Fault Features Extraction and Identification based Rolling Bearing Fault Diagnosis

B Qin¹, G D SUN¹, L Y ZHANG¹, J G WANG¹, J HU ²

¹ School of Mechanical Engineering, Inner Mongolia University of Science & Technology, Baotou 014010
² Chungking Pure-Smart & Technology CO. LTD, Chungking 400030

E-mail:nkdqb@163.com

Abstract. For the fault classification model based on extreme learning machine (ELM), the diagnosis accuracy and stability of rolling bearing is greatly influenced by a critical parameter, which is the number of nodes in hidden layer of ELM. An adaptive adjustment strategy is proposed based on vibrational mode decomposition, permutation entropy, and nuclear kernel extreme learning machine to determine the tunable parameter. First, the vibration signals are measured and then decomposed into different fault feature models based on variation mode decomposition. Then, fault feature of each model is formed to a high dimensional feature vector set based on permutation entropy. Second, the ELM output function is expressed by the inner product of Gauss kernel function to adaptively determine the number of hidden layer nodes. Finally, the high dimension feature vector set is used as the input to establish the kernel ELM rolling bearing fault classification model, and the classification and identification of different fault states of rolling bearings are carried out. In comparison with the fault classification methods based on support vector machine and ELM, the experimental results show that the proposed method has higher classification accuracy and better generalization ability.

1. Introduction
Rolling bearing is playing an increasingly important role in rotary mechanical. However, diverse working conditions, including full of dust, high humidity, high temperature, and unpredictable heavy load etc., may cause the power drive system prone to fault, which is also one of the important sources of system vibration, noise, and lead to catastrophic failure or even huge economic loss. Therefore, timely monitoring of bearing and classifying faults into correct types are of great significance.

Today, data-driven fault monitoring and classification methods have attracted much attention[1]. In fact, the diversity and quality of modeling data influence the effectiveness of the fault classification. After collection of data, several crucial points should be discussed: 1) how to extract the fault features from the collected signal; 2) how to improve the classification accuracy of the fault identification model. For the fault-features extraction, Arnaz et al.[2] proposed a method to construct a mixed domain feature set based on wavelet decomposition. However, the time series is only extracted by wavelet decomposition in the frequency domain and time domain features are neglected. On the contrary, Guo et al.[3] extracted fault-features from frequency domain from six aspects, which include original vibration signal and its spectrum, the filtered, and demodulated signal by wavelet packet etc. To better extract fault features, Cui et al.[4] extracted the signal features by wavelet packet decomposition. In time domain, the time delay of original signal was pretreated by wavelet packet decomposition. Then
in the frequency domain, threshold of signal was constructed and fault features were extracted through spectrum analysis. However, the selection of wavelet-based functions is difficult and subjective.

Recently, intelligent algorithms have been applied to fault classification of bearing, such as artificial neural network\(^5\), fuzzy logic\(^6\), and least squares support vector machines\(^7\) etc. The traditional back propagation neural network (BPNN) faces the problem of slow convergence rate, difficulty of configuring tunable parameters and easily falling into the local optimum. In comparison with BPNN, the generalization performance of support vector machine (SVM) is improved. However, kernel function and its parameters are usually given according to human experiences. Extreme learning machine (ELM)\(^8\) is a newly developed single hidden layer feed-forward neural network, which does not require continuously to adjust parameters of hidden layer. In ELM, iterative parameter optimization process of traditional neural network is replaced by solving linear equation groups and the outputs of the minimum norm least squares solution are employed as the weights of the network. Therefore, the network is trained one-time without iterations. Compared with BPNN and SVM, ELM greatly improve the training speed and generalization ability, which has been successfully applied in the areas such as pattern recognition\(^9\)\textsuperscript{-14}. The idea of introducing ELM into fault classification for bearing is given in this paper. However, the adjustment of the weights of the input and hidden layer adopts random selection method, which can not guarantee the validity of the weight and easily cause the problem of redundant network structure. Because the structure parameters of fault identification model determine the classification ability, how to obtain the optimal structure parameters is the key to improve the classification accuracy of fault identification model.

