

基于 LAL-AdaBoost 算法的空间网架 损伤识别研究

A Damage Identification Approach for Spatial Grid Structures Using LAL-AdaBoost

一级学科：____ 土木工程 ____

研究方向：____ 结构工程 ____

作者姓名：____ 党大智 ____

指导教师：____ 徐 杰 副教授 ____

答辩日期	2021 年 12 月 6 日		
答辩委员会	姓名	职称	工作单位
主席	尹越	副教授	天津大学
委员	芦燕	副教授	天津大学
	金明昌	正高级工程师	天津市房屋鉴定建筑设计院
	王中兴	副研究员	天津大学

天津大学建筑工程学院

二〇二一年十二月

独创性声明

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作和取得的研究成果，除了文中特别加以标注和致谢之处外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得天津大学或其他教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文作者签名： 签字日期： 年 月 日

学位论文版权使用授权书

本学位论文作者完全了解 天津大学 有关保留、使用学位论文的规定。特授权 天津大学 可以将学位论文的全部或部分内容编入有关数据库进行检索，并采用影印、缩印或扫描等复制手段保存、汇编以供查阅和借阅。同意学校向国家有关部门或机构送交论文的复印件和磁盘。

（保密的学位论文在解密后适用本授权说明）

学位论文作者签名： 导师签名：

签字日期： 年 月 日 签字日期： 年 月 日

摘要

以国家鸟巢体育场馆、天津市国家会展中心等为代表的空间结构是现代建筑重要组成部分，这些建筑不仅具有自身的功能性，而且具有重大的经济、文化、政治意义。而结构健康监测（**structural health monitoring**，简称 **SHM**）技术是确保建筑结构在施工以及运营阶段结构安全评估的重要手段，而由于空间结构自身具有杆件众多、结构复杂等特点，针对该类结构的 **SHM** 技术较少，而且实施难度较大。因此针对空间结构开展 **SHM** 研究具有重要的实际意义。

对于空间结构的健康监测相关研究表明，常见的数据驱动的损伤识别方法中常用的损伤指标往往很难通过实测的信号数据进行获取，而且对于数据分析方法的精确度、鲁棒性、泛化性要求较高。基于以上原因，本文提出了一种基于 **Learning Active Learning** 与 **AdaBoost** 相结合的智能算法（**LAL-AdaBoost**）的空间网架损伤识别框架，包括了对于数据的预处理与异常识别、损伤指标特征提取、损伤分析三部分内容，具体工作如下：

提出了一种基于 **LAL-AdaBoost** 算法框架的数据异常识别方法，针对结构健康监测中常见的五种数据异常类型，通过一个网架实验采集数据进行验证其有效性，并对分类精度、主动学习算法的拟合速度等方面进行了分析；提出了一个基于加速度信号频域分析并结合传感器布置的损伤指标，作为机器学习算法的输入使用；对空间网架实验结构使用 **ANSYS** 进行有限元建模，并模拟出了六种不同位置、不同程度的损伤，使用上述的损伤指标作为 **LAL-AdaBoost** 算法的输入，分别进行了四种监测场景的验证：结构出现单一损伤时进行损伤检测、结构出现微小损伤时验证模型灵敏度、多种损伤程度同时存在时模型对于损伤程度判别、不同位置处出现了损伤情况后的损伤定位。结果显示，所提出的损伤识别框架在该网架结构中具有较高的精度与拟合效率。通过在模拟的数据中加入高斯白噪声验证了算法的抗噪性能。最后，通过天津一中预应力网架监测项目的数值模拟算例对本文提出的损伤识别框架进一步验证泛化性与有效性。

关键词：结构健康监测，网架结构，主动学习算法，数据异常识别，损伤识别

ABSTRACT

The spatial structures, represented by the National Stadium (the Bird Nest) and Tianjin National Convention and Exhibition Center, play an important part of modern architectures. These buildings not only have their own functionality, but also have great economic, cultural and political significance. Structural health monitoring (SHM) technology is an important means to ensure the structural safety of building structures during the construction and operation stages. Due to the characteristics of spatial structures, SHM technology for such structures is usually unapplicable. Therefore, it is of great practical significance to carry out SHM research on spatial structures.

For the SHM of spatial structures, the relevant research shows that the commonly used damage indexes in the data-driven damage identification methods often fail to be obtained through the in-situ tests, and the data analysis methods have high requirements for accuracy and robustness. Based on the above reasons, this thesis proposed a damage identification framework especially for spatial structures based on a novel LAL-AdaBoost algorithm, including data preprocessing and anomaly detection, damage features extraction and damage analysis.

Firstly, a data anomaly detection method based on LAL-AdaBoost algorithm was proposed, aiming at detecting five common data anomaly patterns in SHM scenarios, and was verified by dataset collected from a grid experiment. The results showed that the proposed algorithm could perform excellently with high accuracies and fast converging speed. Secondly, a damage index based on frequency-domain analysis considering accelerometers displacement was proposed as the input of machine learning algorithm; ANSYS was used for finite element modeling of the grid experimental structure, and six different damaged scenarios were simulated. Using the above damage indexes as the input for LAL-AdaBoost algorithm, four monitoring scenarios are verified respectively: single damage degrees detection; minor damage degree detection; multiple damage degrees detection; damage location. The results turned out to be accurate for each monitoring scenario. Last, the robustness of the algorithm was verified through anti-noise tests. The proposed framework was then further verified on a SHM project of Tianjin No.1 High School Stadium and the results showed the feasibility of the approach in this paper.

