An Approach For Bearing Fault Diagnosis Based On PCA and Multiple Classifier Fusion

Min Xia[1], Fanrang Kong[1], Fei Hu[1]
1.Department of Precision Machinery and Instrumentation
University of Science and Technology of China
Hefei, China
xiamin09@mail.ustc.edu.cn

Abstract—**The purpose of this paper is to propose a new system, with both high efficiency and accuracy for fault diagnosis of rolling bearing. After pretreatment and choosing sensitive features of different working conditions of bearing from both time and frequency domain, principal component analysis(PCA) is conducted to compress the data dimension and eliminate the correlation among different statistical features. The first several principal components are sent to the classifier for recognition. However, recognition method with a single classifier usually has only a limited classification capability that is insufficient for real applications. An ongoing strategy is the decision fusion techniques. The system proposed in this paper develops a decision fusion algorithm for fault diagnosis, which integrates decisions of multiple classifiers. First, the front four principle components are chosen as input of individual classifier. A selection process of the classifiers is then operated on the basis of correlation measure for the purpose of finding an optimal sequence of them. Finally, classifier fusion algorithm based on Bayesian belief method is applied to generate the final decision. The result of experiments show that this new bearing fault diagnosis system recognize different working conditions of bearing more accurately and more stably than a single classifier does, which demonstrates the high efficiency of the proposed system.**

Keywords-rolling bearing; fault diagnosis; PCA; multiple classifier fusion.

I. INTRODUCTION

Rolling bearing is a significant part widely used in various of rotating machinery and the performance of rolling bearing is directly related to the status of the whole machine. Consequently, many research endeavors have been devoted to the area of condition monitoring and fault diagnosis of rolling bearing and many relevant research papers have been published during the past decades. The traditional methods in this area involve noise analysis, temperature analysis, oil analysis and vibration analysis[1].

Among the methods mentioned above, vibration analysis is the most generally applied one since information about the condition of rolling bearing is generally contained in vibration signal. One issue of this method is the process of exaction and selection of proper features that can correctly represent the condition of bearing. With the development of feature exaction technology in time domain, frequency domain and time-frequency domain, large amount of features can be obtained from the original signal. However, too many feature parameters can increase the cost of computing of classification and sometimes even decrease the accuracy of recognition[2]. Hence, the selection of sensitive features on the basis of feature exaction is of much necessity and forms a challenge to both researchers and engineers .

Principal component analysis(PCA)[3], also known as the Karhunen–Loeve transform, is a basic method in the system of the multivariate analysis techniques. By transforming a complex data set to a simple one with lower dimension, PCA reduce the less significant information in data set for further computing. Because of its excellent capability in extracting relevant information from confusing data sets, PCA has been successfully applied in numerous areas including data compression, feature extraction, image processing, pattern recognition and process monitoring in recent years[4-6].

The other issue of the fault diagnosis system is how to exactly recognize the fault pattern. Recently, many artificial intelligence techniques have been applied in fault diagnosis area such as expert systems, neural networks, supporting vector machine, Gaussian mixture model network and fuzzy logic networks[7-11]. However, it has been detected in some research that an individual decision system with a single classifier can only acquire a limited classification capability that might fail to be sufficient for a particular application.

The application of multiple classifier fusion(MCF) has been investigated in recent years, and much has been achieved in the solution to complex pattern recognition tasks. The MCF can be divided into two categories: the static and the dynamic methods. The static methods are simple, such as majority voting, minimum and maximum, and average[12]. In comparison, the dynamic methods are more accurate with consideration of the information from the training phase of the classifiers, which includes Bayesian belief method, behavior -knowledge space (BKS) and Dempster–Shafer theory[13-15].

In this paper, a new system on the basis of PCA and multiple classifier fusion is explored to be applied to rolling bearings fault diagnosis. First of all, after presentment, several statistic features both in time and frequency domain are exacted from the vibration signal of bearing. Then the PCA process is applied for the obtainment of a lower dimension feature vector with the most representative features. Next, the acquired feature vector is sent to several classifiers to obtain individual classification decisions. With the application of the individual

978-1-4244-8625-0/11/$26.00 ©2011 IEEE

results, the final decision of the fault diagnosis system is calculated by a multiple classifier decision algorithm on the basis of Bayesian belief method. The results of experiment show that the proposed system is effective and stable in bearing fault detection.

