Fault Detection Method for Rolling Mechanical Equipment
Based on EEMD and RF

Yu HUANG, Xiao-lei LI* and Quan-chao JIA
School of Control Science and Engineering, Shandong University, Jinan 250061, China
*Corresponding author

Keywords: Ensemble empirical mode decomposition, Fault detection and diagnosis, Power spectral density, Random forest.

Abstract. Since rolling mechanical equipment is critical to manufacture systems, fault detection and diagnosis for them become more important and urgent with the improving need for safely and economical running. Aiming at streaming vibration signals, a fault detection and diagnosis scheme is proposed with an off-line training stage and on-line detecting stage. After de-noising by wavelet transformation, signals are decomposed into several IMF (Intrinsic Mode Functions) by EEMD (Ensemble Empirical Mode Decomposition) method. Fault features are expressed on definite IMFs rather than in a mixed mode. Based on these IMFs, the fault types of bearings can be represented by PSD (Power Spectral Density) to get good identification ability, and even to get the detailed physical meaning. The RF (Random Forest) method is used for classification of various fault positions and degrees with high efficiency and accuracy. Simulations with real data show the practicability of the fault detection scheme, and the earlier fault detection or fault forecast can be achieved by expanding classification training set for mixed fault data.

Introduction

Mechanical equipment and their attachments generally link with high power drivers and loads, so it will be very dangerous if the system fails. It’s very important to detect the early faults or even to predict the potential malfunction of the system. One of the most efficient ways to detect the fault of mechanical equipment is based on the analysis of vibration signals. There are mainly four steps, that is, signal acquisition, feature extraction, status diagnosis, and status analysis [1]. For the acquired signals, de-noising is an effective method for the improvement of signal quality. Wavelet transformation is a widely applied method for de-noising of frequency signals. Paper [2] studied the wavelet packet de-noising method, and find that it can effectively remove the noise of each band. Paper [3] uses EMD (Empirical Mode Decomposition) method to get the power spectrum of cracks, and effectively identify the gear root faults. Paper [4] proposed an improved Hilbert Huang transform and optimized some shortages of the EMD method. Paper [5] proposed EEMD, an improved EMD method, which optimized the model aliasing phenomenon and eliminate the intermittent phenomenon of the original signal, and it is a major improvement to the EMD method. Paper [6] gets a classification model based on decision tree method. Several trim methods are used to improve the classification accuracy of the decision tree, but it has weak generalization ability. The random forest method may be the best candidate for classification which is studied in this paper.

In this paper, we analyzed the healthy degree of bearing parts of a mechanical facility, to predict its fault position and the erosion degree via monitoring the daily running data. Firstly, we use wavelet algorithm to de-noise the acquired vibration signals and EEMD to separate them into several IMF sets. Based on the PSD analysis of these IMF sets, we get the features of several faults; the Random Forest algorithm is used to classify these faults. Practical data is tested in a streaming mode.
**System Scheme**

It’s very important to detect the fault types and fault degrees in time and to evaluate the health status of some critical parts while the mechanical equipment is running. Since the rolling machines are spinning very fast, the real-time detection is required. Considering the limitation of computing, the fault detection and diagnosis process can be divided into two stages, one is the off-line diagnosis model creation stage, and the other is the on-line fault detection stage, as shown in Figure 1. In the off-line stage, the history data of vibration signals including the normal and abnormal samples are trained. Following the de-noising and feature extraction processes, the RFs are created. In fact, the RFs have the classification information of the normal and all types of abnormal conditions of the system, so it can be the fault detection and diagnosis model. Then, in the online stage, the real-time data is streaming into the system, and after quick preprocessing and decomposition, feature vectors are composed. The created RFs can classify the online feature vectors quickly, then the normal condition, fault type, and fault degrees can be outputted accordingly.

![Figure 1. System scheme for fault detection and diagnosis with off-line training and on-line detections.](image)

**Data Preprocessing**

Original signals need to be de-noised to eliminate the interference of noise, then, the later fault features extraction and fault classification can get better performance.

**Wavelet De-noising.** Wavelet transform is an efficient method for de-noising. For the selection of basis wavelet, the fourth order Daubechies wavelet (db4) has a good de-noising effect and can provide a smooth reconstruction of the signals.

**EEMD and PSD Analysis.** The EMD method is proposed by Norden E. Huang et al [7]. The decomposition is based on a collection of intrinsic mode functions (IMFs), which can give us a full energy-frequency-time distribution of the data. There is a major drawback of EMD called mode mixing of frequent appearance. A single IMF either consists of widely disparate scales signals, or a signal scale resides in different IMF components, and this makes the physical meaning of individual IMF unclear. The ensemble empirical mode decomposition (EEMD) method overcomes the drawback [8]. By adding finite noise, the EEMD eliminated largely the mode mixing problem and preserve physical uniqueness of decomposition.

