Metric Learning Based Rolling Bearing Faults Diagnosis with Curvelet Transform

Ziming Lu¹,²,*, Jiang-Wen Xiao¹,² and Zhengyi Huang¹,²

¹School of Automation, Huazhong University of Science and Technology, Wuhan 430074, China
²Key Laboratory of Image Processing and Intelligent Control (Huazhong University of Science and Technology), Ministry of Education, China

*E-mail: m201672407@hust.edu.cn

Abstract. Rolling bearing faults are among the primary causes of breakdown in mechanical equipment. Aiming at the vibration signals of rolling bearing which are non-stationary and easy to be disturbed by noise, a novel fault diagnosis method based on curvelet transform and metric learning is proposed. This method consists of 3 parts. The first one is feature engineering which includes reshaping the original timing features of rolling bearings, employing curvelet transform to transform reshaped features and making its coefficients as the new features. Curvelet transform can analyse the original signal from many angles. The second one is employing metric learning to map these new features into special embedding space. The last one is applying KNN classifier to detect the rolling bearing faults. Metric learning can effectively improve the performance of KNN by learning a mapping matrix to modify the distribution of samples. The proposed method overcomes the problems such as the subjectivity and blindness of manual feature extraction, poor coupling in each stage and sensitive to the effect of noise. Extensive simulations based on several data-sets show that the our method has better performance on bearing fault diagnosis than traditional methods.

1. Introduction

As an important part of rotating machinery, rolling bearings are widely used in many industrial systems. Rolling bearings have become one of the mechanical parts which runs with a high malfunction rate for its high-speed and high-load working condition [1]. So, it’s extremely important to take measures to prevent the sudden shutdown of equipment caused by rolling bearings from creating huge loss to enterprises.

Fault diagnosis of rotating machinery equipment can be achieved by analysing and processing various information obtained from the operation state of the equipment [2]. Vibration signal is relatively easy to obtain, and contains the main information of equipment operation status. Therefore, vibration signal analysis is one of the most widely used methods in rolling bearing fault diagnosis at home and abroad. With the rapid development of science and technology, bearing fault diagnosis has entered the era of “big data” [3]. Intelligent algorithms, which aim at analysing a large amount of data and automatically providing diagnosis results, have become a new trend in the field of equipment monitoring. According to the references [4-6], the traditional intelligent diagnosis algorithm mainly includes three steps: (1) feature extraction. (2) Feature selection. (3) Pattern recognition. For example, Keskes et al. used wavelet packet technology and support vector machine to diagnose to diagnose
rotor fault of induction motor [7]. Qu et al. used dual-tree complex wavelet packet to overcome the frequency aliasing phenomenon caused by wavelet transform. Also, combined with multi-classifier, the rolling bearings and gears are diagnosed and the result is relatively good [8]. [9] adopts empirical mode decomposition to extract features of rolling bearings under normal and fault conditions, and combining fuzzy C-means clustering to diagnose faults. [10] proves that the features extracted by variational mode decomposition have better noise robustness and good sampling effect than empirical mode decomposition. [11] combined with time scale decomposition method and Tikhonov support vector machine, the results show that it can effectively diagnose bearing faults. [12] uses Hilbert-Huang to analyze vibration signals of rolling bearings, and extracts features by Fourier transform which are used as input of neural network to realize fault diagnosis.

The methods mentioned above have achieved good results, but they still have common inherent shortcomings: (1) The vibration signals collected are often unstable and contain a lot of environmental noise. Without sufficient prior knowledge and engineering experience, manual feature extraction is usually blind, subjective and time-consuming. Moreover, the diagnostic performance of the model is highly dependent on the sensitive features extracted from the data set. (2) Fault diagnosis is divided into three independent stages: signal preprocessing, feature extraction and fault classification, which destroys the coupling relationship of each stage. The matching degree between the extracted features and the applied pattern recognition algorithm is difficult to evaluate, and they are not an organic combination [13]. This results in the loss of some fault information of rolling bearing. In recent years, metric learning has become one of the most active research topics in the fields of computer vision and pattern recognition. This paper uses metric learning to solve the problems discussed above. Its main idea is to use the sample information of the training data set to learn a mapping matrix which can effectively reflect the relationship in the sample space. In the new feature space based on the metric matrix, the same sample distribution is more compact and the different samples are more dispersed. Metric learning is essentially an accurate measurement of sample similarity. In theory, every machine learning can be included in the measurement learning link to improve its performance. Recent studies have shown that learning distance metrics from training data is more effective than extracting features manually [14].

