

2776. Regularized kernel function parameter of KPCA using WPSO-FDA for feature extraction and fault recognition of gearbox

Yan He¹, Zongyan Wang²

School of Mechanical Engineering, North University of China, Taiyuan, China

²Corresponding author

E-mail: ¹zbhy91@163.com, ²wangzongyan91@163.com

Received 12 June 2017; received in revised form 26 September 2017; accepted 15 November 2017

DOI <https://doi.org/10.21595/jve.2017.18755>



Abstract. Gearbox is subject to damage or malfunctions by complicated factors such as installation position and operation condition, meanwhile, accompanied by some nonlinear behaviors, which increase the difficulty of fault diagnosis and identification. Kernel principal component analysis (KPCA) is a commonly used method to realize nonlinear mapping via kernel function for feature extraction. However, choosing an appropriate kernel function and the proper setting of its parameter are decisive to obtain a high performance of the kernel methods. In this paper, we present a novel approach combining PSO and KPCA to enhance the fault classification performance. The standard particle swarm optimization (WPSO) was used to regularize kernel function parameter of KPCA instead of the empirical value. In particular, in view of the thought of Fisher Discriminate Analysis (FDA) in pattern recognition, the optimal mathematical model of kernel parameter was constructed, and its global optimal solution was searched by WPSO. The effectiveness of the method was proven using the Iris data set classification and gearbox faults classification. In the process, gearbox fault experiments were carried out, and the vibration signals in different conditions have been tested and processed, and the fault feature parameters were extracted. At last the analysis results of gearbox fault recognition was obtained by KPCA and compared with PCA. The results show that the separability of failure patterns in the feature space is improved after kernel parameter optimized by WPSO-FDA. The problems of single failure and compound fault recognition have been effectively solved by the optimized KPCA.

Keywords: particle swarm optimization (PSO), kernel principal component analysis (KPCA), kernel parameter, gearbox, nonlinear feature, fault classification.

1. Introduction

In the case of failure of mechanical equipment, there are many faint messages of failure and they are often accompanied by the occurrence of nonlinear behavior, so the fault features extraction and the diverse failure modes recognition become an issue. Signal feature extraction is a critical step to its condition monitoring and fault diagnosis. Principal component analysis (PCA) is a commonly used feature extraction method, but it is only suitable for solving the problem of linear relationship in the data. When the signal characteristics of mechanical equipment have large nonlinear relations, the method does not meet the requirements. Kernel principal component analysis (KPCA) presented by Schölkopf may carry out nonlinear mapping via kernel function [1]. It maps the input vector into a high-dimensional feature space to keep as much information as possible in terms of variance and find directions that have minimal reconstruction error. Fully using ideas of kernel, KPCA has more superiority to solve nonlinear problems.

Currently, KPCA has been widely applied into condition recognition and fault diagnosis, and has obtained good results. Bernal-de-Lázaro [2] used KPCA to improve the detection of small-magnitude faults. Zhang [3] put forward a novel ant-based clustering algorithm using the kernel method. Kuang [4] discussed a novel hybrid KPCA and SVM with GA model for intrusion detection. Deng [5] studied nonlinear process fault pattern recognition using statistics kernel PCA similarity factor. Bernal-de-Lázaro [6] used optimizing kernel methods to reduce dimensionality in fault diagnosis of industrial systems. However, for given data sets, the parameter of kernel

function itself influences the classification of KPCA. Consequently, it is very important to correctly set kernel function parameter (called kernel parameter). At present, it is mainly determined by experiment data or by crossing tests method, which is lack of scientific theoretical basis [7]. So, it is very significant to study kernel parameters optimization method by theory for improving performance of KPCA in feature extraction and fault recognition.

Particle swarm optimization (PSO) [8] is a novel global parallel optimization algorithm, which is inspired by the biological and sociological behavior of animal swarms searching for food. More recently, various PSO and improved PSO algorithms as a type of novel evolutionary computation techniques have been developed and applied to a wide range of optimization problems [9, 10], such as function optimization [11], automatic controlling [12], robotic learning, artificial life [13] and other fields [14, 15]. Compared with genetic algorithm (GA), the PSO algorithm has no complicated evolutionary operators such as crossover and mutation so that it can be computed efficiently and implemented conveniently. Hence, it has been widely used as a more effective and competitive searching method than GA.

This paper investigates how the fault feature extraction technique with KPCA improves the effectiveness of classification algorithms, then a thought of solving the optimization problem of kernel parameter is presented. At first, an optimization mathematical model of kernel parameter based on the thought of Fisher Discriminate Analysis (FDA) is established, and the standard particle swarm optimization algorithm (namely WPSO) is used to search its global optimal solution. Simulation analysis of the Iris data set will verify effectiveness of the method. Then, experiments on gearbox fault are carried out, and vibration signals of six significant measuring points are tested and processed to investigate gearbox fault type, bearing fault type and their combination fault, and frequency domain analysis and power spectral density analysis of their signals are processed. The feature parameters in time domain and frequency domain are collected. The optimization values of kernel parameter between different fault types will be found. Thirdly, KPCA after kernel parameter optimization is applied in gearbox fault classification to maximize the separability of failure patterns in the feature space. In particular, the effectiveness to distinguish the gearbox different fault types with KPCA is analyzed, and the results obtained by KPCA are compared with PCA. Finally, the conclusion and future works are drawn in the last section.

2. Fundamentals of KPCA and kernel parameter optimal model

2.1. Principle of KPCA

The idea of KPCA is to perform PCA in a high-dimensional feature space for acquiring more intrinsic information content of the original data [16], and it is described as follows.

