons demonstrate the feasibility of using HFE and ISVM-BT in the rolling bearing fault diagnosis.

Furthermore, to validate the superiority of HFE, MFE is also performed to analyze the bearing vibration signals under 10 health conditions. For comparison purpose, the MFE over 8 scales are also calculated to construct feather vectors, which are plotted in Fig. 16. Through the same process as the above-mentioned in HFE method, the optimum hierarchical structure of BT can be obtained when the weight $K = 2^{-1}$, and the highest recognition accuracy rate based on HFE and SVM-BT is 98.75%, with 5 testing samples misclassified. The classification results are shown in Fig. 17. It can be found from Fig. 17 that three testing samples with inner race fault are misclassified into different fault degrees and two testing samples with ball fault are misclassified into the other fault type.

Table 2
Hierarchy structure of BT and accuracy rate of different weights *K*.

Weights K	Hierarchy architecture										Accuracy rate (%)
2^{-4}	N	O2	O3	B1	B2	O1	I4	I3	I1	I2	99.25%
2^{-3}	N	O2	O3	O1	B2	B1	I4	I3	I1	I2	99.25%
2^{-2}	N	O2	O3	O1	B1	B2	I4	I3	I1	I2	99.50%
2^{-1}	N	O2	O3	B1	B2	I4	O1	I3	I1	I2	95.50%
2^{0}	O2	N	O3	I4	B1	B2	O1	I3	I1	I2	99.75%
2^{1}	O2	N	O3	I4	B2	B1	I3	O1	I1	I2	100%
2^{2}	O2	N	O3	I4	B2	I3	I1	B1	I2	O1	99.75%
2^{3}	O2	I4	O3	N	I1	I3	B2	I2	B1	O1	99.75%
2^{4}	I4	O2	O3	N	I1	I2	I3	B2	B1	O1	99.50%

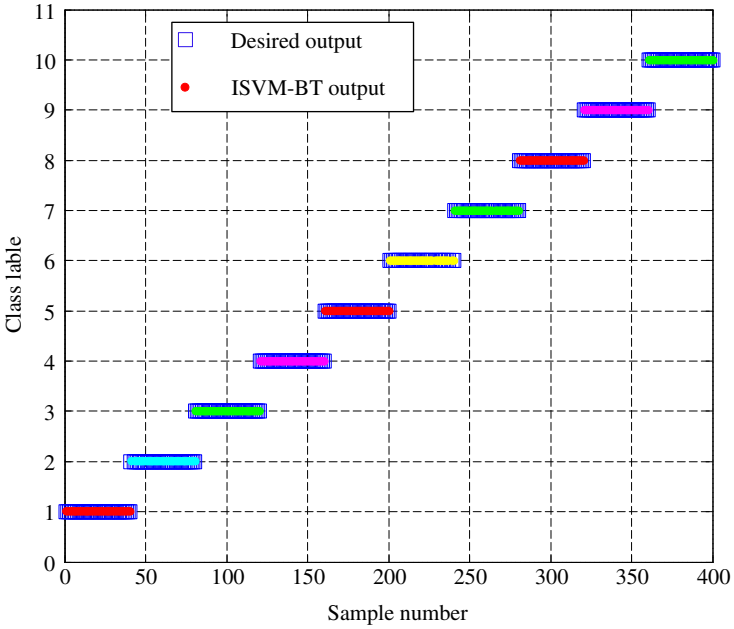

Fig. 13. Classification results of ISVM-BT method.

The comparison results provide compelling evidence that HFE can provide more precise and comprehensive information for assessing complexity than MFE, which is in general agreement with previous theory analysis. This can be explained by the fact that HFE can extract fault information from low to high frequency components of bearing vibration signal, while MFE only considers the low frequency components. Also, in order to verify the superiority of HFE under different multi-fault classifiers, the multi-fault classifiers based on SVM such as: OAO, OAA, Intra-BT, Inter-BT and ISVM-BT are all used to solve the ten-class recognition problem. Besides, the training and testing data are the same in each algorithm. The classification results and parameters [γ, C] of each multi-fault classifier using HFE and MFE are summarized in Table 3 Through comparing the classification results, the conclusions can be got as follows. To begin with, the classification accuracy rate using HFE is higher than that of MFE in