At present, there are few scenarios of applying metric learning to mechanical component fault diagnosis at home and abroad. Based on the classical large margin nearest neighbour algorithm in supervised metric learning, this work utilizes curved wave analysis and achieves the fault diagnosis of rolling bearings. Through the experiments of different fault types and the comparative analysis of the diagnosis results of traditional algorithms, the results indicate that the bearing fault diagnosis based on metric learning has important theoretical and applied values.

2. Related work

2.1. Characteristic and principle of curve wave transform

The second generation curvelet transform (Fast Discrete Curvelet Transform) proposed by Candès et al. in 2002 is a technology developed in recent years. And it is quite favored due to its high computational efficiency, multi-scale characteristics, local filtering function and strong sparse expression ability. curvelet transform, combining the anisotropic characteristics of ridgelet transform and the multi-scale characteristics of wavelet transform, can suppress random noise more effectively and protect the effective signal, so as to achieve better denoising effect. The features extracted in this way can facilitate the latter model being used adaptive learning the strong features.

Fast Discrete Curvelet Transform (FDCT) is proposed by reference [15], which makes curvelet transform easier to use and understand, speeds up the transformation and reduces redundancy. Based on FDCT, this paper establishes the feature engineering of bearing fault diagnosis. The formulas for calculating the curved wave coefficients are as follows:
\[ e^D(j,l,k) = \sum_{n_1, n_2 \in D_j} \hat{f}[n_1, n_2 - n_1 \tan \theta_j] \hat{U}_j[n_1, n_2] e^{2\pi kn_1/L_{1,j} + kn_2/L_{2,j}} \]  \tag{1}

where \( j \) is the scale; \( l \) is the angle; \( L_{1,j} \) and \( L_{2,j} \) are the edge lengths of the parallelogram \( P_{j,l} \);

\[ \hat{f}[n_1, n_2 - n_1 \tan \theta_j] = \hat{f}(2\pi n_1, 2\pi (n_2 - n_1 \tan \theta_j)). \]

According to the formulas above, the steps of the algorithm of FDCT are presented as follows.

1. Getting Fourier array \( \hat{f}[n_1, n_2] \) by carrying Fourier transform on \( f[t_1, t_2] \), \(-n/2 \leq n_1, n_2 \leq n/2\)
2. For any scale \( j \) and angle \( l \), insert \( \hat{f}[n_1, n_2] \) to obtain \( \hat{f}[n_1, n_2 - \tan \theta_j], (n_1, n_2) \in P_j \)
3. Multiply the interpolated object \( \hat{f} \) by the parabolic window \( \hat{U}_j \) to get \( \hat{f}_{j,l}[n_1, n_2] \).
4. Get Curve wave coefficient by doing 2DFFT for each \( \hat{f}_{j,l} \).

The time complexity of the algorithm is \( O(n^2 \log n) \) and the spatial complexity is \( O(n^2) \).

### 2.2. Distance metric learning algorithm based on maximizing classification margin

A good metric learning algorithm ensures not only a high degree of consistency of sample classes in the neighbour, but also makes the edges between classes have a large gap. This idea of adding large intervals in metric learning fits well with support vector machine (SVM). Therefore, metric learning based on SVM emerges, and the relatively mature theory of support vector machine (SVM) is applied to solve the semi-definite programming problem in metric learning. A typical example is the classical Large Margin Nearest Neighbour (LMNN) algorithm.

