Front. Mech. Eng China (2006) 3: 292–298
DOI 10.1007/s11465-006-0022-2

RESEARCH ARTICLE

LI Wei-hua, SHI Tie-lin, YANG Shu-zi

An approach for mechanical fault classification based on generalized discriminant analysis

Abstract To deal with pattern classification of complicated mechanical faults, an approach to multi-faults classification based on generalized discriminant analysis is presented. Compared with linear discriminant analysis (LDA), generalized discriminant analysis (GDA), one of nonlinear discriminant analysis methods, is more suitable for classifying the linear non-separable problem. The connection and difference between KPCA (Kernel Principal Component Analysis) and GDA is discussed. KPCA is good at detection of machine abnormality while GDA performs well in multi-faults classification based on the collection of historical faults symptoms. When the proposed method is applied to air compressor condition classification and gear fault classification, an excellent performance in complicated multi-faults classification is presented.

Keywords gear, air compressor, fault diagnosis, classification, kernel methods, discriminant analysis

1 Introduction

Machine condition monitoring and fault diagnosis have received intensive study for several decades, and various approaches have been taken, such as statistical signal processing, time-frequency analysis, wavelet, and neural networks. Among them, pattern recognition method provides a systematic approach to acquiring knowledge from fault samples. In fact, mechanical fault diagnosis is essentially a problem of pattern classification, in which feature extraction plays an important role.

Generally, a test sample can be classified according to a discriminant function obtained from known samples. However, the processes, which cause vibrations in machinery, are often non-linear, and the strong random noises in an industrial environment degrade the signal-to-noise ratio greatly. It is difficult for LDA (linear discriminant analysis) to recognize the incipient machine fault, especially for the similar faults, it is very hard to separate one from another.

Motivated by successful application of SVM (support vector machine) and KPCA (kernel principal component analysis) in pattern recognition [1, 2], kernel Fisher discriminant analysis, which generalized linear Fisher discriminant to nonlinear one as an effective nonlinear feature extraction method, was proposed by Mika[3, 4]. By kernel trick, most of the linear methods can be introduced into non-linear fields [5]. For solving multi-classification problem, GDA (generalized discriminant analysis), which is a nonlinear discriminant analysis method, is proposed by Baudat [6].

This paper investigates the use of GDA as a powerful tool to process the feature sets in the context of air compressor condition monitoring and gear faults classification. It proceeds as follows. Section 2 presents the basics of GDA. A brief comparison between KDA (kernel-based discriminant analysis) and KPCA is also given in this section. Section 3 focuses on the air compressor experiments, and gear datasets of Laborelec/ULB are also used for and investigation of GDA based fault classification method. The work is concluded in Section 4 with an illustration of the GDA efficiency to classify the data in the goal of gearbox fault classification, and some perspectives are also introduced.

Translated from *Journal of Vibration Engineering*, 2005, 18(2) (in Chinese)

LI Weihua (✉)
College of Automotive Engineering,
South China University of Technology,
Guangzhou 510640, China
E-mail: whlee@scut.edu.cn

SHI Tie-lin, YANG Shu-zi
School of Mechanical Science & Engingeering,
Huazhong University of Science & Technology,
Wuhan 430074, China

2 Generalized discriminant analysis

2.1 Kernel based discriminant analysis

Linear discriminant analysis (LDA) is a classical statistical approach for classifying samples of unknown classes, based on training samples with known classes. By LDA, the datasets were projected from input space onto a new low dimensional space, where the new datasets can be obtained as linear combination of raw data projected on the discriminant direction. In LDA, within-class and between-class scatter are used to formulate criteria for class separability. This method maximizes the ratio of between-class variance to the within-class variance in any particular data set thereby guaranteeing maximal separability.

