

基于随机森林算法的结构损伤识别

■ 谢坤明

(福建省永正工程质量检测有限公司, 福州 350011)

摘 要 传统结构损伤识别需对采集数据进行复杂分析,提取相应力学特征来识别结构的损伤。利用传统方法或人工神经网络模型特征提取过程需消耗大量的计算成本,导致结构健康监测在线损伤识别存在一定的困难。为提高损伤识别的计算效率和自动化程度,提出随机森林算法的损伤识别方法,其特点是通过原始振动信号的简单数据特征提取,准确识别结构的损伤位置和损伤程度。研究设计了5根具有不同脱空状态的钢管混凝土构件,采用随机森林算法对钢管混凝土构件的脱空情况进行了识别。结果表明,随机森林算法可准确识别几种脱空工况,准确率为100%,验证了方法在实际结构损伤识别的应用可行性。

关键词 结构 健康监测 损伤识别 振动响应 机器学习 随机森林

随着结构健康监测系统在土木工程建设及运营过程中的逐渐运用与完善,结构健康监测技术正成为土木工程领域的重要研究热点之一。其中基于动力特性的结构损伤识别方法在试验阶段取得了较好的效果,如基于曲率模态、柔度模态、应变模态等损伤特征的传统识别方法^[1]。然而,由于实际测量数据的误差,这些损伤指标未能在实际监测工作上展现出良好的稳定性。

近年来,结构健康监测系统通过熟练的信息采集技术获得了大量的监测数据。因此,基于大数据挖掘的结构损伤检测技术也正逐渐得到学者重视,其主要思想是利用机器学习对信号特征参数提取,获取结构的敏感特征并完成结构工作状态的分类评估任务。许多学者采用人工神经网络(ANN, Artificial Neural Network)作为结构损伤识别的分类器。Zhongxian Li等^[2]利用ANN中的反馈型神经网络对三跨连续梁的数值模拟模型进行了损伤识别研究,分析结果显示ANN在损伤定位具有有效性但对于小损伤工况下的识别效率不高。Pathirage等^[3]利用基于自编码的结构损伤识别框架,提高了ANN模型在多类损伤下识别的准确率。龙云鹏^[4]利用机器学习算法开展了对混凝土徐变模型的修正和更新;Gui等^[5]提出了一种基于支持向量机SVM方法的框架结构损伤识别方法,结果表明,所提出的方法能有效识别建筑物损坏及完好的2种状态。

随着数据复杂程度的提升,人工神经网络模型由于

所需训练时间及反馈时间较长,不适应于当今及时反应和快速反应的监测要求。而在解决分类问题与回归问题的情况下,集成学习算法相比ANN模型、SVM算法等单个分类器,能有效处理过拟合问题,有效减少了所占用的运行资源。集成分类器通过剔除复杂的优化过程达到灵活处理数据分类的目的。Breiman提出了随机森林(Random Forests)算法^[6]作为bagging集成学习^[7]中性能最优异的算法,其因简单的参数构成及强大的适应能力被应用于不同监测系统的识别任务中。董伟广等^[8]验证了随机森林算法在变压器故障预测上的高准确率。李兵等^[9]将随机森林算法应用于电机轴承信号故障诊断,实验证明了随机森林算法针对振动信号分类问题时的有效性与可行性。

根据以上分析,针对实际施工过程中的钢管混凝土构件中脱空缺陷监测问题,本文提出一种基于随机森林算法的结构损伤检测方法。该方法利用随机森林算法对重组特征属性后的动力试验数据进行分类识别,实验结果表明,该方法在实际结构损伤状态下仍具有识别的有效性。

1 脱空缺陷识别算法原理

1.1 决策树

决策树(Decision Tree)^[10]是基于树结构的一类决策算法,其中树内部结构中的各节点表示单个特征属性划分点,每个分支表示划分后输出结果,树结构末端的每个

叶节点表示每种类别,如图 1 所示。

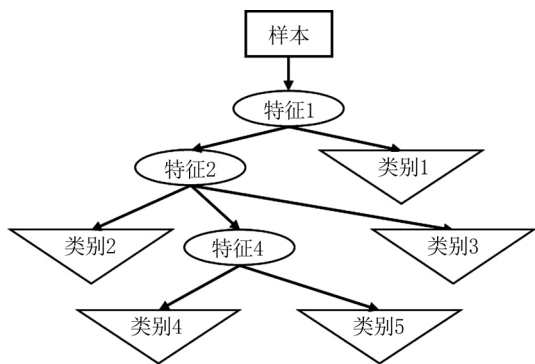


图 1 决策树结构图

单棵决策树的生成过程由收集样本、特征选取、特征分裂 3 个部分组成。其中特征选取标准决定了决策树的模型的优劣, 本文将分类与回归树 (classification and regression tree, CART) 算法^[11]作为特征选取标准, 该算法基于 Gini 指数选择最优特征构建二叉决策树。