Given a set of l training samples $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_l$ in R^N and a nonlinear mapping $\Phi: R^N \rightarrow F$, $\mathbf{x} \rightarrow \Phi(\mathbf{x})$, assuming $\sum_{k=1}^l \Phi(\mathbf{x}_k) = 0$, the following covariance matrix is given by:

$$\mathbf{C}^F = \frac{1}{l} \sum_{j=1}^l \Phi(\mathbf{x}_j) \Phi(\mathbf{x}_j)^T, \tag{1}$$

and it can be found that eigenvalues $\lambda \geq 0$ and eigenvectors $\mathbf{v} \in F \neq \{0\}$ satisfies $\lambda \mathbf{v} = \mathbf{C}^F \mathbf{v}$. Since all solutions \mathbf{v} with $\lambda \neq 0$ lie in the span of $\Phi(\mathbf{x}_1), \dots, \Phi(\mathbf{x}_l)$, there exist coefficients α_i , $i = 1, 2, \dots, l$, such that:

$$\mathbf{v} = \sum_{i=1}^l \alpha_i \Phi(\mathbf{x}_i). \tag{2}$$

Meanwhile, consider the equivalent equation:

$$\lambda(\Phi(\mathbf{x}_k) \cdot \mathbf{v}) = (\Phi(\mathbf{x}_k) \cdot \mathbf{C}^F \mathbf{v}), \quad k = 1, 2, \dots, l. \quad (3)$$

Substituting Eqs. (1)-(2) into Eq. (3):

$$\lambda \sum_{i=1}^l \alpha_i (\Phi(\mathbf{x}_k) \cdot \Phi(\mathbf{x}_i)) = \frac{1}{l} \sum_{i=1}^l \alpha_i \left(\Phi(\mathbf{x}_k) \cdot \sum_{j=1}^l \Phi(\mathbf{x}_j) \right) (\Phi(\mathbf{x}_j) \cdot \Phi(\mathbf{x}_i)), \quad (4)$$

$$k = 1, 2, \dots, l.$$

Define a $l \times l$ matrix \mathbf{K} by $\mathbf{K} = (\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)) = (\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j))$ with $\mathbf{K}(\cdot, \cdot)$ the Mercer kernel; the following compact form is:

$$l\mathbf{K}\boldsymbol{\alpha} = \mathbf{K}^2\boldsymbol{\alpha}, \quad (5)$$

where the squared matrix \mathbf{K}^2 denotes the product of \mathbf{K} with itself. In the following, the following dual equivalent form can be:

$$l\lambda\boldsymbol{\alpha} = \mathbf{K}\boldsymbol{\alpha}. \quad (6)$$

Let $\lambda_1^* \geq \lambda_2^* \geq \dots \geq \lambda_n^*$ denote the eigenvalues of the matrix \mathbf{K} , that is $\lambda_i^* = l\lambda_i$, $i = 1, 2, \dots, l$ where λ_i is the i th eigenvalue of the covariance matrix \mathbf{C}^F . Thus Eq. (6) takes the standard form:

$$\lambda^*\boldsymbol{\alpha} = \mathbf{K}\boldsymbol{\alpha}. \quad (7)$$

Eq. (7) is solved for nonzero eigenvalues λ^* and coefficient eigenvectors $\boldsymbol{\alpha}$ and the solutions $\boldsymbol{\alpha}$ are normalized by requiring that the eigenvectors \mathbf{v}^k of the covariance matrix \mathbf{C}^F are normalized to unit length, that is:

$$(\mathbf{v}^k \cdot \mathbf{v}^k) = 1, \quad k = 1, 2, \dots, p, \quad (8)$$

where it is assumed that the eigenvalues of \mathbf{K} are arranged in decreasing order, with λ_p being the corresponding smallest nonzero eigenvalue. Combining Eqs. (2) and (7), the normalization condition of Eq. (8) is equivalent to:

$$\lambda_k^* (\boldsymbol{\alpha}_k \cdot \boldsymbol{\alpha}_k) = 1, \quad k = 1, 2, \dots, p. \quad (9)$$

Finally, for the extraction of principal components, projections on eigenvectors \mathbf{v}^k : $k = 1, 2, \dots, p$ are computed in the feature space F to acquire the principal components of input data by KPCA. Consequently, let \mathbf{x} be a testing point with an image $\Phi(\mathbf{x})$ in F , then:

$$t(k) = (\mathbf{v}^k \cdot \Phi(\mathbf{x})) = \lambda \sum_{i=1}^l \alpha_i^k (\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x})) = \sum_{i=1}^N \alpha_i^k K(\mathbf{x}_i, \mathbf{x}), \quad k = 1, \dots, p, \quad (10)$$

$$i = 1, \dots, l,$$

where α_i^k is the i th coefficient of eigenvector $\boldsymbol{\alpha}_k$ associated with the k th eigenvalue of the matrix \mathbf{K} , and $t(k)$ is the k th extracted principal component.

2.2. Kernel parameter

KPCA is a kernel-based learning algorithm, it is crucial to choose an appropriate kernel because the geometric structure of the mapped samples is determined by the selected kernel and its parameters. By choosing different forms of kernel functions, KPCA can deal with many nonlinear problems [17, 18]. For kernel method, most of kernel functions researched currently mainly contain different features, such as polynomial kernel function, Radial Basis Function (RBF) kernel function and neural network kernel function. The choice of kernel function has an important influence on the results. Here RBF kernel function is taken as researched objective and it is defined by:

$$k(x, y) = \exp\left(-\frac{\|x - y\|^2}{2w^2}\right), \tag{11}$$

where the parameter w is called width, and indicates the degree of smoothness of the function. The width must be varied very carefully. If w is overestimated, the exponential tends to a linear behavior, and its projection in high-dimensional space loses its ability to separate non-linear data. On the contrary, when w is underestimated, the result is highly sensitive to noise in the information of training. So, the choice of the appropriate function width w is the result of a trade-off between the two, and a scientific method is required to guide through the selection process of the optimal value.

2.3. Establishment of the optimization model of kernel parameter by means of thought of FDA

The main advantage of KPCA stems from the mapping of input data into a high dimensional feature space, which may use linear algorithms despite of the nonlinearity in the original data. In pattern recognition, the idea of FDA is to find an optimal projection direction, along which the input data is projected into a straight line to distinguish better. That is to say, it can make intra class scatter as minimum as possible and the class space as much as possible. Therefore, an optimal method of kernel parameters based on the thought of FDA [19] can be proposed.

The mathematical model of kernel parameter based on Fisher norm is as follows.

Assume that two kinds of characteristic samples in a feature space are $X_1(x_{11}, x_{12}, \dots, x_{1i}), X_2(x_{21}, x_{22}, \dots, x_{2j}), i = 1, 2, 3, \dots, n_1, j = 1, 2, 3, \dots, n_2$.