However, this method fails for a nonlinear problem. Since support vector machine—a learning algorithm proposed by Vapnik [7], which is motivated by theoretical results from statistical learning theory, many researchers began to focus on kernel methods and its potential for the design of new learning algorithms. Among them, GDA has been proposed as a powerful tool to deal with nonlinear problem [6]. The main idea behind kernel methods is to map the input space into a convenient high dimensional feature space *F* in which variables are nonlinearly related to that in the input space. In the new space, one can solve the problem in a classical way such as the *LDA* method, where inner products in the feature space *F* can be computed by a kernel function without knowing the nonlinear mapping explicitly. In fact, GDA is exactly KDA. A short mathematical description follows.

Let $\boldsymbol{x}$ be a sample of $\boldsymbol{X}$ (composed by M samples), which is the datasets of all fault samples in input space, $\boldsymbol{X}_l$ designs subsets of $\boldsymbol{X}$, $\mathbf{X}=\bigcup_{l=1}^{N}\mathbf{X}_l$, where N is the number of fault classes. The cardinality of the subsets $\boldsymbol{X}_l$ is denoted by n_l, $\sum_{l=1}^{N} n_l = M$. By some kind of kernel function, $\boldsymbol{X}$ can be mapped nonlinearly into a high dimensional feature space **F**. Let the image of the sample $\boldsymbol{x}$ be $\varphi(\boldsymbol{x})$. For simplicity, assume for the moment that all the images of $\mathbf{x}_i$ are centered, i.e., $\sum_{i=1}^{M}\varphi(\mathbf{x}_i)=0$, and the covariance matrix of the mapped datasets is denoted as $\bar{\boldsymbol{C}}$, $\bar{c}_l$ designs the mean vector of fault class l.

$$\bar{\mathbf{c}}_l = \frac{1}{n_l}\sum_{k=1}^{n_l}\varphi(\mathbf{x}_{lk}) \tag{1}$$

where $\boldsymbol{x}_{lk}$ denotes the k_{th} sample of fault class l.

Using matrix $\boldsymbol{B}$ to denote the covariance matrix of the fault class centers, which represents the inter-classes scatter in the space **F**

$$\boldsymbol{B} = \frac{1}{M}\sum_{l=1}^{N} n_l \bar{\boldsymbol{c}}_l \bar{\boldsymbol{c}}_l^T \tag{2}$$

And the covariance matrix of all the mapped datasets can be expressed using the class indexes as

$$\bar{\boldsymbol{C}} = \frac{1}{M}\sum_{l=1}^{N}\sum_{k=1}^{n_l}\varphi(\boldsymbol{x}_{lk})\varphi^{\mathrm{T}}(\boldsymbol{x}_{lk}) \tag{3}$$

This represents the total inertia of the data into **F**, usually used to describe the correlation.

According to Vapnik, kernel function can be expressed as inner dot products in feature space. For given fault classes p and q, we express the kernel function by

$$(\boldsymbol{K}_{ij})_{pq} = \varphi^{\mathrm{T}}(\boldsymbol{x}_{pi})\varphi(\boldsymbol{x}_{qj}) \tag{4}$$

where $\boldsymbol{K}$ is an $(M \times M)$ matrix defined on the class elements by $\boldsymbol{K}_{pq} = (\boldsymbol{k}_{ij})_{i=1,\cdots,n_p,\, j=1,\cdots,n_q}$. $\boldsymbol{K}_{pq} = (k_{ij})_{i=1,\cdots,n_p,\, j=1,\cdots,n_q}$ is an $(n_p \times n_q)$ matrix constructed by inner dot products in feature space.

It can be seen that, the total inter-class scatter and the within-class scatter can be measured by the matrices $\bar{\boldsymbol{C}}$ and B respectively. As such for the LDA, the purpose of the GDA method is to maximize the inter-classes scatter and minimize the within-classes scatter. So, the discriminant function can be noted as

$$J_F = \frac{\boldsymbol{v}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{v}}{\boldsymbol{v}^{\mathrm{T}}\bar{\boldsymbol{C}}\boldsymbol{v}} \tag{5}$$

where $\bar{\boldsymbol{C}}$ is invertible. It was shown that this maximization is equivalent to eigenvalue resolution of the equation

$$\lambda\bar{\boldsymbol{C}}\boldsymbol{v} = \boldsymbol{B}\boldsymbol{v} \tag{6}$$

This can be rewritten as

$$\lambda\boldsymbol{v} = \bar{\boldsymbol{C}}^{-1}\boldsymbol{B}\boldsymbol{v} \tag{7}$$

where λ is the maximum eigenvalue of matrix $\bar{\boldsymbol{C}}^{-1}\boldsymbol{B}$, and $\boldsymbol{v}$ eigenvectors correspondingly.