KEY WORDS: structural health monitoring, grid structure, active learning, data anomaly detection, damage identification

目 录

第 1 章 绪论.....	1
1.1 研究背景.....	1
1.2 国内外研究现状.....	2
1.2.1 空间结构的损伤识别研究现状.....	2
1.2.2 机器学习算法在 SHM 中的应用现状.....	5
1.2.3 现有研究存在的主要问题.....	8
1.3 本文研究方法.....	9
第 2 章 主动学习理论与集成学习算法.....	11
2.1 LAL 主动学习算法.....	11
2.1.1 主动学习概述.....	11
2.1.2 主动学习质询策略.....	12
2.2 AdaBoost 算法.....	18
2.2.1 集成学习算法.....	18
2.2.2 AdaBoost 算法.....	21
2.3 本章小结.....	25
第 3 章 健康监测数据预处理与异常识别.....	27
3.1 信号频域分析方法介绍.....	27
3.1.1 快速傅里叶变换.....	27
3.1.2 小波包分解.....	28
3.2 数据异常检测方法.....	30
3.2.1 常见时域加速度数据异常类型.....	30
3.2.2 基于 LAL-AdaBoost 算法的数据异常识别框架.....	33
3.3 基于 LAL-AdaBoost 算法的空间网架数据异常识别研究.....	34
3.3.1 实验设计与数据采集.....	35
3.3.2 构造异常数据类型与方法.....	38
3.3.3 分类结果与讨论.....	40
3.4 本章小结.....	47
第 4 章 损伤指标与损伤识别方法研究.....	49
4.1 基于动力测试的损伤指标.....	49
4.2 基于 LAL-AdaBoost 算法的损伤识别框架.....	50

4.3 基于 LAL-AdaBoost 算法的空间网架结构损伤识别研究	52
4.3.1 健康与损伤数据库建立	52
4.3.2 损伤诊断研究	55
4.3.3 损伤定位研究	62
4.4 本章小结	74
第 5 章 总结与展望	75
5.1 全文工作总结	75
5.2 后续工作展望	76
参考文献	77
发表论文和参加科研情况说明	83
致谢	85

第 1 章 绪论

1.1 研究背景

空间结构是近 20 年来世界范围内发展最快的一种结构类型，其设计建造往往是衡量一个国家或地区的经济水平与建筑工程技术水平的标准^[1]。近年来，随着经济发展与建筑技术的进步，我国建成了一大批知名的大跨空间结构作为国际性的会场、场馆，如国家体育场“鸟巢”，国家游泳中心“水立方”，老山自行车馆，天津文化中心等。这些建筑不仅具有自身的功能性，而且往往会成为各地区甚至全国的地标性建筑，具有重大的经济、文化、政治意义。中国地震局于 2018 年 7 月提出实施《国家地震科技创新工程》，其中“韧性城乡”计划要求城市具有“韧性”，即震后快速恢复能力^[2]。以体育馆、航站楼等作为代表的大跨空间结构在震后作为快速安置点、交通枢纽，在“韧性城乡”体系中发挥着重要作用。因此，保障大跨空间结构的安全，加强其日常运行维护具有重要意义^[3]。



图 1-1 天津一中预应力空间网架结构^[3]

结构的损伤导致的失效会对生命、财产带来严重威胁，因此需要工程人员对结构的运营状况进行及时的评估以及判断^[4]。结构健康监测（structural health monitoring，简称 SHM）技术是确保大型复杂工程结构安全建造和健康运营的一个重要保障手段。SHM 是指利用传感器对结构的损伤、老化进行判断，对结构整体情况进行实时监测，并对结构的可靠性、耐久性进行综合评估，在结构状态

异常时发出预警，为结构的运营维护提供参考参考^[5,6]。而常见的空间结构（如网壳结构、网架结构、悬索结构、膜结构等）具有体型复杂、覆盖面积大、构件众多等特点，这其实对 SHM 系统的设计提出了很高的要求。理论上，为了获取最全面的结构响应信息，应给空间结构安装大量的传感器，以最大程度的获取所有节点位置的响应信息来对结构可能出现的损伤进行分析；但这也往往伴随着庞大的数据量所带来的巨大的数据处理与分析压力。尤其是在长期监测过程中，如何基于这些监测数据对结构的状态进行分析以及如何合理妥善的处理监测数据成为 SHM 研究的核心内容^[6,7]。在有限的监测条件下，合理利用传感器采集的信息进行结构分析以及损伤识别，是对于空间结构的 SHM 亟待解决的课题。

1.2 国内外研究现状

1.2.1 空间结构的损伤识别研究现状

结构健康监测在土木工程领域的发展已经有了几十年的历史，但是由于土木结构和机械结构不同，其损伤发生的频率相比于机械结构要低很多，而且土木工程结构由于自身的特殊性一般允许结构带有轻微损伤时继续工作，不像机械结构那样对损伤的判断那么界限分明，所以土木工程领域的结构健康监测和损伤识别最初发展十分缓慢。在上世纪四十和五十年代，结构损伤识别主要注重于损伤的修复，而损伤监测方面主要依靠目测，这样只能解决一些破坏程度比较大、损伤面积较为明显的损害，其作用有限，不能及时诊断出结构内部出现的一系列损伤；到二十世纪六十和七十年代，对于土木工程以及基础设施结构的 SHM 技术开始重视现场监测，有很多监测方法运用到了施工现场或建成结构的损伤检测中；八十年代以来，结构健康监测领域开始慢慢地完善起来，针对于损伤检测相继制定了各种规范，如抗震设计规范中的最大层间位移角、爆破规程里的地表峰值振速等损伤判据，为结构状态判断提供了参考。过去的三十年中，使用振动监测数据进行的损伤评估一直备受关注，有学者已经开发出通过动力监测的加速度数据来定位损坏并评估结构的严重性^[4,8]。

然而，由于空间结构自身的特殊性（体型复杂，覆盖面积大，构件众多、数据传输压力大、模态振型密集、模型分析较为复杂等方面的难题），特别针对于空间结构的损伤识别定位方法较少，而且损伤定位通常比较难度大。目前对于空间结构的 SHM 解决方案可分为基于模型的与数据驱动的损伤识别两条思路，在下文中作者将分别介绍。

基于模型的 SHM 技术的核心在于基于有限元模型的结构分析过程。结构损

伤监测与有限元分析的结合在近十几年得到了迅速地发展,因为这种方法具有无损检测的优点,而且有限元模型的分析在土木工程领域得到了充分的认可。这种基于模型的分析方法不但可以利用在现场利用检测工具与设备对结构进行监测分析数据,而且还可以结合计算机有限元分析对结构的损伤情况进行更加细致的分析评估,这样可以发现现场检测中无法发现的损伤^[9]。借助有限元模型来模拟加速度响应,可以对其他结构中常用的损伤识别方法在大跨度结构中的优缺点进行讨论,如增强坐标模态,损伤指标方法,模态曲率方法和模态柔韧性指标方法的四种不同的损伤识别方法^[10]。

