

基于随机森林的建筑结构损伤识别方法^{*}

周绮凤¹, 杨小青¹, 周青青¹, 雷家艳²

(1. 厦门大学自动化系 厦门, 361005) (2. 厦门大学建筑与土木工程学院 厦门, 361005)

摘要 针对利用分类器对建筑结构进行损伤识别的问题, 引入一种新的组合分类器算法——随机森林, 提出基于小波包分解和随机森林的结构损伤识别方法。首先, 采用小波包对结构在不同损伤程度和位置上的振动加速度信号进行分解, 得到各个频带上的总能量; 然后, 利用各频带上能量值存在着差异性作为输入到分类器的特征向量; 最后, 训练随机森林模型并对建筑结构的损伤位置和损伤程度进行识别。应用该方法对一座8层剪切型钢框架结构进行损伤判别, 并与BP神经网络和支持向量机方法进行对比, 结果表明该方法具有较好的识别精度与稳定性。

关键词 损伤识别; 随机森林; 小波包分解; 加速度信号

中图分类号 TP391; O327

引言

建筑在长期的使用过程中, 经历老化或是受到地震、台风等因素的影响, 结构会受到不同程度的损伤。目前数据挖掘中的很多方法被应用到建筑结构损伤识别领域^[1-6], 方法主要有神经网络和支持向量机等机器学习算法。

将一种泛化性好、抗噪性强且稳定性高的机器学习算法——随机森林(RF), 应用到上述领域。随机森林是Breiman于2001年提出的一个新的组合分类器^[7-9]。首先采用Bagging方法制造有差异的训练样本集, 并以分类回归树作为元分类器, 当构建单棵树时, 采用类似随机子空间划分的策略, 随机地选择特征对内部节点进行属性分裂。这种“双随机”的策略在各子分类器之间形成较大的差异性, 使随机森林具有优越的分类性能, 成为最成功的集成学习方法之一。

损伤识别中, 提取敏感信号有利于提高识别的精度, 而传统的信号处理方法难以提取对损伤敏感的特征。小波分析是一种全新的时频分析方法, 对非平稳信号具有宽频响应的特点, 在低频处有较高的频率分辨率, 在高频处有较高的时间分辨率, 适合分析非平稳信号^[10]。而小波包分解技术是从小波分析延伸出来的一种对信号进行更加细致的分析和重构的方法, 将频带进行多层次划分, 对多分辨分析没有

细分的高频部分进一步分解, 这些分解频带信号都具有一定的能量, 不同损伤的频带能量分布不同, 所以用能量的相对变换可以反映损伤类型。

将小波包变换与随机森林方法相结合, 进行损伤识别, 并对一座8层剪切型钢框架结构进行损伤诊断。

1 小波包分解及能量特征提取^[11]

小波包具有多维多分辨率分析的特点。频带进行多层次划分, 多分辨分析中没有细分的高频部分可进一步分解, 根据被分析信号的特征, 自适应地选择相应频段, 与信号频谱相匹配, 从而提高时频分辨率。

小波包分析算法和重构算法为

$$\begin{cases} d_k^m = \frac{1}{2} \int_L d_L^m g_{L-2k} \\ d_k^{2m+1} = \frac{1}{2} \int_L g_L^m h_{L-2k} \end{cases} \quad (1)$$

$$d_k^m = \int_L d_L^{2m} h_{k-2L} + \int_L d_L^{2m+1} g_{k-2L} \quad (2)$$

其中: h_k, g_k 为正交共轭滤波器系数。

对信号进行 N 层小波包分解时, 得到从低频到高频 2^N 个频带成分的特征信号 $D_{Nj}, j = 1, 2, 3, \dots, 2^N$ 。

计算小波包能量包括对小波包分解系数进行重构, 提取各频带范围的信号, 计算各频带信号的能量。

* 福建省自然科学基金资助项目(编号: 2011J01373); 中央高校基本科研业务费专项资金资助项目(编号: 2010121065)