The mean vectors of them are respectively as follows:

$$\mu_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} \Phi(x_{1i}), \tag{12}$$

$$\mu_2 = \frac{1}{n_2} \sum_{j=1}^{n_2} \Phi(x_{2j}). \tag{13}$$

The square of the distance between classes is:

$$D_b = \|\mu_1 - \mu_2\|^2 = (\mu_1 - \mu_2)^T(\mu_1 - \mu_2) = \frac{1}{n_1^2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} k(x_{1i}, x_{1j}) - \frac{2}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} k(x_{1i}, x_{2j}) + \frac{1}{n_2^2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} k(x_{2i}, x_{2j}). \tag{14}$$

The square of the inner-class discrete degree X_k is:

$$D_{w1} = \sum_{i=1}^{n_1} \|\Phi(x_{1i}) - \mu_1\|^2 = \sum_{i=1}^{n_1} \Phi(x_{1i})^T \Phi(x_{1i}) - n_1 \mu_1^T \mu_1 = \sum_{i=1}^{n_1} k(x_{1i}, x_{1i}) - \frac{1}{n_1} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} k(x_{1i}, x_{1j}), \tag{15}$$

$$D_{w2} = \sum_{i=1}^{n_1} \|\Phi(x_{2i}) - \mu_2\|^2 = \sum_{j=1}^{n_2} k(x_{2j}, x_{2j}) - \frac{1}{n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} k(x_{2i}, x_{2j}). \tag{16}$$

According to idea of FDA, the optimized objective function by PSO is set up [20]:

$$fitness = f(w) = \frac{D_{w1} + D_{w2}}{D_b}. \tag{17}$$

The minimal value of $f(w)$ can be solved to realize that the between-class scatter is the maximal whereas the within-class scatter is minimal. The parameter w^* is set as the minimum point of Fisher discriminate function. Many experiments in references proved that for entirely non-linearly separable problem, the minimum point w^* exists, while for linearly separable or almost linearly separable problem, the value of $f(w)$ is sharply decreased and then becomes steady with w changing from the large to the little. At the same time the parameter w can be w^* , and $f(w)$ can change to be steady in the end. For the multiclass problems, they can be transformed into two-class problems.

3. Particle swarm optimization algorithm

3.1. Principle of standard PSO

In the PSO algorithm, the process of searching for the best solution is imitating the movement and flocking of birds. Each particle stands for a potential solution which consists of a position and a velocity. At each iteration, the particle adjusts the velocity and position based on itself optimal location and the global optimal location in population. The equations WPSO are written as follows [8, 9]:

$$v_{id}(t + 1) = \omega v_{id}(t) + c_1 r_1 (p_{id} - x_{id}(t)) + c_2 r_2 (p_{gd} - x_{id}(t)), \tag{18}$$

$$x_{id}(t + 1) = x_{id}(t) + v_{id}(t + 1), \quad i = 1, 2, \dots, n, \tag{19}$$

where d is the dimension of search space, n is the total number of particles, $\mathbf{x}_i = (x_{i1}, \dots, x_{id})$ and $\mathbf{v}_i = (v_{i1}, \dots, v_{id})$ denote the position and velocity of the particle, respectively; p_{id} is current optimal position of the particle i , and p_{gd} is whole population's optimal position; c_1 and c_2 are acceleration constants with positive values, and usually set as 0-2; r_1 and r_2 are random numbers within the range [0 1]; v_{id} is current flying velocity of the particle i within the range $[-v_{max}, v_{max}]$, where v_{max} is maximum particle velocity, being a non-negative number; ω is called inertia weight, which decreases linearly with the iteration:

$$\omega = \omega_{max} - \frac{\omega_{max} - \omega_{min}}{T} t, \tag{20}$$

where ω_{max} is the maximum and ω_{min} is the minimum of ω , and typically, $\omega_{max} = 0.9-1.4$, $\omega_{min} = 0.4$, t is current iteration number, and T is the maximum number of iterations.

3.2. The steps optimized by WPSO-FDA for best solution

According to Eqs. (17) (18) and (19), the optimized steps and processes of the kernel parameter are as follows:

- 1) Calculate the square of the distance between the two classes D_b and the square of intra-class scatter D_{w1} , D_{w2} by Eqs. (14), (15) and (16);
- 2) Construct Fisher discriminate function $f(w)$ by Eq. (17), and regard it as a fitness for WPSO;
- 3) Give initial parameters such as the value range of the parameter $\omega(\omega_{min}, \omega_{max})$, population size m of WPSO, maximum iterative step T , constants c_1 and c_2 , and maximum velocity limit v_{max} ;
- 4) Randomly create initial population w , then calculate the value of individual fitness f_p of each particle and the global fitness value of the whole population f_g ;
- 5) Carry out the particle's velocity $v_{id}(t)$ and location $x_{id}(t)$ updating according to Eqs. (18) and (19);
- 6) Judge whether the termination condition of evolution is satisfied. If $t < T_{max}$, steps (4)-(6) will be repeated until termination condition is satisfied. Or else, the current solution will be output as the optimal solution w^* , and the evolutionary computation is ended.

3.3. Simulation analysis of Iris data set by KPCA optimized by WPSO-FDA

The Iris data set is a typical test case for many statistical classification techniques, and it is selected in this paper as the test data for simulation analysis. In the data set, each group of data contains four properties of an Iris flower, i.e. sepal length, sepal width, petals length and petals width. Among three different types of flowers, the first and the other two types of flowers are linearly separable, but the latter two types of flower are nonlinear [20, 21]. 50 sets of data selected from each flower were used to simulation analysis. So, 150 groups of data from the three types of flowers can be divided into training set and test set, each containing 75 groups of data.

In the classification process of the Iris data by KPCA, it firstly maps the input vector into a high-dimensional feature space. Then, the mapping data in the feature space are analyzed, and their projections on the nonlinear principal component are obtained. This method firstly computes a vector in feature space called the first kernel principal component (KPC1) used to describe the maximum change direction of the data, and the second kernel principal component (KPC2) is orthogonal to the first principal component and describes the maximum of residual change direction. KPCA projections for Iris data with different kernel parameter width w are shown in Fig. 1(a), (b), (c) and (d). When w is too small, as shown in Fig. 1(a) and (b), where w is set to 0.387 and 0.707, respectively, the projections appear "and" phenomenon, namely they have mixed together and cannot be separated into samples 2 and 3. In Fig. 1(c), when w is 1.41, the latter two classes aliasing are less than that in Fig. 1(a) and (b), but it is still difficult to distinguish between them. In Fig. 1(d), when w is 2.499, the output response range of the sample widens for the group, but there are some samples belonged to the 2nd type and the 3rd crossing together. So, it is necessary to optimize the regular parameter w by WPSO-FDA.