To solve this equation, we have to find eigenvalues λ and eigenvectors $\boldsymbol{v}$, which indicate discriminant directions for faults classification. When knowing the normalized vectors $\boldsymbol{v}$, we then compute projections of a test sample $\boldsymbol{z}$ by

$$\boldsymbol{v}^{\mathrm{T}}\varphi(\boldsymbol{z}) = \sum_{p=1}^{N}\sum_{q=1}^{n_p}\boldsymbol{\alpha}_{pq}K(\boldsymbol{x}_{pq},\boldsymbol{z}) \tag{8}$$

where $\varphi(\boldsymbol{z})$ is the image of $\boldsymbol{z}$ in feature space.

Notice that the conclusion above is obtained based on the assumption that all observations in feature space are centered, which may not be exactly. Therefore, we need to replace the eigenvectors $\boldsymbol{v}$ and kernel matrix $\boldsymbol{K}$ with $\tilde{\boldsymbol{v}}$ and $\tilde{\boldsymbol{K}}$ respectively.

$$\tilde{\boldsymbol{V}}^{\mathrm{T}}\varphi(\boldsymbol{z}) = \sum_{p=1}^{N}\sum_{q=1}^{n_p}\tilde{\boldsymbol{\alpha}}_{pq}\tilde{K}(\boldsymbol{x}_{pq},\boldsymbol{z}) \tag{9}$$

where $\tilde{\boldsymbol{K}} = \boldsymbol{K} - 1_M\boldsymbol{K} - \boldsymbol{K}1_M + 1_M\boldsymbol{K}1_M$. For more detail information about the calculation of $\tilde{\boldsymbol{v}}$ and $\tilde{\boldsymbol{K}}$, please refer to[1–3,7].

If we select a polynomial kernel function of degree 1, GDA is LDA in fact, therefore, LDA is a special example of nonlinear discriminant analysis. It should be pointed out that all the fault samples mentioned above are feature samples preprocessed and not the raw vibration signals.

Different type of kernels allows the construction of Polynomial Classifiers, Radial Basis Function Classifiers, and Neural Networks Classifiers, and their examples are given below:

Polunomial kernel: $k(x, y) = [(x \cdot y) + 1]^d$, where d is the polynomial degree;

RBF (radial basis function) kernel: $k(x, y) = \exp\left[-\|x - y\|^2 / 2\sigma^2\right]$, where the parameter σ has to be chosen;

Neural network type sigmoid kernel: $k(x, y) = \tanh[v(x \cdot y) + c]$, where the parameter c has to be chosen.

According to Mercer theorem, if k is a continuous kernel of a positive integral operator, there exists a mapping into a space where k acts as a dot product. Other kernels can be used, provided that they satisfy the Mercer theorem. In this paper, classical Gaussian kernel (RBF) is used as

$$K(\boldsymbol{x}, \boldsymbol{y}) = \exp(-\|\boldsymbol{x} - \boldsymbol{y}\|^2 / 0.7) \qquad (10)$$

2.2 Relation to KPCA

One of the powerful techniques for extracting features from high dimensional data sets is KPCA (*kernel principal component analysis*), which is similar to KDA. Here we point out the close relationship between the KDA and the KPCA. Both of them employ the technique called "kernel tricks" to realize the nonlinear mapping of input dataset to a high dimensional feature space where nonlinear features are obtained linearly. The purpose of KPCA is an optimal lower dimensional representation rather than discrimination; the features captured by KPCA are effective for data representation and reconstruction. Data representation means compressing data along the principal directions (direction of principal component) with little loss of structural information. However, the direction for representation may not be the best discriminative direction, however, sometimes it is. After capturing the data structure information, further classification methods may also need, KPCA based classifying methods are presented by Scholköpf [1], which have been used in gearbox diagnosis [8]. On other hand, KDA has been shown to provide a better performance than LDA and KPCA in classification problems, and it can find the directions along which the classes can be optimally differentiated.