To solve the above mentioned problems, a comprehensive and effective fault feature extraction and classification algorithm is proposed in this paper. First, to overcome the influence of disturbances, morphological average filtering algorithm is given to filter the collected signals and then the mode function (MF) is obtained through vibrational mode decomposition (VMD)\(^15\) after de-noising. Mode function presents fault features and MF components that have large correlation coefficients are used to calculate energy index in time domain. Besides, on the other hand, entropy of these MFs are obtained by Permutation Entropy (PE)\(^16\) analysis, which can be used to obtain the singular values. The matrix of PE is reconstructed high dimension feature vector used for classification. Third, the inner product of Gauss kernel function is utilized to express the ELM output function, which adaptively determine the number of the hidden layer nodes. At last, the fault classification model is developed by these feature vectors and the K-ELM algorithm.

The remainder of this paper is organized as follows. In Section 2, the extraction of multiple feature vectors are given. Fault classification method is formulated based on the enhanced ELM algorithm is proposed in Section 3. Experiment results clearly demonstrate the effectiveness and feasibility of the proposed method in the last section.

### 2. The extraction of feature vectors

The specific processes of feature vectors extraction are shown in Fig. 1. First, rolling bearing vibration signals are de-noised by morphological averaged filter. After de-noising, feature extraction contains two sections: 1) employ VMD to get intrinsic mode components (MFs) that have large correlation coefficients; 2) calculate permutation entropy using the obtained MFs. Through the above two steps, the calculated feature matrix is treated as feature vectors for bearing fault classification.
2.1. De-noising of original signal

Because data collected from signal gathering system usually contain noise, here mathematical morphology (MM) and average filtering algorithm are used to filter it. The basic idea of MM is to use some structural elements that have certain shapes to measure and extract images corresponding to the shape and achieve the purpose of image analysis. Based on the geometric characteristics of the signal, MM based average filter can cope with the nonlinear signal noise by morphological operations between structural elements and the original signal. The proposed filter inherits the advantages of MM, including simple operation, processing the local properties of signal in the time domain, does not have amplitude attenuation and phase shift in the waveform. Therefore, it is helpful for the processing of mechanical fault signals.

Opening and closing are the basic operations of MM, as shown in Eqs. (1) and (2), as follows,

\[(f \circ g)(n) = (f \Theta g + g)(n)\]
\[(f \bullet g)(n) = (f + g \Theta g)(n)\]

where \(\Theta\) is erosion operation presenting the relationship of \((f \Theta g)(n) = \min[f(n+m) - g(m)], n,m \in [0,N-1]\), \(\bullet\) is dilation operation having the relationship of \((f \oplus g)(n) = \max[f(n-m) + g(m)], n,m \in [0,N-1]\) and \(n > m\).

The linear combinations of Eqs. (1) and (2) can be used to construct the average filter (AVG), as shown in Eq. (3). It eliminates positive and negative impulse of the signal and can smooth the signal and reduce the signal noise.

\[\text{AVG}(f) = (f \cdot g + f \circ g) / 2\]

2.2. Extraction of feature vectors

2.2.1. Calculation of MFs components based on VMD

Figure 1 Flow chart of the extraction of feature vectors.
However, harsh operation conditions of rolling bearing, its vibration signal always contains process disturbances, including the resonance, external noise, etc. Then after de-noising using Eq. (3), Variational Mode decomposition\cite{15} is employed to extract the intrinsic characteristics of signals, which attempts to obtain these central frequencies and intrinsic mode functions centered on those frequencies concurrently using alternate direction method of multipliers (ADMM)\cite{18}.

Here, note the de-noised signal still as \( f \) that is represented as Eq. (4) through ADMM,

\[
\begin{equation}
\min_{\{u_k, \omega_k\}} \left\{ \sum_k \left| \left( e^{j\omega_k t} E_k(t) \right) \right| \right\}
\end{equation}
\]

\[
\text{s.t.} \sum_k u_k = f
\]

The objective function is defined as an unconstrained optimization problem as Eq. (4) by means of augmented Lagrangian Multiplier method.

\[
L(\{u_k\}, \{\omega_k\}, \lambda) = \sum_k \left| \left( e^{j\omega_k t} E_k(t) \right) u_k(t) \right| e^{-j\omega_k t}
\]

\[
+ \left\{ f(t) \cdot \sum_k u_k(t) \right\}^2 + \lambda(t) \cdot f(t) \cdot \sum_k u_k(t)
\]

In ADMM philosophy, we solve for one variable at a time assuming all others are known. So the formula for updating \( u_k \) at the \( n+1 \) iteration is as follows. The convergence of the algorithm is given in\cite{19}.

