An intelligent fault diagnosis method of rolling bearings based on regularized kernel Marginal Fisher analysis

Li Jiang\textsuperscript{1,2}, Tielin Shi\textsuperscript{1,2,3} and Jianping Xuan\textsuperscript{1,2}

\textsuperscript{1} School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, Peoples R China.
\textsuperscript{2} State Key Laboratory for Digital Manufacturing Equipment and Technology, Huazhong University of Science and Technology, Wuhan 430074, Peoples R China.
\textsuperscript{3} Corresponding author (T. L. Shi). Tel/fax: +86 2787556835.

Abstract. Generally, the vibration signals of fault bearings are non-stationary and highly nonlinear under complicated operating conditions. Thus, it’s a big challenge to extract optimal features for improving classification and simultaneously decreasing feature dimension. Kernel Marginal Fisher analysis (KMFA) is a novel supervised manifold learning algorithm for feature extraction and dimensionality reduction. In order to avoid the small sample size problem in KMFA, we propose regularized KMFA (RKMFA). A simple and efficient intelligent fault diagnosis method based on RKMFA is put forward and applied to fault recognition of rolling bearings. So as to directly excavate nonlinear features from the original high-dimensional vibration signals, RKMFA constructs two graphs describing the intra-class compactness and the inter-class separability, by combining traditional manifold learning algorithm with fisher criteria. Therefore, the optimal low-dimensional features are obtained for better classification and finally fed into the simplest K-nearest neighbor (KNN) classifier to recognize different fault categories of bearings. The experimental results demonstrate that the proposed approach improves the fault classification performance and outperforms the other conventional approaches.

1. Introduction

Within the last decade, rotating machinery in modern industry has been moving toward high speed, super precision and automation. Their structure becomes more complicated and potential faults are more difficult to detect, especially under multiple operating conditions [1]. In rotating machinery, the root cause of faults is often damaged rolling bearings [2]. Hence, it’s a big challenge to detect bearing faults and decrease possible production loss. Although it’s sufficient to recognize faults by the visual inspection of vital diagnostic information from the measured signals in some cases, many signal processing techniques available today require plenty of expertise to apply them successfully. Thus, there is a demand for intelligent techniques to allow relatively unskilled operators making reliable decisions without the requirement of a diagnosis specialist to examine data and diagnosis problems [3].
The essence of intelligent fault diagnosis is pattern recognition. Feature extraction plays an important part in pattern recognition. In most cases, the features of complexity equipments are high-dimensional to gain comprehensive fault information. Thus, it’s worthwhile exploring how to extract optimal features for improving classification performance and simultaneously decreasing feature dimension. Generally, the vibration signals of fault bearings are non-stationary and highly nonlinear. However, some classical feature extraction techniques for dimensionality reduction, such as Principal component analysis (PCA) [4] and linear discriminant analysis (LDA) [5], are linear, which fail to fully capture nonlinear characteristics of fault bearings. Furthermore, LDA has limitation owing to the Gaussian distribution assumption of the dataset.

Since three papers [6-8] were published in science, manifold learning has become a hot point. Based on the assumption that the original high-dimensional data lies on a low-dimensional manifold, manifold learning aims to discover the intrinsic structure of nonlinear data and simultaneously project the high-dimensional data in ambient space into a low-dimensional feature space. Kernel Marginal Fisher analysis (KMFA) [9] is one of such typical manifold learning techniques, which has successfully applied to face recognition and human gait recognition. It not only preserves intrinsic manifold structure of the raw data, but also utilizes class label information together with the optimal margin Fisher criteria. However, KMFA tend to perform poorly because of over-fitting when labeled data is inadequate and the dimension of data is too high. To avoid small sample size problem in KMFA, we propose regularized KMFA (RKMFA), which introduces a regularization term to KMFA. Based on RKMFA, an intelligent fault diagnosis method is put forward introducing KMFA to mechanical fault diagnosis field.

The rest of this paper is organized as follows. Section 2 briefly introduces RKMFA algorithm. In section 3, the fault diagnosis procedure based on RKMFA algorithm is proposed. The proposed method is applied to the detection of rolling bearings faults in Section 4. Finally, the conclusions are drawn in Section 5.