虽然借助有限元模型的方法可以预测目标结构的静动力响应并分析其在不同情况下的力学行为,但再精确的有限元模型与原结构总会存在差异,这主要是由于数值模型和实际结构在材料、几何特性以及边界条件等方面存在误差。因此需要对有限元模型进行修正和更新,这其实是一个求解反问题的过程。近年来关于有限元模型更新的方法主要有运用环境振动测试结果更新模型^[11,12]、模态柔度矩阵更新模型^[13]、基于贝叶斯理论的模型更新法^[14]等。虽然借助有限元更新的方法可以直观的会对结构损伤进行定位以及量化损伤的程度,但是在实际过程中,有限元模型的分辨率受有限元网格密度定义,更新后的信息受实验结果控制;同时,模型的参数受到诸多理论假设限制。因此,需要建立更精密的有限元模型^[11,15],合理的设置分辨率^[12],贴近结构工作状态的模型实验^[13],借助高性能的传感器,利用更加先进的增强、更新算法,限制逆问题求解设置中的未知数与方程的比率,并防止算法对优化参数施加不切实际的假设,在统计模式识别方法的帮助下处理模型输出数据^[16]。

相比较有限元模型而言,实际工程应用中,使用数据驱动下的损伤识别方法,即利用数据分析的方法对模型精度要求更低,因此更易实现与操作。数据驱动的损伤识别方法需要从结构的监测数据中提取出对于损伤较为敏感的一系列指标来进行分析,其中频域参数以及模态参数已被广泛用于损伤检测,例如固有频率、模态振型、应变模态、频率响应函数、局部模态刚度、模态阻尼等。模态参数的特点是对结构损伤有较高的灵敏性,往往可以较为准确地反映损伤的产生。以模态柔度为例, Pandey 等最先提出了运用模态柔度差作为损伤指标^[17];之后学者提出了识别精度更高的模态柔度改变率和振型曲率改变率^[18]、MFC(模态柔度曲率差)^[19,20]等;但是这些指标在复杂的框架结构尤其是多跨框架结构的损伤识别中表现不够精确,因此不能直接适用于大跨建筑结构;在二维、三维结构的损伤识别上,使用模态柔度差曲率精度更高,可以有效识别桁架组成的网架体系损伤^[21]。

然而这些模态指标在实际工程中大多都不方便测得,易受环境影响,鲁棒性较差。文献^[17-19]中涉及的基于模态信息的指标仅仅在理想的有限元算例基础上做

了初步的验证,即使加入了一定的白噪声依然缺乏可信性以及实用性。为了考虑环境温度影响,黄江等提出了基于振动模态参数与突变指数结合的损伤识别方法^[22]。也有学者采用小波包变换和希尔伯特-黄变换等方法滤波^[23-25],或主成分分析^[26,27]的技术,一定程度上消除了环境和操作对模态特征的影响,仅在响应测量的基础上提取了对损伤敏感的特征,从而无需直接测量环境参数。当环境参数能够测量时,可以通过构建基于多项式混沌扩展的方法,以建立用于损伤检测的鲁棒状态指示器^[27]。将多项式混沌扩展与独立成分分析算法相结合,可用于获得输入(即环境参数)和输出(即自然频率特征)输入之间的直接关系或数值模型(例如,AR 自回归模型)的参数化表示的训练过程。在进一步的相关工作中,文献^[28]发现能够使用神经网络模拟温度对梁桥的模态频率和形状的影响。与模态形状有关的广泛使用的度量标准包括使用 MSC(模态形状曲率)。Shokrani 等^[29]介绍了一个在环境和操作参数可变的情况下检测和定位损坏的框架。该方法在模拟的四跨桥模型上得到了验证。首先利用主成分分析将环境影响与结构破坏分开;其中在第一阶段中,固有频率用作检测潜在损坏的特征,而在第二阶段中, MSC 用于定位。但是, MSC 自身方法存在许多缺点,包括依赖于所考虑的模式数量,数值微分问题以及需要用于确保准确性(特别是对于较高模式)的密集传感器网络。土木工程结构自身较低的频率模式降低了损坏检测的分辨率,并且受到环境和操作条件的严重影响,而自身较高的频率模式则更难提取且变化较大,因此检测损坏的可靠性较差^[30]。柔韧性矩阵同样属于模态特征,尽管与刚度矩阵的形式类似,但是不一定能直接与损伤量化和局部化联系起来,因此应用相对较少^[31]。类似地, FRF(频率响应函数)能够根据现场测试数据确定不同严重程度的不同破坏情景;然而, FRF 方法只能通过一定的传感器布局来检测损坏,这样的损伤检测方法降低了所提出方法的可行性^[32],但在 2D 主成分分析方法的数据缩减技术以及人工神经网络的帮助下, FRF 数据可以在 20%的噪声水平下正常工作^[30,33]。

下图总结了上述数据驱动的损伤识别的通用流程(图 1.2)。该流程主要包括数据采集、特征提取、损伤识别三个步骤,近年来关于损伤识别方法的创新主要集中在后两个步骤中。对 SHM 数据的预处理与异常检测过程确保了数据的可靠性,为后续的分析过程打下基础;而损伤分析部分则侧重于提取对于损伤及其损伤位置、程度敏感的特征(如上文提出的加速度信号的频域特征等),运用高效的分析手段(有限元模型模拟响应、机器学习模型拟合结构响应变化规律等),完成对于结构发生损伤后的预警系统的建立。