The final algorithm for VMD\cite{15} as follows.

**First**, initialize \( \{\hat{u}_k^0\}, \{\hat{\omega}_k^0\}, \hat{\lambda}^0 \). \( n \leftarrow 0 \)

**Repeat,**

\[
\begin{align*}
& n \leftarrow n + 1 \\
& \text{For } k \leftarrow 1:K \\
& \text{Update } \hat{u}_k \text{ for all } \omega \geq 0 \\
& \hat{u}^{n+1}_k \leftarrow \frac{\hat{f}(\omega) \cdot \sum_k \hat{u}_k(\omega) + \hat{\lambda}(\omega)}{1 + 2\alpha(\omega - \omega_k)^2}
\end{align*}
\]

**Update** \( \omega_k \):

\[
\omega^{n+1}_k = \int_0^\infty \omega |\hat{u}(\omega)|^2 d\omega
\]

**End**

**Dual ascent for all** \( \omega \geq 0 

\[
\hat{\lambda}^{n+1}(\omega) \leftarrow \hat{\lambda}^n(\omega) + \tau \left( \hat{f}(\omega) - \sum_k \hat{u}^{n+1}_k(\omega) \right)
\]

**Until convergence:** \( \sum_k \left( ||\hat{u}^{n+1}_k - u_k^n||^2 / \sum_k ||u_k^n||^2 \right) < \varepsilon \).

**2.2.2. Permutation entropy**

With the superiority, permutation entropy (PE)\cite{16} has been valid used in many spheres, such as fast calculation, robustness, invariance to nonlinear monotonous transformations, and so on. By comparing the size of the adjacent data, dynamic mutation of the time series are tested by PE. In the process of the above, specific values of data operation will not be involved, which avoid the interference of noise and reduce the computational complexity. When the rolling bearing failure occurs, the nonlinear
vibration of fault source aroused, meanwhile the complexity of the bearing vibration signal will be changed. Thus, PE can be used to measure the complexity of the signal. The basic principle of PE is briefly described as follows, more detailed information about PE can be found in Refs. [16] and [33].

For a time series \( \{x(k), k=1,2,...,N\} \), an \( D \)-dimensional delay embedding vector at time \( i \) is defined as:

\[
X_i^D = [x(i), x(i+\tau), x(i+2\tau), \ldots, x(i+(D-1)\tau)]
\]

(9)

where \( i=1,2,\ldots,N \), \( D \geq 2 \) represents the embedded dimension and \( \tau \) represents the time delay.

Then, denote the symmetric group of order \( D! \) as \( S_D \), which is the group of all the permutations of length \( D \). The symbols in \( S_D \) can be set as \( \pi=(i_1, i_2, \ldots, i_D) \). Then, we say that \( X_i^D \) has a permutation \( \pi_i \), if and only if \( \pi_i \) is unique in \( S_D \) and satisfies:

\[
x(i+(j_1-1)\tau) \leq x(i+(j_2-1)\tau) \leq \ldots \leq x(i+(j_D-1)\tau)
\]

(10)

where \( 1 \leq j \leq N-(D-1)\tau \).

For each permutation \( \pi_i \), the relative frequency can be obtained by:

\[
P(\pi_i) = \frac{\text{Num}\{X_i^D\}}{N-(D-1)\tau}
\]

(11)

where \( 1 \leq j \leq N-(D-1)\tau \).

Then, according to Shannon’s entropy theory, the PE is defined as:

\[
H_p(D) = -\sum_{\pi \in S_D} p(\pi_i) \ln(p(\pi_i))
\]

(12)

Obviously, \( 0 \leq H_p(D) \leq \ln(D!) \), where the lower bound is attained for an increasing or decreasing sequence of values, and the upper bound for a completely random system where all of the possible \( D! \) permutations appear in the time series with the same probability \( 1/\ln(D!) \). For convenience, Eq. (11) is always normalized onto the interval of \([0, 1]\) by \( \ln(D!) \), and denoted as:

\[
H_p = H_p(D) / (\ln(D!)) = -\frac{1}{\ln(D!)} \sum_{\pi \in S_D} p(\pi_i) \ln(p(\pi_i))
\]

