ISBN 978-981-14-1684-2
Proceedings of 2019 the 9th International Workshop on Computer Science and Engineering
(WCSE 2019 SUMMER)
Hong Kong, *15-17 June*, *2019*, *pp. 358-362*
doi: 10.18178/wcse.2019.06.054

Fault Diagnosis of Rolling Bearings Under Variable Load Conditions Based on Multi-domain Features and Random Forests

Xiaoming Xue [1 +], Quanping Sun [1], Suqun Cao [1], Xuecheng Wang [1] and Yanxia Zhuang [2]

[1] Faculty of Mechanical and Material Engineering, Huaiyin Institute of Technology, Huai'an, 223003, China
[2] Automotive Engineering College, Huaian Vocational College of Information Technology, Huai'an, 223003, China

Abstract. Vibration signals of rolling bearings collected under variable load conditions often have complex dynamic properties which pose a huge challenge for its effective fault diagnosis. To solve this problem, a novel fault diagnosis method based on multi-domain features and random forests is proposed in this paper. In features extraction, the fast ensemble empirical mode decomposition method is first used to decompose the original signals into a collection of intrinsic mode functions (IMFs). After signal decomposition, the singular values of the matrix formed by the row vectors of IMFs can be obtained by singular value decomposition.
On the other hand, to obtain a comprehensive description about vibration signals, the statistical analysis method and Fourier transform are employed to extract 10 time domain features and 10 frequency domain features. As for the automatic diagnosis of bearing faults, a novel combined classifier algorithm named as random forests is used to classify the multi faults under different load conditions. Finally, the proposed method is evaluated by experiments with 10 fault types and some comparative studies are also given. The experimental results indicate its effectiveness and robustness for rolling bearing fault diagnosis under variable load conditions.

Keywords: rolling bearing, fault diagnosis, variable load conditions, fast EEMD, random forests.

1. Introduction

Rolling bearings are the key components and the most vulnerable parts of rotating machinery. Its operation condition plays an important role on maintaining the stability and reliability of equipment [1]. With the development of computer technology, signal processing technology and artificial intelligence, the intelligent fault diagnosis of bearings has been concerned by more and more researchers in recent years [1-3]. It can be seen from the previous studies that the most work was focused on the fault diagnosis problem with constant operating conditions. However, in reality, the working condition is variable and even the faults can occur under this situation, which will lead the characteristic property of vibration signals more complex and pose a significant challenge for the fault diagnosis [4].

The fault diagnosis of rolling bearings under variable load conditions is still a typical pattern recognition problem. As for features extraction, traditional signal processing and features extraction techniques may not do well on describing the non-stationary and nonlinear characteristics of vibration signals. Recently several time-frequency analysis methods, included short time Fourier transform (STFT), wavelet transform (WT), local mean decomposition (LMD) and empirical mode decomposition (EMD), have been proposed to analyzing non-stationary signals [5]. Among these methods, EMD has been widely used in the fault diagnosis field due to its fine adaptivity and time-frequency resolution. However, some shortcomings of EMD, such as the mode mixing and the end effects, have hindered its further applications. To eliminate the mode mixing problem of EMD, a noise assisted data analysis method, ensemble EMD (EEMD), was raised by Wu [6].

+ Corresponding author. Tel.: + 8615261747255; fax: +860517 83559196.
E-mail address: xue_xiaoming@foxmail.com.

Based on the EEMD method, a fast EEMD is proposed by Wang to further improve the computation performance of EEMD [7]. The FEEMD method is with all the advantages of EEMD and can effectively resolve the high time consumption problem.

After features extraction, the final faults identification is realized by establishing a classifier based on machine learning methods, such as artificial neural network (ANN), extreme learning machine (ELM) and support vector machine (SVM). Although these methods have been successfully applied in many fault diagnosis fields, they also have some obvious drawbacks. For example, the parameter optimization is a tricky problem for all these three methods, and ANN and ELM also suffer from the trouble of convergence rate. To enhance the generalization performance and the overfitting inhibiting ability of single decision tree, Leo Breiman proposed one combined classifier algorithm namely random forests (RF) [8]. By introducing the concepts of random attributes and random samples, the method performs good generalization and accuracy in dealing with the small sample size and high dimensional features problem. In recent years, the RF method has been used in many fields such as speech recognition [9] and face recognition [10]. However, the research of RF about its application in the field of rolling bearings fault diagnosis is rarely reported.

