Gear Fault Classification Using Genetic Programming and Support Vector Machines

Jianping Xuan[1], Hanhong Jiang[2], Tielin Shi[1], and Guanglan Liao[1]

[1] School of Mechanical Science and Engineering
Huazhong University of Science and Technology
Wuhan,Hubei 430074,China
{jpxuan, tlshi,g.l.liao}@mail.hust.edu.cn

[2] Department of Electrical Engineering
Naval University of Engineering
Wuhan,Hubei 430033,China
E-mail:jhh2004@263.net

Abstract

Gears are critical elements in complex machinery. An approach is proposed to solve classification of multiple gear conditions. These conditions include the normal, slightly spalling, moderate spalling, severe spalling, one worn tooth, and two worn teeth conditions. The power spectral density (PSD) of the vibration signals of gearbox casing is introduced to construct some original features. The PSD is estimated by the periodogram from which the features are evaluated. To get useful features from the original ones, which are high dimensional, genetic programming (GP) is used to select them and reduce the feature dimension. The new features provide more sensitive information for a classifier. The classifier is based on support vector machines (SVMs), which have the ability of multi-class classification and good generalization with the one against all method. The approach presents its effectiveness by experiment. The results show that the original features extracted from estimation of the PSD of the raw data have certain abilities of classification, and that the SVMs can classify correctly the 6 conditions above and give the classification rates with high accuracy.
Keyword: Gears, Periodogram, Genetic programming, Support vector machines