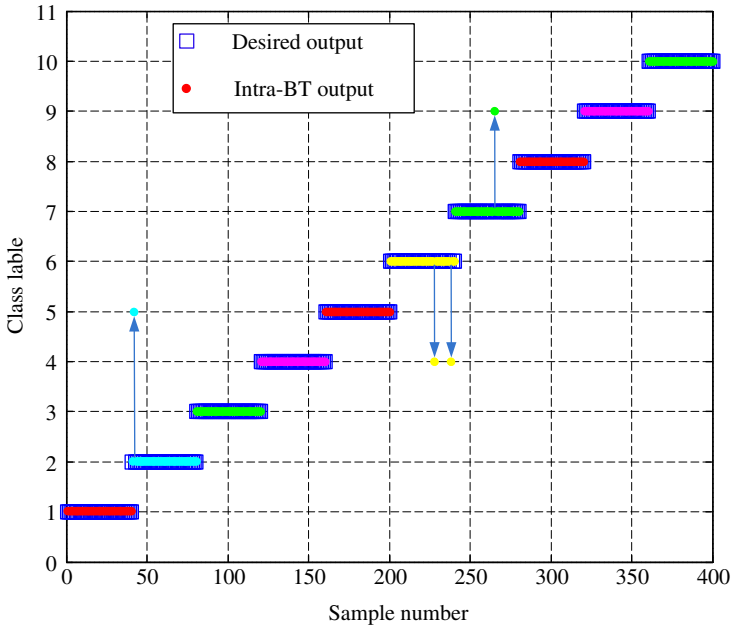

Fig. 14. Classification results of intra-BT method.

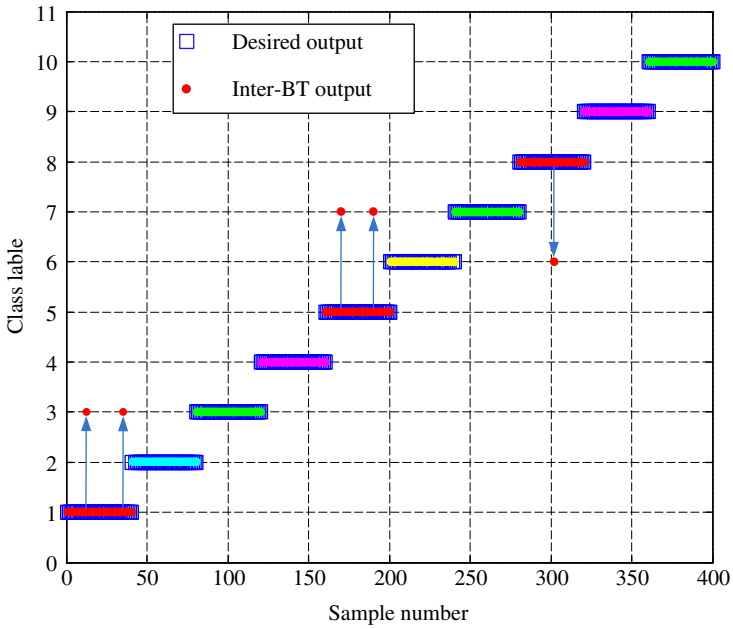

Fig. 15. Classification results of inter-BT method.

each multi-fault classifier, which reinforces superiority of HFE over MFE; secondly, of all the multi-fault classifiers, ISVM-BT has the highest classification accuracy rate, which further highlights the advantages of ISVM-BT classification performance. Hence, the results rule out the possibility that the above advantages of HFE, LS and ISVM-BT are a result of occasionality. The comparisons confirm the effectiveness of the fault diagnosis method based on HFE, LS and ISVM-BT.

In addition, the problem about the number of inputs of ISVM-BT needs to be addressed here. The total testing accuracy rates with different number of inputs are listed in Table 4. As can be seen, 4 or 5 is the suitable number to conduct the fault identification with a

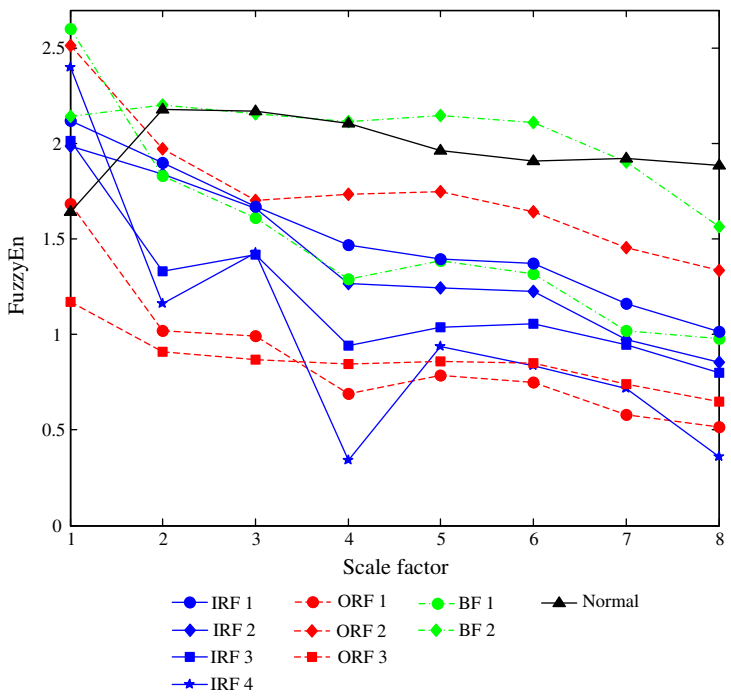

Fig. 16. MFE values over 8 scales for analyzing 10 health bearing conditions with the average of twenty trails.