In this paper a novel fault diagnosis model based on multi-domain features and random forests is proposed to solve the rolling bearing fault diagnosis under different load conditions. To extract more comprehensive features, the fast ensemble empirical mode decomposition method is first used to decompose the original signals into a collection of intrinsic mode functions (IMFs), then the singular values of the matrix formed by the row vectors of IMFs can be obtained by singular value decomposition. As a complementary description about vibration signals, the statistical analysis method and Fourier transform are employed to extract 10 time domain features and 10 frequency domain features. Then the random forests method is used to classify the multi faults under variable load conditions. Finally, the proposed method is evaluated by experiments with 10 fault types and some comparative studies are also given.

The rest of this paper is organized as follows. A brief description about the multi-domain features extraction is given in Section 2. In Section 3, the random forests theory is reviewed. Then the proposed diagnosis model based on multi-domain features and random forests is presented in Section 4. In Section 5, the proposed model is employed to the fault diagnosis of rolling bearings under variable load conditions. Finally, the conclusions are drawn in Section 6.

2. Multi-domain Features Extraction

For the multi-fault diagnosis problem, especially under variable load conditions, the vibration signals can't be fully characterized only by the single domain features, which can result in the final failure of fault diagnosis. Hence, different signal processing techniques and feature extraction methods are employed in this work to extract time domain features, frequency domain features and time-frequency domain features.

In general, the time domain analysis method and the frequency domain analysis method are used to describe the statistical property of the signal series. Here, 10 features in time domain and 10 features in frequency domain are extracted. The detailed description about these parameters can be found in [11].

For time-frequency features extraction, FEEMD is used to decompose the original vibration signals into a series of narrowband components and then singular value features can be obtained by performing singular value decomposition on the matrix which is composed of IMFs.

Finally, the multi-domain features of one signal sample can be obtained and denoted as $W_i=(T_{i1}, T_{i2}, \ldots, T_{i10}, F_{i1}, F_{i2}, \ldots, F_{i10}, S_{i1}, S_{i2}, \ldots, S_{iL})$, where T is the time domain feature, F is the frequency domain feature and S is the multiscale singular value feature.

3. Random Forests

Essentially, the random forests method is a novel combined classifier algorithm consisting of multiple CART decision trees, where the model is built based on the Bagging method with the introduction of random attributes. Given the original dataset $\boldsymbol{X}$ with M column size, the construction steps of the classifier $\{h(\boldsymbol{a},\boldsymbol{\theta}_k), k=1,2,\ldots,K\}$ are given as follows. K is the total number of decision trees, $\{\boldsymbol{\theta}_k\}$ denotes the random vector set with independent and identical distribution and $\boldsymbol{a}$ represents the testing sample.

- Through random sampling of $\boldsymbol{X}$, get one self-service sample set $\boldsymbol{\theta}_k$ with the returned sampling way, where the sampling process is named as the Bootstrap method. In addition, $\boldsymbol{\theta}_k$ is required to have the same size of $\boldsymbol{X}$.
- For each sample subset $\boldsymbol{\theta}_k$, during the fission process of each branch node, a new data set with m dimensional features randomly selected from M dimensional features are employed to train the CART decision tree. In general, m is equal to $\sqrt{M}$. According to the principle of minimization node impurity, one feature is selected from the m features as the classification attributes and none of trees are pruned.
- Repeat steps (1) and (2) until K decision trees are obtained and then the RFs model can be established by integrating these binary trees, where K=500~1000.
- For one test sample, the final classification result can be obtained by using the following formula

$$c = \arg\max_c \left(\frac{1}{T}\sum_{k=1}^{T} I(h(\boldsymbol{a},\boldsymbol{\theta}_k) = c)\right) \tag{1}$$

  where $I(\cdot)$ is the indicator function and c represents the sample type with the most votes.

4. The Proposed Method

The detailed description about the proposed method is given as follows.