II. PRNICIPAL

A. Pretreatment and Feature Exaction

Measurement error in the process of data acquiring is inevitable because of the influence of environment and sensors. As a result, pretreatment is usually operated to reduce this influence. In this paper, Mean-Variance standardization[16] method is applied which transforms the original sample signals to be with zero means and unit variances.

Each sample of the vibration signal $y(y_1, y_2 \ldots y_n)$ is n points discrete data set with one-dimension. And the mean value of the sample date is :

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \tag{1}$$

Sample variance is calculated in the following equation :

$$S = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2, \tag{2}$$

Then the sample date after Mean-Variance standardization is given by the following equation :

$$y_i' = \frac{y_i - \bar{y}}{\sqrt{S}}, \tag{3}$$

After pretreatment, feature extraction is essential for estimating the conditions of observed system. Statistical parameters in time domain and frequency domain are generally used to represent average properties of acquired data. In this paper, eleven features in time domain are selected, which include absolute mean value, crest, root mean square, square root amplitude, skewness, kurtosis, shape factor, crest factor, pulse factor, clearance factor and kurtosis factor. And five features in frequency domain: mean energy, variance, crest, center frequency and kurtosis are calculated from the sample date. All the sixteen features exacted are calculated according to reference[1] and are listed in TABLE I.

TABLE I. STATISTIC FEATURES

Features of the vibration signal		
Time domain		***Frequency domain***
absolute mean	shape factor	mean energy
crest	crest factor	variance
root mean square	pulse factor	crest
square root	clearance factor	center frequency
skewness	kurtosis factor	kurtosis
kurtosis		

B. PCA Process

Though the more features selected from original data, the more information about working condition might be maintained, the computing cost in their calculation and in the following recognition process will obviously increase. In addition, some of the features exacted may be closely relevant to others that reduce the significance of large amount of computing. Fortunately, PCA provides a means to the identification of the most representative features and the decrease of the dimension of the feature space.

The N dimension matrix $X(x_1, x_2, \ldots, x_N)$ is the feature vector matrix consisting of the N features acquired in the feature exacting process and its column vectors are the d-dimension feature vectors. According to reference[3], the covariance matrix of X is calculated via the following equation :

$$R_X = \sum_{i=1}^{N}(x_i - \bar{x})(x_i - \bar{x})^T, \tag{4}$$

where N stands for the number of samples and $\bar{x}$ is the mean vector of each feature vector that is given by :

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i, \tag{5}$$

The eigenvalue of R_x can be calculated by the following equation :

$$\lambda v = R_x v, \tag{6}$$

where λ is the eigenvalue of R_x and v is the eigenvector of R_x. At most d eigenvalue $\lambda_i (i = 1, 2, \ldots, d)$ are calculated and $\lambda_1 > \lambda_2 > \ldots > \lambda_d$, corresponding to eigenvector $v_i (i = 1, 2, \ldots, d)$. Then the principal component on v_i can be obtained by projecting sample x_j to eigenvector v_i :

$$y_{ij} = v_i^T (x_j - \bar{x}), \tag{7}$$

After the PCA process, most information of the original signal is remained in the first several principal components. Therefore the properties of the original date can be approximately represented by the first $m (m < d)$ principal components.

C. Multiple Classifier Fusion(MCF)

1).Classifier selection based on correlation:

In the process of MCF, a proper method for classifier selection is crucial because combination of different classifiers can significantly influence the fusion accuracy. Therefore, which one of the classifiers to be chosen among many classifiers is an issue before a final fusion strategy is applied. In recent years, classifier selection technique is ongoing and most of the selection methods are based on statistic theory such as Q statistic, generalized diversity and agreement[17]. It has been detected that the dependency among classifiers can affect the fusion results. An effective method for classifier selection has been proposed by Goebel through the calculation of correlation degree of different classifiers[18].The correlation degree is given by:

$$\rho_n = \frac{nN^f}{N - N^f - N^r + nN^f}, \tag{8}$$

where n is the number of classifiers, N represents the total number of experiments samples, N^f is the number of samples that are misclassified by all classifiers and N^r means those samples which are correctly classified by all classifiers.