收稿日期: 2010-12-13; 修改稿收到日期: 2011-03-21

量。各频带信号的能量为

$$E_{N_j} = \int |D_{N_j}(t)|^2 dt = \sum_{k=1}^n |d_{jk}|^2 \quad (3)$$

其中: d_{jk} 为重构信号 D_{N_j} 的第 k 个离散点的幅值; n 为重构信号 D_{N_j} 的离散点的个数 $j = 0, 1, \dots, 2^N - 1$; $k = 1, 2, \dots, n$ 。

2 随机森林原理

随机森林是由许多单棵分类回归树(CART)组合而成的,最后由投票法决定分类结果。Breiman 采用 Bagging 和 Randomization 相结合的方法,在保证单颗树分类效能的同时,减少各分类树之间的相关度,提高了组合分类器的性能。

2.1 Bagging 方法形成新的训练集

假设原始训练集的样本数为 N , Bagging 方法有放回地随机从原始训练集中抽取 N 个样本,组成一个新的训练集。通过简单计算得知,每次产生的新训练集中有近 37% 的数据可能未被选中,这部分数据称为袋外数据(OOB),可以用来作为测试数据对该分类树的泛化性能进行估计,该方法称为“Out-Of-Bag Estimation”。选择 Bagging 方法生成新的训练集有两个优点: a. 可以用 OOB 估计计算泛化误差、各分类树的效能、分类树之间的相关度以及各输入特征的重要性; b. Bagging 方法与 Randomization 方法相结合能增加随机森林的分类准确性。

2.2 Randomization 方法

随机森林的重要特征是针对树的内部节点随机地选择特征进行分裂,用 CART 方法生成单颗分类树。每棵树任其发展,不需要剪枝,直至叶子节点。这样可以增加单颗树的分类效能,同时增加各分类树之间的差异性。随机选择特征分裂有两种方式: a. Forest-RI 先确定用于每次分裂的候选特征的个数 F , 然后随机地从特征全集中选出 F 个特征,再根据最优分裂准则对节点进行分裂; b. Forest-RC 随机选出 L 个特征,再随机地选择系数对其进行线性组合,生成 F 个新特征,然后根据最优分裂准则对节点进行分裂。

Dietterich 通过试验证明^[12], Bagging 和 Randomization 都能有效降低噪声的影响;因此,两者的结合使得随机森林具有良好的容忍噪声能力。

2.3 随机森林的泛化误差

为表示决策树节点特征的随机向量,基于 X 和 Θ 的分类器输出记为 $h(X, \Theta)$ 。定义随机森林的间隔函数为 $mr(X, Y) =$

$$P_{\Theta}(h(X, \Theta) = Y) - \max_j P_{\Theta}(h(X, \Theta) = j) \quad (4)$$

间隔函数 $mr(X, Y)$ 表示样本数据被分对与分错的概率之差。间隔函数的值越大,表明分类器的泛化性能越好。

组合分类器 $\{h(X, \Theta)\}$ 的总体分类效能 s 定义为

$$s = E_{X,Y} mr(X, Y) \quad (5)$$

若用 $\bar{\rho}$ 表示各个分类树之间相关度的平均值,则可得到随机森林的泛化误差 PE^* 的上界

$$PE^* \leq \bar{\rho}(1 - s^2)/s^2 \quad (6)$$

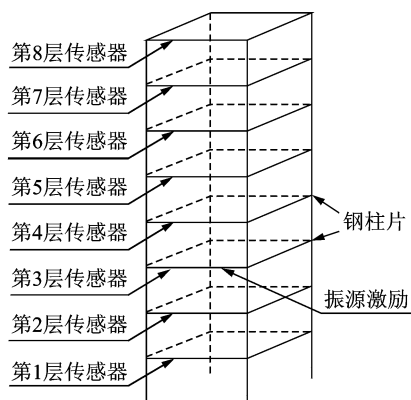
显然,为使组合分类器能达到好的泛化性能,应尽量增大单颗分类树的效能,减小分类树之间的相关性。可以证明:假定 $s > 0$, 当森林中的分类树足够多时,随机森林的泛化误差几乎处处收敛于一个有限值;因此,随着森林中分类树数目的增长,随机森林算法并不会导致过拟合。