对于给定训练集 D 的 Gini 指数定义为:

$$Gini(D)=1-\sum_{k=1}^k\left(\frac{|C_k|}{|D|}\right)^2 \quad (1)$$

其中, k 表示含 k 个样本类别, C_k 表示在样本 D 中类别 k 类样本所占比例。

样本 D 在通过特征 X_i 节点时, 样本集 D 都将被划分为 D_1 和 D_2 2 个数据集, 其中特征 X_i 在集合 D 中的 Gini 指数定义为:

$$Gini(D, X_i)=\frac{|D_1|}{|D|}Gini(D_1)+\frac{|D_2|}{|D|}Gini(D_2) \quad (2)$$

通过选择 $Gini(D, X_i)$ 为最小时的特征 X_i 及其对应切分点最为最优特征与最优切分点, 根据切分点将输入该节点的数据集按该特征划分为 2 个数据集并分配到至下 2 结点, 重复计算 Gini 指数完成针对于给定训练集 D 的单棵决策树的生成。

1.2 Bagging 集成学习

在机器学习中, 单个弱分类器适应性差, 易受数据集影响并造成过拟合现象的发生。为解决这一问题, Bagging 集成学习法在训练阶段利用自助采样法对训练集进行有放回的随机抽样, 将随机抽样得到 n 个子训练集训练并生成对应弱分类器, 随后, 该法通过对 n 个分类器的并行组合完成 1 个强分类器的构建, 如图 2 所示。在测试阶段, 该强分类器将测试数据放入各分类器中, 采用投票方式得到最终预测结果。

1.3 随机森林算法

在分类算法的实际应用过程中, 为解决决策树模型中常出现的过拟合问题, 随机森林算法基于决策树模型和集成学习思想, 利用 Bootstrap 重采样法对每次抽取数

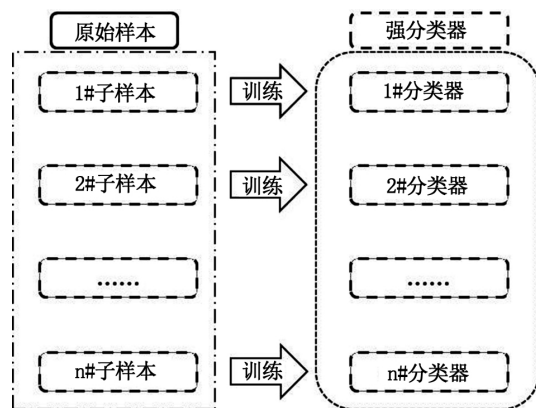


图 2 集成学习方法原理

据进行决策树模型的训练与产生, 通过随机抽取样本和随机选择分类特征增强对样本的扰动, 以及模型泛化能力, 提高随机森林模型的准确度。算法具体实现步骤如下:

(1) 对样本数 M 的原始数据集, 通过 Bootstrap 重采样方法随机且有放回地选择 N 个样本作为单个决策树的训练样本。其中每个样本含 S 个特征类别, 即构成 $N \times S$ 的子样本矩阵。

(2) 从生成训练子集随机挑选 s 个特征类别作为决策树的各个节点, 利用对样本及样本属性的 2 次随机抽取, 减弱各决策树之间的相关性, 其中 s 应小于 S 。

(3) 重复步骤 (1)、(2) n 次, 生成 n 棵决策树, 组成 RF 模型。

(4) 将测试集中的各样本分别代入 RF 模型中的 n 棵决策树中进行预测分类并得到 n 个预测结果, 根据投票机制得到测试样本通过 RF 模型后分类后的预测结果。

基于随机森林算法的结构损伤识别流程如图 3 所示, 具体如下:

