Measurement 45 (2012) 1515–1526

Contents lists available at SciVerse ScienceDirect

Measurement

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

Denoising with advanced stepwise false discovery rate control and its application to fault diagnosis

Du Wenliao [a,b], Li Yanming [a], Yuan Jin [a], Liu Chengliang [a,*]

[a] State Key Laboratory of Mechanical System and Vibration, Shanghai Jiao Tong University, No. 800 Dongchuan Road, Shanghai 200240, China
[b] School of Mechanical and Electronic Engineering, Zhengzhou University of Light Industry, No. 5 Dongfeng Road, Zhengzhou 450002, China

ARTICLE INFO

Article history:
Received 23 June 2011
Received in revised form 13 February 2012
Accepted 20 February 2012
Available online 3 March 2012

Keywords:
Wavelet denoising
Multiple hypotheses testing
Advanced FDR procedure
Roller bearing
Fault diagnosis

ABSTRACT

To extract defect features from the signal with background noise for fault diagnosis, a novel approach is proposed by using advanced false discovery rate procedure (AFDR). The main idea is based on controlling false discovery rate (FDR) through combination of all three stepwise procedures (step-up, step-down, step-up-down) and estimation of the number of true null hypotheses. The AFDR procedure differs from the standard FDR procedure in two respects, i.e., enhancing the efficiency by reducing the number of tested hypotheses and improving the power. The proposed procedure consists of two main steps: firstly, the signal is represented more parsimoniously in wavelet domain; secondly, a most appropriate stepwise FDR procedure is selected according to the character of wavelet coefficients. Both the numerical simulation results and the experimental results for bearing defect diagnosis show that the proposed approach is a competitive shrinkage method compared with other popular techniques.

1. Introduction

Safety and reliability of mechanical system are essential for industry. The faults may break down the machine or deteriorate its performance. Failure detection has therefore received considerable attentions. Since vibration signals contain the dynamic characteristics of the machine condition, an increasing level in the vibration signature generally indicates a potential failure. The vibration signature contains a variety of information on many components and structures, such as gear meshing frequencies, bearings characteristic frequencies, and structural resonances [1]. Vibration-based analysis has been the most popular approach to detect the mechanical defect, especially for rotating machine. Al-Raheem et al. utilized the autocorrelation of vibration signal for rolling element bearing fault diagnosis [2]. Liu et al. proposed an extended wavelet spectrum analysis technique to assess bearing health conditions [3]. Wang et al. employed the dual-tree complex wavelet transform to vibration signal for rotating machinery fault diagnosis [4]. However, the vibration signals are often contaminated and the characteristic frequency is always immersed in noise. Thus, the contaminated signal denoising is one of important topics in signal processing field [5]. Due to its good localization and multi-resolution features in the time–frequency domain, the wavelet transform has been widely used [6,7].

The selection of threshold has an important impact on the denoising effect in wavelet denoising procedure. It is therefore desirable to develop a suitable threshold selection method. Since the seminal paper by Donoho and Johnstone [8], various alternative data-adaptive wavelet thresholding estimators have been developed [9–11]. Moreover, various Bayesian approaches for nonlinear wavelet thresholding have also been proposed [12–15]. However, a flexible threshold, based on a different selection of the significance level, is needed for the same wavelet coefficients in some situations, while the popular approaches cannot meet this requirement. From the statistical viewpoint, thresholding, as pointed out by Donoho and Johnstone [8], is closely

* Corresponding author. Tel.: +86 21 34204350; fax: +86 21 34206529.
E-mail addresses: dwenliao@zzuli.edu.cn (W. Du), chlliu@sjtu.edu.cn (C. Liu).

0263-2241/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.measurement.2012.02.027

related to hypotheses testing. The traditional concern in multiple hypotheses testing is to control the familywise error rate (FWER), which is too severe with lower power and reduces the probability of identifying real effects. Fortunately, the false discovery rate (FDR) procedure developed by Benjamini and Hochberg [16] controls the proportion of errors among rejected tests and provides a less conservative approach.

The wavelet thresholding based on multiple hypotheses testing process has been reported before [17–21], and Dudoit and van der Laan [22] provided a good review of the multiple hypotheses. Moreover, Efron [23] and Abramovich et al. [24–27] give a more intensive discussion about this issue. In the spirit of Tamhane's step-up-down procedure controlling FWER [28], we propose a novel wavelet denoising method by controlling false discovery rate with advanced stepwise procedure (AFDR), which integrates step-up, step-down and step-up-down procedures and enables to set the significance level flexibly according to the denoising result.