2. Regularized Kernel Marginal Fisher Analysis

Based on the graph embedding framework, KMFA is developed to extract low-dimensional features from high-dimensional dataset [9]. Given a dataset $X = \{x_1, x_2, ..., x_N\}, x_i \in \mathbb{R}^D$, where $N$ denotes the total sample number and $D$ is the feature dimension. The class label of the feature sample $x_i$ is $l(x_i) \in \{1, 2, ..., N_c\}$, where $N_c$ is the number of classes.

2.1. Graph embedding framework

Let $G = \{X, S\}$ be an undirected weighted graph, called intrinsic graph, with vertex set $X$ and similarity matrix $S \in \mathbb{R}^{N \times N}$. Each element of the symmetric matrix $S$ measures its similarity relationship between vertex pairs. The diagonal matrix $D$ and the Laplacian matrix $L$ of the graph $G$ are defined as

$$D_{ii} = \sum_{j \neq i} S_{ij}, L = D - S, \forall i$$

Suppose $G_p = \{X, S^p\}$ is a penalty graph, whose vertex set $X$ is the same as that of $G$. However, the similarity matrix $S^p$ of $G_p$ represents the suppressed similarity characteristics.

The low-dimensional representations of the vertices are denoted by a vector $Y = [y_1, y_2, ..., y_N], y_i \in \mathbb{R}^d (d << D)$. The similarity preserving criterion of a graph is represented as

$$Y^* = \arg \min_{Y \in \mathbb{R}^{N \times d}} \sum_{i,j} \left\| y_i - y_j \right\|^2 \mathcal{S}_{ij} = \arg \min_{Y \in \mathbb{R}^{N \times d}} Y^tL^tLY$$

where $c$ is a constant and $B$ is a constraint matrix. Generally, $B$ is a diagonal matrix for scale normalization and may also be the Laplacian matrix of the penalty graph $G_p$.

The linear technique may fail to discover the intrinsic geometry when the data manifold is highly nonlinear. So, a natural way is to apply the kernel trick to the graph embedding in order to obtain the
nonlinear embedding. Assuming the original data space $\mathbb{R}^D$ is mapped into a reproducing kernel Hilbert space (RKHS) $\mathbf{H}$ through a kernel function $\phi$. Let $\phi(\mathbf{X})$ denote the data matrix in RKHS, $\phi(\mathbf{X}) = [\phi(x_1), \phi(x_2), ..., \phi(x_N)]$. Define the kernel function $k(x,y) = \phi(x) \cdot \phi(y)$, and the kernel Gram matrix is $\mathbf{K}$ with $K_{ij} = k(x_i, x_j)$. We seek a transformation matrix $\mathbf{W}_\phi$, so the kernelization of graph embedding is expressed as

$$\mathbf{W}^*_\phi = \arg \min_{\mathbf{W}_{\phi}} \sum_{i \neq j} \left( \mathbf{W}^T_{\phi} \phi(x_i) - \mathbf{W}^T_{\phi} \phi(x_j) \right)^2 S^*_{ij}$$

where the weight $S^*_{ij}$ represents similarity relationship between $\phi(x_i)$ and $\phi(x_j)$.

By reproducing kernel theory, the solution to the optimization problem in equation (3) is a linear combination of $\phi(x_i)$ [10]. Accordingly, there exists a vector $\mathbf{a} = [a_1, a_2, ..., a_N]^T$ satisfying

$$\mathbf{W}_\phi = \sum_{i=1}^{N} a_i \phi(x_i) = \phi(\mathbf{X}) \mathbf{a}$$

2.2. Kernel Marginal Fisher analysis

The specific procedure of KMFA algorithm can be described as follows:

Step 1: Constructing two neighborhood graphs.

We construct two graphs defined on local neighborhood according to the adjacency relationship of data points in the nonlinear feature space. The one is intrinsic graph reflecting the intra-class compactness and connecting each data point with its neighboring points in the same class. The other is penalty graph illustrating the inter-class separability and connecting the marginal points in different classes.