(1) 对脱空圆钢管试件采取动力测试, 通过敲击获取在各测点在不同损伤工况下的原始加速度响应并建立原始数据集。

(2) 将数据集按测点位置及脱空状态进行人为标记及划分训练集和测试集。

(3) 将训练集代入随机森林算法中进行分类训练并得到随机森林模型。

(4) 将测试集代入随机森林模型中, 利用投票机制得到概率最大类别并作为输出结果。

2 圆钢管混凝土脱空缺陷识别试验

2.1 圆钢管混凝土振动试验

文献[12]设计了针对圆钢管混凝土脱空缺陷的振动试验, 本次算例以该试验为研究对象, 利用 5 根圆钢管混

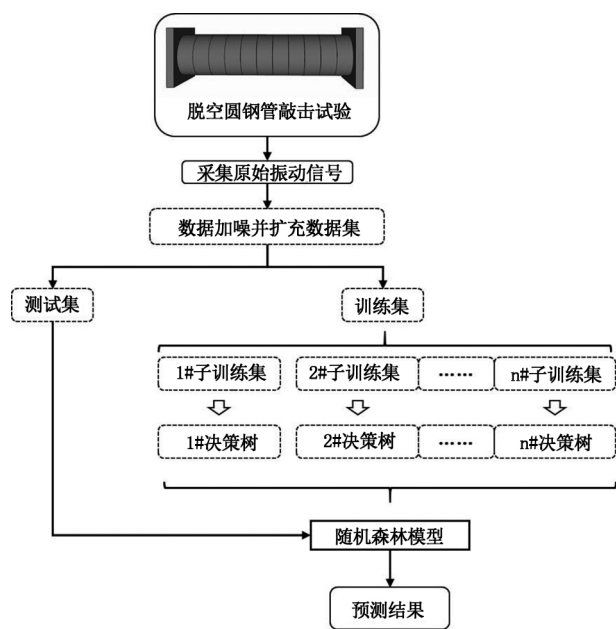
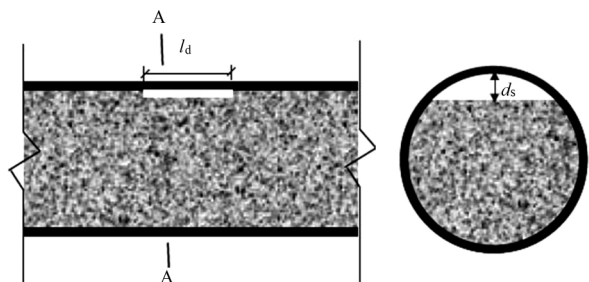


图3 基于随机森林算法的结构损伤识别流程图

混凝土试件进行钢管混凝土振动试验,包括健康状态试件1根,带冠形脱空试件4根,试件缺陷形式如图4所示,其中试件总长 $L=1200\text{ mm}$,试件直径 $D=150\text{ mm}$,局部脱空圆钢管混凝土试件脱空值 $d_s=8\text{ mm}$ 。



(a)局部脱空圆钢管混凝土试件 (b) A-A脱空部位截面

图4 圆钢管混凝土冠形脱空缺陷形式示意图

如图5所示,将钢管均匀分成11个单元,其标号为C0T-C4T并按顺序排列,其中,试件C0T是作为参考的无脱空试件;C1T、C2T为2个不同位置的1处脱空;C3T、C4T均有脱空。为设置实际脱空工况,笔者采用3D打印技术制作完成脱空模具,将模具与钢管内壁粘贴后浇筑混凝土,以此模拟实际工程中存在不同数量的局部脱空缺陷情况。试件脱空工况设置如表1所示。

将钢管混凝土试件平放在地上,两端各垫2个橡胶支座。在每个单元的上部中点设置测点,采用力锤敲击跨中单元,每个单元中心所布置的传感器记录测点受激励后的加速度时程信号。力锤激励的时间通常为几十毫秒,若不能完整地采集到其激励时程信号,在后续的数据中可能会出现严重的偏差。此外,钢管混凝土的刚度较大,为充分地记录到信号特征,将敲击力和加速度的采集



图5 试件脱空位置示意图

表1 试件脱空工况设置

工况号	脱空单元	脱空类型
1	3	无脱空
2	6	无脱空
3	9	无脱空
4	6	单处脱空
5	9	单处脱空
6	6,9	两处脱空
7	3,6,9	三处脱空

频率设置为 20000 Hz 。根据加速度的变化情况,将记录时间设置为 2 s ,以确保能在振动结束之后还能采集。



图6 钢管混凝土振动实验装置图

2.2 随机森林算法应用

通过以上试验所测得的钢管混凝土表面加速度曲线,笔者采集 0.25 s 内的数据进行处理,即以 5000×1 的数据长度作为单条样本,数据所处环境干扰,对试验敲击所产生的数据添加 10% 的高斯白噪声以达到扩充和丰富数据集的目的,建立 1000 条样本的数据集。

对训练集和测试集利用时序数据构造新特征,取各条样本均值、标准差、方差、最小值、最大值、中位数、偏度和峰态系数作为 8 个特征属性。将 1000×5000 的数据矩阵转化为 1000×8 的维度,时域特征提取运行时间为 21.07 s 。