KPCA is a powerful technique for extracting nonlinear features and reducing dimensionality, and it has been widely used in unsupervised learning. While KDA introduces the discriminative criterion of maximization between-class scatter and minimization within-class scatter, provides directions for separation, it can find applications for supervised learning.

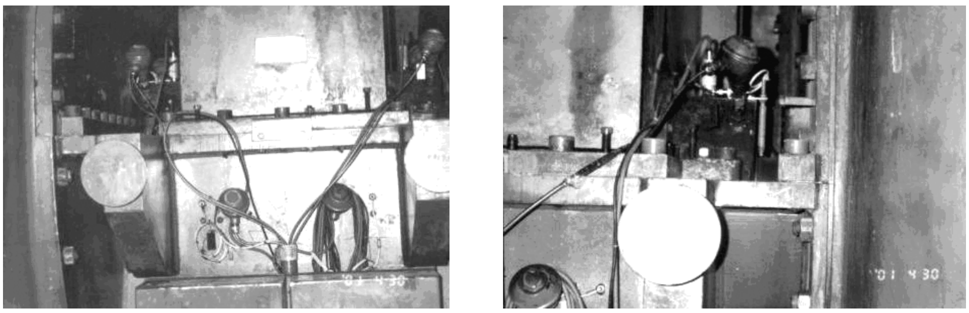
3 Experiments analysis

3.1 Air compressor condition monitoring

Air compressor is one of classical rotation machinery, which has many types of failure mode. It is very hard to separate the fault with features in frequency domain if it performs well in time domain with information optimization, such as compressor surge and steam pressure oscillation. As is well known, the compressor surge is a very harmful fluid dynamic effect for working media with heavier molecular weight, for example, CO_2 or SO_2. Serious surge may result in catastrophic damage of the compressor rotor, which is difficult to extract frequency features. However, with some statistical time domain features, such as variance, skewness, kurtosis and crest factor, nonlinear discriminant analysis can be used to diagnose the faults effectively.

The practical signals were measured from a large air compressor sets in an oxygen manufacturing plant. Two eddy-current displacement vibration amplitude transducers were arranged orthogonally near the case of bushes to pick up the vibration signals of the axis, and phase reference was also used to ensure that the data acquisitions always start for the same pairs of meshing teeth at precisely the same stage. The test rig layout is shown in Fig. 1, and more information can be found in [9].

After overhaul, the air compressor was running again. During the running, compressor surge occurred, the vibration of the shaft increased rapidly, and the shaft orbit was in disorder as shown in Fig. 2. Besides, steam pressure oscillation happened with another air compressor in several months, the vibration amplitude of the shaft near the motor obviously increased, and the orbit shows as square shaped, maybe because the stator and the rotor are not homocentric. Because these two compressors have the same construction, we will show the effectiveness of kernel based nonlinear discriminant analysis on classifying these conditions together.



(a) (b)

Fig. 1 Sensors arrangement for air compressor monitoring. (**a**) Layout of eddy-current transducers. (**b**) Layout of phase reference

The shaft orbits in different operating conditions are shown in Fig. 2. The left column in the figure is that of the shaft side near the motor while the right one is that of the shaft far from the motor. It can be seen that, when the compressor was running in normal condition, the vibration amplitude fluctuated within ±10 μm, and the orbits took the shape of a circle. When compressor surge occurred, the vibration increased to 15 μm, the left orbit changed into the shape of a rectangle, and the right one was in disorder. When steam pressure oscillation happened, the vibration increased rapidly to ±20 μm, the left orbit got shaped as a square, and the right one as an irregular circle.