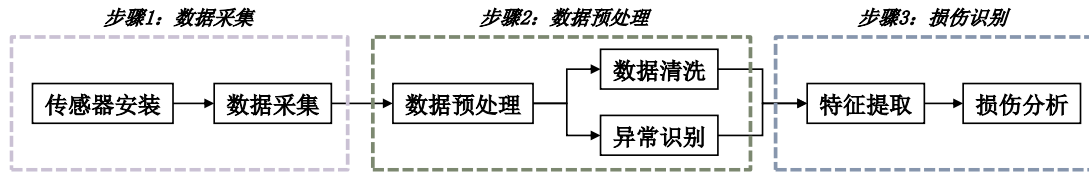


图 1-2 数据驱动的结构损伤识别流程图

1.2.2 机器学习算法在 SHM 中的应用现状

机器学习的研究方法在在结构健康监测数据处理于数据分析中也十分重要。机器学习是一门多学科交叉专业，涵盖概率论、统计学知识，近似理论知识和复杂算法知识，利用数值手段从现有数据中总结经验。目前机器学习在 SHM 中的应用按照实现的检测功能可大致分为两种：基于机器学习方法的 SHM 数据处理和基于机器学习算法的结构损伤监测。

1.2.2.1 基于机器学习方法的 SHM 数据处理技术研究现状

对于空间结构而言，为采集整体监测数据一般需要布置大量的多类型的传感器。同时布置多种传感器系统可以提高 SHM 系统的鲁棒性，但也带来了数据处理的难题。除优化布置传感器系统外，数据融合在一定程度上可以提高基于数据的损伤识别的效率和精度^[34]。目前，数据融合主要方法有贝叶斯理论的数据融合、DS 证据（Dempster-Shafer evidence）理论、模糊理论的数据融合等^[35]。相对于多损伤定位保证标准法和频率改变损伤探测法，实验证明数据融合的方法对于结构的多损伤识别效果更为出色^[35,36]。此外，也有学者发现数据融合技术与香农熵结合的损伤识别可以提高精度，无论是贝叶斯理论还是 DS 证据理论都提高了 SHM 系统的工作效率与精确度^[37]。

压缩采样（Compressive Sampling，简称 CS）方法通过根据数据信号密度来确定采样频率的方式而不是以信号最高频率的两倍来采样数据，这种信号采样技术能够显著减少数据量。CS 方法对于数据驱动的损伤识别的另一个应用是丢失数据的恢复，尤其是在无线传感器网络中^[38]。理论上只要监测数据具有一定的稀疏化特征，CS 方法对其数据恢复精度较好。此外，运用机器学习方法^[39]可以实现压缩采样数据的高精度重组。除应用于对信号处理外，CS 技术也用于系统识别任务，如结构模态识别、结构损伤识别和荷载识别。就其本质而言，压缩采样方法可视为是主成分分析（Principal Component Analysis，简称 PCA）的一种，只是具体算法不同于常见的 PCA 算法。对于结构模态识别，参照频谱分析的原理，当采用合适的基底进行表达时，模态参数直接从压缩测量中识别；但是，对

于结构损伤识别和载荷识别,结构损伤和载荷分布的空间稀疏性常被利用于解决识别中涉及的优化问题^[25,40]。

机器学习算法在 SHM 监测数据的异常处理技术上也有诸多应用。对于空间结构而言,庞大的传感器系统带来了大量的监测数据,其中出现的诸多数据异常很大程度上对于后续的数据分析以及损伤识别过程带来了严重的干扰。为了解决这一问题,Bao^[41]等人提出了使用计算机视觉的方法来对数据异常类型进行分类研究,通过学习某一悬索桥监测数据来训练一个卷积神经网络 (Convolutional Neural Network); Tang^[42]等人通过时频域特征提取的方法分析了异常数据与正常数据的特征,并使用了深度学习的方法对异常模式进行分类;面对如此庞大的数据集,有学者^[43]进一步提出了使用数据压缩的方法进行快速识别,在保证精度的前提下提高数据异常识别的效率。

1.2.2.2 基于机器学习算法的结构损伤监测研究现状

对于空间结构的健康监测而言,其杆件众多且体型复杂,在有限的预算下,选择大跨空间结构较为关键的构件进行监测尤为重要。一般来说,无论是传统的壳体结构、网架结构,还是较为特殊的悬索结构、充气结构等,建筑主体大多由钢结构构件组成,因此对于大跨空间结构的局部构件健康监测可以参考钢结构健康监测的方法与对象。而对于钢结构而言,其连接部位一旦出现损伤就会对相邻构件甚至整个结构造成安全隐患。以大跨空间结构中常用的螺栓球节点为例,一般工程上可以用扭矩扳手法或电阻应变片测试法来解决螺栓的松动和预应力损失等问题^[44],但对于大跨空间结构而言此类传统方法显得过于繁琐,在实际应用中受到诸多限制。近年来一些新兴的无损探伤技术被引入到了 SHM 中,比如超声导波检测方法具有无损性、敏感性、检测速度快、传播距离远且能量衰减小等优点^[45],一定程度上适用于大跨结构的螺栓群监测,因为采集到的声信号可以使用以迁移学习为代表的机器学习算法进行拟合与处理,从而实现对于损伤的预测工作^[46,47];运用小波包分解法和短时傅里叶变换可以对螺栓的扭矩进行不同层次的能量表征,从而实现对螺栓群的健康检测^[48];此外,声发射技术 (Acoustic Emission, 简称 AE 技术) 也广泛应用于大跨空间结构结构的损伤识别中,其不仅应用于混凝土结构的探伤^[49],在钢结构中 AE 技术也可以通过机器学习的手段提取对损伤敏感的动力时频数据进行识别 (如功率谱密度函数等指标),如螺栓松动、锈蚀等破坏。这些技术都具有无损性特点,即检测损伤过程中结构仍可以继续正常工作,且不会对原结构造成不可逆的破坏。

