Mechanical Systems and Signal Processing 101 (2018) 292–308

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp

Sparsity guided empirical wavelet transform for fault diagnosis of rolling element bearings

Dong Wang [a], Yang Zhao [a], Cai Yi [a,b,*], Kwok-Leung Tsui [a], Jianhui Lin [c]

[a] Department of Systems Engineering and Engineering Management, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong, China
[b] School of Automobile and Transportation, Xihua University, Chengdu, China
[c] State Key Laboratory of Traction Power, Southwest Jiaotong University, Chengdu, China

ARTICLE INFO

Article history:
Received 16 April 2017
Received in revised form 22 August 2017
Accepted 27 August 2017

Keywords:
Empirical wavelet transform
Sparsity
Bearing fault diagnosis
Fourier segments
Multiple bearing defects
Railway axle bearings

ABSTRACT

Rolling element bearings are widely used in various industrial machines, such as electric motors, generators, pumps, gearboxes, railway axles, turbines, and helicopter transmissions. Fault diagnosis of rolling element bearings is beneficial to preventing any unexpected accident and reducing economic loss. In the past years, many bearing fault detection methods have been developed. Recently, a new adaptive signal processing method called empirical wavelet transform attracts much attention from readers and engineers and its applications to bearing fault diagnosis have been reported. The main problem of empirical wavelet transform is that Fourier segments required in empirical wavelet transform are strongly dependent on the local maxima of the amplitudes of the Fourier spectrum of a signal, which connotes that Fourier segments are not always reliable and effective if the Fourier spectrum of the signal is complicated and overwhelmed by heavy noises and other strong vibration components. In this paper, sparsity guided empirical wavelet transform is proposed to automatically establish Fourier segments required in empirical wavelet transform for fault diagnosis of rolling element bearings. Industrial bearing fault signals caused by single and multiple railway axle bearing defects are used to verify the effectiveness of the proposed sparsity guided empirical wavelet transform. Results show that the proposed method can automatically discover Fourier segments required in empirical wavelet transform and reveal single and multiple railway axle bearing defects. Besides, some comparisons with three popular signal processing methods including ensemble empirical mode decomposition, the fast kurtogram and the fast spectral correlation are conducted to highlight the superiority of the proposed method.

1. Introduction

Rolling element bearings are widely used in various industrial machines, such as electric motors, generators, pumps, gearboxes, railway axles, turbines, and helicopter transmissions [1]. Fault diagnosis of rolling element bearings is an emerging topic and it is beneficial to preventing any unexpected accident and reducing economic loss. When there is a defect on the surface of an outer race or inner race, impacts generated by rollers striking the defect excite the resonant frequencies

* Corresponding author at: Department of Systems Engineering and Engineering Management, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong, China

E-mail addresses: dongwang4-c@my.cityu.edu.hk (D. Wang), yangzhao9-c@my.cityu.edu.hk (Y. Zhao), justin.yi@163.com (C. Yi), kltsui@cityu.edu.hk (K.-L. Tsui), lin13008104673@126.com (J. Lin).

http://dx.doi.org/10.1016/j.ymssp.2017.08.038
0888-3270/

of a structure and then repetitive transients are observed in a time domain [2]. Repetitive transients are relevant to a shaft rotation frequency, the geometry of a bearing and the location of the defect. As a result, a ballpass frequency on outer race, a ballpass frequency on inner race, a fundamental cage frequency and a ball spinning frequency can be used to identify various bearing defects [3]. Because bearing fault related frequencies are often overwhelmed by heavy noises and other strong low-frequency vibration components, signal processing methods are required to reveal bearing defects. Two main approaches are spectral correlation [4] and envelope analysis [2,5]. The spectral correlation aims to build a spectral frequency and cyclic frequency plane to locate resonant frequency bands and then identify bearing defect frequencies. Even though this approach is powerful, it costs much calculation time and computer memory, and it is only suitable to off-line bearing fault diagnosis in the past years prior to the recent development of the fast spectral correlation proposed by Antoni et al. [6] for cyclic spectral analysis of bearing fault signals. The envelope analysis aims to determine resonant frequency bands and use a band-pass filter to retain one of the resonant frequency bands for further envelope analysis with demodulation. The main problem in the envelope analysis is how to establish the resonant frequency bands. Because Randall et al. [7] clarified that squared envelope analysis is equivalent to the integration of spectral correlation over spectral frequency, for the rest of this paper, the authors mainly focus on reviewing some popular signal processing methods relevant to the envelope analysis, especially band-pass filtering for envelope analysis with demodulation. For the state of the art relevant to spectral correlation, please refer to the recent work done by Antoni et al. [6].