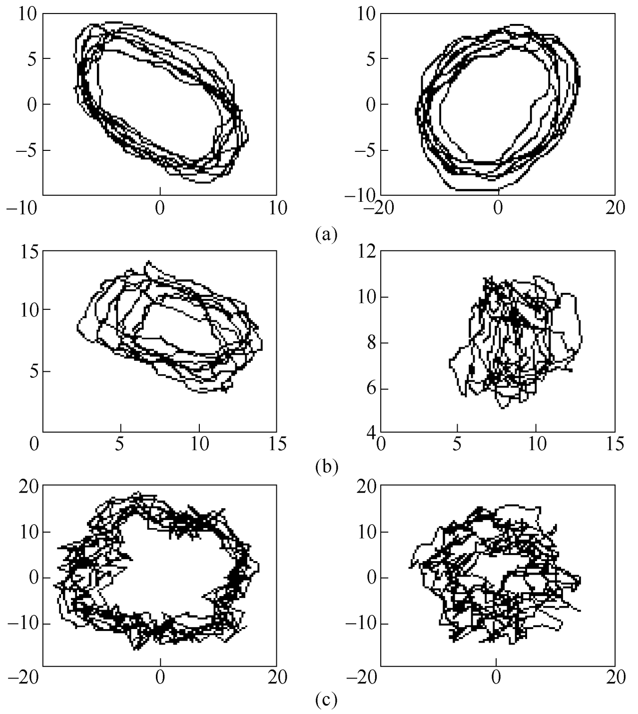


Fig. 2 Shaft orbits of different air compressor running conditions. (**a**) Normal condition. (**b**) Compressor surge. (**c**) Steam pressure oscillation

Typical dataset of the three conditions, giving 60 datasets per condition, are used to verify the proposed method. Every dataset has two channel data acquired by two eddy-current sensors, three time domain features, including *variance*, *skewness,* and *kurtosis*, which were used as raw feature sets for further analysis. These three features correspond to zero-lag 2nd, 3rd, and 4th cumulants of the vibration signals, which are immune to additive Gaussian noise. Therefore, there is a 6D feature vector describing the state for a dataset. All the datasets are pre-processed to get the raw feature sets, and then the feature sets are centered and normalized to construct the input space of a 180×6 matrix. Then they are divided into two groups; one group of 30 feature sets per condition is used for training, and the other one of 30 feature sets per condition for testing.

The classification results are presented in Fig. 3, where (a) is the result obtained by LDA, (b) is that by KPCA, and (c) by KDA respectively. The kernel functions adopted by KPCA and KDA are the same one as Eq.(10). Obviously, LDA could not separate the condition of compressor surge from that of steam pressure oscillation, while the latter two performed well. Thus it can be known that the feature samples of compressor conditions satisfy the requirements of data representation and discrimination simultaneously. However, the situation is not always in this manner, such as the following example of gear failure experiments.