The main obstacle in wavelet denoising with step-up-down procedure is the absence of knowledge of the start point of the testing. To the best of our knowledge, the previous step-up-down approaches rarely pointed out the start point of the test explicitly. As the step-up-down procedure answers the question "Can at least r hypotheses be rejected?" substantially [28], it is reasonable that the test starts with $r = m_1$, which is the number of false null hypotheses. As $\hat{m}_1 = m - \hat{m}_0$, here, $\hat{m}_0$ is the estimation number of true null hypotheses. The estimation of m_0, should be as accurate as possible, and it is of great importance in AFDR. Several schemes have been reported to estimate m_0 in the literatures [21,29–32]. We denote these different methods as CSM (wavelet coefficients smoothness based approach) [21], LSL (lowest slope based approach) [29], FRR (fixed rejection region based approach) [30,31], PQU (quantile of p-values based approach) [32], respectively. In this paper, we also propose a method to determine the start point of the test according to the wavelet coefficients changing along the different scales. This approach is based on wavelet coefficients variance (CVA) for the estimation of m_0, and all the above estimators can be used into the AFDR procedure and then each of them will be compared in terms of performance.

The paper is organized as follows. Section 2 describes the advanced stepwise false discovery rate procedure. In this section, we also focus the estimation of the number of false null hypothesis, which is a critical factor for AFDR procedure. In Section 3, the adaptive wavelet thresholding based on AFDR procedure is introduced, which integrates step-up, step down and step-up-down procedures. Section 4 demonstrates the experimental results using the proposed method to simulation signals and roller bearings with an inner and an outer race fault, respectively. Section 5 contains a conclusion.

2. The advanced stepwise false discovery rate procedure

2.1. Background on stepwise false discovery rate procedure

The multiple hypotheses testing is more complicated than the single hypothesis testing. Table 1 gives the different results in m numbers of hypotheses testing. The classical method is to control the FWER, such as Bonferroni multiple hypotheses procedure. In many cases, especially when the number of tested hypotheses is large, controlling the FWER is too conservative and hampers its application. Benjamini and Hochoberg [16] introduced the false discover rate (FDR) criteria to measure the error rate of total test. The FDR is the proportion of rejected null hypotheses which are erroneously rejected. This can be described as

Table 1
The possible outcomes from m hypothesis testings.

	Declared null-significant	Declared significant	Total
True null hypothesis	U	V	m_o
Alternative hypothesis	T	S	m_1
Total	W	R	m

$$\mathrm{FDR} = E(V/R|R > 0)\Pr(R > 0). \tag{1}$$

In this definition, when $R = 0$, FDR = 0, which is mathematically meaningful; when $m_0 = m$, FDR = FWER; $m_0 < m$, FDR < FWER. If the procedure can control the FWER, it also can control the FDR; on the contrary, if the procedure can control the FDR, it cannot be sure to control the FWER. So the FDR procedure is more positive.

The FDR controlling procedures are typically stepwise in nature where the ordered p-values $p_{(1)} \leqslant p_{(2)} \leqslant \cdots \leqslant p_{(m)}$ are in effect compared with a series of properly chosen critical values. The step-up FDR procedure (BH procedure) presented by Benjamini and Hochberg [16] adopted the procedure mentioned by Simes in 1986 [33], which started with testing the least significant hypothesis with the largest p-value $p_{(m)}$ and continued with decreasing p-values until the first rejection of the null hypothesis. The critical constant for each p-value is $c_{(i)}=(i/m)q$, where, q is the significant level. Benjamini and Hochberg [29] later pointed out that the above step-up FDR procedure was also conservative when false null hypothesis existed. They proposed an adaptive FDR procedure based on the estimator $\hat{m}_0$ of the number of true null hypothesis and the critical constant is $c_{(i)} = (i/\hat{m}_0)q$. In addition, Benjamini and Liu [34] proposed a step-down FDR procedure, which started with $p_{(1)}$ and continued with increasing p-values until the first acceptance. They also verified the procedure was more powerful than the step-up one in the circumstances of small samples with most false hypothesis. The critical constant series are defined as: $c_{(i)} = 1 - \left[1 - \min\left(1, \frac{m}{m-i+1}q\right)\right]^{\frac{1}{m-i+1}}$. Further, Troendle [35] verified that both the step-up procedure and the step-down procedure can control the FDR at a required level if an appropriate critical constant was selected. Inspired by Tamhane's adaptive stepwise procedure controlling the FWER [28], in 2002, Sarkar [36] further introduced a step-up-down procedure to control the FDR and proved that the step-up procedure and the step-down procedure were consistent essentially.