From intrinsic graph, the intra-class compactness is described by

$$\mathbf{S}_c = \sum_{i} \sum_{(i \in N_k(i) \ or \ j \in N_k(i))} \left( \mathbf{W}_\phi^T \phi(x_i) - \mathbf{W}_\phi^T \phi(x_j) \right)^2 S^*_{ij} = 2\mathbf{a}^T \mathbf{K} \left( \mathbf{D}^* - \mathbf{S}^* \right) \mathbf{K} \mathbf{a}$$

where, $N_k(i)$ denotes the index set of $k$ nearest neighbors of the sample $\phi(x_i)$ in the same class, the diagonal matrix $\mathbf{D}^* = (D^*_{ij})_{N \times N} = \sum_{i \neq j} S^*_{ij}$ and the entries in the similarity matrix $\mathbf{S}^* \in \mathbb{R}^{N \times N}$ are defined as

$$S^*_{ij} = \begin{cases} 1, & \text{if } i \in N_k(i) \text{ or } j \in N_k(i) \\ 0, & \text{else} \end{cases}$$

From penalty graph, the inter-class separability is depicted by

$$\mathbf{S}_p = \sum_{i} \sum_{(i,j) \in P_{k}(\phi(x_i)) \ or \ (i,j) \in P_{k}(\phi(x_j))} \left( \mathbf{W}_\phi^T \phi(x_i) - \mathbf{W}_\phi^T \phi(x_j) \right)^2 S^*_{ij} = 2\mathbf{a}^T \mathbf{K} \left( \mathbf{D}^* - \mathbf{S}^* \right) \mathbf{K} \mathbf{a}$$

where, $P_{k2}(c)$ denotes the index set of data pairs which are $k$ nearest vertex pairs among the set $\{(i,j)\ | \ d(x_i) \neq c \}$, the diagonal matrix $\mathbf{D}^p = (D^p_{ij})_{N \times N} = \sum_{i \neq j} S^p_{ij}$ and the similarity matrix $\mathbf{S}^p \in \mathbb{R}^{N \times N}$ with elements

$$S^p_{ij} = \begin{cases} 1, & \text{if } (i,j) \in P_{k2}(\phi(x_i)) \text{ or } (i,j) \in P_{k2}(\phi(x_j)) \\ 0, & \text{else} \end{cases}$$

Step 2: Finding the optimal mapping direction $\mathbf{a}^*$ with Marginal Fisher Criterion:
2.3. Regularized kernel Marginal Fisher analysis

The optimal problem in equation (9) comes down to the maximum eigenvalue solution to the generalized eigenvalue decomposition problem:

$$\mathbf{K}(\mathbf{D}^p - \mathbf{S}^p)\mathbf{a} = \lambda \mathbf{K}(\mathbf{D}^c - \mathbf{S}^c)\mathbf{a}$$

(10)

To get a stable solution of the eigenvalue problem in equation (10), the matrix $\mathbf{K}(\mathbf{D}^c - \mathbf{S}^c)$ is required to be non-singular. However, it might be singular when the sample size is relatively small. To avoid the small sample size problem, the common solution is first projecting the original dataset to a PCA subspace before feature extraction [11]. However, this unsupervised preprocessing method may lose some useful discriminant information [12]. Motivated by [13], we incorporate the manifold structure of the samples as the regularization term, so as to preserve the local geometric structure of the data manifold. Thereby, the other regularized version of equation (9) is described as

$$\mathbf{a}^* = \operatorname*{arg\,max}_{\mathbf{a}} \frac{\mathbf{a}^T \mathbf{K}(\mathbf{D}^p - \mathbf{S}^p)\mathbf{a}}{(1 - \beta)\mathbf{a}^T \mathbf{K}(\mathbf{D}^c - \mathbf{S}^c)\mathbf{a} + \beta J(\mathbf{a})}$$

(11)

where $0 \leq \beta \leq 1$ controls the smoothness of estimator.

The regularization term $J(\mathbf{a})$ is defined as

$$J(\mathbf{a}) = \frac{1}{2} \sum_{i \neq j} \left\| y_i - y_j \right\|^2 \mathbf{S}_{ij} = \mathbf{a}^T \mathbf{K}(\mathbf{D} - \mathbf{S})\mathbf{a} = \mathbf{a}^T \mathbf{KLK}\mathbf{a}$$

(12)

where $\mathbf{L}$ and $\mathbf{D}$ is defined in equation (1), and $\mathbf{S}_{ij} = 1$, if vertex $i$ is the nearest neighbor of vertex $j$ or vertex $j$ is the nearest neighbor of vertex $i$, otherwise $\mathbf{S}_{ij} = 0$.