(13)

The main idea of the PE algorithm is to match the candidate time series onto a symbolic sequence which describes the relationship between present values and a fixed number of equidistant values at a given past time. To calculate the value of PE only needs relationship between any two sample points in the time series. This is a highly favourable characteristic which makes PE avoid noise interference and has strong robustness with low computational complexity. Moreover, PE measures the departure of the time series from a complete random one: the smaller the value of the PE, the more regular the time series is. This implies that PE value increases with the irregularity or randomness of the time series\(^{[20]}\). The changes of PE value reflect and amplify the tiny changes of the time series.

By processing each group of the signal under different status according to the above steps, we obtain these high dimensional singular value matrixes as feature vectors to classify fault of rolling bearing. And feature vectors are adopted to train classification model of rolling bear based on kernel extreme learning machine, which will be given in the following section.

3. Identification Model Based Kernel Extreme Learning Machine

The accuracy of fault classification depends on the intelligent model used in the process of machine learning methods. Compared with the algorithm of BPN, SVM, ELM only needs to set the number of nodes for hidden layer to train the network and it has the advantages of high efficiency, fast learning speed and the unique solution. However, two structure parameters of ELM, i.e. input weights and hidden layer threshold, are randomly given and they will inevitably lead to poor accuracy. Having the advantages of dynamic control of global and local search conversion and avoids falling into local optimum, ELM output function is expressed by the inner product of Gauss kernel function to adaptively determine the number of the hidden layer nodes. Therefore, Gauss kernel function conveyed ELM is used in the construction of a rolling bearing fault classification model to improve
the precision and generalization ability. The processes of intelligent fault classification are shown in Fig. 2 concretely.

![Flowchart of intelligent fault classification](image)

**Figure 2 Flow chart of intelligent fault classification.**

### 3.1. The establishment of fault classification model

In this part, the fault classification model is proposed based on Kernel-ELM. Only requiring the number of neurons in hidden layer, ELM randomly generates connection weights and threshold of hidden layer neurons between the input layer and hidden layer and it can obtain the unique optimal solution.

Assuming the number of samples is \( N \), i.e. \( (x_i, y_i) \in [1, 2, \ldots, N] \), the number of nodes of hidden layer is \( H \) and the activation function is \( g(x) \), the mathematical model of ELM is below,

\[
y_i = \sum_{j=1}^{H} \beta_j g(w_j x_i + b_j)
\]  

(14)

where \( w_j = [k_{j1}, k_{j2}, \ldots, k_{jm}] \) is the connection weights vector between the input node and the \( j^{th} \) node of hidden layer, \( b_j \) is threshold of the \( j^{th} \) node in hidden layer.

In Eq. (14), a feed-forward neural network model of single hidden layer is developed, of which the output is close to zero error,

\[
\sum_{i=1}^{N} \| y_i - t_i \|^2 = 0
\]

(15)

Sequentially, parameters \( w_j, b_j, \beta_j \) satisfy the relationship of Eq. (16),

\[
\sum_{j=1}^{H} \beta_j g(w_j x_i + b_j) = t_i, i = 1, 2, \ldots, N
\]

(16)

And Eq. (14) can be further simplified as \( \mathbf{H} \beta = \mathbf{T} \), where,

\[
\mathbf{H} = \begin{bmatrix}
g(k_{i1}, b_{i1}, x_1) & \ldots & g(k_{iL}, b_{iL}, x_1) \\
\vdots & \ddots & \vdots \\
g(k_{i1}, b_{iN}, x_N) & \ldots & g(k_{iL}, b_{iL}, x_N)
\end{bmatrix}_{N \times L}, \quad \beta = \begin{bmatrix}
\beta_1^T \\
\vdots \\
\beta_L^T
\end{bmatrix}_{L \times 1}, \quad \mathbf{T} = \begin{bmatrix}
t_1^T \\
\vdots \\
t_N^T
\end{bmatrix}_{N \times m}.
\]

\( \mathbf{H} \) is the output matrix of hidden layer, and \( \mathbf{H}(i, j) \) stands for the output of the \( i^{th} \) training data in the \( j^{th} \) hidden node.