LMNN is proposed by [16]. The main idea of this algorithm is to improve the performance of K-nearest neighbour algorithm by maximizing the marginal distance of different classes. Moreover, the algorithm does not need to assume that the data need to obey a certain distribution. In order to achieve the goal of the algorithm, a special three-dimensional loss function is constructed. On the one hand, if a sample is too far to the same sample, it will be punished; on the other hand, if a sample is too close to the same sample, it will also be punished. The algorithm is as follows:

\[
\{(x_i, y_i)\}_{i=1}^{n} \text{ represents a training set consisting of } n \text{ samples, in which input } x_i \text{ corresponds to the category label } y_i. \text{ The scalar } y_{ij} \in \{0,1\} \text{ is used to denote whether the class labels } y_i \text{ and } y_j \text{ match, and a linear transformation } L: R^d \rightarrow R^d \text{ is usually needed to measure the following distance:}
\]

\[
d_M(x_i, x_j) = (x_i, x_j)^T M (x_i - x_j) = \|L(x_i - x_j)\|^2, \tag{2}
\]

where \( M = I^T L I \).

For any input sample \( x_i \), its target neighbours are defined as an input sample that meets the following conditions: 1) The label \( y_i \) which has the same category as \( x_i \); 2) The input sample which has a minimum distance with \( x_i \) is calculated by formula (2). For the sample \( x_i \), \( k \) target neighbours can be specified. In the absence of prior knowledge, Euclidean distance is usually used to calculate \( k \) nearest neighbours. By using scalar \( \eta_{ij} \in \{0,1\}, y_{ij} \in \{0,1\}, \text{ and } x_j \) is the target neighbour of \( x_i \) when \( \eta_{ij} = 1, x_i \) and \( x_j \) belong to the same category when \( y_{ij} = 1. \) Binary groups \( y_{ij} \) and \( \eta_{ij} \) are fixed, that is, they will not change during the learning process.
According to the idea of maximizing classification margin, we can get the objective function as follows:

\[ \varepsilon(L) = \sum_{ij} \eta_{ij} \left[ L(x_i - x_j)^2 + c \sum_{ij} \eta_{ij} (1 - y_{ij}) \left[ 1 + \|L(x_i - x_j)\|^2 - L(x_i - x_j) \|^2 \right] \right] \]  

(3)

where \( [z]_+ = \max(z, 0) \) and \( c \) is a normal number.

The first item on the right side of the equal sign in equation (3) adjusts the distance between all input samples and the target neighbours. By minimizing the first item, the distance between the input sample and the target neighbour can be reduced as small as possible. The second item adjusts the marginal distance among different categories and maximizes the marginal distance by minimizing the item. In order to facilitate solving the problem in a larger feasible domain, introduce slack change \( \xi_{ijl} \) is introduced, then equation (3) is transformed to solve the following semi-definite programming problem:

\[
\begin{align*}
\min M & : d_M(x_i, x_j) + c \sum_{ij} (1 - y_{ij}) \xi_{ijl} \\
\text{s.t.} & \quad \forall (i, j, l), d_M(x_i, x_j) - d_M(x_i, x_j) \geq 1 - \xi_{ijl} \\
& \quad \forall (i, j, l), \xi_{ijl} \geq 0 \\
& \quad M \succeq 0
\end{align*}
\]  

(4)

Then obtaining the metric matrix \( M \) is obtained by solving the above formula using semi-definite programming algorithm.

2.3. Modeling process

The whole model is divided into three modules: feature engineering module, metric learning module and classification module. Feature Module: First, reshape the original timing features. The purpose of this is to facilitate the analysis of curvelet transform. Then, the features after reshaping are curved wave analysed, and the results are combined to form new features. Metric learning module: This module is mainly responsible for mapping the combined features into a new embedded space. The specific selection of metric learning algorithm is LMNN. Classification module: Classify the sample set on the basis of embedded space. Since the distribution of samples has been modified by metric learning to make it easier to distinguish, this paper chooses simple KNN as the classifier in this module. Figure 1 illustrates the process of this method.

![Figure 1](image.png)

Figure 1. The modelling process.

3. Experiment and result analysis

In order to demonstrate the feasibility and effectiveness of the proposed method, two other traditional methods are also used to analyse the same data set for comparison, including BP neural
network (BPNN) and SVM.