除了对于钢结构节点的监测,也有诸多学者对其他关键构件进行了监测研究,如基于机器学习的状态分析法可以确定荷载作用下拉索的拉应力比率^[50]等。近年

来,也有学者提出使用计算机视觉的方法对于大跨空间结构的表观损伤进行识别工作: Yeum 等^[33]提出了一种基于视觉的桥梁裂缝自动检测方法,利用目标检测和分组技术提取可能损坏区域的图像; Zauri 等^[51]提出了一种基于视频图像和常规传感器网络数据的监视系统,以识别在美国佛罗里达州的可移动钢梁桥上可能造成的损坏; Valenca^[52]提出了一种基于图像处理和激光扫描的混凝土桥梁裂缝自动检测方法,利用地面激光扫描技术捕获桥梁的几何信息,对从桥梁采集的图像进行校正; Lin 等^[53]提出了一种基于深度卷积神经网络的深度学习方法,以在仿真中实现对简单支撑梁的损伤检测,并在单损伤与多损失状态下均取得了较好的识表现。这种图像识别的方法更适合于结合无人机、监控摄像头、智能手机等设备辅助拍摄采集数据^[54],而不适用于对于结构整体的监测。而且对于空间结构来说,其自身体量较大,要获取完整的数据集需要对于图像拍摄角度、拍摄距离、覆盖面积等提出较高的要求。因此基于动力测试的损伤指标被提出来用于进行空间结构的损伤监测,而空间结构自身的结构特点决定了这些损伤指标不仅需要对于损伤敏感,而且对定位、程度都要有对应的映射关系来描述损伤。

1.2.2.3 主动学习与集成学习算法的应用现状

作者在以上几个小节总结的机器学习算法之所以可以在 SHM 中成功应用,归结于机器学习算法自身的优秀性能。高效的机器学习算法不但可以保证精确度,而且在运算时可以保证效率。近年来,诸多新型算法逐渐被开发出来并被应用到了各个领域中进行模式识别、状态识别、结构优化等工作。其中,主动学习与集成学习的提出从算法结构上颠覆了传统的监督学习流程,通过优化迭代结构将分类器的性能进行提升。

主动学习(Active Learning, 简称 AL)算法的核心思想是从数据样本的特性入手,通过特定的质询策略在样本中筛选出有价值的数据点进行质询标签的过程,从而达到节省时间与节约算力的目标。目前对于主动学习的创新性研究主要集中在质询策略的选择上,如随机选取样本、基于不确定性的质询方法、投票委员会方法^[55]、基于图密度方法的质询策略^[56]、考虑等级权重的多种质询混合方法^[57]等,然而这些新型的主动学习算法大多都没有被应用于土木结构的健康监测领域进行模式识别等验证有效性。然而与此同时,AL 算法已经被广泛地应用于机械结构领域的健康监测中。Bull 等^[58]提出了将 AL 算法用于半监督的结构健康监测中,通过基于聚类算法的质询策略对某个飞机机翼结构的损伤监测数据进行分析,取得了较为理想的精度;之后 Bull 等^[59]又将不确定性原理加入到了主动学习框架中,形成了一个基于贝叶斯理论的损伤识别框架并通过 Z24 桥梁数据集、某铸钢构件的声发射检测数据集验证了分类效果;Hughes 等人^[60,61]对于 Bull 等提出

的主动学习框架进行了改进，引入了风险决策算法；Zhao 等^[62]运用高斯混合模型进行主动学习算法创新，相比较其他常见的质询策略效果更好。主动学习由于其优秀的性能和创新性的样本筛选模式，在未来将被广泛地应用于各个领域的数据挖掘与数据处理中。

集成学习算法是一种将多个分类器整合于一体的元算法，作为一种非常成熟的判别式算法，其历史要远比主动学习更悠久。集成学习旨在将多个弱分类器整合成一个强分类器，取长补短，从而达到更优秀的分类或回归结果。最常使用的几种集成学习算法包括随机森林算法、AdaBoost 算法、XGBoost 算法、Stacking 算法、Blending 算法等。尽管结构不一致，但是这些集成学习算法的目的是一致的——集成弱分类器减弱残差并提高整体分类器的分类精度。其中，AdaBoost 算法^[63]由于其精炼巧妙的结构被广泛地应用到了各个领域，如车辆识别^[64]、预测山体滑坡^[65]等。在 SHM 领域，AdaBoost 算法也已经有了一定的应用：Furuta 等^[66]最先提出使用 AdaBoost 算法来进行结构响应的拟合过程，从而弱化对于有限元模型的依赖性；随后该学者又进一步提出可以使用 AdaBoost 算法进行桥梁监测，通过使用模糊神经网络作为弱分类器构造的 AdaBoost 算法可以拟合车辆荷载等环境输入从而预测结构的响应以及安全状态^[67]。由于 AdaBoost 算法自身具有对样本的适应性，可以在不同样本集上进行高效率拟合而避免常见的过拟合与欠拟合过程。从目前来说，该算法在 SHM 中的应用较少，仍具有较大的潜力为 SHM 数据挖掘做出突破性的成果。

1.2.3 现有研究存在的主要问题

经过数十年的发展历程，土木工程结构的 SHM 技术取得了较大的发展。然而，对于空间结构而言，其 SHM 技术的应用还比较受限制，缺乏具有针对性的方法来进行损伤识别技术。针对目前的研究现状，可以归纳出以下三点主要存在的痛点问题：

(1) 对于大跨空间结构而言，复杂的传感器系统带来的庞大数据需要一个快速的异常诊断框架来剔除掉其对后续结构分析的影响。目前针对该方面的研究主要采用了深度学习的方法，虽然精度较高，但是需要先对数据集进行人工手动贴标签来标定样本的异常类型，而且训练时间长，缺乏泛用性。

(2) 之前学者提出了诸多基于时频域的损伤指标，并在多种结构类型上实现了有效性的验证。然而，空间结构体型复杂，特征维度高，缺少一种考虑传感器空间分布的损伤指标用于损伤分析。

(3) 计算机视觉与深度学习已经广泛地应用于结构表面损伤识别，但是要想

发掘结构内部出现的损伤，仍需将结构作为一个整体来进行研究，并形成完整的损伤识别框架。此外，深度学习算法在进行拟合过程中由于神经元数量多、结构复杂，容易出现过拟合的现象，而且参数控制难度大，泛用性较差。

1.3 本文主要研究内容

本文针对现有研究的现状，本文以空间网架结构为研究对象，结合了 Learning Active Learning 主动学习框架与 AdaBoost 分类器（称为 LAL-AdaBoost 算法）构造了一个完整的结构损伤识别框架，包括了健康监测中关于数据预处理部分，并提出了一个适用于空间网架结构的损伤指标辅以损伤识别过程。通过使用 LAL-AdaBoost 算法对 SHM 分类问题中的优化，不但有效地节省了赋予样本标签的繁琐工作，而且有效地避免了不均衡的数据集对于分类问题产生的偏差。