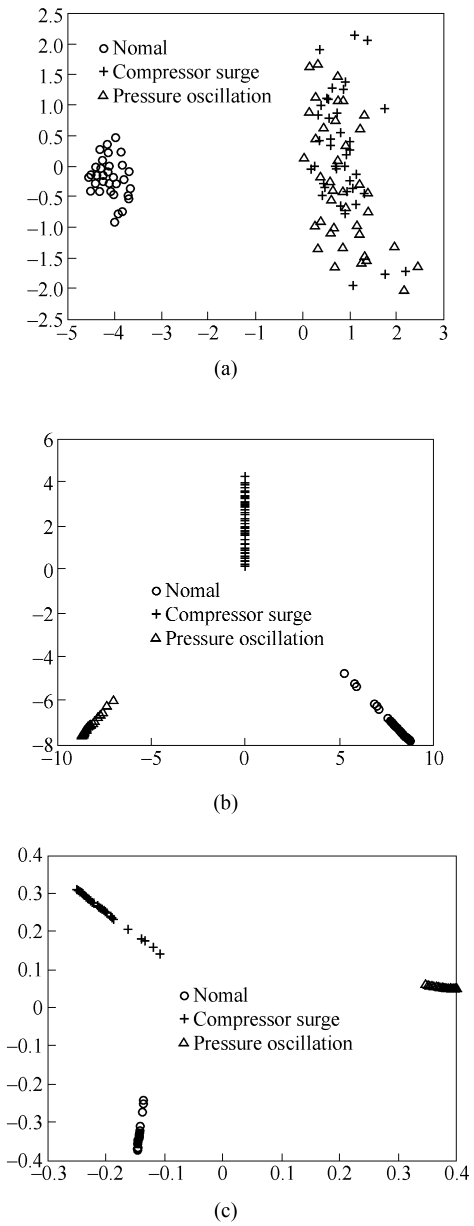


Fig. 3 Classification of air compressor running conditions. (**a**) Classification using LDA. (**b**) Classification using KPCA. (**c**) Classification using KDA

3.2 Gear faults classification

The experiments were carried out on a helical gear train. The gear train and test rig are shown in Fig. 4. The input torque

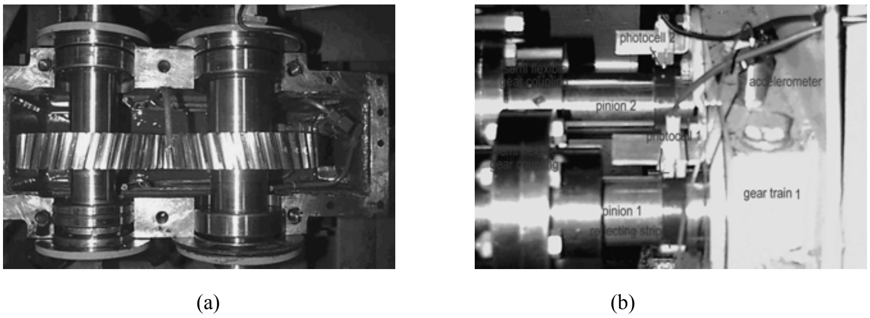


Fig. 4 Helical gear train and test rig. (**a**) Gear train. (**b**) Test rig

was 80 Nm, the input rotating speed on 41 teeth gear-shaft was 10Hz, and the sampling rate was 10 kHz. Acoustic responses are measured by accelerometers mounted on the casing of the gear trains next to the (healthy) bearings supporting the gears. Using two phase references and the appropriate pick-ups (Here infrared Hamatsu photocells), one ensures that the data acquisitions always start for the same pairs of meshing teeth at precisely the same meshing stage. Characteristics of the gears are as follows: gear ratio 41: 37, 5 mm normal module, angle: 20 °, and distance between shaft centerlines: 200 mm. Material: mild steel to easily introduce surface defects. For more details, please refer to [10].

The basic period corresponds to 37 revs of the 41 teeth pinion. In a second series of tests, the gear train was made of spur gears, and typical raw signals of different conditions were measured (normal condition, pinion with small spall condition, pinion with severe spall condition, and pinion wear condition), giving 74 datasets(74 revs) per condition. Period averaging was performed before signal analysis. The

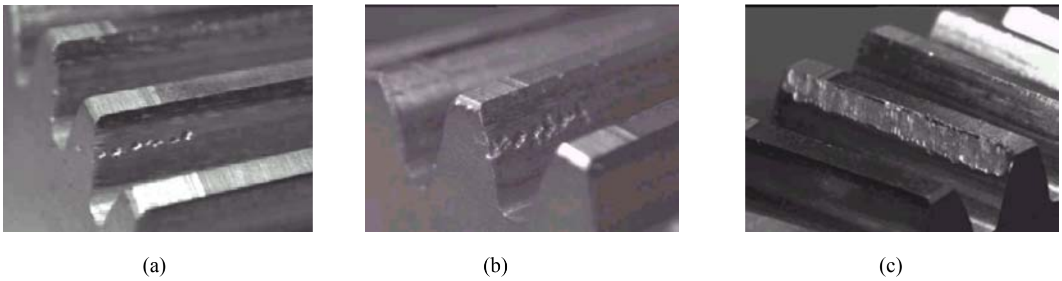


Fig. 5 Gear trains with defects. (**a**) Small spall. (**b**) Severe spall. (**c**) Wear on one tooth profile