3.1. Experimental data
The experimental vibration data of rolling bearings used in this study is from Case Western Reserve University lab. The experimental setup is shown in Fig.6.205-2RS JEM SKF deep groove ball bearings are arranged with a single point of failure on the inner ring, outer ring and rolling elements, with fault diameters of 0.18 mm, 0.36 mm, 0.53 mm, and 0.71 mm. This experiment uses the vibration signal collected by the drive end of the 12 kHz sampling frequency at 1797 rpm. There are 12 cases, including four types of bearing health conditions: signal under normal conditions, outer ring fault, inner ring fault and rolling fault. Each state contains 117 samples, each containing 1024 samples. Each sample is divided into training set and test set according to 80% and 20%. The details of these 12 status data are listed in Table 1, and the original time domain part waveform diagram is shown in Figure 2.

| Bearing operating condition | Fault diameter | Outer race fault orientation | The number of training/testing samples | Condition label |
|-----------------------------|----------------|------------------------------|----------------------------------------|-----------------|
| Normal                      | 0              | -                            | 93/24                                  | 1               |
| Ball                        | 0.007          | -                            | 93/24                                  | 2               |
| Ball                        | 0.014          | -                            | 93/24                                  | 2               |
| Ball                        | 0.021          | -                            | 93/24                                  | 2               |
| Inner race                  | 0.007          | -                            | 93/24                                  | 3               |
| Inner race                  | 0.021          | -                            | 93/24                                  | 3               |
| Inner race                  | 0.028          | -                            | 93/24                                  | 3               |
| Outer race                  | 0.007          | Vertical @3:00               | 93/24                                  | 4               |
| Outer race                  | 0.007          | Center @6:00                 | 93/24                                  | 4               |
| Outer race                  | 0.014          | Center @6:00                 | 93/24                                  | 4               |
| Outer race                  | 0.021          | Vertical @3:00               | 93/24                                  | 4               |
| Outer race                  | 0.021          | Center @6:00                 | 93/24                                  | 4               |

Figure 2. Vibration signals of the four of the twelve bearing operating conditions.
3.2. Result analysis
We use metric learning to change the distribution of samples. In order to visualize the effect of metric learning, we set the dimension of embedded space to 2, and the visualization results are shown in Figure 3. From the Figure 3 it can be seen that metric learning can aggregate samples of the same category and separate samples of different categories. The sample distribution can effectively improve the performance of KNN classifier. In the embedded space, KNN classifier is used to classify the samples. At the same time, we choose BPNN and SVM as the comparison algorithm. Table 2 lists the experimental results of each algorithm. Through this table, we can see that the correctness of BPNN and the algorithm in the training set is very high. But in the test set, the result of BPNN is obviously lower than that of this algorithm, which shows that the generalization ability of this algorithm is better than that of BPNN. Analysing the experimental results of SVM, the accuracy of the algorithm in training set and test set is far lower than that of the other two algorithms. This is because the dimension of the sample has 1000 dimensions, and the SVM model can’t effectively handle high-dimensional samples. Through systematic simulation experiments, it is shown that the proposed algorithm can effectively detect the fault of rolling bearings and has good generalization ability.

4. Conclusion
Aiming at the problem of fault diagnosis of rolling bearing in mechanical equipment, this paper proposes a smart diagnosis algorithm based on curvelet transform and metric learning. The proposed method effectively solves the shortcomings of traditional fault diagnosis algorithms. The main reason is that metric learning can effectively learn the essential features from the input data and improve the performance of the KNN by learning a mapping matrix to modify the distribution of the samples. Ten trials are carried out for diagnosing each bearing data set. The comparison result indicates that the proposed method shows higher accuracy than traditional methods. The result also proves that the proposed method have good generalization ability and robustness.

Figure 3. Figure with short caption (caption centred).
Table 2. Diagnosis results of different methods.

|               | Training samples | Testing samples |
|---------------|------------------|-----------------|
|               | Mean             | Variance        | Mean             | Variance        |
| Method1       | 0.92722          | 0.0245          | 0.84269          | 0.03501         |
| Method2       | 0.54103          | 0.04297         | 0.45169          | 0.04931         |
| Method3       | 0.94321          | 0.02313         | 0.90137          | 0.03902         |

Remarks: Method1-BPNN; Method2-SVM; Method3-The proposed method.