- Given a training sample set with different fault types and different loads, let $P_i = (p_{i1}, p_{i2,} \cdots, p_{in}),\ i = 1,2,\cdots,N$ denotes one signal sample. According to the mentioned method in section 2, obtain the corresponding features in time domain and frequency domain, respectively denoted as $(T_{i1}, T_{i2}, \cdots, T_{i10})$ and $(F_{i1}, F_{i2}, \cdots, F_{i10})$.
- Decompose the original vibration signals into a collection of IMF components by FFEMD and form the matrix $[imf_{i1}, imf_{i2}, \cdots, imf_{iL}]$, where imf_{ij} is the j-th decomposition component of P_i and L is total number of IMFs. Then perform singular value decomposition on $[imf_{i1}, imf_{i2}, \cdots, imf_{iL}]$ and obtain the singular value features $(S_{i1}, S_{i2}, \cdots, S_{iL})$ of P_i.
- Through features integration, for each signal sample, the multi-domain features can be constructed as $(T_{i1}, \cdots, T_{i10}, F_{i1}, \cdots, F_{i10}, S_{i1}, \cdots S_{iL})$.
- Calculate the multi-domain features of all training samples and normalize the characteristic matrix. Then train the RFs based diagnosis model by using the feature dataset, where the parameters of RFs are set to default values.
- For testing sample set, obtain multi-domain features and recognize the fault types based on the established RFs model.

5. Experiments and Analysis Results

5.1. Experimental data

The vibration signals were collected from rolling element bearings installed at the drive end by a data acquisition instrument with 16 channels, where the experimental platform is constructed by Case Western Reserve University [12]. The sample frequency is 48 kHz. In order to investigate variable working conditions, 3 fault types with different defect size were considered under different motor loads of 0~3hp. A detailed description about the fault conditions of rolling element bearings is displayed in Table 1, where " √ " denotes this kind of fault condition is considered. As illustrated in Table 1, included the normal condition under different loads, vibration signals of 10 fault conditions were collected for diagnosis analysis. Here, three cases with different training set and testing set were employed to validate the proposed method. The data construction of these three cases is listed in Table 2.

Table 1: Fault information under different motor loads of 0~3hp.

Defect size / Fault type	0.007in	0.014in	0.021in
Inner race fault	√	√	√
Ball fault	√	√	√
Outer race fault	√	√	√

Table 2: Sample proportion and size of training and testing set of three cases.

	Training set		Testing set	
Case 1	20%	633	80%	2532
Case 2	40%	1266	60%	1899
Case 3	60%	1899	40%	1266

5.2. Diagnosis analysis based on the proposed method

Considering the nonstationarity and nonlinearity of the vibration signals, FEEMD is used to decompose the original signals into a series of IMFs. Fig. 1 displays the decomposition results of one sample with inner race fault, where the number of IMF components is set to 10. Obviously, the mono components in different narrow frequency band can be effectively separated. In addition, the decomposition speed of FEEMD is much faster than that of EEMD. After signal decomposition, we can get the singular value features of each sample. Then the statistical analysis method and Fourier transform are used to get 10 time domain features and 10 frequency domain features. Finally, for each sample, the dimension size of the multi-domain features is 30.

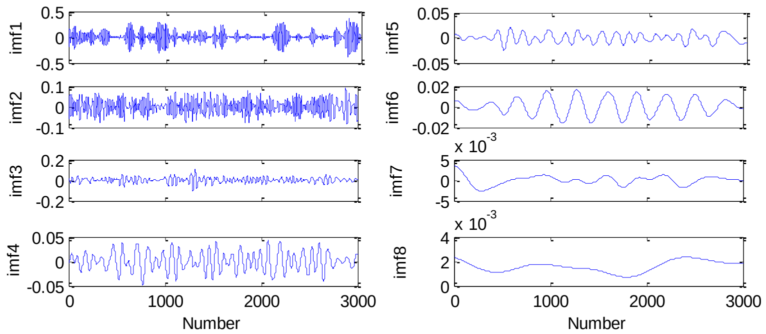

Fig. 1: The FEEMD decomposition results of one vibration signal with inner race fault.

After features extraction, the features set of training set is employed to train the RFs based recognition model. In order to illustrate the classification ability of multi dimension features, features dataset only with time domain (TD) feature, frequency domain (FD) feature, multiscale singular value (MSV) feature or multi-domain (MD) feature is respectively considered for diagnosis. Table 3 lists the diagnosis results with different features based on the RFs model. As displayed in Table 3, each domain features have some capacity in terms of the representation of signal properties. But the multi-domain features perform the best recognition ability for the three cases.

Table 3: Diagnosis accuracy of testing set with different domain features of Case 1, 2 and 3.

	TD	FD	MSV	MD
Case 1	93.21%	83.69%	95.54%	**96.45%**
Case 2	94.10%	86.15%	96.16%	**97.21%**
Case 3	94.63%	87.91%	96.68%	**97.79%**

Table 4: Diagnosis accuracy of testing set with different classifiers.