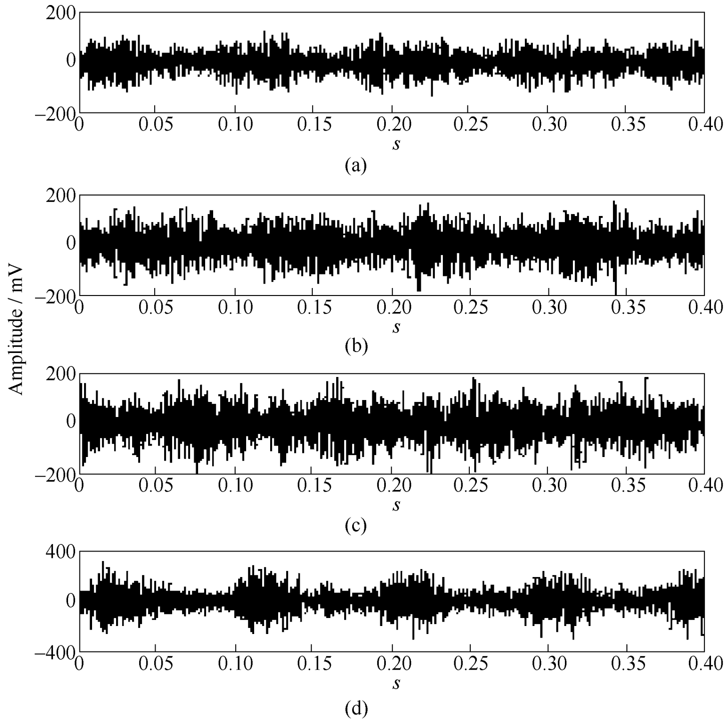


Fig. 6 Time domain waveform of gear vibration signals. (**a**) Normal. (**b**) Small spall. (**c**) Severe spall. (**d**) Wear

following are the defective gear trains:
Train1: Small spall on gear profile made of mild steel (Fig. 5(a)).
Train2: More severe spall on gear profile (Fig. 5(b)).
Train3: Wear on one tooth profile of the 41 teeth pinion. It is limited to the addendum (Fig. 5(c)).

Vibration signals with time synchronous averaging are displayed in Fig. 6. Obviously, there are distinct differences between the vibration signal of the gear with wear and those under the other two defects, but it is very difficult to separate the signal under small spall from that under severe spall conditions.

Ten feature parameters of each raw signal were computed, and then 74 feature sets were divided into two groups. One group of 37 feature sets per condition is used for training, and the other one of 37 feature sets per condition is used for testing. Then these features were normalized and trained to set up the classifying model and the discriminative subspace, which can separate one condition from another one. Here, RBF kernel function as Eq. (10) was employed.

In the three dimensional discriminative subspace, feature sets of gearbox conditions are classified into four clusters, and result of KPCA based method is shown in Fig. 7(a), and that of KDA based one in Fig. 7(b).

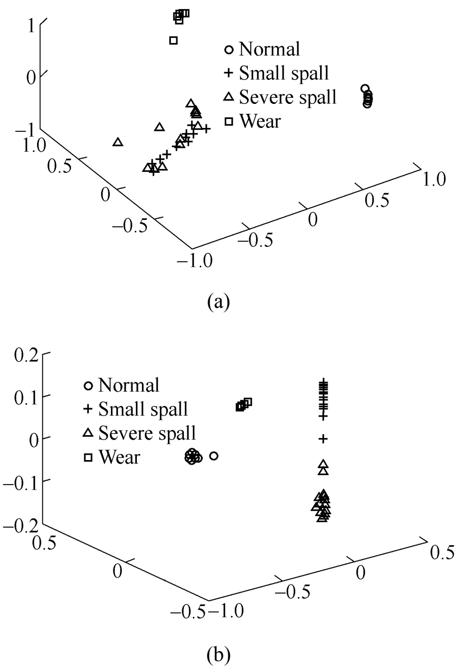


Fig. 7 Gear fault classification. (**a**) Result of KPCA based method. (**b**) Result of KDA based method