In general, lower ρ_n represents that the classifiers are more independent which contributes to a better classification performance. The classifier selection method on the basis of the principle of correlation measurement can be generalized as follows: First of all, the classifier with the best performance according to a proper measure such as accuracy ratio is selected as the first classifier of the team. Then the correlation degree between the first classifier and other classifiers are calculated using Eq.(8) and the classifier with lower correlation degree is chosen as the next one. The calculation of the correlation degree between the fixed classifiers and those in the waiting list is repeatedly conducted until all the classifiers are located. Then the optimal classifiers sequence is obtained.

2).Decision fusion based on Bayesian belief method

According to the output of a classifier, methods of classifiers fusion can be divided into three types: abstract type, rank type and measurement type[19]. Among them, the required information for a classification increases in sequence and the measurement style contains the most information that usually generate the best results. However, the probability evaluating how much the input subjects to each class is seldom available[17]. Therefore, the abstract type classifier fusion algorithms is widely investigated and applied. The main methods of the abstract type involve voting, Bayesian, and behavior-knowledge space(BKS). Voting may be the simplest and most popular one because voting strategies in this method are easier to be understood such as majority, unanimity and Borda count. However, consideration of different characters of each classifier is neglected that can decrease the accuracy of recognition ratio. And one of the significant limitations of BKS method is that it requires large numbers of training data which is always a problem in real application. In this paper, Bayesian belief method[20] is applied to the decision fusion process. In this method, belief measure of recognition is calculated for each classifier after the confusion matrix is established and the highest combined belief measure is chosen as the final decision.

For a pattern recognition problem with M classes $s_i (i = 1,...,M)$ and K classifiers $c_k (k = 1,...,K)$, the error of kth classifier can be represented by a two-dimensional confusion matrix $PT_k (k = 1,...,K)$ using Eq.(9):

$$PT_k = \begin{bmatrix} n_{11} & n_{12} & \cdots & n_{1M} \\ n_{21} & n_{22} & \cdots & n_{2M} \\ \vdots & \vdots & \vdots & \vdots \\ n_{M1} & n_{M2} & \cdots & n_{MM} \end{bmatrix}, \tag{9}$$

where the element n_{ij} illustrates the input samples from class s_i while assigned to class s_j by classifier c_k.

After acquiring the confusion matrix, belief measure of recognition of each classifier can be calculated by the belief function as follows:

$$Bel(x \in s_i / c_k(x)) = P(x \in s_i / c_k(x) = j_k), \tag{10}$$

where $i, j = 1,...,M$ and

$$P(x \in s_i / c_k(x) = j_k) = n_{ij}^{(k)} / \sum_{i=1}^{M} n_{ij}^{(k)}, \tag{11}$$

The final belief measure of the multiple classifier system is obtained by combining the belief measures of all fusion classifiers:

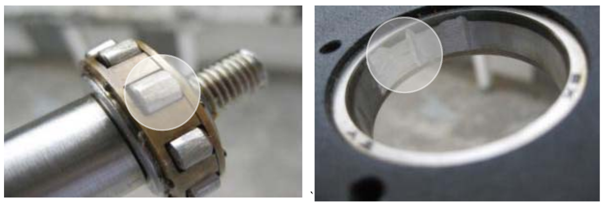
$$Bel(i) = P(x \in s_i) \frac{\prod_{k=1}^{K} P(x \in s_i / c_k(x) = j_k)}{\prod_{k=1}^{K} P(x \in s_i)}, \tag{12}$$

The highest combined belief measure $Bel(i)$ is chosen as the final classification decision.

III. EXPERIMENT RESULT AND DISCUSSION

The experiment was conducted on the YVS-2 vibration analysis platform shown in Fig.1. Vibration signal of bearing N203 with three working conditions was measured by this system, which include normal, rolling element wearing defect and outer race wearing defect. The defects of bearing are shown in Fig.2. The motor speed was set to be 1750rpm and the sampling frequency was 32768 Hz. The recording time was 10000 Ms. Waveforms of the collected vibration signal are shown in Fig.3. For each condition 100 samples were measured, and 60 of them were used for training parameters of the classifiers and the other 40 for the test.