According to the steps optimized by WPSO-FDA described in section 3.2, the kernel parameter w has been searched, and its evolution curve is displayed in Fig. 2(a). The width w achieves the optimized value w^* as being 1.064 after iteration steps and the fitness function $f(w)$ becomes steady. Iris data set classification by KPCA with w^* is displayed in Fig. 2(b). It can be seen that the three types flowers projections are in a different location, which is related to the three different categories.

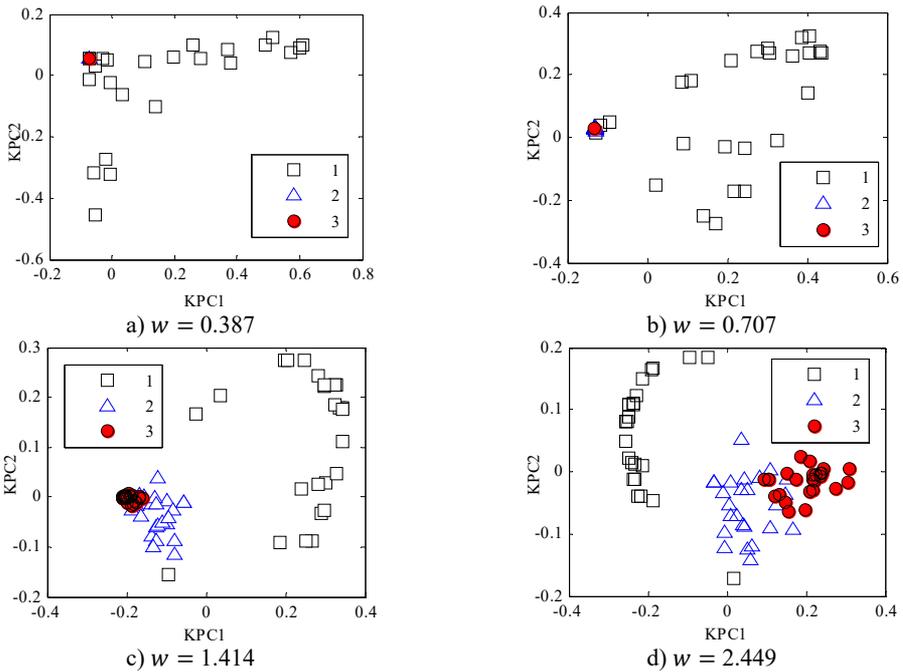


Fig. 1. KPCA projection diagram for Iris data with different width w

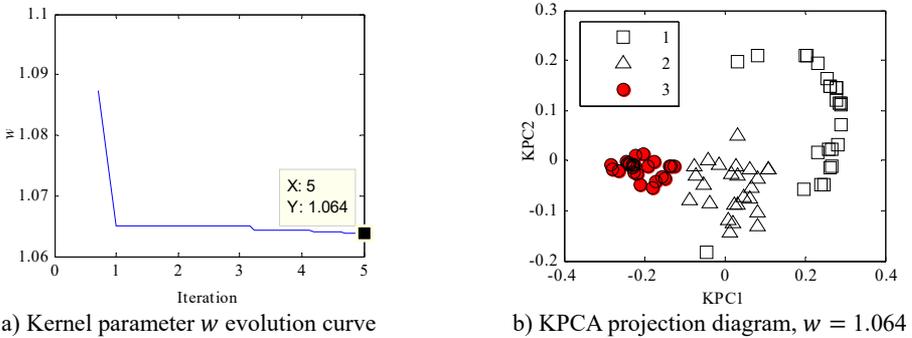


Fig. 2. KPCA analysis results for Iris data

4. The gearbox fault diagnosis experiment and features extraction

Gearbox is a part of common mechanical equipment. Due to the poor working environments, it is subject to damage or malfunctions. It is complicated that gearbox feature extraction and fault recognition because the internal components such as gears, bearings, and shafts are influenced by complicated factors such as installation position and operating condition, among which the relationship between the fault and symptom is a nonlinear mapping relation. Gearbox fault experiments were carried out, and the vibration signals in different conditions have been tested and processed. KPCA was used to extract nonlinear characteristics of gearbox vibration signal and to identify failure pattern [22, 23].

4.1. Fault simulation experiment of gearbox

The gearbox faults simulation experiment was carried out in the vibration lab of North University of China. The gearbox (JZQ-250 type) shown in Fig. 3 is composed of two pairs of gears, three shafts (input shaft, intermediate shaft and output shaft), three pairs of bearings and

housing. Motor and magnetic powder brake are coupled with input shaft and output shaft, respectively. The experimental setup was composed of 36 channels dynamic signal data acquisition analysis system from Beijing Oriental Vibration Noise Institute, three-way accelerometer (piezoelectric), laptop computers, as shown in Fig. 4.

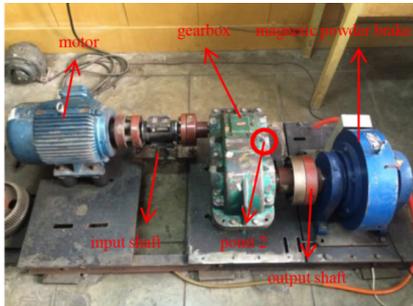


Fig. 3. The gearbox fault diagnosis experimental system



Fig. 4. The gearbox fault diagnosis test system

4.2. Faults of gearbox set up and its characteristic parameters

In the gearbox fault diagnosis experiment, the following simulation failures were set up besides normal working condition of gearbox: bearing outer ring crack, bearing cage fracture, gear broken tooth and gear broken tooth combined with bearing outer ring fracture. The three fault types are shown in Fig. 5.



a) Gear broken tooth



b) Bearing cage fracture



c) Bearing outer ring crack

Fig. 5. The fault tested in the experiment

4.3. Vibration analysis of gearbox in different conditions

In this experiment, the input shaft speed was set to 1200 r/min, and the sampling frequency was 5120 Hz. Vibration signals of the normal working condition of the gearbox (A mode), bearing outer ring crack in intermediate shaft (B mode), bearing cage fracture (C mode), gear broken tooth (D mode), and gear broken tooth combined with bearing outer ring fracture (E mode) were acquired. The bearing characteristic frequencies of the five working states above are listed in Table 1 and include the rotating frequency of bearings mounted on three shafts, and the bearing theoretical failure frequency in every shaft are listed Table 2. The vibration signals under five conditions have been collected and processed. Their frequency spectrum curves are shown in Fig. 6, and their power spectrum density (PSD) curves are shown in Fig. 7.