5. Conclusion
Aiming at the problem of fault diagnosis of rolling bearing in mechanical equipment, this paper proposes a smart diagnosis algorithm based on curvelet transform and metric learning. The proposed method effectively solves the shortcomings of traditional fault diagnosis algorithms. The main reason is that metric learning can effectively learn the essential features from the input data and improve the performance of the KNN by learning a mapping matrix to modify the distribution of the samples. Ten trials are carried out for diagnosing each bearing data set. The comparison result indicates that the proposed method shows higher accuracy than traditional methods. The result also proves that the proposed method have good generalization ability and robustness.

Acknowledgment
This work is supported by the National Natural Science Foundation of China under Grants 61773172 and 61572210, the Natural Science Foundation of Hubei Province of China (2017CFA035) and the 111 Project on Computational Intelligence and Intelligent Control under Grant B18024.

Reference
[1] Shao H, Jiang H, Wang F, et al. Rolling bearing fault diagnosis using adaptive deep belief network with dual-tree complex wavelet packet[J]. Isa Transactions, 2017:187-201.
[2] Lei Y, Zuo M J, He Z, et al. A multidimensional hybrid intelligent method for gear fault diagnosis[J]. Expert Systems with Applications, 2010, 37(2):1419-1430.
[3] BUXTON B, GOLDSTON D, DOCTOROW C, et al. Big data: science in the petabyte era [J]. Nature, 2008, 455(7209):8-9.
[4] Wei Z, Wang Y, He S, et al. A novel intelligent method for bearing fault diagnosis based on affinity propagation clustering and adaptive feature selection[J]. Knowledge-Based Systems, 2017, 116(C):1-12.
[5] Ding X, He Q, Luo N. A fusion feature and its improvement based on locality preserving projections for rolling element bearing fault classification[J]. Journal of Sound and Vibration, 2015, 335:367-383.
[6] Li W, Zhu Z, Jiang F, et al. Fault diagnosis of rotating machinery with a novel statistical feature extraction and evaluation method[J]. Mechanical Systems and Signal Processing, 2015, 50-51:414-426.
[7] Keskes H, Braham A, Lachiri Z. Broken rotor bar diagnosis in induction machines through stationary wavelet packet transform and multiclass wavelet SVM[J]. Electric Power Systems Research, 2013, 97(97):151-157.
[8] Qu J, Zhang Z, Gong T. A novel intelligent method for mechanical fault diagnosis based on dual-tree complex wavelet packet transform and multiple classifier fusion[J]. NEUROCOMPUTING, 2016, 171(C):837-853.
[9] Chuan Z, Xing W U, Chang L, et al. Rolling Bearing Fault Diagnosis Based on EMD and Fuzzy C Means Clustering[J]. Journal of Kunming University of Science & Technology, 2009.
[10] Changliang L, Yingjie W U, Chenggang Z. Rolling Bearing Fault Diagnosis Based on
Variational Mode Decomposition and Fuzzy C Means Clustering[J]. Proceedings of the CSEE, 2015.

[11] Yongqi C, Yiming Z, Yang C. Fault diagnosis of roller bearing based on intrinsic time-scale decomposition[J]. Journal of Electronic Measurement and Instrumentation, 2015.

[12] Unal M, Onat M, Demetgul M, et al. Fault diagnosis of rolling bearings using a genetic algorithm optimized neural network[J]. Measurement, 2014, 58:187-196.

[13] Heng L I, Qin Z, Xianrong Q, et al. Fault diagnosis method for rolling bearings based on short-time Fourier transform and convolution neural network[J]. Journal of Vibration and Shock, 2018.

[14] Zheng-Yi Huang, Qing-Qing Mu, Mian Hu, Ying Zou, Jiang-Wen Xiao. Regression-based metric learning [C]. 2018 37th Chinese Control Conference (CCC), 2018: 1934-1768.

[15] Emmanuel J. Candès, Donoho D L. New tight frames of curvelets and optimal representations of objects with piecewise C2 singularities[J]. Communications on Pure & Applied Mathematics, 2010, 57(2):219-266.

[16] Weinberger K Q, Blitzer J, Saul L K. Distance Metric Learning for Large Margin Nearest Neighbor Classification[M]. JMLR.org, 2009.