Fig.1. YVS-2 System for bearings vibration signal collection



(b) Rolling element wearing defect (b) Outer-race wearing defect

Fig.2. Defect types of N203 bearing

After data acquisition, pretreatment and feature exacting processes were exerted. Sixteen features selected both in time and frequency domain were calculated after Mean-Variance standardization. These features are shown in Fig.4, where the abscissa indicates the number of samples in three working conditions with one hundred each. The PCA process was then utilized to reduce dimension of the feature space, and the first six principal components are shown in Fig.5.

Next, six commonly used classifiers were utilized for individual classification including Prazen classifier, quadratic discriminant classifier(QDC), Levenberg-Marquardt trained feed-forward neural net classifier(LMN), Gaussian mixture model(GMM), k nearest neighbors(k-NN) and supporting vector machine(SVM). The relevant parameters setup for these classifiers are listed in TABLE II.

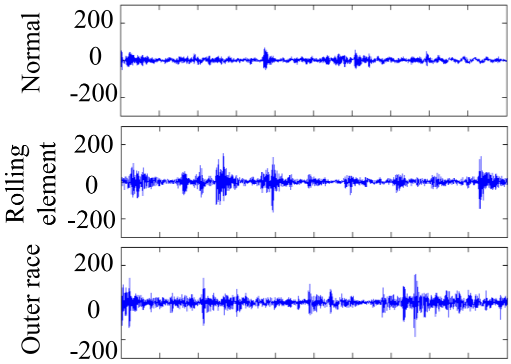


Fig.3. Waveforms of one group of vibration signal

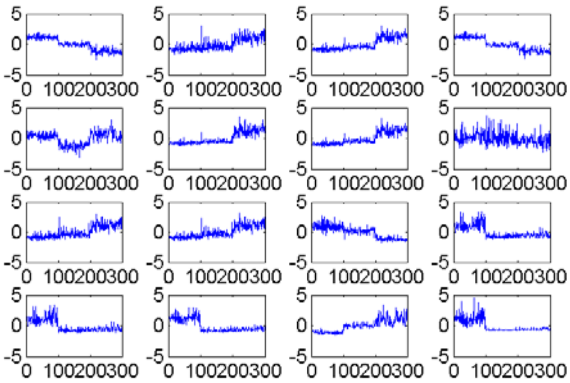


Fig.4. Sixteen features calculated in time and frequency domain

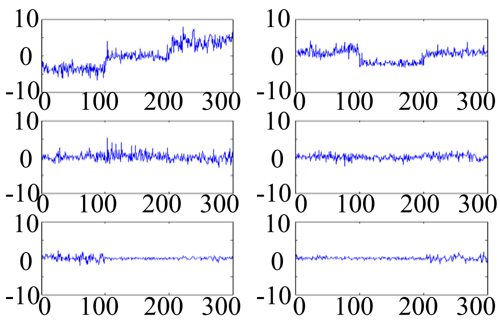


Fig.5. First six principal components

In this experiment, the classification accuracy is evaluated by the ratio of number of the samples classified correctly to the total samples. The training accuracy of the six classifiers is shown in Fig.6. It can be seen that k-NN classifier has the highest classification accuracy 0.933 followed by Prazen. QDC and LMN are of the same accuracy that is in the middle. The accuracy of GMM and SVM locates behind indicates that they are not as well performed as the former ones. However, in this experiment these suboptimum classifiers are still utilized because the classifiers are usually predetermined while the measured signals are probably of significant difference in real application. According to the individual accuracy, k-NN was selected as the best single classifier and determined as the first classifier in the optimal team of classifiers.

TABLE II. PARAMETER OF INDIVIDUAL CLASSIFIER

Classifier	***Parameter setup***
LMN	Number of neurons=5;Epochs=50
k-NN	k=3
SVM	Linear kernel function Euclidean distance type

TABLE III. RESULT OF OPTIMAL CLASSIFIER SEQUENCE

Number of Classifies	***Sequence of classifiers***
1	k-NN;
2	k-NN; GMM
3	k-NN; GMM; QDC
4	k-NN; GMM; QDC; Prazen
5	k-NN; GMM; QDC; Prazen; SVM
6	k-NN; GMM; QDC; Prazen; SVM; LMN

Then the decision fusion algorithm based on Bayesian belief method was utilized to enhance the diagnosis performance. The fusion system consists of two parts: classifier selection and classifier fusion. In the step of classifier selection, the correlation degree of different number of classifiers were calculated and the selection results is shown in TABLE III.