Seen from Fig. 6(a)-(e), there are no big differences among the five frequency spectrum curves of vibration signal, and the amplitudes in characteristic frequencies are large, namely their vibrations are strong at this frequency. The frequency spectrum curves of B and C mode are very similar, so they cannot be used to differentiate between fault modes. In fact, the vibration signal on each measured point is not from a single component, which is the sum of different vibration signals from shafts, bearing and meshing gears in the gearbox. Therefore, it can be concluded that

the curves only show the intrinsic characteristics of gearbox itself. From Fig. 7(a)-(e), the PSD curves in the case of failure are consistent with the frequency spectrum curves of the gearbox. Since the fault modes of gearbox cannot be determined to judge the gearbox fault status accurately only through the frequency curves, the further extraction of feature parameters in the time and frequency domain of vibration signals is required.

Table 1. The bearing characteristic frequencies

Installation location	Bearing type	Outer ring	Inner ring	Cage	Rolling element
Input shaft	6406	40.95	79.05	6.83	28.32
Intermediate shaft	6406	17.81	34.37	2.97	12.31
Output shaft	6312	6.0	9.46	0.74	3.91

Table 2. The bearing theoretical failure frequency in every axis

Frequency	One times / Hz	Two times / Hz	Three times / Hz
Input shaft	20	40	60
Intermediate shaft	8.69	17.38	26.09
Output shaft	1.93	3.86	5.79
Gear _(1,2) meshing	600	1200	1800
Gear _(3,4) meshing	156.25	312.5	468.7