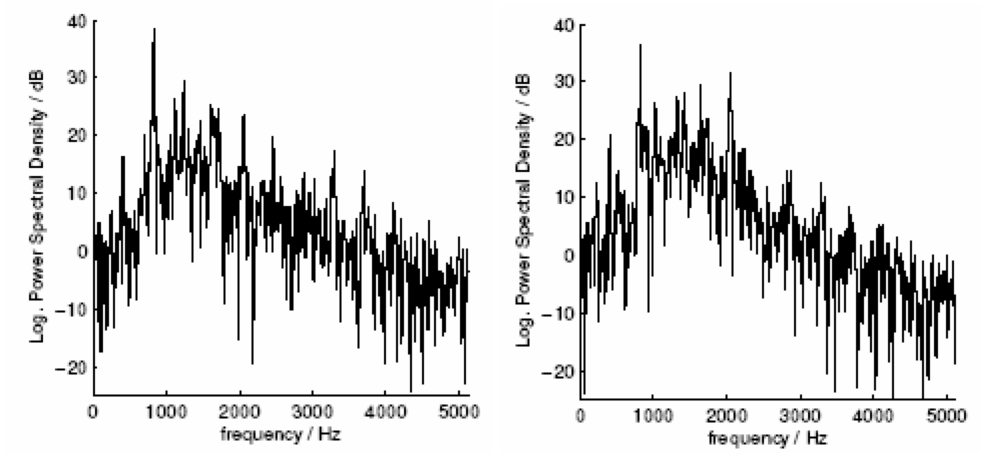
I. Introduction

The omnipresence of gears in rotating machinery has made the study of vibration a more interesting subject. One study found that 60% of gear box damage is due to faults in the gears [1]. In engineering, they show a considerable number of different forms of damage. The common types of gear damage mainly consist of pitting, scuffing, spalling, cracking and wear [2]. One of the major reasons for gear faults is excessive vibration. Vibration can be thought of as a ratio of the forces acting on the gear to its dynamic stiffness. The backlash, the error of the gear transmission, the unbalanced inertia mass, the time varying mesh stiffness of tooth, the friction between tooth faces and the time varying support stiffness of geared system, change the ratio, i.e., vibration which characteristics can reflect symptoms of a lot of faults or defects. The benefit of using vibration analysis for their monitoring and diagnosis has been demonstrated to be successful since the early time because of the ease of measurement. The approaches of gear vibration analysis are mainly subdivided into three categories according to analysis domains. They are time domain, frequency domain and time-frequency domain. The time domain methods include time synchronous average and statistical analysis. The former is a signal averaging process over a large number of cycles, synchronous with the running speed of a specific shaft in the gearbox. It can remove not only the background noise but also periodic events that are not exactly synchronous with the gear being monitored [3,4]. The latter consists of many descriptive statistics such as sample skewness, kurtosis and so on. The frequency domain methods include spectral analysis based ones such as power spectral density (PSD) and Cepstrum analysis [5], and higher-order statistics and spectra [9]. The time-frequency domain methods are composed of the short-time Fourier transform, Wigner-Ville, Cohen distribution and wavelet analysis [6-8]. Feature extraction methods play an important role in machine condition monitoring and fault diagnosis, from which the diagnostic information can be obtained. Through gear vibration analysis, a lot of features are acquired, and the next step is optimization and classification. The task of classification is to find a rule, which, based on external observations, assigns an object to one of several classes. A lot of techniques have been proposed to carry out these tasks. They are artificial neural networks [10], genetic programming [11], and support vector machine [14]. GP evolves a population of computer programs, which are possible solutions to a given optimization problem, using the Darwinian principle of survival of the fittest. It uses biologically inspired operations like reproduction, crossover and mutation. Each program or individual in the population is generally represented as a tree composed of functions and data/terminals appropriate to the problem domain. Chen proposes a fault diagnosis method for rolling bearings in an unsteady operating condition using instantaneous power spectrum and genetic programming [12]. Zhang [13] uses GP to detect faults in rotating machinery with rolling bearings. As a powerful machine learning approach for classification problems, support vector machine is known to have good generalization ability. SVMs can trace their roots back to statistical learning theory, as introduced by Vapnik in the late 1960s. It was not until the early 1990s that the techniques used for SVMs began to emerge, and become practical, with the increased computing power available [14]. In this paper, we introduce PSD of the vibration signals of gearbox casing to construct the original spectral features. The specific characteristics of the vibration spectrum that are associated with common gear damage conditions are well known, for example, uniform wear tends to show up at the tooth-meshing frequency and its harmonics. The typical use of these features would be to determine when a gear should be taken out of operation in the presence of deteriorating fault conditions. Reliable diagnostics of deteriorating conditions may however be more problematic in the presence of simultaneous multiple fault conditions. To address these problems, genetic programming is used to reduce the feature dimension from the originals. The new features provide more sensitive information for a classifier. The classifier is based on SVMs.

II. Construction and Calculation of the Original Spectral Features

A. ***Power Spectral Density***

The PSD of stationary random process is the Fourier transform of its correlation sequence. The PSD S(f) is estimated by the periodogram in this paper. $S(f)=\frac{1}{N}|F(f)|^2$ where $F(f)$ is the Fourier transform of the vibratory signal in various conditions and N is the length of samples .These conditions include normal, one slightly spalling tooth, one moderate spalling tooth, one severe spalling tooth, one worn tooth, and two worn teeth conditions denoted by normln, spal1n, spal2n, spal3n, wear1n, wear2n for short, respectively. Fig.1 shows the periodograms of two signals of gearbox vibration. Fig.1(a) is the periodogram of one moderate spalling tooth condition signal; Fig.1(b) is the periodogram of one worn tooth condition signal.



(a) One moderate spalling tooth fault (b) One worn tooth fault

Fig. 1. The periodogram of signals of two fault conditions

B. *Original Spectral features*

Seven feature parameters are usually used for construction of features in the frequency domain [12], which are shown as follows:

$$X_1 = \sqrt{\frac{\sum_{i=1}^{K} f_i^2 \cdot S(f_i)}{\sum_{i=1}^{K} S(f_i)}} \quad (1)$$

$$X_2 = \sqrt{\frac{\sum_{i=1}^{K} f_i^4 \cdot S(f_i)}{\sum_{i=1}^{K} f_i^2 \cdot S(f_i)}} \quad (2)$$

$$X_3 = \frac{\sum_{i=1}^{K} f_i^2 \cdot S(f_i)}{\sqrt{\sum_{i=1}^{K} S(f_i) \sum_{i=1}^{K} f_i^4 \cdot S(f_i)}} \quad (3)$$