3. The proposed fault diagnosis method based on RKMFA

Based on the assumption that different fault signals of rolling bearings have different manifold structures [14]. Therefore, the fault classification of rolling bearings is multiple manifolds learning problem in principle. The fault data in the same class resides on a sub-manifold and the defective samples indifferent classes are distributed on different sub-manifolds. Based on RKMFA, we propose a new approach to intelligent fault diagnosis of rolling bearings. The method of fault diagnosis based on RKMFA is stated as follows:

Step 1: Collecting data from monitoring equipments and acquiring high-dimensional signal samples generated from the raw vibration signals.

Step 2: Normalizing each signal sample to zero mean and unit variance, and constructing a high-dimensional pattern space.

Figure 1. The flow chart of the proposed fault diagnosis procedure.
Step 3: Implementing feature extraction on the original pattern space with RKMFA and exploiting their intrinsic manifold structures for different sub-manifolds, by constructing two graphs defined on local neighborhood. Thus, the high-dimensional signal samples are projected into a low-dimensional feature space, and simultaneously obtain the optimal mapping direction.

Step 4: Carrying out pattern classification on the low-dimensional features with the simplest K-nearest neighbor (KNN) classifier, and finally obtaining diagnosis results.

The implementation process of the proposed strategy is illustrated in figure 1.

4. Application experiments and results

In order to show how the above procedure works and validate the effectiveness of the new approach, an experiment study of bearing fault diagnosis is put forward.

4.1. Data acquisition

The raw vibration signals of fault bearings were obtained from the bearing data center [15]. The bearing data has been validated in many researches [2] and become a standard data set of rolling bearings. The bearings were installed in a motor driven mechanical system. An accelerometer is mounted on the motor housing. The vibration signals were collected at 12000 samples per second by using a 16 channel DAT recorder. The test bearings are deep groove ball bearings whose type is 6205-2RS JEM SKF. Single point faults were introduced to the test bearings using electro-discharge machining. The fault diameter is 0.021 inches (1inch=25.4 mm) and fault depth is 0.011 inches. Four datasets (inner race fault, ball fault, outer race fault and normal condition) are obtained under 3 hp loads with the motor speed of 1730rpm. Each signal sample was acquired with 1024 data points. So the feature dimension of each signal sample in pattern space is 1024.

4.2. Experiments and results analysis

4.2.1. Classification accuracy comparison of different fault diagnosis techniques. In order to exhibit the superiority of fault diagnosis based on RKMFA, we also perform experiment on KMFA. In the comparative experiment, 70 signal samples per class are selected randomly as the training set, and the training set consists of 20 signal samples per class. For intuitional display, the first two principal components of the mapping results are plotted after feature extraction with these methods. It can be observed that from figure 2, KMFA couldn’t recognize ball fault and outer race fault for the training dataset. Figure 3 shows that RKMFA obtains a fairly good and clear separation of the mapping results on the training dataset as well as the testing dataset for different fault categories.

Figure 2. The mapping results with KMFA: (a) the training dataset, (b) the testing dataset.
To compare the classification results more objectively, we feed the low-dimensional feature vectors into KNN classifier as the final evaluation criteria. Here, the leave-one-out procedure is used for the proper selection of parameter $K$. In the following experiments, if not explicitly stated, the classification is based on 1-nearest neighbor classifier with Euclidean distance. The digital experiment results are shown in table 1. It can be seen that RKMFA has the highest classification accuracy compared with KMFA.

Table 1. The classification accuracy (%) of each method with KNN classifier.

| Fault type          | Inner race fault | Ball fault | Outer race fault | Normal |
|---------------------|------------------|------------|------------------|--------|
| KMFA                | 100              | 100        | 100              | 55     |
| RKMFA               | 100              | 100        | 100              | 100    |

From figure 2, figure 3 and table 1, it proves that the method of fault diagnosis based on RKMFA excavates the underlying fault information embedded in the high-dimensional nonlinear data and successfully distinguishes different fault categories of bearings.

4.2.2. Classification accuracy comparison for different training sample sizes. For assessing the effect of different training sample sizes upon classification performance based on RKMFA, the experiments are also conducted with KMFA, PCA and LDA. We vary the training sample size when the testing sample size per class is 20.