在训练阶段按照 $6:2:2$ 的比例将数据集划分为训练
(下转第104页)

在于健康监测系统采用的设备复杂，前期投入的成本较高；但从整个监测周期看，总成本与常规监测成本相持平。随着今后大规模的应用，成本将逐年下降。

5 结语

永安火电厂大桥由于交通环境的变化在 1 年半时间内承受比以往更大的交通量和交通荷载，为了保证桥梁的健康安全采用远程实时健康监测系统。该系统集成了传感器系统、数据采集与传输系统、数据管理系统、桥梁分析与预警系统等主要功能模块，相比传统的常规监测技术优势明显。经过 1 年半的监测实践，该系统数据采集反应灵敏，数据传输平稳，数据分析和预警功能正常，较

(上接第 93 页)

集、验证集、测试集。选择决策树个数为 50 生成随机森林模型。利用 bootstrap 重采样技术对样本进行迭代训练，得到用于圆钢管混凝土脱空缺陷识别的随机森林模型，训练过程运行时间为 0.09 s。

在测试阶段，将测试集代入随机森林模型进行预测，模型反馈时间为 0.01 s，准确率为 100%。

表 2 模型运行时间

步骤	运行时间/s
特征提取	21.07
训练模型	0.09
测试模型	0.01
合计	21.17

3 结论

本文提出了一种基于随机森林算法的结构损伤识别方法。以圆钢管混凝土脱空缺陷试验作为研究对象，采集表面加速度建立数据集，通过随机森林算法建立用于圆钢管混凝土脱空缺陷的识别模型，输入测试信号完成损伤识别，得到以下结论：

(1) 基于随机森林算法的结构损伤识别模型适配性较强，硬件要求低。模型完成及测试时间反馈时间较短，反馈时间分别为 21.97s 及 0.01s，能基本满足实时检测基本要求。

(2) 在实际测试中识别效率较高，准确率为 100%，具备一定的抗噪性能。

好地完成桥梁健康监测工作，并在今后的运营中继续发挥作用。

参考文献

- [1]武智霞,韩鹏,郑树泉,等.桥梁健康监测及养护平台设计与实现[J].计算机应用与软件,2018(5):109-114.
- [2]任普,丁幼亮,李亚东,等.基于大数据的桥梁健康监测数据存储及预警方法[J].科学技术与工程,2019(19):266-270.
- [3]郑毛祥.高速铁路无线桥梁健康监测系统研究[J].铁道工程学报,2017(34):69-72.
- [4]曹素功,黄立浦,张勇,等.桥梁健康监测数据的结构化存储与分析[J].岩石力学与工程学报,2019(39):102-108.

参考文献

- [1]朱宏平,余璟,张俊兵.结构损伤动力检测与健康监测研究现状与展望[J].工程力学,2011,28(2):1-17.
- [2]Zhongxian Li,Xiaoming Yang.Damage identification for beams using ANN based on statistical property of structural responses [J].Computers & Structures,2008(86):64-71.
- [3]Chathurdara Sri NadithPathiragea, Jun Li, Ling Li, et al.Structural damage identification based on autoencoder neural networks and deep learning[J].Engineering Structures,2018(172):13-28.
- [4]龙云鹏.基于机器学习的混凝土徐变模型研究[D].北京:北京交通大学,2019.
- [5]GuoqingGui, Hong Pan,Zhibin Lin, et al.Data -Driven Support Vector Machine withOptimizationTechniques for Structural Health Monitoring and Damage Detection [J].Design Optimization and Applications in Civil Engineering, 2017,21(2):523-534.
- [6]Breiman L.Random forests[J].Machine Learning,2001,45(1):5-32.
- [7]Breiman L.Bagging Predictors[J].Machine Learning,1996,24(2):123-140.
- [8]董伟广,钟建伟,张钦惠,等.基于数据挖掘技术和随机森林算法对变压器的故障诊断[J].电网运维,2020(3):54-56.
- [9]李兵,韩睿,何怡刚,等.改进随机森林算法再轴承故障诊断钟的应用[J].中国电机工程学报,2020,40(4):1310-1319.
- [10]唐华松,姚耀文.数据挖掘中决策树算法的探讨[J].计算机应用研究,2001(8):18-19,22.
- [11]刘勇洪,牛铮,王长耀.基于 MODIS 数据的决策树分类方法研究与应用[J].遥感学报,2005,9(4):405-412.
- [12]廖飞宇,王静峰.钢管混凝土的脱空问题——检测、设计和加固[M].北京:中国建筑工业出版社,2020.