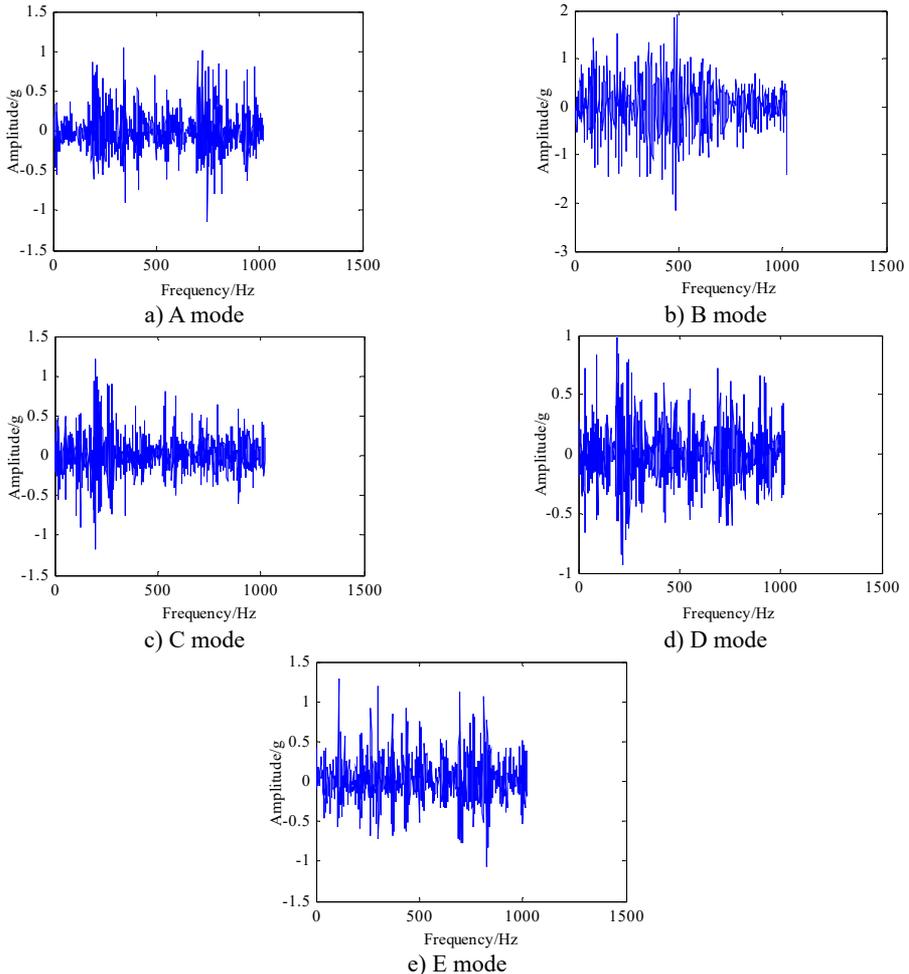


Fig. 6. Frequency spectrum curves of vibration signals

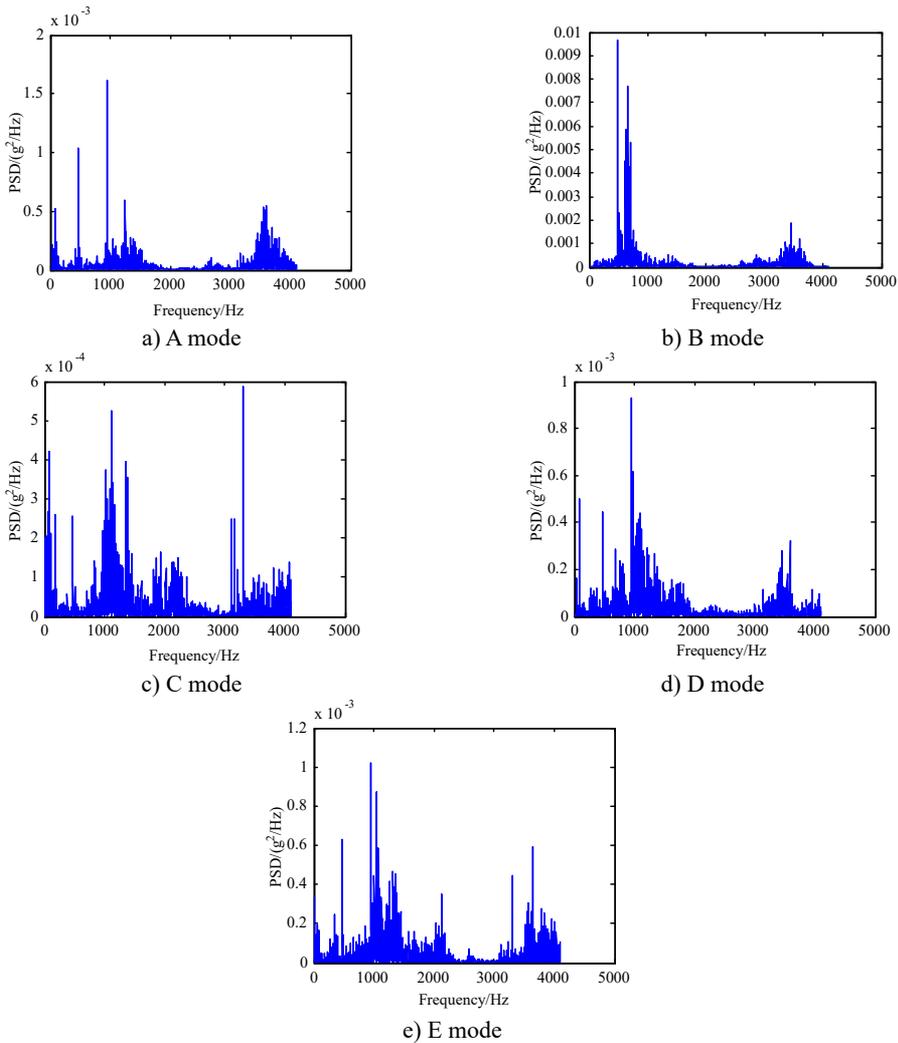


Fig. 7. Power spectrum density curves of vibration signal

4.4. Construction of gearbox feature parameter set

The collected signals were processed, and in particular, the data of the measuring point 2 shown in Fig. 3 were analyzed. The original feature parameters set extracted from time domain and frequency domain are summarized in Table 3. The time-domain features include dimensional and dimensionless parameters, such as the mean value, skewness, kurtosis, waveform index and peak index. The frequency-domain features include only dimensional parameters such as frequency domain variance and correlation factor. 60 groups of feature samples were extracted under five working conditions, and half of them are training samples and the others are test samples. They were applied to KPCA for fault identification and classification of gearbox [24, 25].

5. Kernel parameters optimization and faults classification of gearbox

5.1. Kernel parameters optimization based on WPSO-FDA

When KPCA is applied, the kernel parameter w of the Radial Basis Function is a key to

maintain good performance of nuclear study. Excessively large width w will bring about the result that the influence spheres of a sample may be so large that irrelevant training samples will interfere the new test samples to make the right judgment; while too small value will result in the consequence that nuclear learning with only memory is unable to make a judgment to new sample [7]. Therefore, for training samples, kernel parameters optimization based on WPSO-FDA was conducted, and the parameter settings of WPSO are summarized in Table 4. The process of fitness function optimization with iteration under different conditions of gearbox is shown in Fig. 8. The width values w^* of RBF between different fault states were obtained by WPSO-FDA optimization, and the optimization results are shown in Table 5. From Fig. 8 the width w can be stabilized at an optimum value within the 50 iterations of fitness function evaluation. For example, for bearing outer ring crack and gear broken tooth, optimal parameter values w^* can be stabilized at a width of 2.660. The optimal values are used in the parameters setting of KPCA for gearbox fault classification.

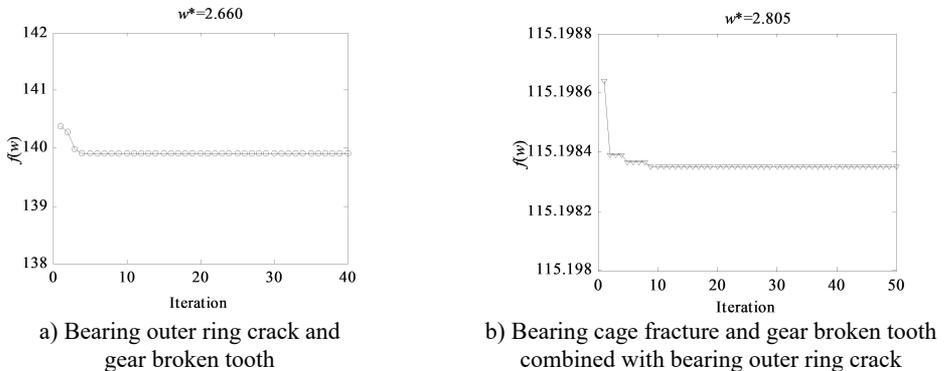


Fig. 8. Fitness function evolution curves based on WPSO

Table 3. Feature parameters in time-domain and frequency-domain

Domain category	Dimensional parameter	Dimensionless parameter
Time-domain	Mean value, root mean square, variance, kurtosis, skewness, mean square value, root square amplitude, absolute average amplitude	Waveform index, peak index, margin index, kurtosis index, skewness index, bias-normal distribution index, eight-order moments coefficient, sixteen-order moment coefficient
Frequency-domain	Correlation factor, mean-square spectrum, harmonic factor, power and gravity centre index, frequency domain variance, origin moment of spectrum	-

Table 4. Parameters set in WPSO

WPSO Parameters	m	ω_{max}	ω_{min}	c_1	c_2	T	v_{max}
	20	1.2	0.4	2	2	60	1

Table 5. The optimization results of Kernel parameter in five models samples

Fault model	w^*
Bearing outer ring crack and gear broken tooth (B and D)	2.660
Bearing outer ring crack and bearing cage fracture (B and C)	0.556
Bearing cage fracture and gear broken tooth combined with bearing outer ring crack (C and E)	2.805
Bearing outer ring crack and gear broken tooth combined with bearing outer ring fracture (D and E)	1.078

5.2. Fault classification of KPCA based on WPSO-FDA

First, 30 groups of original feature sets drawn from normal condition were trained, then tests

were carried through 30 groups of data from four fault modes such as bearing outer ring crack, bearing cage fracture, gear broken tooth and gear broken tooth combined with bearing outer ring fracture. Finally, for the same fault modes, the fault recognition results obtained by KPCA and PCA were analyzed, as shown in Fig. 9 and Fig. 10, respectively.