本论文主要内容如下：

1. 将 Learning Active Learning 主动学习框架与 AdaBoost 分类器进行结合，构造出了一个具有优秀性能的分类器模型，并用于后续研究。
2. 对健康监测数据的预处理及异常识别研究：(1) 通过小波包分解将时域信号各个频段的小波包能量提取出来并归一化为能量占比特征向量，结合 LAL-AdaBoost 算法构造出了一个数据异常识别框架；(2) 设计了一个空间网架实验结构并进行动力测试数据的采集工作，并通过随机函数的方法使用实测数据构造出了一批具有五种加速度时域异常的数据库，用于后续的数据异常验证工作；(3) 在空间网架异常数据集上验证了提出的数据异常识别框架的可行性，通过对比实验验证了该算法对于不均衡数据集的有效性。
3. 损伤指标与损伤识别算法：(1) 根据数据的时频域分析结合传感器布置构造出了一种针对于空间结构的损伤指标，可转化为一维向量的形式作为 LAL-AdaBoost 算法的输入；(2) 使用 ANSYS 对空间网架实验结构进行有限元建模，以高斯白噪声作为输入模拟出了包含 6 种工况的数据集，并分别将某个单元内的随机杆件进行一定程度的截面刚度折减来模拟可能出现的损伤工况；(3) 分别分析了四种结构损伤情况下 LAL-AdaBoost 算法的识别精度与拟合速度，单一程度损伤监测、微小程度损伤的判断、多种损伤程度、损伤定位研究；(4) 时域数据中加入了高斯白噪声验证算法的抗噪性。(5) 通过天津一中预应力网架监测项目的数值模拟算例对本文提出的损伤识别框架进一步验证泛化性与有效性。

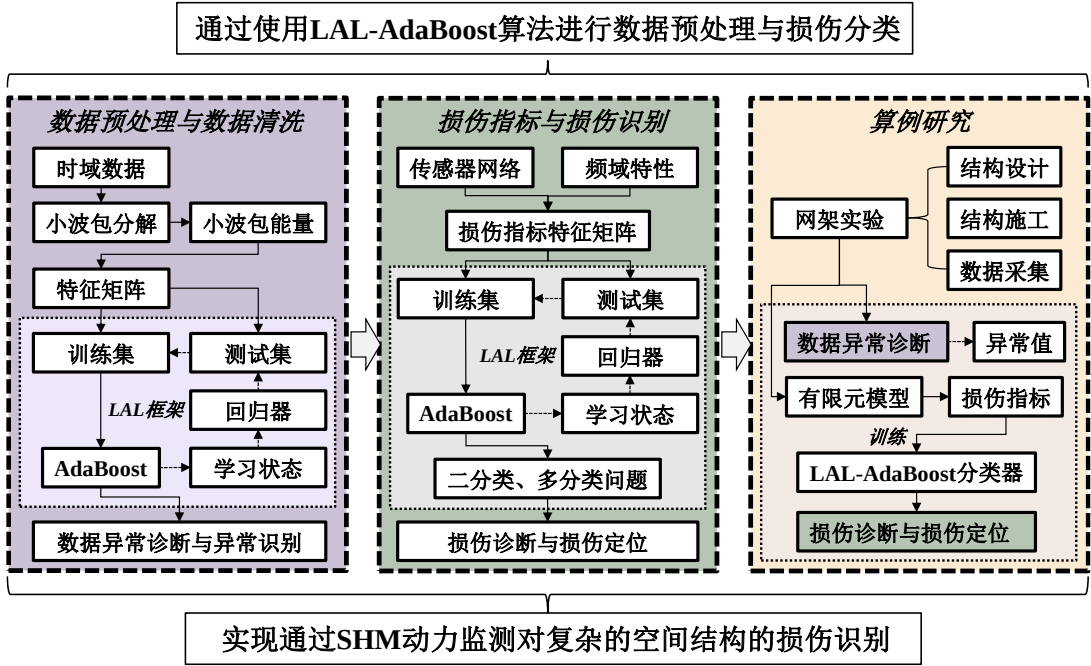


图 1-3 本文研究技术路线图

第 2 章 主动学习理论与集成学习算法

在本章中，作者将介绍本论文使用的主要核心算法和计算模型框架。2.1 节主要对主动学习的框架做详细介绍，并介绍了包括本文所用的 Learning Active Learning 算法等几个较为常用的主动学习质询策略，这些策略将在之后的章节中分别进行对比实验效果；2.2 节主要对于集成学习算法进行简要介绍，并针对本论文使用到的集成学习算法 AdaBoost 详细介绍，引出了改进的多分类 AdaBoost.SAMME 算法并与传统的 AdaBoost 做了对比。最后，作者对于将 Learning Active Learning 算法与 AdaBoost.SAMME 算法相结合的分类器框架在 SHM 中应用潜力进行了简要分析。

2.1 LAL 主动学习算法

2.1.1 主动学习概述

在计算科学领域中，根据机器学习模型对于样本标签（labels）的依赖性，可以将机器学习算法分为有监督、半监督和无监督的机器学习算法三类。其中，无监督学习算法（聚类算法，超像素分割算法等）可以在未知样本标签的情况下实现样本的分类或回归运算；而有监督的机器学习算法充分考虑了样本所属的标签后可以优化目标函数，因此一般认为其性能优于无监督的机器学习算法。近几十年许多性能优秀的有监督学习算法被应用在各个研究领域中，如 SVM（支持向量机）、CNN（卷积神经网络）、Decision Tree（决策树）以及以 AdaBoost、XGBoost 等为代表的集成学习等。但是这些算法的性能很大程度上依赖于样本的标签，对于缺失部分标签或者样本中存在有噪声标签的情况，这些机器学习算法无法发挥出理想的性能。对于大部分实际问题而言，给庞大的样本赋予精确而全面的标签往往是非常不现实的，尤其是对于大型、复杂的结构而言，健康监测的时间一般上百甚至上千小时，是无法对于每个时间段的数据进行分析并贴上标签的。事实上，出于控制成本的考虑，大部分的健康监测数据集中都不包含标签，因为标签的获取对于结构工程师和专家的精力和经验都要求较高。