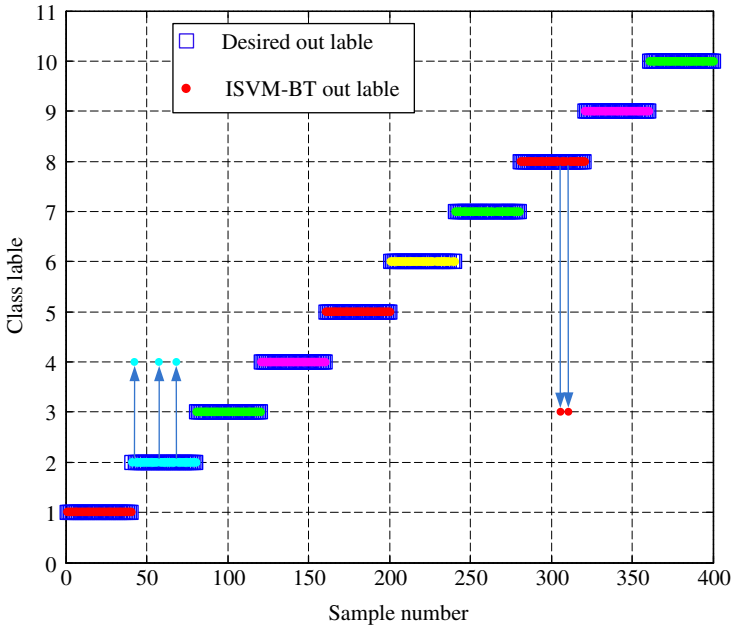

Fig. 17. Classification results of ISVM-BT using MFE method.

higher accuracy rate while smaller or larger number can both result in a lower accuracy rate, which is due to the fewer features with fewer fault information while more features with information redundancy.

Eventually, we need to discuss the calculation efficiency of HFE and MFE methods, as a comparison, we take the same data with length $N = 2048$ to test the calculation efficiency. The computing time of HFE is 2.16 s, while that of MFE is 22.56 s, which is more than 10 times than HFE. The comparison result indicates that HFE is a suitable and effective bearing feather extraction method, which can not only provide more sensitive information but also enhance the calculation efficiency appreciably.

7. Conclusions

Focus on the non-linear and non-stationary characteristics of rolling bearing vibration signal, a novel feature extraction algorithm based on HFE, LS and ISVM-BT is proposed in this paper. Firstly, HFE is used to extract fault features to form feature vectors. Then LS method is introduced to rank the feature vectors and refine the feature vectors with most important information. Lastly, ISVM-BT classifier is employed to recognize the various fault categories. For comparison purpose, HFE is compared with MFE by analyzing the experimental signal, the comparison results demonstrate that HFE can extract more fault information and present better divisibility than MFE. Also the ISVM-BT classifier is contrasted with OAO, OAA, Intra-BT, Inter-BT by using the same training and testing data. By comparing the recognition results, it indicates that ISVM-BT has superiority of high classification accuracy. Finally, the experimental rolling bearing fault diagnosis shows that the proposed approach based on HFE and ISVM-BT has superior performance in identifying different types and severities of rolling bearings. The proposed method is promising and should be applied in fault diagnosis of other mechanical equipment, and further work should focus on the determination of the optimum weight K in ISVM-BT.

Acknowledgments

The research is supported by National Natural Science Foundation of China (no. 11172078) and 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.

Table 3
Optimal parameters and recognition accuracy of each algorithm.

Methods	OAO	OAA	Intra-BT	Inter-BT	ISVM-BT
	[γ,C] R(%)	[γ,C] R(%)	[γ,C] R(%)	[γ,C] R(%)	[γ,C] K, R(%)
HFE	(1.5,13.2) 98.50	(4.2,15.7) 98.25	(2.6,14.4) 99.00	(1.0,15.1) 98.75	(1.3,8.4) 2^{1}, 100
MFE	(2.5,25.2) 97.50	(1.3,23.6) 97.00	(3.3,14.2) 98.25	(1.4,15.2) 97.75	(0.8,14.0) 2^{-1}, 98.75

Table 4
The accuracy rates of the ISVM-BT outputs with different number of inputs.