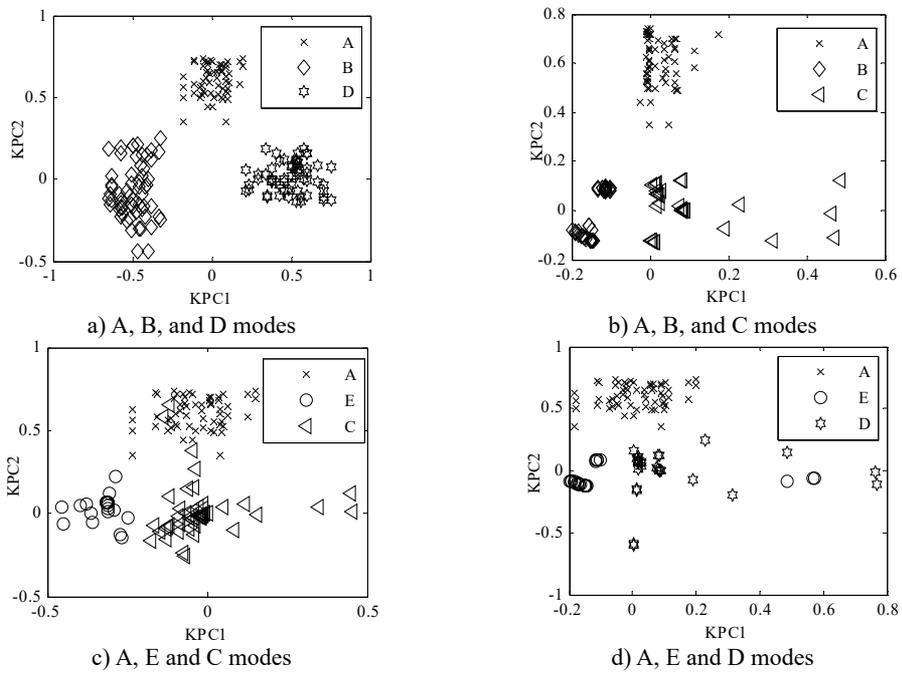


Fig. 9. KPCA projection in different conditions

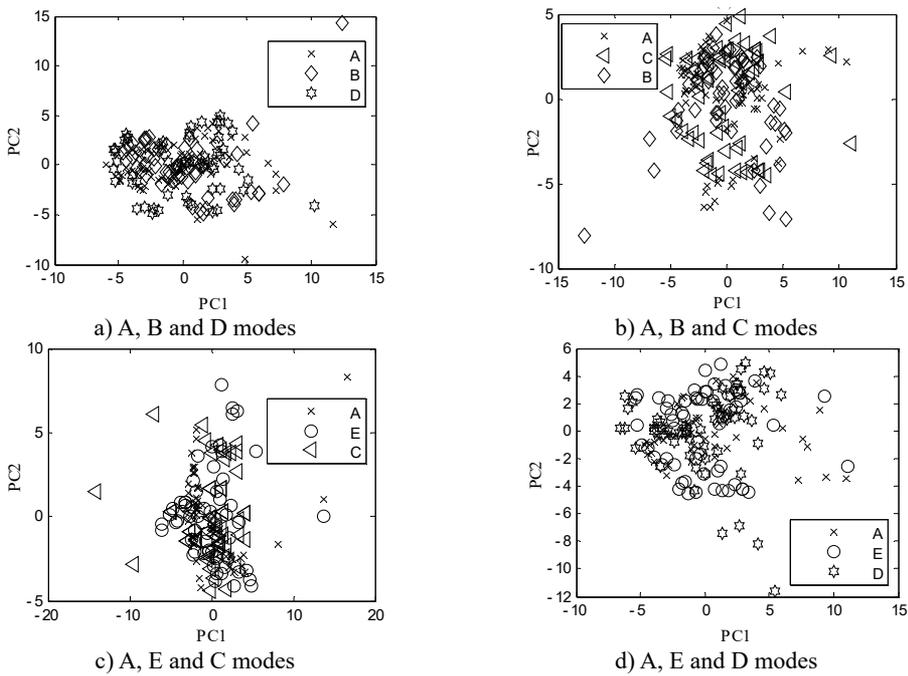


Fig. 10. PCA projection in different conditions

The KPCA projections from the first kernel principal component (KPC1) to the second kernel principal component (KPC2) in feature space are displayed in Fig. 9, and the PCA projections in original space are shown in Fig. 10. Observed from the Fig. 9, KPCA projections can clearly distinguish the various fault modes. The principal component projections from five modes conditions show obviously self-aggregation behaviours. Especially, the normal condition (A mode) can focus on a centre, while the bearing cage fracture (C mode) are relatively dispersed, but far from the normal cluster centre, apparently belong to different classes. From Fig. 9(c), there are a few crossing points between D mode and E mode, as a whole, the intra-class space of KPCA is little, whereas the space of different classes of KPCA is much. From Fig. 10, the PCA projections are mixed together, and normal condition cannot be clearly distinguished with the fault modes, and several faults cannot be separated from each other due to their nonlinearity. Therefore, the KPCA performance optimized by WPSO-FDA has obviously improved in the non-linear analysis, and accurately identifies not only single fault modes of gearbox but also the compound fault.

6. Conclusions

KPCA is usually used to feature extraction for fault recognition and classification because it extracts the principal components by adopting a nonlinear kernel method and remain good divisibility. But How to reasonably set the kernel function parameter is the key to obtain good performance. This paper presented a systematic method to optimize a RBF kernel parameter of KPCA through WPSO and used example analysis of gearbox fault diagnosis to demonstrate its effectiveness. It is relying on a theoretical model of kernel parameters established by means of thought of FDA, and the WPSO is applied to optimize it for reducing the kernel parameter blindness in a set. The classification results for Iris data set show that the method improved the KPCA classification performance. At last, the KPCA is used multi-fault classification in gearbox. The results show that fault classifications are effective by KPCA after the kernel parameter is optimized by WPSO-FDA, that is to say, within-class scatter of feature samples in every condition realizes as minimum as possible and between-class scatter of feature samples in different condition types as much as possible. It is concluded that non-linear mapping performance of KPCA after its kernel parameter regularized by WPSO-FDA is greatly improved in nonlinear feature extraction and the proposed method fits for single mechanical fault classification and compound fault classification. However, how to effectively address a large number of diversity compound faults diagnosis of mechanical equipment in real applications still requires further work.