The goal of adjustment is to find a set of optimal parameters \( w_j, b_j, \beta_j \) that make the \( \| (\mathbf{H} \beta)^T - \mathbf{T} \| \) minimum.

### 3.2. The establishment of fault classification model

The weights of input layer and thresholds of hidden layer might be zero which result into some hidden layers might be failed. Thus, the number of hidden layer nodes has to be increased to obtain high classification accuracy. However, it may lead to poor adaptability and low generalization capacity for testing data. The Gauss kernel function as Eq. (17) by means of inner product is employed to express the input weights and threshold of hidden layer of ELM to achieve fast and high accuracy convergence. In this way, the classification accuracy and generalization ability will be improved. The specific process is shown in Fig. 3.
\[ k(x,x_j) = \exp\left(-\frac{\|x-x_j\|^2}{\sigma^2}\right) \] (17)

**Figure 3** The flow chart of Kernel-ELM.

The hidden node output function \( g(x) \) of ELM algorithm is unknown and its kernel function can be used to express \( g(x) \) in the form of inner product. However, by using specific kernel functions \( K(x,x_i) \) as Eq. (18) we can easily find value of the output function and the hidden layer nodes can be adaptively determined. In the above process, the concrete form on hidden node output function \( g(x) \) didn't given directly.

\[ K_{ELM} = HH^T \cdot K_{ELM} = g(x_i) \cdot g(x_j) = K(x_i,x_j) \] (18)

To sum up, K-ELM can be summarized as below: Given a containing \( L \) samples of the training set \((x_i,y_i)\), \( x \) represents an input vector, \( y \) represents the corresponding output, \( i=1,2,\ldots,L \), and the kernel function \( K(x,x_i) \). And then the output equation as:

\[
 f(x) = g(x)H^T \left( \frac{1}{\epsilon} + HH^T \right)^{-1}T
 = \left[ K(x,x_1) \ldots K(x,x_L) \right] \left( \frac{1}{\epsilon} + \Omega_{\text{ELM}} \right)^{-1}T
\] (19)

4. **Rolling bearing fault classification based VMD, PE and K-ELM**

The actual bearing vibration signals are usually complex, which produced by multiple sources. Although PE can be used to measure the complexity of the signal, it can not identify the bearing failure under different working condition. The vibration signal is adaptive decomposed into a series of different time scales of PE component by VMD. Above all, the entropy of PE component can be used to indicate the intrinsic characteristics of the vibration signals in different time scales, which provided
accurate information for fault diagnosis. In addition, Gauss kernel function conveyed ELM is used in the construction of a rolling bearing fault classification model to improve the precision and generalization ability. Process of bearing fault diagnosis is shown in Fig. 4 based on the VMD, PE and K-ELM.

![Diagram](image)

Figure 4 The flow of bearing fault diagnosis based on the VMD, PE and K-ELM.

The specifics are:

1. Sample $k$ times respectively in accordance with given sampling frequency $f_s$ under normal, inner ring, outer ring and rolling elements of rolling bearing four different conditions, totaling gained $4k$ signals eventually. A total of $k$ signals, $m$ signals were randomly selected as training sample of each state, the others ($n$ signals) as testing sample sets.

2. Decompose above any signal by VMD, and choose four PE component ($PE_1,.., PE_4$) in front which contains the main failure.

3. Construct high dimensional feature vector $T$. First, according to Eq. (9) confirm time delay $\tau = 1$ and embedded dimension $D=4$. And then, calculate normalized PE component $PE_1,.., PE_4$.

   \[
   T = [PE_1, PE_2, \cdots PE_4]
   \]  

(20)

4. Repeat Steps (2) and (3), construct high dimensional feature vector sets, including $4k$ signals.

5. The high dimensional feature vector sets are used as the input of K-ELM algorithm to establish the kernel function extreme learning machine rolling bearing fault classification model. Meanwhile, four kinds of status corresponding fault category label is defined. If the output value is 1, it stands for normal status. Similarly, inner ring fault, outer ring fault and rolling body fault are identified as the value of 2, 3 and 4 respectively.

5. Results and discussions

5.1. Data acquisition and extraction of feature vectors

At last, the experimental results on the Case Western Reserve University bearing data set verify the feasibility and validity of the proposed method in this paper. Bearing designation of testing is deep groove ball bearings 6205-2RS. Usually, there are three types of fault: rolling body fault, inner ring
fault and outer ring fault. Combined with the normal status, four kinds of signal can be collected. For each operation status, 30 times of experiment are performed. Each experiment contains 5120 data points.