$$X_4 = \frac{\sigma}{\bar{f}} \quad (4)$$

$$X_5 = \frac{\sum_{i=1}^{K} ((f_i - \bar{f})^3 \cdot S(f_i))}{\sigma^3 \cdot K} \quad (5)$$

$$X_6 = \frac{\sum_{i=1}^{K} ((f_i - \bar{f})^4 \cdot S(f_i))}{\sigma^4 \cdot K} \quad (6)$$

$$X_7 = \frac{\sum_{i=1}^{K} \sqrt{(f_i - \bar{f})} \cdot S(f_i))}{\sqrt{\sigma} \cdot K} \quad (7)$$

where

$$\bar{f} = \frac{\sum_{i=1}^{K} f_i \cdot S(f_i)}{\sum_{i=1}^{K} S(f_i)} \quad (8)$$

and

$$\sigma = \sqrt{\frac{\sum_{i=1}^{K} (f_i - \bar{f})^2 \cdot S(f_i)}{K}} \quad (9)$$

Here, K is the number of spectrum lines used to calculate the features, fi the frequency of the ith spectrum line of the periodogram and S(fi) the magnitude of the ith spectrum line.

C. Genetic Programming (GP)

GP starts with lots of randomly created computer programs. This population of programs is progressively evolved over a series of generations. The evolutionary process uses the Darwinian principle of natural selection (survival of the fittest) and analogs of various naturally occurring operations, including crossover, mutation, gene duplication, and gene deletion. GP mainly consists of construction of terminals, functions (operators) and initial population, fitness evaluation, selection, crossover, mutation and reproduction.

GP uses the following processes to solve problems:

(1) Generate an initial population of random compositions of the functions and terminals of size N.

(2) Perform the following substeps:

(a) Run each program in the population and assign it a fitness value.

(b) Create a new population of computer programs by applying the genetic operators. The new computer programs are chosen from the population with a probability based on fitness. A genetic operator is selected probabilistically. Every operation only uses one of the three operators as follows:

i. Reproduction: Reproduce an existing program (an individual) by copying it into the new population.

ii. Crossover: Create two new computer programs from two existing programs by genetically recombining randomly chosen parts of two existing programs using the crossover operation applied at a randomly chosen crossover point within each program.

iii. Mutation: Create one new computer program from one existing program by mutating a randomly chosen part of the program.

(3) Continue Step (2), until the new population gets N solutions.

(4) Step (2)-(3) are repeated till a desired solution is achieved. Otherwise, terminate the GP operation after a predefined number of generations.

Construction of Terminals and Functions

We use X1, X2, X3, X4, X5, X6, X7 as terminals and use plus (+), minus (-), times (*), mydivide (protected division function), mysqrt (protected square root function), mylog10 (protected common logarithm function) as functions (operators). These functions and terminals satisfy the closure and sufficiency properties. The closure property demands that the function set is well defined and closed for any combination of arguments that it may encounter. On the other hand, the sufficiency property requires that the set of functions and the set of terminals be able to express a solution of the problem [11].

Fitness Function

Fitness function is defined as the ratio of the minimum value of between-class scatter square against the mean value of within-class scatter. In calculation process, the population evolves along the direction of best fitness, which maximizes the fitness function. The new features provide more significant information for classification.

$$fitness = \frac{Min(D_{ij}^{b})^{2}}{\overline{D^{w}}} \tag{10}$$

where D_{ij}^{b} is the between-class scatter between class i and j ,Min means searching the least between-class scatter and $\overline{D^{w}}$ the mean value of within-class scatter.