The first popular signal processing method is spectral kurtosis formulated by Antoni [8] and its realization called the fast kurtogram [9]. The main idea of spectral kurtosis is to design some predefined band-pass filters and use kurtosis to pick up the most informative signal among all signals processed by the band-pass filters. Following this idea, some variants including the improved kurtogram [10], the enhanced kurtogram [11], the sparsogram [12], the infogram [13], multiscale clustering grey infogram [14], etc., were proposed to realize the same idea. Given a fixed bandwidth, the adaptive spectral kurtosis called Protrugram [15] was proposed to measure kurtosis from a frequency domain. To fuse some predefined band-pass filters and make spectral kurtosis adaptive, a window construction process by superposition was proposed by Wang and Liang [16]. More improvements on spectral kurtosis and their applications to bearing fault diagnosis were reviewed by Wang et al. [17]. One interesting recent work done by Borghesani et al. [18] revealed that kurtosis of an analytical signal obtained by Hilbert transform is proportional to the sum of the amplitudes of the squared Fourier spectrum of the squared envelope, which indicates that the spectral kurtosis has a very close relationship with cyclostationary analysis. This discover is consistent with the early conclusion made by Randall et al. [7].

The second popular signal processing method is wavelet transform [19–21] and its variants including multiwavelet transform [22], etc. This is because wavelet transform can be regarded an inner product operator [23] between an artificial wavelet mother function and repetitive transients caused by a bearing defect. According to the convolution theorem, wavelet transform can be also regarded as band-pass filtering by using an artificial mother function, which accelerates the calculation time of wavelet transform [24]. If the scale of wavelet transform is properly chosen [25], one of the resonant frequency bands of a structure is retained for envelope analysis with demodulation. Consequently, besides selection of wavelet filters [26,27], optimization of a wavelet filter is another hot topic [28–31]. Recently, Wang and Tsui connected spectral kurtosis based fast fault detection methods with wavelet transform and proposed dynamic Bayesian wavelet transform [32]. The main idea of dynamic Bayesian wavelet transform is to use spectral kurtosis based fast fault detection methods to initialize the state space model of wavelet parameters and then use dynamic Bayesian inference to find the optimal wavelet transform for extraction of repetitive transients, such as extraction of bearing fault signals.

The third popular signal processing method is empirical mode decomposition and its variants [33,34]. Empirical mode decomposition initially proposed by Huang et al. [35] is an adaptive signal processing method, which completely depends on the local maxima and minima of the amplitudes of a temporal signal. To solve the mode mixing problem, an improvement called ensemble empirical mode decomposition was proposed by Wu and Huang [36]. More improvements on empirical mode decomposition and their applications to bearing fault diagnosis were reviewed by Lei et al. [34]. Empirical mode decomposition and its variants can be regarded as adaptive band-pass filters [37]. The main problem of empirical mode decomposition and ensemble empirical mode decomposition for bearing fault diagnosis is that the frequency band of an intrinsic mode function (IMF) is often too wide so that heavy noises and other unwanted strong vibration components overwhelm the resonant frequency bands of a structure. To solve this problem, preprocessing methods including wavelet packets [38,39], spectral kurtosis [40], etc. are often required. Another main problem of empirical mode decomposition and ensemble empirical mode decomposition for bearing fault diagnosis is extensive calculation time. Unlike the aforementioned spectral kurtosis and wavelet transform, empirical mode decomposition and ensemble empirical mode decomposition cannot utilize the fast Fourier transform and the convolution theorem to accelerate their calculation time. As the length of a signal increases, searching for the local maxima and minima of the amplitudes of a signal is extensive. Besides, many iterations are used in empirical mode decomposition and ensemble empirical mode decomposition to achieve the definition of an IMF.

Inspired by wavelet transform and the adaptiveness of empirical mode decomposition, recently, Gilles [41] proposed empirical wavelet transform. The main idea of empirical wavelet transform is to determine Fourier segments of a signal and then design a series of wavelet filters to decompose the signal to several sub-signals. If Fourier segments are automatically determined, empirical wavelet transform becomes adaptive and it owns the characteristics of wavelet transform and the adaptiveness of empirical mode decomposition. One empirical rule for making empirical wavelet transform adaptive is to rely on the local maxima of the amplitudes of the Fourier spectrum of a signal and then use the center of the locations of two adjacent local maxima to determine Fourier segments of the signal [42]. Since heavy noises and other unwanted strong