3 结构的损伤诊断

3.1 8层剪切型钢框架结构

8层剪切型钢框架结构如图1所示,每层钢框架平面边长为 $350 \text{ mm} \times 250 \text{ mm}$, 层高为 200 mm 。结构损伤表现为某一层间刚度的减少,试验中通过变换不同重量的钢柱片模拟刚度的降低。激振点位于第3层框架处,振源是利用 RIGOL 公司生产的 DG-1022 信号发生器产生的白噪声激励。在激励的作用下,模型会出现轻微的摆动。每层框架均布设一个 PCB 传感器,用以采集响应时程。设定采样频率设为 128 Hz , 采样时间为 40 s 。

结构以钢柱片发生损伤情况作为研究对象,共



有10种损伤情况:

- 1) 第3层框架发生25%的损伤;
- 2) 第3层框架发生50%的损伤;
- 3) 第5层框架发生25%的损伤;
- 4) 第5层框架发生50%的损伤;
- 5) 第7层框架发生25%的损伤;
- 6) 第7层框架发生50%的损伤;
- 7) 第3层和第5层框架同时发生25%的损伤;
- 8) 第3层和第7层框架同时发生25%的损伤;
- 9) 第5层和第7层框架同时发生25%的损伤;
- 10) 8层全部无损伤。

3.2 损伤特征提取

考虑环境噪声因素的影响,对每一种损伤情况的振源激励设定不同的白噪声水平。小波包损伤特征提取的步骤为:

1) 利用正交小波为小波基函数,将每层振动的加速度信号 $x(t)$ 进行5层小波包分解,分别得到从低频到高频32个频率成分的信号特征。

2) 根据式(3)计算各频带的能量,利用各频带的能量为元素构造一个特征向量 S

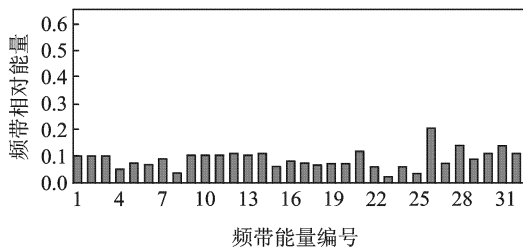
$$S = [E_{50}, E_{51}, \dots, E_{5j}, \dots, E_{531}] \quad (7)$$

分别对每一层数据得到的特征向量进行归一化处理,作为训练和测试分类器性能的损伤样本。

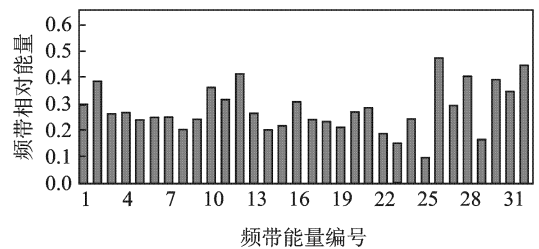
对每一种损伤情况做10组试验,其中每组振源激励加入不同强度的白噪声,每一层的加速度传感器记录各层的加速度变化值。这样,由每个加速度传感器采集的10种损伤情况的加速度响应样本共计100个。对每个样本按照以上步骤进行损伤特征提取得到一个32维的特征向量。

由第2层加速度传感器得到第1,2,3,4和7种和第10种部分损伤情况的小波包相对能量分布如图2所示。

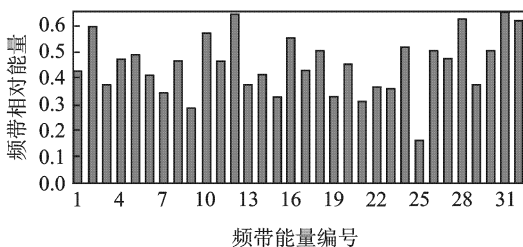
由图2可以看出,在不同的损伤状况下,同1层的加速度信号经过小波包分解后,各频带的能量大小和分布不同,而且差异较明显。由此说明振动信号经过小波包分解得到的特征向量对损伤信号敏感,有利于分类器对损伤情况的识别。