D. Support Vector Machines

SVM is a powerful methodology for solving problems in nonlinear classification, which has also led to many other recent developments in kernel based learning methods in general. SVMs have been introduced within the context of statistical learning theory and structural risk minimization. We used the one-against-all method for SVM multiclass classification [15]. The method sets up k SVM models where k is the number of classes. The ith SVM is trained with all of the examples in the ith class with positive labels, and all other examples with negative labels. Thus given training data (x_1,y_1),…,(x_m,y_m),where $x_i \in R^n, i = 1,2,\cdots,m$ and $y_j \in 1,2,\cdots,k$ is the class of x_i, the ith SVM solves the following optimization problem:

$$\min\left[\ \frac{1}{2}\left\|\mathbf{w}^i\right\|^2 + C\sum_{j=1}^{m}\xi_j^i\ \right]$$

$$s.t.\quad \left(\mathbf{w}^i\right)^T \varphi\left(\mathbf{x}_j\right) + b^i \geq 1 - \xi_j^i \quad \text{if} \quad y_j = i$$

$$\left(\mathbf{w}^i\right)^T \varphi\left(\mathbf{x}_j\right) + b^i \leq -1 + \xi_j^i \quad \text{if} \quad y_j \neq i$$

$$\xi_j^i \geq 0,\ j = 1,2,\cdots,m \qquad (11)$$

where the training data are mapped to a higher dimensional space by the function ф and C is the penalty parameter. ξ is a slack variable, w a weight,b a threshold. In practice, a high noise level may cause a large overlap of the classes, so one introduces slack variables ξs in order to relax the constraints y_i· ((w · x_i) + b) ≥1; i = 1,…,m.

The kernel function of the SVMs is a Gaussian basis function of the form, $K(x, y) = \exp(-\gamma(x - y)^2)$, where $\gamma = \sigma^2 / 2$ and σ is global basis function width.

We say x is in the class which has the largest value of the decision function

$$\arg\max_{i=1,\cdots,k}\sum_{j=1}^{m}\alpha_j^i K(x, x_j) + b^i \qquad (12)$$

III. Experimental Results and Analysis

Extensive experiments were conducted on a helical gear train having a gear ratio of 41:37 and a module of 5mm. An accelerometer was installed to measure the vibration signals generated on the axes of the 41 teeth pinion. A phase reference and appropriate pickups were used to ensure that the vibration signal was acquired when the same pairs of teeth were meshing. Different vibration data sets were collected when the helical gear train was working at the normal, one slightly spalling tooth, one moderate spalling tooth, one severe spalling tooth, one worn tooth, and two worn teeth conditions. A total of 37 groups of data were collected for each operating condition with a sampling rate of 10 kHz. The original spectral features are calculated by the equation (1)-(9), then genetic programming is used to construct the new ones and reduce the dimensionality for them.

Table 1. GP parameters setup

Generation No.	100	Prob. crossover	0.5
Max depth of tree	7	Prob. mutation	0.5
Population size	100	Selection	roulette
Generative method	Full	Survival	keepbest

We used Gplab toolbox. Table 1 gives the GP parameters setup for feature construction. 100 generations with 100 individuals in each generation were used. The tree depth is limited to a maximum of 7 levels. Generative method is the Full one which shows trees initialized with this method will be perfectly balanced with all the branches of the same length. The probabilities of crossover and mutation both are 0.5. Selection method is roulette, which means the probability of an individual being selected as a parent is proportional to its fitness value. Survival method is 'keepbest', i.e., the best individual from both parents and children is kept for the new population. The tree representations of two new features (Xn1 and Xn2) are acquired after two runs of GP are implemented.