It is clear that clusters of normal condition and wear condition are well separated, and KDA based method outperforms KPCA based one for distinguishing the cluster of small spall condition from that of severe spall one. Because the features extracted by KPCA are not sensitive to the difference between these two conditions, the classification result is influenced greatly by data structure. The three features picked up by KDA are very discriminative to different classes, even these four conditions. The reduction in dimension is also obtained by the projection of the feature sets onto a smaller subspace defined by the selected discriminative features. In the subspace, clusters are identified, and each cluster represents a particular condition. The selected features are determined by the first three eigenvalue of kernel matrix (0.999 285, 0.996 720 and 0.841 989 descending). According to kernel discriminant analysis, the greater the eigenvalue is, greater is the between-class scatter and smaller is the within-class scatter, which means that the samples within a same group cluster together closely and samples of different groups separate from each other as far as possible. For KDA, the percentage of correct classification is 95.59 %.

4 Conclusions

KPCA, which is similar to KDA, is one of the powerful techniques for extracting features from high dimensional data sets. The purpose of KPCA is an optimal lower dimensional representation rather than discrimination. The features captured by KPCA are effective for data representation and reconstruction, and its ability for classification is influenced greatly by data structure. It is suitable for failure detection using unsupervised KPCA learning algorithm. On the other hand, KDA has been shown to provide a better performance than LDA and KPCA in classification problems The KDA method is suited to capture the discriminated features, at the same time similar faults were classified effectively. Experimental results indicate that the proposed method is sensitive to different running conditions of air compressor and gearbox, and also demonstrate the superiority of KDA over KPCA for gear fault classification. Because it introduces the discriminative criterion of maximization of the between-class scatter and minimization of the within-class scatter, it is able to find the directions along which the classes can be optimally differentiated. By calibration, this model can be applied to complicated fault classification.

Acknowledgements This research was supported in part by the National Natural Science Foundation of China (No. 50475095), the Natural Science Foundation of Guangdong Province (Nos. 05300143, 04020082), and the Open Research Foundation of State Key Laboratory of Vibration, Shock and Noise (No.VSN-2004-03). We are also grateful to Laborelec/ULB, BELG for making the Gear datasets available for public research.

References

1. Schölkopf B, Smola A, Müller K R, Nonlinear component analysis as a kernel eigenvalue problem, Neural Computation, 1998, 10: 1299–1319
2. Mika S, Rätsch G, Weston J, Schölkopf B, et al, Invariant feature extraction and classification in kernel spaces, Advances in Neural Information Processing Systems, 2000(12): 526–532
3. Mika S, Rätsch G, Weston J, et al, Fisher discriminant analysis with kernels, Neural Networks for Signal Processing IX, 1999,

41–48
4. Mika S, Smola A, Schölkopf B, An improved training algorithm for kernel fisher discriminants, In Proceedings AISTATS, Morgan Kaufmann, 2001, 13: 98–104
5. Roth V, Steinhage V, Nonlinear discriminant analysis using kernel function, Advances in Neural Information Proceeding Systems, 2000, 12: 568–574
6. Baudat G , Anouar F, Generalized discriminant analysis using a kernel approach, Neural Computation, 2000, 12(10): 2385–2404
7. Vapnik V N, An overview of statistical learning theory, IEEE Trans, on Neural Networks, 1999, 10(5): 988–999
8. Li W H, Liao G L, Shi T L, KPCA and its application in gear fault diagnosis, Chinese Journal of Mechanical Engineering, 2003, 39(8): 65–70(in Chinese)
9. Li W H, Mechanical fault feature extraction and classification based on kernel methods, PhD. Dissertation, Wuhan: Huazhong University of Sci &Tech, 2003 (in Chinese)
10. Leduc V, Kanny G, Moneret-Vautrin D A, et al, Diagnostic de l'allergie aux graines de Sésame: choix de la matière première et optimisation du procédé de fabrication, Revue Francaise d'Allergologie et d'Immunologie Clinique, 2001, 41(1): 131