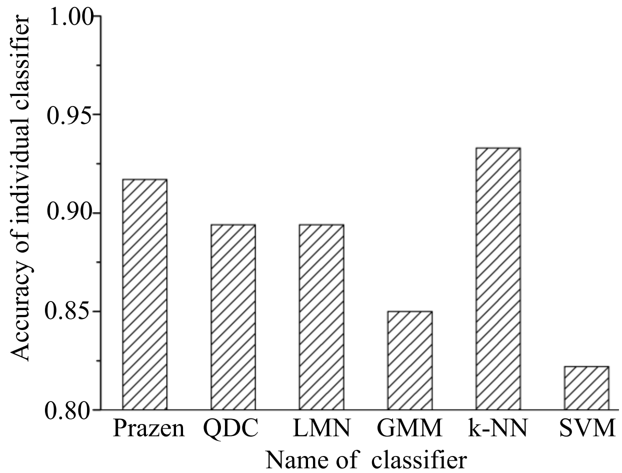


Fig.6. Individual classification accuracy

On the basis of the selected sequence of classifiers, the individual classification decisions were fused via Bayesian belief method. The result is shown in Fig.7.

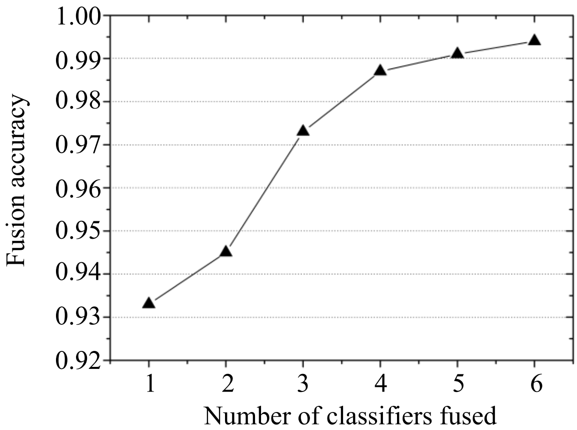


Fig.7. Fusion accuracy of multiple classifiers

As shown in Fig.7, fusion accuracy gradually increases with the augment of classifier number. And the speed of the enhancement in fusion accuracy is initially sharper and then turns to be more stable. When the number of classifiers is six, the highest fusion accuracy of this experiment is 0.994 which is higher than any of the individual classification accuracy. In addition, when singly utilized, even some of the classifiers selected in this experiment fail to perform as successfully as others do, they do contribute in the fusion process that demonstrates the significance of classifier selection and the excellent capacity of this fault diagnosis system with the application of multiple classifier fusion.

IV. CONCLUSION

In this paper, a fault diagnosis system on the basis of PCA in the feature exaction phase and MCF in the classification phase is proposed. After pretreatment, representative statistic features from time and frequency domain are calculated from the original vibration signal of bearing. The features obtained are then compressed and selected via PCA, outputting the most sensitive four principal components to the following classification phase. The reduce of the dimension of features saves the waste of computing in following steps and maintains sufficient information about bearing conditions for recognition. The classifiers are then utilized individually to recognize the working conditions on the basis of the principal components and the final decision is generated by a multiple classifier fusion method based on Bayesian belief. The efficiency of this proposed system was confirmed by the experiment of the recognition of the three defect type of bearing.

REFERENCES

[1] K. X. Cheng and Q. C. LI, Machine condition monitoring and fault diagnosis technology. Beijing: Beijing Science and Technology, 1991.

[2] A. Malhi and R. X. Gao, "PCA-based feature selection scheme for machine defect classification," Ieee Transactions on Instrumentation and Measurement, vol. 53, pp. 1517-1525, 2004.

[3] Q. B. He, Z. H. Feng, and F. R. Kong, "Principal component representations for machine noise monitoring," Guangxue Jingmi Gongcheng/Optics and Precision Engineering, vol. 14, pp. 1093-1099, 2006.