	Case 1		Case 2		Case 3	
	Accuracy	Time/s	Accuracy	Time/s	Accuracy	Time/s
MD-SVM	96.37%	50.32	97.00%	153.53	98.10%	298.74
MD-ELM	80.25%	0.01	83.52%	0.01	85.23%	0.04
MD-RF	96.45%	0.81	97.21%	1.77	98.14%	2.81

The diagnosis results listed in Table 3 indicate that the proposed model based on multi-domain features and RF is in high precision and good convergence. Furthermore, in order to further illustrate the performance of the proposed model, the training set with multi domain features was employed to train the SVM and ELM based fault classifiers respectively. The diagnosis results obtain by different recognition models of the three cases are given in Table 4. As displayed in Table 4, the advantages of the ELM based model mostly accrue to its high computing efficiency, but its diagnosis accuracy is relatively low. Although both the SVM model and the RF model can obtain good diagnosis results with high accuracy, the SVM model is too time intensive. In conclusion, dealing with multi faults diagnosis of rolling bearings under different load conditions, the proposed method not only can extract full and effective characteristic information of vibration signals, but also shows excellent performance in diagnosis accuracy and efficiency.

6. Conclusions

Features extraction and high demand of diagnosis accuracy and efficiency still pose a great challenge for the multi-fault diagnosis problem of rolling bearings. In this work, a novel diagnosis model based on multi-domain features and RF is proposed to solve this problem. The multi-domain features integrated with time domain features, frequency domain features and time-frequency domain features can effectively and fully represent the characteristic information of vibration signals. Meanwhile in order to improve the diagnosis accuracy and efficiency, the RF algorithm is introduced to construct the recognition model. Finally, the proposed model is validated by the experiments with multi faults under variable load conditions. The analysis results demonstrate its effectiveness and robustness for this kind of diagnosis problem and the model is suitable for practical applications. Furthermore, the rolling bearings fault diagnosis with unknown load conditions will become our next study emphasis.

7. Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 51709122) and the open fund for Jiangsu Key Laboratory of Advanced Manufacturing Technology (Grant No. HGAMTL-1713).

8. References

[1] X. Zhang, Y. Liang, J. Zhou, et al. A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM. *Measurement*, 2015, 69: 164-179.

[2] B. Li, M. Y. Chow, Y. Tipsuwan, et al. Neural-network-based motor rolling bearing fault diagnosis. *Industrial Electronics IEEE Transactions on*, 2002, 47 (5): 1060-1069.

[3] Y. Yang, D. J. Yu, J. Cheng. A roller bearing fault diagnosis method based on EMD energy entropy and ANN. *Journal of Sound & Vibration*, 2006, 294 (1): 269-277.

[4] J. Shi, M. Liang, Y Guan. Bearing fault diagnosis under variable rotational speed via the joint application of windowed fractal dimension transform and generalized demodulation: A method free from prefiltering and resampling. *Mechanical Systems & Signal Processing*, 2016, s68–69 (6): 15-33.

[5] K. H. Hui, L. M. Hee, M. S. Leong, et al. Time-Frequency Signal Analysis in Machinery Fault Diagnosis: Review. *Advanced Materials Research*, 2014, 845: 41-45.

[6] Z. H. Wu, N. E. Huang. Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in Adaptive Data Analysis*, 2011, 1 (1): 41.

[7] Y. H. Wang C. H. Yeh, H. W. V. Young, et al. On the computational complexity of the empirical mode decomposition algorithm. *Physica A: Statistical Mechanics and its Applications*, 2014, 400: 159-167.

[8] A. Cutler, D. R. Cutler, J. R. Stevens. Random Forests. *Machine Learning*, 2004, 45 (1): 157-176.

[9] J. Xue, Y. Zhao. Random Forests of Phonetic Decision Trees for Acoustic Modeling in Conversational Speech Recognition. *IEEE Transactions on Audio Speech & Language Processing*, 2008, 16 (3): 519-528.

[10] G. Fanelli, M. Dantone, J. Gall, et al. Random Forests for Real Time 3D Face Analysis. *International Journal of Computer Vision*, 2013, 101 (3): 437-458.

[11] X. Xue, J. Zhou. A hybrid fault diagnosis approach based on mixed-domain state features for rotating machinery. *Isa Transactions*, 2017, 66: 284-295.

[12] M. Luo, C. Li, X. Zhang, et al. Compound feature selection and parameter optimization of ELM for fault diagnosis of rolling element bearings. *Isa Transactions*, 2016, 65: 556-566.