为了解决样本标签有限的问题，有学者提出了主动学习的框架，通过特定的质询策略来训练一个高效率、高精度的机器学习模型，在有标签样本数量十分有限的情况下实现分类或回归任务。通常来说，一个完整的主动学习框架包含以下 5 个实现步骤：

- (1) 通过特定的质询准则或策略，在现有的无标签样本中筛选出有价值的样本或对于后续分类判定最模糊的样本；
- (2) 将筛选出来的样本发送给专家(oracle & annotator)进行诊断并赋予标签，反馈给原始数据库；
- (3) 使用这一批有了新标签的样本和原本样本库中含原始标签的样本进行混合，共同作为机器学习分类器模型的输入；
- (4) 训练模型；
- (5) 将剩余样本作为测试集和验证集，带入训练好的模型输出最终结果。

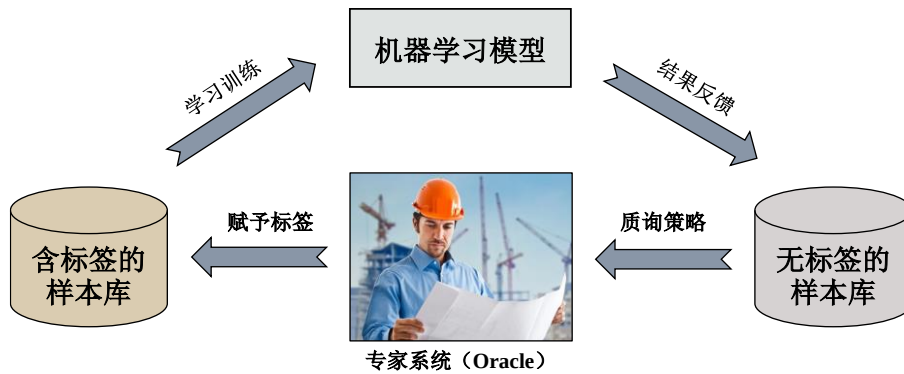


图 2-1 主动学习流程框图

可以从该流程框架中得知，一个高效率、高精度的主动学习框架主要取决于两个关键因素：合理的质询策略、高性能的分类器。其中，作者将在 2.1.2 中介绍几种较为常见的质询策略，对于本文采用的分类器则将在 2.2 中详细介绍。

2.1.2 主动学习质询策略

所有的主动学习算法都需要对于未标记标签的样本集进行评估其信息量，这些实例可以从新生成，也可以从给定分布中采样。许多学者在文献中已经提出了许多较为常见的策略。本小节中，作者将介绍两种最为常见的质询策略的运算原理，分别为较为常用的不确定采样（Uncertainty Sampling，简称 US）策略以及学习主动学习（Learning Active Learning，简称 LAL）策略。对于一个主动学习模型，其目标是使用人工智能算法选取特定的数据点来进行质询并贴标签，相比较传统的机器学习需要人工全部判定数据所属标签的方法要更加经济、省时省力。不同的主动学习模型之间的差别主要在于其质询算法的不同，下面为了介绍分类算法的流程，不妨以一个最简单的二分类问题为例来阐述其原理：

- (1) 假设目标数据集为 $Z = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$ ，其中 $x_i (i = 0, 1, \dots, N)$

是一组维度为 d 的特征向量, $y_i (i = 0, 1, \dots, N)$ 定义为每一组特征向量对应的标签, 满足 $y_i \in \{0, 1\}$;

(2) 定义一个任意的机器学习分类器算法 f 用于训练样本以及预测测试集的样本工作, 即 $f(x_i) = \hat{y}_i$;

(3) 在进行迭代之前将数据按照 7: 3 的比例划分为训练集 D 与测试集 D_t , 并在 D 中进一步划分初始标记的样本集 L_t 以及无标签的样本集 U_t 使得 $L_t \cup U_t = D$ 始终成立, 其中 t 为迭代次数, 此处由于循环未开始, 设 $t = 0$;

下面分别介绍 US 策略以及 LAL 策略的计算原理。

2.1.2.1 US 策略

基于 US 策略的主动学习方法是在文献中最为常见的策略之一^[68]。该方法基本构想是将对于该分类器而言最没把握的样本进行筛选而质询。比如说, 对于一个基于概率估计模型的二分类问题, 当某一个样本的后验阳性概率接近 0.5 时, 可以认为该样本是具有质询价值的。US 策略的核心在于对于不确定程度的测量, 较为常见的指标主要有最小可信度 (Least Confidence)、间隔法 (Margin)、熵方法 (Entropy) 等。

以最小可信度方法 (简称为 LC 法) 为例, 假设一个分类问题中的样本存在满足下式的点:

$$x_{LC}^* = \arg \min_x P_f(\hat{y} | x) = \arg \max_x 1 - P_f(\hat{y} | x) \quad (2-1)$$

即, 该点在机器学习模型 f 中的后验概率达到最小, 通俗的讲机器学习对于该点错分的概率最大, 即最无把握。我们可以认为满足上式条件的样本点一般位于两类样本交汇处, 通过选择边界处的样本点可以达到划定不同类别样本的界限, 从而达到仅标记少部分样本的情况下达到一个较好的分类效果。

但是在一些特殊情况下, US 方法会存在一定的局限性。由上式得知, US 方法倾向于选择分类器最没把握的样本点用于质询, 当两类或者多类的样本点数量相对均衡时, 该方法效果较好, 可以精确地锁定分类边界。但是当样本数量不均衡时, 该方法会存在一定的偏向性。Konyushkova 等^[69]在相关研究中列举了两个最简单的二分类模型来解释这一现象。在第一个二分类模型中, 随机生成了两组二维高斯云分布的数据样本, 记为 Z , 这两类的样本数量一致, 且具有不同的均值相同的方差; 第二个实验则生成了两组数量不相等的高斯云分布样本集 Z_t , 其中一类样本的数量是另一类样本的两倍, 方差与均值设置为与第一个实验一致。实验结果如下图所示, 在第一个实验中 US 方法表现良好, 而在第二个实验中则出现了较大的偏差。

