Mechanical Systems and Signal Processing 72-73 (2016) 65–79

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

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

Detection and diagnosis of bearing faults using shift-invariant dictionary learning and hidden Markov model

Haitao Zhou [a,*], Jin Chen [a], Guangming Dong [a], Ran Wang [b]

[a] Fundamental Science on Vibration, Shock & Noise Laboratory, Shanghai Jiao Tong University, Shanghai 200240, China
[b] College of Logistics Engineering, Shanghai Maritime University, Shanghai 200135, China

ARTICLE INFO

Article history:
Received 14 July 2015
Received in revised form
5 October 2015
Accepted 19 November 2015
Available online 2 December 2015

Keywords:
Shift-invariant dictionary learning
Sparse representation
Adaptive feature extraction
Rolling element bearing
Hidden Markov model
Double impulses

ABSTRACT

Many existing signal processing methods usually select a predefined basis function in advance. This basis functions selection relies on a priori knowledge about the target signal, which is always infeasible in engineering applications. Dictionary learning method provides an ambitious direction to learn basis atoms from data itself with the objective of finding the underlying structure embedded in signal. As a special case of dictionary learning methods, shift-invariant dictionary learning (SIDL) reconstructs an input signal using basis atoms in all possible time shifts. The property of shift-invariance is very suitable to extract periodic impulses, which are typical symptom of mechanical fault signal. After learning basis atoms, a signal can be decomposed into a collection of latent components, each is reconstructed by one basis atom and its corresponding time-shifts. In this paper, SIDL method is introduced as an adaptive feature extraction technique. Then an effective approach based on SIDL and hidden Markov model (HMM) is addressed for machinery fault diagnosis. The SIDL-based feature extraction is applied to analyze both simulated and experiment signal with specific notch size. This experiment shows that SIDL can successfully extract double impulses in bearing signal. The second experiment presents an artificial fault experiment with different bearing fault type. Feature extraction based on SIDL method is performed on each signal, and then HMM is used to identify its fault type. This experiment results show that the proposed SIDL-HMM has a good performance in bearing fault diagnosis.

1. Introduction

The unexpected machinery failure may lead to great economic loss and personal safety threat. During recent decades, machinery maintenance based on condition monitoring is becoming more and more important in predicting equipment failure and avoiding unexpected machinery failure. As an important component in many rotating machines, bearing condition monitoring has received a great deal of attention from researchers [1].

The main procedure of machinery fault diagnosis usually consists of three main steps, i.e., data acquisition, feature extraction, and diagnostic identification. Feature extraction is used to extract the useful diagnostic information from the acquired raw signal, which enables the following classifiers to recognize the fault type correctly. Many feature extraction techniques have been proposed including three main categories: time–domain analysis, frequency–domain analysis, and

* Corresponding author. Tel.: +86 21 3420 6095x215.
E-mail address: zhouht1988@163.com (H. Zhou).

http://dx.doi.org/10.1016/j.ymssp.2015.11.022

time–frequency analysis [2,3]. The mostly used features in time domain are traditional statistical parameters such as root mean square (RMS), impulse factor, crest factor, kurtosis factor and so on. These parameters are evaluated as fault severity metrics, and proved to be very effective to monitor machine operation condition in [4,5]. The spectral analysis can identify the main frequencies and give a better interpretation of machinery signal [6]. However, the disadvantage of spectrum analysis lies on its insufficiency of processing non-stationary signal, which is usually the case of fault signal. Thus, feature extraction techniques based on time–frequency analysis have been developed recently, such as short-time Fourier transformation (STFT) [7], Wigner–Ville distribution (WVD) [8] and empirical mode decomposition (EMD) [9]. Wavelet transformation (WT) with its extensions is a new kind of time–frequency analysis techniques, which can represent a signal in time–frequency distribution diagrams with multi-resolution. It has been widely applied in machinery fault detection and diagnosis [10–12]. Some other fault feature extraction techniques are also developed including cyclostationary analysis [13,14], Spectral kurtosis [15,16] and so on.