Table 2. The classification accuracy (%) with KNN as a function of training sample size per class.

| Training sample size | 10   | 20   | 40   | 50   | 70   | 90   |
|----------------------|------|------|------|------|------|------|
| PCA                  | 86.25| 85.00| 86.25| 90.00| 92.50| 93.75|
| LDA                  | 81.25| 82.50| 82.50| 83.75| 87.50| 90.00|
| KMFA                 | 77.50| 83.75| 83.75| 85.00| 88.75| 93.75|
| RKMFA                | 100.00| 100.00| 100.00| 100.00| 100.00| 100.00|

Table 2 reveals that the more training samples are involved, the higher the classification accuracy is. However, when the training sample size per class is smaller than 40, the classification accuracy of the other three methods is much lower than that of RKMFA. By comparison, the classification rate of
RKMFA has the least fluctuation for different training sample sizes. Therefore, RKMFA can achieve the highest recognition rate in case of small sample sizes.

5. Conclusions

A new feature extraction method called regularized Kernel Marginal Fisher analysis (RKMFA) is first proposed. It’s computational efficiently due to turning complicated procedure of feature extraction into simple generalized eigenvalue decomposition problem. An intelligent approach to fault diagnosis based on RKMFA is presented, for the sake of removing the redundant information and acquiring optimal low-dimensional features for better classification. The proposed method applies to highly nonlinear faulty bearings and effectively extracts low-dimensional nonlinear features embedded in the original high-dimensional pattern space. Its application to fault classification of rolling bearings demonstrates its excellent fault classification performance and superiority over the other traditional approaches, even for small sample size. Therefore, it’s a powerful method for fault classification of rolling bearings.

Acknowledgements

This research is supported by the National Natural Science Foundation of China under Grant No. 51075161. The authors also thank the free download of the rolling element bearing fault datasets provided by Case Western Reserve University, Rockwell Science Office of Naval Research and CVX.

References

[1] Lei Y G, He Z J, Zi Y Y and Chen X F 2008 New clustering algorithm-based fault diagnosis using compensation distance evaluation technique Mech. Syst. Signal Process. 22 419–35
[2] Yang H Y, Mathew J and Ma L 2005 Fault diagnosis of rolling element bearings using basis pursuit Mech. Syst. Signal Process. 19 341–56
[3] Lei Y G, He Z J, Zi Y Y and Hu Q 2007 Fault diagnosis of rotating machinery based on multiple ANFIS combination with GAS Mech. Syst. Signal Process. 21 2280–94
[4] Jolliffe I T 1986 Principal component analysis (Berlin: Springer)
[5] Fukunaga K 1990 Introduction to statistical pattern recognition 2nd ed. (New York: Academic Press)
[6] Roweis S T and Saul L K 2000 Nonlinear dimensionality reduction by locally linear embedding, Science 290 2323–6
[7] Seung H S and Lee D D 2000 The manifold ways of perception Science 290 2268–9
[8] Tenenbaum J B, Silva V de and Langford J C 2000 A global geometric framework for nonlinear dimensionality reduction Science 290 2319–23
[9] Yan S C, Xu D, Zhang B Y, Zhang H J, Yang Q and Lin S 2007 Graph embedding and extensions: A general framework for dimensionality reduction IEEE Trans. Pattern Anal. Mach. Intell. 29 40–51
[10] Smola A J and Scholkopf B 2002 Learning with kernels (Boston: MIT Press)
[11] He X F, Yan S C, Hu Y X, Niyogi P and Zhang H J 2005 Face recognition using Laplacianfaces IEEE Trans. Pattern Anal. Mach. Intell. 27 328–40
[12] Dai D Q and Yuen P C 2003 Regularized discriminant analysis and its application to face recognition Pattern Recogn. 36 845–7
[13] Cai D, He X F and Han J W 2007 Semi-supervised discriminant analysis IEEE Int. Conf. on computer vision. (Rio de Janeiro, Brazil, 14–21 October 2007) pp 222–8
[14] Li M, Xu J W, Yang J H, Yang D B and Wang D D 2009 Multiple manifolds analysis and its application to fault diagnosis Mech. Syst. Signal Process. 23 2500–9
[15] Loparo K 2011 Bearings vibration data set Case Western Reserve University