Acknowledgement

This work was supported by The Key Research and Development Program of Shanxi Province under Grant No. 2016-33.

References

- [1] Twining C. J., Taylor C. J. The use of kernel principal component analysis to model data distributions. *Pattern Recognition*, Vol. 36, Issue 1, 2003, p. 217-227.
- [2] Bernal De Lázaro J.-M., Llanes Santiago O., Prieto Moreno A., et al. Enhanced dynamic approach to improve the detection of small-magnitude faults. *Chemical Engineering Science*, Vol. 146, 2016, p. 166-179.
- [3] Zhang L., Cao Q. A novel ant-based clustering algorithm using the kernel method. *Information Sciences*, Vol. 181, Issue 20, 2011, p. 4658-4672.
- [4] Kuang F., Xu W., Zhang S. A novel hybrid KPCA and SVM with GA model for intrusion detection. *Applied Soft Computing Journal*, Vol. 18, 2014, p. 178-184.
- [5] Deng X., Tian X. Nonlinear process fault pattern recognition using statistics kernel PCA similarity factor. *Neurocomputing*, Vol. 121, 2013, p. 298-308.

- [6] **Bernal De Lázaro J. M., Prieto Moreno A., Llanes Santiago O., et al.** Optimizing kernel methods to reduce dimensionality in fault diagnosis of industrial systems. *Computers and Industrial Engineering*, Vol. 87, 2015, p. 140-149.
- [7] **Wang X., Qiu J., Liu G.** Kernel function optimization in kernel principle component analysis and its application to feature extraction of gear faults. *Journal of Vibration, Measurement and Diagnosis*, Vol. 27, Issue 1, 2007, p. 62-64, (in Chinese).
- [8] **Kennedy J., Eberhart R.** Particle swarm optimization. *IEEE International Conference on Neural Networks*, Vol. 4, 1995, p. 1942-1948.
- [9] **Shi Y., Eberhart R.** A modified particle swarm optimizer. *Applied Mathematics and Computation*, Vol. 189, Issue 5, 1998, p. 69-73.
- [10] **Gou J., Lei Y.-X., Guo W.-P., et al.** A novel improved particle swarm optimization algorithm based on individual difference evolution. *Applied Soft Computing*, Vol. 57, 2017, p. 468-481.
- [11] **Delgarm N., Sajadi B., Kowsary F., et al.** Multi-objective optimization of the building energy performance: A simulation-based approach by means of particle swarm optimization (PSO). *Applied Energy*, Vol. 170, 2016, p. 293-303.
- [12] **Eswaran T., Kumar V. S.** Particle swarm optimization (PSO)-based tuning technique for PI controller for management of a distributed static synchronous compensator (DSTATCOM) for improved dynamic response and power quality. *Journal of Applied Research and Technology*, Vol. 15, Issue 2, 2017, p. 173-189.
- [13] **Rastegar S., Araújo R., Mendes J.** Online identification of Takagi-Sugeno fuzzy models based on self-adaptive hierarchical particle swarm optimization algorithm. *Applied Mathematical Modelling*, Vol. 45, 2017, p. 606-620.
- [14] **Beigyand S. D., Abdi H., La Scala M.** Hybrid gravitational search algorithm-particle swarm optimization with time varying acceleration coefficients for large scale CHPED problem. *Energy*, Vol. 126, 2017, p. 841-853.
- [15] **Ansari Z., Almasganj F.** Implementing PCA-based speaker adaptation methods in a Persian ASR system. *5th International Symposium on Telecommunications*, Vol. 32, 2010, p. 769-774.
- [16] **Schölkopf B., Smola A., Müller K.-R.** Nonlinear component analysis as a Kernel eigenvalue problem. *Neural Computation*, Vol. 10, Issue 5, 1998, p. 1299-1319.
- [17] **Zhang Y., Zuo H., Bai F.** Classification of fault location and performance degradation of a roller bearing. *Journal of the International Measurement Confederation*, Vol. 46, Issue 3, 2013, p. 1178-1189.
- [18] **Shao R., Hu W., Wang Y., et al.** The fault feature extraction and classification of gear using principal component analysis and kernel principal component analysis based on the wavelet packet transform. *Measurement*, Vol. 54, 2014, p. 118-132.
- [19] **Pan H., Wei X., Huang J.** Fault feature extraction based on KPCA optimized by PSO algorithm. *IEEE International Conference on Industrial Informatics*, 2010, p. 102-107.
- [20] **Wei X., Pan H., Wang F.** Feature extraction based on kernel principal component analysis optimized by particle swarm optimization algorithm. *Journal of Vibration, Measurement and Diagnosis*, Vol. 29, Issue 9, 2009, p. 162-166, (in Chinese).
- [21] **Xiao Y., Feng L.** A novel neural-network approach of analog fault diagnosis based on kernel discriminant analysis and particle swarm optimization. *Applied Soft Computing*, Vol. 12, Issue 2, 2012, p. 904-920.
- [22] **Shi H., Liu J., Zhang Y.** An optimized kernel principal component analysis algorithm for fault detection. *Compute*, Vol. 42, Issue 6, 2009, p. 846-851.
- [23] **Zhu Z. B., Song Z. H.** A novel fault diagnosis system using pattern classification on kernel FDA subspace. *Expert Systems with Applications*, Vol. 38, Issue 6, 2011, p. 6895-6905.
- [24] **Reddy K. N., Ravi V.** Differential evolution trained kernel principal component WNN and kernel binary quantile regression: application to banking. *Knowledge-Based Systems*, Vol. 39, 2013, p. 45-56.
- [25] **Cheng G., Chen X., Li H., et al.** Study on planetary gear fault diagnosis based on entropy feature fusion of ensemble empirical mode decomposition. *Measurement*, Vol. 91, 2016, p. 140-154.



Yan He is a Ph.D. candidate in mechanical design and theory at North University of China. Her current research interests include fault diagnosis, and digital design and manufacture.



Zongyan Wang received Ph.D. in aerospace engineering from Northwestern Polytechnical University, Xi An, China, in 1998. Now he works at North University of China. His current research interests include digital design and manufacture, computer integrated manufacturing system.