The number of inputs	2	3	4	5	6
Accuracy rate	100%	97.8%	100%	100%	98.2%

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.mechmachtheory.2015.11.010.

References

[1] 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.
[2] R. Yan, X. Gao, Approximate entropy as a diagnostic tool for machine health monitoring, Mech. Syst. Signal Process. 21 (2007) 824–839.
[3] 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.
[4] J. Zheng, J. Cheng, Y. Yang, A rolling bearing fault diagnosis method based on multi-scale fuzzy entropy and variable predictive model-based class discrimination, Mech. Mach. Theory 78 (2014) 187–200.
[5] H. Liu, M. Han, A fault diagnosis method based on local mean decomposition and multi-scale entropy for roller bearings, Mech. Mach. Theory 75 (2014) 67–78.
[6] D. Logan, J. Mathew, Using the correlation dimension for vibration fault diagnosis of rolling element bearing-I, Mech. Syst. Signal Process. 10 (1996) 241–250.
[7] M. Pincus, Approximate entropy as a measure of system complexity, Proc. Natl. Acad. Sci. 88 (1991) 2297–2301.
[8] L. Zhang, G. Xiong, H. Liu, Bearing fault diagnosis using multi-scale entropy and adaptive neuro-fuzzy inference, Expert Syst. Appl. 37 (2010) 6077–6085.
[9] J.S. Richman, J.R. Moorman, Physiological time-series analysis using approximate entropy and sample entropy, Am. J. Physiol. Heart Circ. Physiol. 278 (2000) 2039–2049.
[10] M. Costa, A.L. Goldberger, C.K. Peng, Multiscale entropy analysis of complex physiologic time series, Phys. Rev. Lett. 89 (2002) 068102.
[11] S.D. Wu, C.W. Wu, K.Y. Lee, S.G. Lin, Modified multiscale entropy for short-term time series analysis, Physica A 392 (2013) 5865–5873.
[12] Y. Jiang, C.K. Peng, Y.S. Xu, Hierarchical entropy analysis for biological signals, J. Comput. Appl. Math. 236 (2011) 728–742.
[13] K.H. Zhu, X.G. Song, D.X. Xue, A roller bearing fault diagnosis method based on hierarchical entropy and support vector machine with particle swarm optimization algorithm, Measurement 47 (2014) 669–675.
[14] W. Chen, W. Zhuang, Z. Wang, Measuring complexity using FuzzyEn, ApEn, and SampEn, Med. Eng. Phys. 31 (2009) 61–68.
[15] J.D. Zheng, J.S. Cheng, Y. Yang, A rolling bearing fault diagnosis approach based on LCD and fuzzy entropy, Mech. Mach. Theory 70 (2013) 441–453.
[16] M. Hu, H.L. Liang, Adaptive multiscale entropy analysis of multivariate neural data, Biomed. Eng. 59 (2012) 12–15.
[17] X. He, D. Cai, P. Niyogi, Laplacian Score for Feature Selection, Advances in Neural Information Processing System2005 1–6.
[18] Vladimir Vapnik, The Nature of Multiscale Entropy Theory, The Nature of Statistical Learning Theory 2000, pp. 132–158.
[19] S. Cheong, S.Y. Lee, Support vector machines with binary tree architecture for multi-class classification, Neural Inf. Process. Lett. Rev. 2 (2014) 47–51.
[20] K. Stefan, L. Personnaz, G. Dreyfus, Single-layer learning revisited: a stepwise procedure for building and training a neural network, Neurocomputing (1990) 41–50.
[21] S.S.Y. Ng, P.W. Tse, K.L. Tsui, A one-versus-all class binarization strategy for bearing diagnostics of concurrent defects, Sensors 14 (2014) 1295–1321.
[22] C. John, C. Nello, S.T. John, Large Margin DAGs for Multiclass Classification in nips, 1999 547–553.
[23] M. Gjorgji, G. Dejan, Evaluation of distance measures for multi-class classification in binary SVM decision tree, Artif. Intell. Soft Comput. (2010) 437–444.
[24] F.M. Tang, Z.D. Wang, J.Y. Chen, On multiclass classification methods of support vector machines, Control Decis. 20 (2005) 746–749.
[25] S.F. Yuan, F.L. Chu, Multi-class fault diagnosis based on support vector machines with sequenced binary tree architecture, J. Vib. Shoch 28 (2009) 51–55.
[26] M. Gjorgji, G. Dejan, Hybrid decision tree architecture utilizing local SVMs for multi-label classification, Hybrid Artif. Intell. Syst. (2012) 1–12.
[27] Bearing Data Center, Case Western Reserve University, http://csegroups.case.edu/bearingdatacenter/pages/download-data-file.