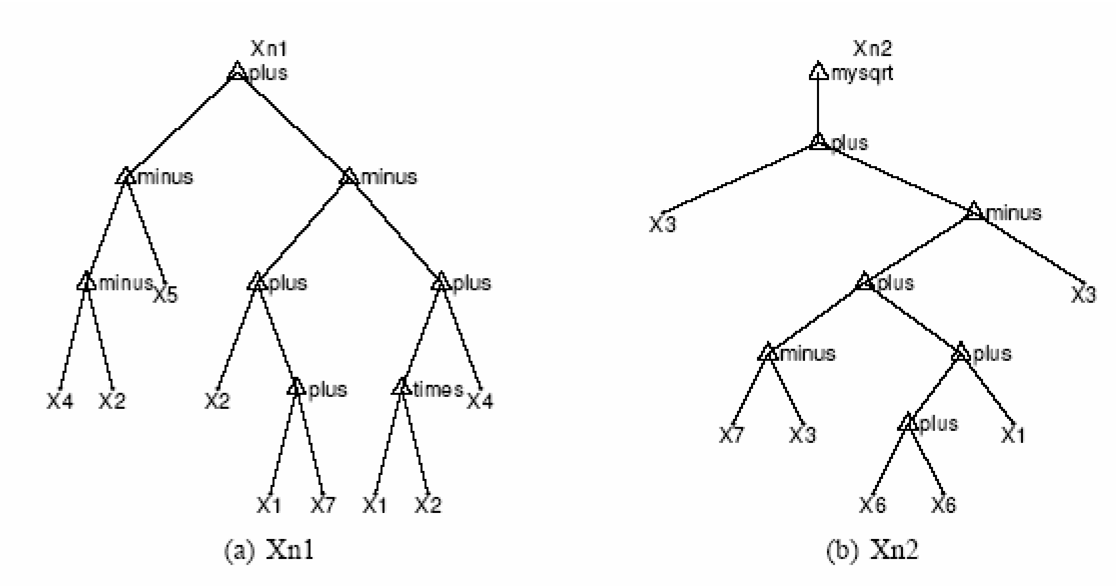


Fig. 2. Tree representation of two new features

Fig.2 show the tree representation of two new features. Through two runs of GP, 7 dimension features reduce to two dimension ones. The new feature Xn1 (Fig.2(a)) and Xn2 (Fig.2(b)) are as follows respectively,

Xn1=plus(minus(minus(X4,X2),X5),minus(plus(X2,plus(X1,X7)),plus(times(X1,X2),X4)))

Xn2=mysqrt(plus(X3,minus(plus(minus(X7,X3),plus(plus(X6,X6),X1)),X3)))

Table 2 gives the SVM parameters setup for training.

Table 2. SVM training parameters setup

Bound on the lagrangian multipliers	C=100
Kernel	gaussian
Kernel parameter	$\gamma = 0.3$
Tolerance of termination criterion	$\varepsilon = 0.001$

Table 3.The confusion matrix of the train set

Gear condition	Normal	Spall1	Spall 2	Spall3	Wear1	Wear2
Normal	1	0	0	0	0	0
Spall1	0.0667	0.9333	0	0	0	0
Spall2	0	0	1	0	0	0
Spall3	0	0	0	1	0	0
Wear1	0	0	0	0	1	0
Wear2	0	0	0	0	0	1

Table 3 is the confusion matrix of the train set. The train classification rate is 98.9% for 30 groups of train data. The confusion matrix M(i,j) is defined as the conditional probability P(Cj,Ci) which represents the probability of a class C that should belong to the class i is assigned by the SVMs to the class j. The row labels denote true conditions and the column labels indicate predicted conditions. For test samples, the results show that the test classification rate is 100% for 7 groups of test data.

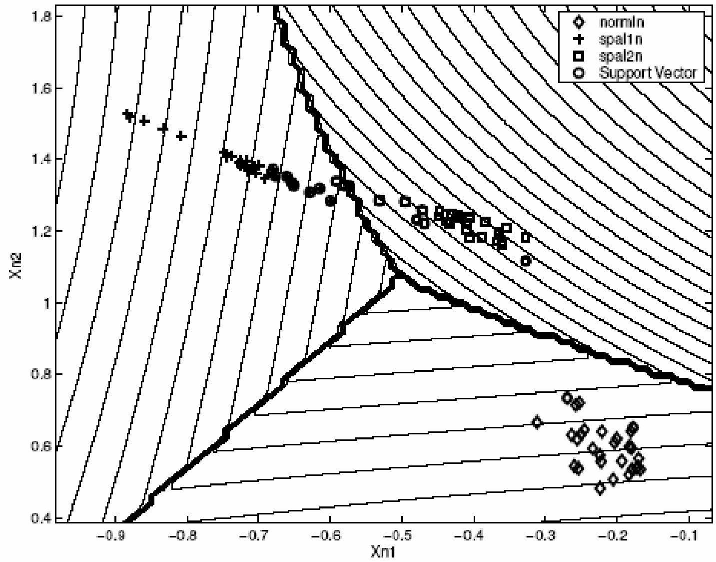


Fig. 3. The First 3 Conditions Classification Results i.e., normal, one slightly spalling tooth, one moderate spalling tooth conditions denoted by normln, spal1n, spal2n, for short, respectively.