First, morphological average filter is used to de-noise the above signal. The linear structural element is selected, and each structural element value is 0, namely $g=\{0,0,0\}$. According to the determined structural elements, implement the four states signals’ noise reduction by morphological average filter. Then, VMD is used to decompose the state sample under different status. According to the rule given in Subsection 2.2.1, four MFs will be retained. The decomposition of one experiment under inner ring fault status is shown in Fig. 5, and part extraction of de-noising signal feature vectors based VMD and PE in four cases are shown in Table 1.

![Figure 5 The results of VMD for inner ring fault condition.](image)

| CONTION            | $PE_1$  | $PE_2$  | $PE_3$  | $PE_4$  |
|--------------------|---------|---------|---------|---------|
| Normal             | 0.2585  | 0.4668  | 0.4979  | 0.6823  |
|                    | 0.2715  | 0.4475  | 0.4953  | 0.6679  |
|                    | 0.2817  | 0.4655  | 0.4882  | 0.6855  |
|                    | 0.2752  | 0.4509  | 0.4970  | 0.6817  |
| Inner ring fault   | 0.3721  | 0.4583  | 0.489   | 0.6083  |
|                    | 0.3683  | 0.4558  | 0.4914  | 0.5556  |
|                    | 0.3063  | 0.4587  | 0.4936  | 0.5579  |
|                    | 0.3296  | 0.4686  | 0.4982  | 0.5564  |
| Outer ring fault   | 0.3156  | 0.4102  | 0.4756  | 0.5616  |
|                    | 0.3674  | 0.4089  | 0.4826  | 0.5616  |
|                    | 0.3267  | 0.4085  | 0.4959  | 0.5496  |
|                    | 0.3186  | 0.4083  | 0.4921  | 0.5518  |
| Rolling body fault | 0.2054  | 0.3333  | 0.4783  | 0.5703  |
|                    | 0.2135  | 0.3478  | 0.4758  | 0.6298  |
|                    | 0.2510  | 0.3855  | 0.4787  | 0.5888  |
|                    | 0.2210  | 0.3479  | 0.4836  | 0.6742  |

5.2. The improvement of classification model
Totally, thirty experiment data are conducted under each condition. Twenty of them are used as training data and the remaining ten are used as testing data. Using the value of PE as the input of the inputs of the K-ELM algorithm, the fault classification model is developed.

For the testing samples, fault classification accuracy of K-ELM model is 100%, which is a high accuracy and is shown in Fig. 6. The value of y-axis stands for the different operation status. If the value is 1, it stands for normal condition. Similarly, inner ring fault, outer ring fault and rolling body fault are identified as the value of 2, 3 and 4 respectively. To better illustrate the performance of the proposed method, SVM and ELM are employed for comparison. The fault classification model accuracy of SVM is 87.5% and ELM is 95%, which are shown in Figs. 7 and 8. The comparison results are shown in Table 2 and the proposed method has a higher classification accuracy.

Table 2 Comparisons of SVM, ELM and K-ELM.

| Algorithm | Accuracy (%) |
|-----------|--------------|
| SVM       | 100 90 60 100 |
| ELM       | 100 100 80 100 |
| K-ELM     | 100 100 100 100 |

Figure 6 Classification of testing data based on K-ELM.
6. Conclusions
To solve the existing problems of fault features extraction and fault classification method of rolling bearing, several crucial points are considered in this paper. First, because rolling bearing works under a complex environment, the collected vibration signal needs filter before using. To effectively remove noise, a morphological average filtering algorithm is proposed. Then VMD is performed on the filtered data to obtain high dimensional feature vectors. Then, these feature vector sets are used as
inputs for fault modelling. Finally, the fault classification model is developed based on enhanced extreme learning machine, which output function is expressed by the inner product of Gauss kernel function to adaptively determine the number of the hidden layer nodes. In comparison with fault classification methods based on support vector machine and extreme learning machine, the experimental results show that the proposed method has higher classification accuracy and better generalization ability.