Unfortunately, these popular feature extraction techniques presented above also have suffered problems. For example, FFT-based methods are not able to reveal the inherent information of nonstationary signal; and interference terms will appear on the time–frequency plane for the WVD method. Compared with traditional representation approaches, sparse decomposition provides much more succinct representations aiming to seek sparsest or nearly-sparsest representations of signal over particular over-complete dictionaries. It has been successfully applied on image processing [17,18] widely. Recently, sparse representation has been introduced into machinery fault detection. Liu [19] shows that matching pursuit (MP) works better than continuous wavelet transformation and envelope spectrum in bearing fault detection by employing MP with time–frequency atoms. Yang et al. [20] proved that basis pursuit (BP) can represent time–frequency features with fine resolution and sparsity thus rendering easier interpretation of bearing vibration signal. Feng [21] made a comparison of four sparse representation algorithms in the analysis of vibration signals from both healthy and faulty gearbox. Shi et al. [22] analyzed the transient vibration signal of rotating machinery in run-up stages based on adaptive time–frequency decomposition. Cui [23] proposed an adaptive matching pursuit method by using a new dictionary according to the characteristics of bearing fault.

Considering the computational efficiency, predefined dictionaries are usually used in the methods reviewed above. However, the performance of sparse representation in terms of the approximately quality and the sparsity of the coefficients depends not only the signal itself but also on the redundant dictionary. Besides that, in order to select proper predefined dictionary, we must have a priori knowledge about the target signal. Dictionary learning for sparse signal representations has drawn attentions from researches for its ability of learning dictionary. Related algorithms are proposed such as maximum likelihood (ML) [24], the method of optimal directions (MOD) algorithm [25], K-SVD [17] and so on. The advantage of learned dictionary resides in excellent interpretation of high-level structures in original signal. Shift-invariant dictionary implies that the same basis atom can be used to represent a feature and its related versions occurring in different positions within the signal [26–28]. This property of shift-invariance is very suitable to extract periodic impulses, which is typical symptom of mechanical fault signal. Liu [29] built a redundant dictionary and extracted sparse features by merging all learned sub-dictionary from each class of bearings data. Then the sparse features generated by counting the nonzero coefficients are used to realize bearing fault classification. Chen [30] proposed a new impulse extraction method based on adaptive dictionary learning, and used this method to detect gearbox fault. Tang [31] employed shift-invariant dictionary learning method to decompose a signal into a series of latent components, and detect the bearing or gear fault. Zhu [32] presents a new approach for cutting force denoising in micro-milling condition monitoring by using dictionary learning method.

Hidden Markov model (HMM) is an effective pattern recognition method which has drawn a lot of attentions from researchers. As a dual random process, hidden Markov model has two kinds of stochastic variable, namely hidden state and observation vector. The state sequence is unobservable, but can be estimated by the observation sequence. HMM has been widely used in speech recognition [33], visual recognition [34], fault diagnosis [35–37] and so on. This paper mainly introduces an effective approach based on shift-invariant dictionary learning and HMM for machinery fault diagnosis. The SIDL method is employed to learn basis atoms to capture the underlying structures in the noisy signal. Based on these basis atoms, the signal is decomposed into a collection of latent components, and the energy of each latent component is computed to form sparse feature set. Finally HMM is used to identify the bearing fault type.

This paper is set out as follows. The shift-invariant dictionary learning method with its algorithm is introduced in Section 2. Feature extraction based on shift-invariant dictionary learning is also presented in this section. Section 3 briefly reviews the principles of hidden Markov model. The bearing fault diagnosis model based on shift-invariant dictionary learning and HMM in presented in Section 4. In Section 5, a simulation example with double-impulse structure is analyzed using the SIDL method. And an experiment of bearing with specific notch size is also presented in this section. In Section 6, an experiment of bearing is conducted to verify the ability of the proposed SIDL-HMM approach in recognizing different fault types. Conclusions are made in Section 7.