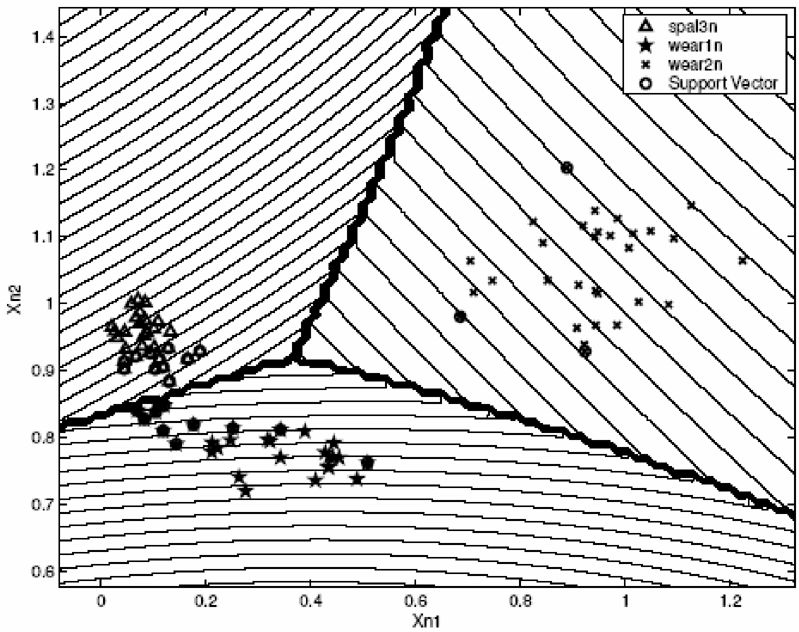


Fig. 4. The Last 3 Conditions Classification Results, i.e., one severe spalling tooth, one worn tooth, and two worn teeth conditions denoted by spal3n, wear1n, wear2n for short, respectively.

Fig. 3 and Fig. 4 show classification results based on SVMs. Fig.3 shows results of the first three conditions, i.e., normal (diamond mark), slightly spalling (plus mark), moderate spalling (square mark). Fig.4 shows results of the last three conditions, i.e., severe spalling (triangle mark), one worn tooth (pentagram mark), and two worn teeth conditions (x mark). The circle mark represents the support vector. The contours show highly generalized ability. The results show that the SVMs can classify correctly the 6 conditions above and give the classification rates with high accuracy.

IV. Conclusion

In this paper, we present an effective method for classifying gear damage conditions. These conditions include the normal, one slightly spalling tooth, one moderate spalling tooth, one severe spalling tooth, one worn tooth, and two worn teeth conditions. The PSD of the vibration signals of gearbox casing is introduced to construct some original spectral features. The PSD is estimated by the periodogram from which the features are evaluated. To get useful features from the original spectral ones which are high dimensional, genetic programming is used to select them and reduce the spectral feature dimension along the direction of best fitness which maximizes the fitness. The new features provide more significant information for a classifier. The classifier is based on support vector machines with multi-class classification ability. The results show the effectiveness of the approach. Because real faults are various and complicated, experiments on large problems need be further done for practical use.

Acknowledgment

The authors would like to acknowledge the financial support provided by the National Key Basic Research Special Found of China (Grant No.2005CB724101, No. 2003CB716207), the National Natural Science Foundation of China (Grant No 50405033) and Wuhan University of Science and Technology of HuBei Province Key Mechanical Transmission Lab Foundation (Grant No.2003A15).

