Mechanical Systems and Signal Processing 46 (2014) 82–90

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

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

Structure damage detection based on random forest recursive feature elimination

Qifeng Zhou [a,b,*], Hao Zhou [a,b], Qingqing Zhou [a,b], Fan Yang [a,b], Linkai Luo [a,b]

[a] Department of Automation, Xiamen University, 361005 Xiamen, Fujian, China
[b] Center for Cloud Computing and Big Data, Xiamen University, 361005 Xiamen, Fujian, China

ARTICLE INFO

Article history:
Received 8 May 2012
Received in revised form
20 July 2013
Accepted 23 December 2013
Available online 31 January 2014

Keywords:
Damage detection
Wavelet packet decomposition
Random forest
Recursive feature elimination

ABSTRACT

Feature extraction is a key former step in structural damage detection. In this paper, a structural damage detection method based on wavelet packet decomposition (WPD) and random forest recursive feature elimination (RF–RFE) is proposed. In order to gain the most effective feature subset and to improve the identification accuracy a two-stage feature selection method is adopted after WPD. First, the damage features are sorted according to original random forest variable importance analysis. Second, using RF–RFE to eliminate the least important feature and reorder the feature list each time, then get the new feature importance sequence. Finally, *k*-nearest neighbor (KNN) algorithm, as a benchmark classifier, is used to evaluate the extracted feature subset. A four-storey steel shear building model is chosen as an example in method verification. The experimental results show that using the fewer features got from proposed method can achieve higher identification accuracy and reduce the detection time cost.

1. Introduction

The performance of civil structures deteriorates during their service life [1]. Damage, which includes changes of the boundary conditions and system connectivity or the reduction of stiffness etc., is defined as material or geometric property changes in the structural system. In order to extend the life of facilities more attention should be given to infrastructures and buildings.

Research in vibration-based damage identification has rapidly expanded over the last few years. In general, the associated signal analysis and pattern recognition algorithm are the two key factors in the development of vibration-based damage identification methods. Researchers have proposed a variety of damage detection methods. Especially, many data mining methods have been applied to this field, such as artificial neural network (ANN), support vector machine (SVM) and genetic algorithm (GA) etc. [2–6].

Feature selection, also known as variable selection, plays an important role in machine learning and data mining tasks. It is the process of selecting a subset of relevant features for use in model construction. Usually we only obtain the original acceleration signal in vibration-based damage identification methods. How these signals are converted into the input features of classifier is a key former step in structural damage detection. The mainly used signal processing approaches include Fourier transform, wavelet transform and wavelet packet transform. Among them, WPD as a more elaborate space-time processing

* Corresponding author at: Xiamen University, Department of Automation, Siming South Road, Xiamen 361005, Fujian, China. Tel/fax: +86 592 2580036.
E-mail addresses: zhouqf@xmu.edu.cn (Q. Zhou), zhoudahao2008@163.com (H. Zhou), dzqq0815@126.com (Q. Zhou), yang@xmu.edu.cn (F. Yang), luolk@xmu.edu.cn (L. Luo).

0888-3270/$ - see front matter © 2014 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ymssp.2013.12.013

method has been attracting much attention in the field of signal processing in recent years [7–9]. The energy of sub-bands and corresponding coefficients of WPD are selected as features. The number of layers of WPD determines the dimension of the feature vectors, and the outcome of feature selection will directly influence the accuracy and time cost of damage identification. However, the present studies simply use all of the WPD sub-bands, but ignore that assessing whether they are beneficial for the final detection.

This paper discusses a feature extraction method by integrating the WPD technology and RF–RFE approach. WPD is used to convert the acceleration signal into the initial energy feature set while RF–RFE is a method to eliminate the least important features and reorder the new feature importance sequence. The excellent performance and effectiveness of this method have been proved by the experiments on a four-storey steel shear benchmark building model.

2. Wavelet packet transform and damage features generation

2.1. Wavelet packet transform

Wavelet packet transform is an important generalization of wavelet transforms. Wavelet packet function can offer more flexibility than wavelet in representing different types of signals, and has broad prospect of application in the field of civil engineering [10–12].

2.2. Damage features generation

For N levels of decomposition, the WPD produces 2^N different band signals $S_{Nj}(j=1,2,\ldots,2^N)$ from the low frequency to the high frequency. The energy of each frequency band signal is defined as:

$$E_{Nj} = \int |S_{Nj}(t)|^2 dt = \sum_{k=1}^{n} |x_{jk}|^2 \tag{1}$$

where x_{jk} is the amplitude of kth discrete point in reconstructed signal S_{Nj} and n is the number of discrete point.

E_{Nj}, reflecting the original signal energy in selected frequency band and certain time interval, is significant for the analysis of non-stationary and time-varying signal [16]. The damage feature vector $T=(E_{N1},E_{N2},\ldots,E_{N2^N})$ is constructed according to the sequence $\{E_{Nj}|j=1,2,\ldots,2^N\}$. By normalizing T, we get the final damage feature vector $T'=(E'_{N1},E'_{N2},\ldots,E'_{N2^N})$.

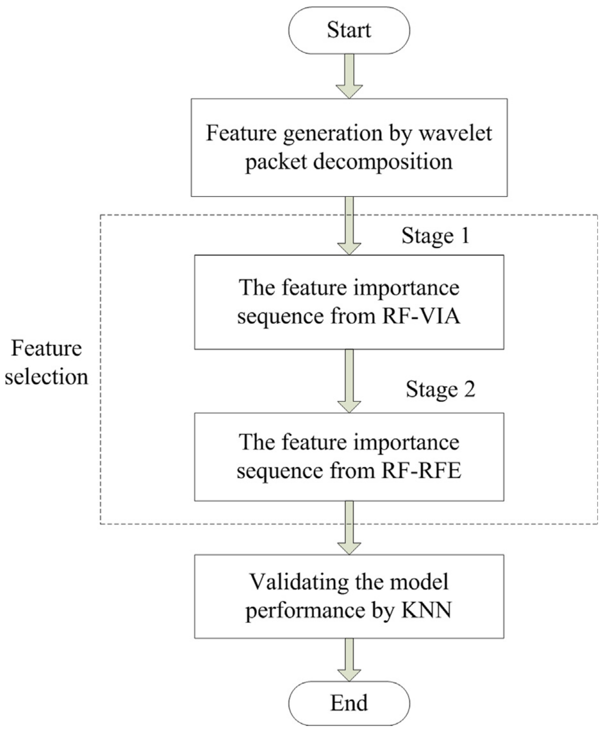

Fig. 1. RF–RFE damage identification model.