As for the EEMD result of a signal, it’s hard to tell the fault features directly or which component may have more fault features. Thus, the power spectral density analysis has been done.
**Power Spectral Density Analysis.** The power spectral density describes the power distribution of a signal over frequency. For the actual acquired vibration signal, the average density of vibration energy on various frequencies can be read from the PSD. For the fault detection of mechanical equipment, the PSD shows the definite frequency feature of each fault can be important features for classification. In this research, the EEMD separates the energy distribution of original signal into several IMFs first, and the PSD helps to observe features of these distributions.

**Pure Faults Ample Analysis.** The selection of features is an important factor that influences the classification algorithms efficiency. The features of PSDs based on IMFs not only have good identification degrees, but also have definite engineering significance. Figure 2 shows the PSD analysis of the faults at outer race, inner race, and ball of the test bearing.

![PSD Analysis of Faults](image)

Figure 2. PSD illustrations of different fault features.

The main differences between each fault are the strength and characteristic frequencies of the PSDs for the corresponding IMF components. Different types of fault have obvious difference between the strength and characteristic frequencies. Different degree faults of the same type have more difference in PSD strength. As for the ball fault, the characteristic frequencies are 78Hz, 146Hz, 159Hz, and 166Hz. As for the inner race fault, the characteristic frequencies are 53Hz, 87Hz, 181Hz, and 196Hz. As for the outer race fault, the characteristic frequencies are 48Hz, 78Hz, and 123Hz to 143Hz. Also, the strength of PSD of outer race is weaker than that of inner race. As for the normal signal, the characteristic frequencies of PSDs based on IMFs are 134Hz and 264Hz. So, the features have good distinguish degrees for each fault.

**Mixed Faults Sample Analysis.** Pure fault samples generally can get good classification results. But during the real-time detection, the signal is streamed into the sample data-window, so for a period of time especially the fault just happened, the classification will be disordered since the sample data will be mixed with normal data and fault data. The detection will be delayed until the sample data-windows is filled with fault data.

In this research, since the width of the sample data-window is 1500, we set 100 as a step, that is the detection algorithm compute once for 100 new data enter the data-windows and 100 old data leave the data-window. Figure 3 shows the progress of the inner-race fault, which there are 40% of the fault.
data streaming into the data-window. During the experiment, we find that the fault feature starts to appear when there are 20% of the fault data in the data-window, and the detector becomes sensitive to the fault data when the fault data reaches 30%. For 40% of the fault data, the detector can detect the fault definitely. So, based on the EEMD, the PSD features have definite physical concept and good classification ability.

![PSD illustrations of mixed faults comparing of normal sample and inner fault sample.](image)

**Figure 3.** PSD illustrations of mixed faults comparing of normal sample and inner fault sample.

**Fault Classification Based on RF**

**The Random Forest**

Random forest (RF) is an efficient and simple way for the fault classification of mechanical equipment. Random forest [9] is a combination of tree classifiers. The method samples the dataset with replacement to create new datasets, and uses them to fit different trees. These generated trees vote for the most popular class they predicted as the result. Theoretical studies have proved that the RF has high prediction accuracy and has good tolerance to outliers and noises [10].

**Experiment Analysis**

**Pure Fault Classification Test.** We label the data first. For the fault degree of 7 mils damage, we label the ball fault as 1, the inner-race fault as 2, and the outer-race fault as 3. For the fault degree of 21 mils damage, we label the ball fault as 4, the inner-race fault as 5, and the outer-race fault as 6. The normal data is labeled as 0. We construct data sets with 120000 sample points for each fault type, and we use half of them as training set and half as test set. We use 40 samples per fault for training, each sample with 1500 sample points. The model predicts well with zero misclassified samples.

**Real-time Streaming Data Simulation.** We mix the normal data and fault data to create a real-time data stream. Based on the former analysis, the fault feature cannot be detected when the proportion of fault data is below 20%, so the mixed fault data percentage is from 20% to 80%. These samples are trained and added to new labels. During the test, we set the detection data feeding speed from 500/step, 300/step, 200/step to 100/step. Figure 4 and Figure 5 show that when the feeding speed is 100/step, the detection will be delayed for 1 iteration, in comparison, there is no delay for feeding speed faster than 200/step. As we can see, the algorithm can detect the fault happening in 25ms (about 200-300 fault data points) and can classify the fault types and fault damage degrees correctly.

**Summary**

For the fault detection and diagnosis of mechanical equipment, frequency features are definite based on vibration signals. The EEMD method is an efficient way to decompose the de-noised vibration signal into several IMFs. Based on these IMFs, fault type of bearings can be represented by PSDs which have good identification abilities. Due to the variety of fault positions and degrees, more data...
samples are required for the training of classifications. Then, the RF is a good classification algorithm for high dimension samples with high efficiency and accuracy. Simulations with real data show the applicability of the proposed fault detection scheme. With the expansion of classification training for mixed fault data, the earlier fault detection or fault forecast can be achieved.

Figure 4. Fault detection with data feeding speed 100/step  Figure 5. Fault detection with data feeding speed 200/step.

References