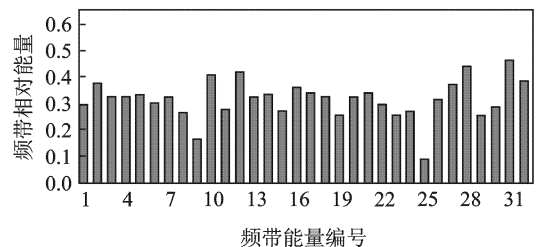
(a) 第3层发生25%的损伤



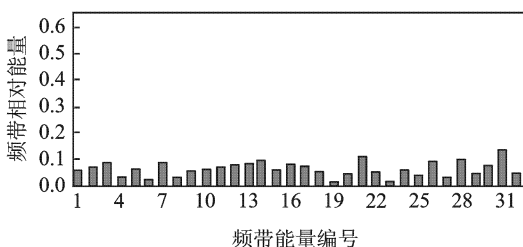
(b) 第3层发生50%的损伤



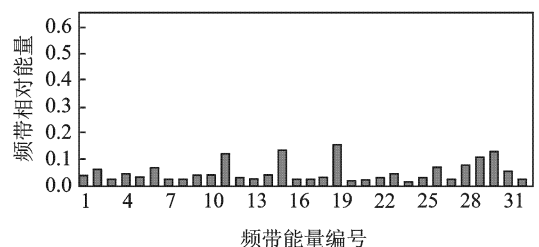
(c) 第5层发生25%的损伤



(d) 第5层发生50%的损伤



(e) 第3层和第5层同时发生25%的损伤



(f) 8层全部无损伤

图2 第2层6种损伤情况下小波包相对能量分布

3.3 基于随机森林的损伤识别

每一层传感器收集到 100 个样本, 包含 10 种损伤情况, 每种损伤情况含有 10 个样本。实验时, 对每一层数据, 每次随机选取每种损伤情况样本的 70% 作为训练随机森林的数据, 剩下的 30% 作为测试数据。进行多次试验, 调整模型参数。在随机森林中, 分类树的数目设为 2 000, 每个内部节点的候选特征数取 3。对 2 层数据全部做训练和测试, 测试准确率取 10 次平均, 结果如表 1 所示。

表 1 给出了利用 BP 神经网络和支持向量机 (SVM) 作为分类器对各层数据的损伤识别准确率。各分类器模型的参数均是进行多次试验调整后选取较优的。其中 BP 模型采用一个隐层, 隐层节点数设为 21, 训练代数设为 100, SVM 采用高斯径向基核函

数, C 和 γ 通过 5 折交叉验证获得。由表 1 可以看出, 振动信号经过 5 阶小波包变换后, SVM 的识别效果在每一层数据的准确率上都要明显高于神经网络。随机森林在第 1 层、第 4 层、第 5 层、第 6 层、第 7 层和第 8 层数据的判准率上好于 SVM, 并且这 6 层的准确率都达到较高的数值。在第 1 层和第 3 层数据, SVM 的分类效果要稍好于随机森林。

试验过程中, 神经网络模型对参数的改变较为敏感。在多次试验中, 其准确率波动比较大。如图 3 (a), (b), (d) 所示的部分数据层中, 分别用 3 种模型做 10 次随机试验, 神经网络模型准确率变化比较大且不稳定; 支持向量机的结果相对好一些; 而随机森林最好, 准确率最稳定。图 4 表明 3 种模型在各层数据中准确率的标准差。