References

[1] Allianz Versicherungs-AG, *Handbook of Loss Prevention*, Springer-Verlag, Berlin, 1978.
[2] A. N. Michel and R. K. Miller,*Gears and Their Vibration*, Marcel Dekker,INC., 1983.
[3] P. D. Mcfadden,“A Technique for Calculating the Time Domain Averages of the Vibration of the Individual Planet Gears and Sun Gear in an Epicyclic Gearbox”, Journal of Sound and Vibration, vol. 144,1991,pp. 163-172.
[4] P. D. Mcfadden, “Interpolation Techniques for Time Domain Averaging of Gear Vibration, Mechanical Systems and Signal Processing”, vol. 3,1989, pp. 87-97.
[5] R. B. Randal, “A New Method of Modeling Gear Faults”, Journal of Mechanical Design, vol. 104,1982, pp. 259-267.
[6] W. J. Wang and P. D. Mcfadden, “Early Detection of Gear Failure by Vibration Analysis I. Calculation of the Time-Frequency Distribution”, Mechanical Systems and Signal Processing, vol. 7,1993, pp. 82-192.
[7] W. J. Wang and P. D. Mcfadden, “Application of Wavelet to Gearbox Vibration Signals for Fault Detection”, Journal of Sound and Vibration, vol. 192,1996, pp. 816-828.
[8] Wuxing Lai, Jianping Xuan, Tielin Shi and Shuzi Yang, “Research of Vigner-Ville Time Frequency and Application in Detecting Gear Pinion Fault”, Journal of Vibration Engineering, vol. 16,2003, pp. 247-250.
[9] Guicai Zhang, Tielin Shi, Jianping Xuan and Shuzi Yang, “ Classification of Gear Faults Using RBF Network Combined with Higher-order Statistics”, China Mechanical Engineering, vol. 11,1999, pp. 1250-1252.
[10] A.J. Hoffman, N.T. van der Merwe, “The Application of Neural Networks to Vibrational Diagnostics for Multiple Fault conditions”, Computer Standards and Interfaces, vol. 24,2002, pp.139-149.
[11] J. R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. Cambridge, MA: MIT Press,1992.
[12] Chen P., Taniguchi M., Toyota T., et al, “Fault Diagnosis Method for Machinery in Unsteady Operating Condition by Instantaneous Power Spectrum and Genetic Programming”, Mechanical Systems and Signal Processing, vol. 19, 2005, pp. 175-194.
[13] Liang Zhang, Lindsay B. Jack, Asoke K. Nandi, “Fault Detection Using Genetic Programming”, Mechanical Systems and Signal Processing, vol. 19,2005, pp. 271-289.
[14] Vapnik V. , *The Natural of Statistical Theory*. New York : Springer Verlag,1995.
[15] Chih-Wei Hsu and Chih-Jen Lin, “A Comparison of Methods for Multiclass Support Vector Machines”, IEEE Transactions on Neural Networks, vol. 13,2002, pp. 415-425.

Jianping Xuan received the Ph.D degree from Huazhong University of Science and Technology in 1999. Since November 2001, he has been an associate professor in School of Mechanical Science and Engineering at Huazhong University of Science and Technology. His major research interests include intelligent computing, condition monitoring, fault diagnosis, machine learning, wavelet analysis, and time series analysis.

HanHong Jiang received the Ph.D degree from Huazhong University of Science and Technology in 1999. Since 1990, he has been with the Naval University of Engineering, Wuhan, where he is currently an associate professor in the College of electrical and Information Engineering. His research interests include machinery remote diagnosis, intelligent diagnosis.

Tielin Shi received the Ph.D degree from Huazhong University of Science and Technology (HUST) in 1991. Now, he is a professor and vice dean of School of Mechanical Science and Engineering at HUST. He is currently studying stroboscopic interferometer, MEMS, information technology, structure dynamic diagnosis and reliability evaluation. Awards received include the first award of state education commission science and technology progress (1995), the first award of state mechanical industry commission science and technology progress (1996), the second award of state education commission science and technology progress (1994), the second award of national science and technology progress (1997) and the second award of national excellent science and technology works (1995).

Guanglan Liao received the Ph.D degree in Mechanical Science and Engineeering from Huazhong University of Science and Technology, Wuhan, P.R. China. His research focuses on condition monitoring, fault diagnosis, neural networks, and MEMS.