[4] M. Gonzalez-Audicana, J. L. Saleta, R. G. Catalan, and R. Garcia, "Fusion of multispectral and panchromatic images using improved IHS and PCA mergers based on wavelet decomposition," Ieee Transactions on Geoscience and Remote Sensing, vol. 42, pp. 1291-1299, 2004.

[5] I. Y. Tumer and E. M. Huff, "Analysis of triaxial vibration data for health monitoring of helicopter gearboxes," Journal of Vibration and Acoustics-Transactions of the Asme, vol. 125, pp. 120-128, 2003.

[6] B. L. Zhang, M. Y. Fu, and H. Yan, "A nonlinear neural network model of mixture of local principal component analysis: application to handwritten digits recognition," Pattern Recognition, vol. 34, pp. 203-214, 2001.

[7] T. Xu, X. Pei, and X. Wei, "The study of expert system in rolling bearing faults diagnosis," 2010 International Conference on Educational and Network Technology (ICENT 2010), pp. 338-41, 2010.

[8] V. Crupi, E. Guglielmino, and G. Milazzo, "Neural-network-based system for novel fault detection in rotating machinery," Journal of Vibration and Control, vol. 10, pp. 1137-1150, 2004.

[9] A. Widodo and B. S. Yang, "Support vector machine in machine condition monitoring and fault diagnosis," Mechanical Systems and Signal Processing, vol. 21, pp. 2560-2574, 2007.

[10] T. Marwala, U. Mahola, and F. V. Nelwamondo, "Hidden Markov models and Gaussian mixture models for bearing fault detection using fractals," 2006 International Joint Conference on Neural Networks, pp. 3237-42, 2007.

[11] P. Jayaswal, S. N. Verma, and A. K. Wadhwani, "Application of ANN, Fuzzy Logic and Wavelet Transform in machine fault diagnosis using vibration signal analysis," Journal of Quality in Maintenance Engineering, pp. 190-213, 2010.

[12] L. O. Jimenez, A. Morales-Morell, and A. Creus, "Classification of hyperdimensional data based on feature and decision fusion approaches using projection pursuit, majority voting, and neural networks," Ieee Transactions on Geoscience and Remote Sensing, vol. 37, pp. 1360-1366, 1999.

[13] N. Perrot, G. Trystram, D. LeGuennec, and F. Guely, "Sensor fusion for real time quality evaluation of biscuit during baking. Comparison between Bayesian and fuzzy approaches," Journal of Food Engineering, vol. 29, pp. 301-315, 1996.

[14] A. Raudys and F. Roli, "The Behavior Knowledge Space fusion method: Analysis of generalization error and strategies for performance improvement," Multiple Classifier Systems, Proceeding, vol. 2709, pp. 55-64, 2003.

[15] S. Challa and D. Koks, "Bayesian and Dempster-Shafer fusion," Sadhana-Academy Proceedings in Engineering Sciences, vol. 29, pp. 145-174, Apr 2004.

[16] P. Li, F. R. Kong, and L. Dang, "A New Fault Diagnosis System for Train Bearings Based On PCA and ACO," Proceedings of 2010 International Conference on Logistics Systems and Intelligent Management, Vols 1-3, pp. 526-530, 2010.

[17] G. Niu, T. Han, B. S. Yang, and A. C. C. Tan, "Multi-agent decision fusion for motor fault diagnosis," Mechanical Systems and Signal Processing, vol. 21, pp. 1285-1299, 2007.

[18] K. Goebel and W. Yan, "Choosing classifiers for decision fusion," in Proceedings of the Seventh International Conference on Information Fusion, FUSION 2004, June 28, 2004 - July 1, 2004, Stockholm, Sweden, 2004, pp. 563-568.

[19] L. Xu, A. Krzyzak, and C. Y. Suen, "METHODS OF COMBINING MULTIPLE CLASSIFIERS AND THEIR APPLICATIONS TO HANDWRITING RECOGNITION," Ieee Transactions on Systems Man and Cybernetics, vol. 22, pp. 418-435, May-Jun 1992.

[20] L. Lam and C. Y. Suen, "Optimal combinations of pattern classifiers," Pattern Recognition Letters, vol. 16, pp. 945-954, 1995.