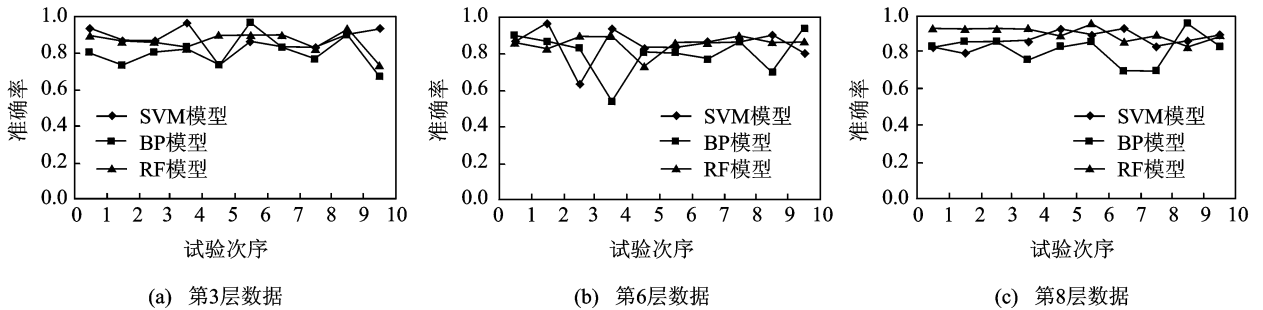


图 3 数据在 3 种模型准确率的分布

表 1 损伤预测准确率比较

分类器模型	第 1 层	第 2 层	第 3 层	第 4 层	第 5 层	第 6 层	第 7 层	第 8 层
BP 网络	0.813 4	0.796 7	0.780 0	0.816 7	0.826 7	0.800 0	0.780 0	0.823 3
SVM	0.920 0	0.943 3	0.873 3	0.896 7	0.920 0	0.850 0	0.900 0	0.870 0
随机森林	0.923 3	0.940 0	0.866 7	0.913 3	0.955 6	0.860 0	0.909 5	0.910 3

由图 3、图 4 可知, 随机森林模型具有较好的稳定性和较高的准确率, 支持向量机模型次之, 而神经网络模型的稳定性较差。试验结果表明, 小波包变换结合随机森林的损伤识别方法具有较强的稳定性和分类能力。

4 结束语

随机森林是一个组合分类器算法, 能有效处理含噪声的数据。该算法对于其泛化性有严格的数学证明, 不会过拟合。将其用于识别建筑结构损伤, 通过与 BP 网络、支持向量机方法对比, 结果表明, 小波包分解结合随机森林的损伤识别方法用在 8 层剪切型钢框架结构上, 具有较好的损伤识别精度和稳定性。

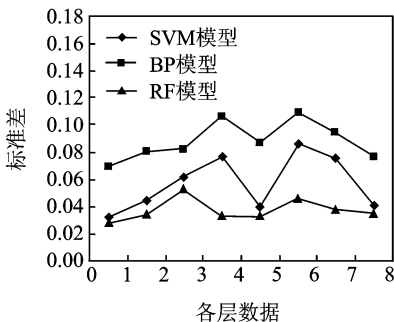


图 4-3 种模型在各层数据中准确率的标准差

参 考 文 献

[1] Hojjat A. Counterpropagation neural networks. in Electronic Publishing House. All rights reserved. <http://www.cnki.net>

structural engineering [J]. Journal of Structural Engineering, 1995, 121(8): 1205-1212.

[2] Kaminski P C. The approximate location of damage through the analysis of natural frequencies with artificial neural networks [J]. Journal of Process Mechanical Engineering, 1995, 209: 117-124.

[3] 何浩洋, 闫维明, 彭凌云. 基于支持向量机的钢筋混凝土桥梁损伤识别[J]. 公路交通科技, 2008, 2(3): 65-69.

He Xiaoxiang, Yan Weiming, Peng Lingyun. Damage identification of reinforced concrete bridge based on support vector machine[J]. Journal of Highway and Transportation Research and Development, 2008, 2(3): 65-69. (in Chinese)

[4] 陈素文, 李国强. 人工神经网络在结构损伤识别中的应用[J]. 振动、测试与诊断, 2001, 21(2): 116-124.

Chen Suwen, Li Guoqiang. Application of artificial neural networks to damage identification of structure [J]. Journal of Vibration, Measurement & Diagnosis, 2001, 21(1): 116-124. (in Chinese)

[5] 何浩祥, 闫维明, 周锡元. 小波支持向量机在结构损伤识别中的应用研究[J]. 振动、测试与诊断, 2007, 27(1): 53-57.

He Haoxiang, Yanweiming, Zhou xiyuan. Application of wavelet support vector machine to structural damage detection [J]. Journal of Vibration, Measurement & Diagnosis, 2007, 27(1): 53-57. (in Chinese)

[6] 赵学风, 段晨东, 刘义艳, 等. 基于小波包变换的支持向量机损伤诊断方法[J]. 振动、测试与诊断, 2008, 28(2): 104-107.