References
[1] Yan Qin, Chunhui Zhao, Furong Gao. An Iterative Two-Step Sequential Phase Partition (ITSPP) Method for Batch Process Modeling and Online Monitoring. AIChE Journal, 2016, 62(7): 2358-2373
[2] Arnaz M, Robert X G. PCA-Based Feature Selection Scheme for Machine Defect Classification. IEEE Transactions on Instrumentation and Measurement, 2008, 53(6): 1517-1525
[3] Leiya Guo, Hezheng Jia, Ziyun Yang. Application of an intelligent classification method to mechanical fault diagnosis. Expert Systems with Applications, 2009, 36: 9941-9948.
[4] Cui Lingli, Kang Chenhui. The enhanced resonance demodulation method based on delay correlation and wavelet packet coefficient entropy threshold. Mechanical Engineering, 2010, 46(20): 53-57
[5] González-Grimaldo, Raymundo A, Cuevas-Tello, Juan C. Analysis of time series with artificial neural networks. 7th Mexican International Conference on Artificial Intelligence, MICAI 2008, 131-137
[6] Rubens N O. The Application of fuzzy Logic to the construction of the ranking function information retrieval systems. Computer Modelling and New Technologies, 2006, 10(1): 20-27
[7] Yang Jie, Ma Jun. A sparsity-based training algorithm for Least Squares SVM. 5th IEEE Symposium on Computational Intelligence and Data Mining, CISM 2014, 345-350
[8] Huang GB, Zhu QY, Siew CK. Extreme learning machine: theory and applications. Neurocomputing, 2006, 70(1): 489-501
[9] Mohammed A A, Minhas R, Jonathan Wu Q M, Sid-Ahmed M A. Human face recognition based on multidimensional PCA and extreme learning machine. Pattern Recognition, 2011, 44(10-11): 2588-2597
[10] Van Heeswijk M, Miche Y, Lindh-Knuutila T, Hilbers P A, Honkela T, Oja E, Lendasse A. Adaptive ensemble models of extreme teaming machines for time series prediction. 2009, Lectures Notes in Computer Science, 5769: 305-314
[11] Tian HX, Mao ZZ. An ensemble ELM based on ModiFed AdaBoost RT algorithm for predicting the temperature of molten steel in ladle furnace. IEEE Transactions on Automation Science and Engineering, 2010, 7(1): 73-80
[12] Yuan Q, Zhou W D, Li S F, Cai D M. Epileptic EEG classification based on extreme learning machine and nonlinear features. Epilepsy Research, 2011, 96(1-2): 29-38
[13] Akyildiz I F, Weilian S, Sankarasubramaniam Y, et al. Wireless sensor networks: A survey. Computer Networks, 2002, 38(4): 393-422
[14] Jing Wang, Guanghua Xu, Qing Zhang, et al. Application of improved morphological filter to the extraction of impulsive attenuation signals. Mechanical Systems and Signal Processing, 2009, 23(1): 236-245
[15] Konstantin Dragomiretskiy and Dominique Zosso. Variational Mode Decomposition. IEEE Transactions on Signal Processing. vol. 62, no.3. p.531-544. Feb.2014
[16] X. Zhang, Y. Liang, J. Zhou A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM Measurement, 69 (2015), pp. 164-179
[17] Jing Wang, Guanghua Xu, Qing Zhang, et al. Application of improved morphological filter to the extraction of impulsive attenuation signals. Mechanical Systems and Signal Processing, 2009, 23(1): 236-245
[18] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato and Jonathan Eckstein. Distributed
Optimization and Statistical Learning via the Alternating Direction Method of Multipliers.
Foundations and Trends in Machine Learning. vol. 3, no.1. p.1-122. 2010

[19] S.-D. Wu, P.-H. Wu, C.-W. Wu, J.-J. Ding, C.-C. Wang Bearing fault diagnosis based on multi scale permutation entropy and support vector machine Entropy, 14 (2012), pp. 1343-1356

[20] N. Nicolaou, J. Georgiou, Detection of epileptic electroencephalogram based on permutation entropy and support vector machines, Expert Syst. Appl.39 (2012) 202-209

Acknowledgments
This research is supported in part by the National Natural Science Foundation of China (No. 51565046), Innovation Fund of Inner Mongolia University of Science and Technology (No. 2015QDL12), and Natural science Fund of Inner Mongolia (No. 2016MS0543).