Zhao Xuefeng, Duan Chendong, Liu Yiyan, et al. Diagnosis of structure damage by support vector machine based on wavelet packet transform [J]. Journal of Vibration, Measurement & Diagnosis, 2008, 28(2): 104-107. (in Chinese) [7] Breiman L. Bagging predictors [J]. Machine Learning, 1996, 24(2): 123-140.

[8] Breiman L, Friedman J H, Olshen J H, et al. Classification and regression trees [M]. Cole: Wadsworth & Brooks, 1984: 1-70.

[9] Breiman L. Bagging predictors [J]. Machine Learning, 1996, 24(2): 123-140.

[10] 何正嘉, 孟庆丰, 赵纪元. 机械设备非平稳信号的故障诊断原理及应用[M]. 北京: 高等教育出版社, 2001: 2-105.

[11] 徐章遂. 故障信息诊断原理及应用[M]. 北京: 国防工业出版社, 2001: 52-56.

[12] Dietterich T. An experimental comparison of three methods for constructing ensembles of decision tree bagging boosting and randomization [J]. Machine Learning, 1998, 26: 1-22.



第一作者简介: 周绮凤, 女, 1976年10月生, 博士、副教授。主要研究方向为模式识别及其应用。曾发表《内嵌空间排序支持向量机及其在文本检索中的应用》(《信息与控制》2010年第39卷第5期)等论文。

E-mail: zhouqf@xmu.edu.cn

Fault Feature Extraction Method for Rotor System Under Radial Impact Based on BF-HHT

Xiong Xin, Yang Shixi, Gan Chunbiao

(The State Key Lab of Liquid Power Transmission and Control, Zhejiang University Hangzhou, 310027, China)

Abstract As radial impact signals are often composed of rich frequency components continuously distributed in a wide region, accurately extracting the characteristic frequencies of vibration signals of a rotor system under radial impact is of great significance for fault diagnosis. The customary Hilbert-Huang transform (HHT) cannot be directly used to decompose such signals into several mono-frequency intrinsic mode functions (IMFs). Because of the existence of multi-frequency IMFs, the estimations of instantaneous frequencies for these signals are usually inaccurate. A new feature extraction method is developed in this study, i. e., the bandpass filtering HHT (BF-HHT), which can accurately extract the characteristic frequencies of radial impact signals. Firstly, frequencies with main power are collected through calculating power spectrum densities of the signals. Secondly, a set of bandpass filters with centering frequencies selected from the first step are designed to filter the original signals. Finally, the time-frequency characteristics are extracted from the filtered signals by HHT. Both the customary HHT and the BF-HHT are used to analyze the signals collected from rotor test bench and industrial sites. Comparisons between the results from both the methods show that, the BF-HHT can be used to extract the characteristic frequencies of radial impact signals, meanwhile, it can explore the instantaneous characteristic frequencies.

Keywords bandpass filtering, Hilbert-Huang transform, rotating machinery, radial impact, time-frequency feature

Damage Identification Technique for Building Structure Based on Random Forest

Zhou Qifeng¹, Yang Xiaoping¹, Zhou Qingqing¹, Lei Jiayan²

(1. Department of Automation, Xiamen University Xiamen, 361005, China)

(2. School of Architecture & Civil Engineering, Xiamen University Xiamen, 361005, China)

Abstract For the use of classifier in the structure damage diagnosis, a structural damage identification method is proposed based on wavelet packet decomposition and random forest which is a new combination classifier algorithm. Wavelet package decomposition is used to decompose the vibration acceleration signals of building structure with different damage degrees and locations. The energy sequences at different bands of frequencies decomposed by the wavelet packet decomposition are inputted to classifier as feature vectors. A random forest model is trained and used to identify the location and degree of injury. The method is applied for damage identification of an eight-story shear steel frame model. Experimental results show that the method has good recognition accuracy and stability compared with BP neural network and support vector machines.

Keywords damage identification, random forest, wavelet package decomposition, acceleration signal