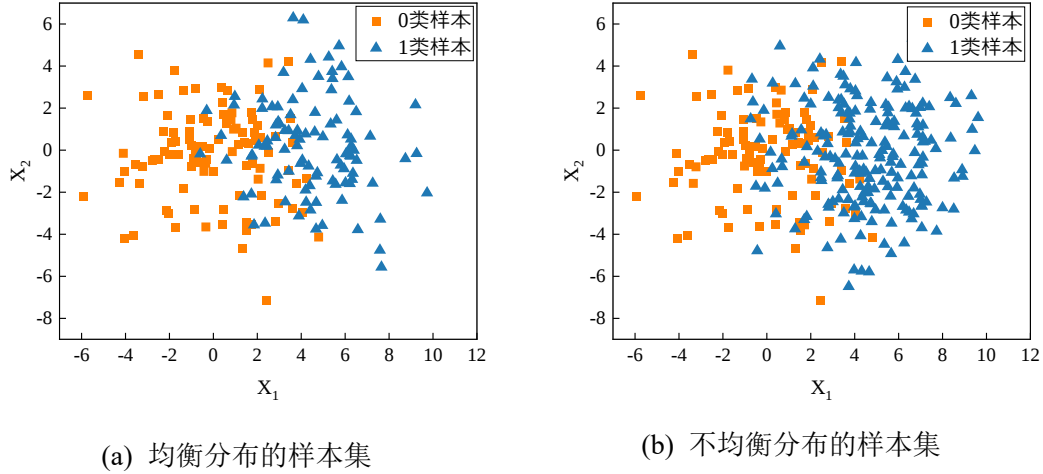


图 2-2 二分类实验。

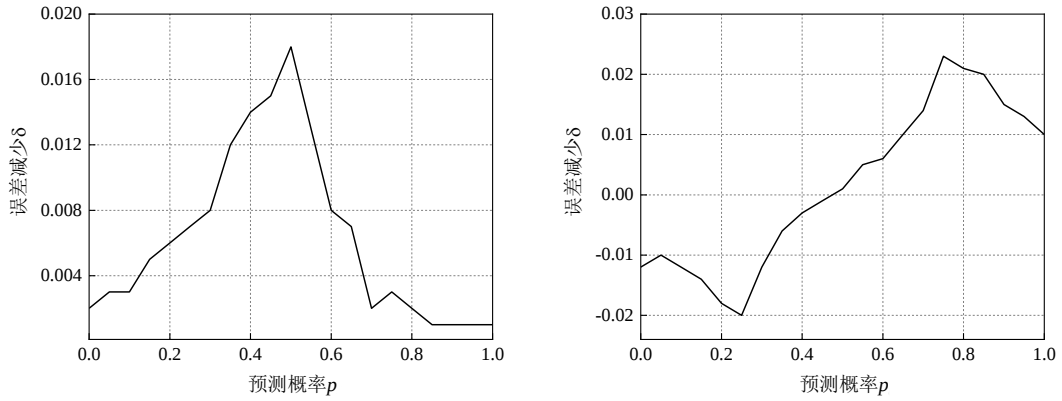
具体的计算过程如下：

- (1) 在初始过程中，首先使用已经具有标签的数据集 L_0 训练一个初始的分类器，该分类器可记为 f_0 ，对应的损失函数记为 l_0 ；
- (2) 在无标签数据集 U_0 中分别抽出一个新样本点 $\mathbf{x}$ 并赋予其标签 y ，并形成新的数据集 $L_x = L_0 \dot{\cup} \{\mathbf{x}, y\}$ ；
- (3) 使用新数据集 L_x 重新训练分类器，得到 f_x 及其对应的损失函数 l_x ，对比质询前后的两个分类器损伤函数从而得到损失缩减值 $d_x = l_0 - l_x$ ；
- (4) 有放回地重复步骤(3)，可以得到完整的样本点损失缩减的后验概率分布情况，可以将 d_x 与 0-1 损失函数的形式结合起来研究 US 方法的有效性：

$$L(y, f(x)) = \begin{cases} 1, & y = f(x) \\ 0, & y \neq f(x) \end{cases} \quad (2-2)$$

如此重复实验，可以得到预测样本点的标签的概率函数 $p(x)$ ，表示对于某一个样本点 $\mathbf{x}$ 而言其在被分类器判定为 $y = 1$ 的概率值；

- (5) 最后将 d_x 与 $p(x)$ 绘制在同一个坐标系下，观察结果。



(a) 均衡数据集结果

(b) 不均衡数据集结果

图 2-3 US 方法预测曲线。(左：两类样本数量一致，右：两类样本数量不一致)

可以在图中看出，在第一组实验中， d_x 的峰值与 $p_0 = 0.5$ 的样本点基本重合，可以解释为：US 方法会自动选择后验概率介于两类标签之前的样本点，并认为其对于整体错误率的降低最为有利，而对于数据样本比例均衡的数据集 Z 而言，其预测错误率减少最为明显的样本点恰好也是后验概率接近于两者之间的样本点，从而达到了有效质询样本点的效果。然而反观第二组实验，US 算法依然会选择后验概率介于两类标签之前的样本点，即 $p(x) \approx 0.5$ ，然而我们通过重复实验可知，这些样本点并不是最佳选择。由于不均衡样本集 Z 中分布情况不对称，0 类样本数量是 1 类样本数量的两倍，分类器会受影响而忽略弱势样本点对于整体精度的影响，可以认为错误分类 1 类样本的代价相比较于错误分类 0 类样本要小。这时，主动学习质询算法的引导作用将会被体现出来，这里先抛开分类器自身的分类性能不谈，如果质询算法可以自动选取偏向于 1 类样本的数据点而适当削弱 0 类样本在监督学习中的主导作用，那么分类错误率的降低值将会提升明显。显然，US 方法在此过程中无法自动选择出合适的样本点，抑或是合适的样本边界进行质询，从而导致分类器存在偏差。

因此，我们在进行主动学习质询过程时，需要有一个可以预测未来迭代过程中可能产生的错误率降低值（即 d_x ）的运算，从而对于质询样本点的策略进行及时引导和修缮。基于此思想，一种可以在每次迭代时通过运算学习过程中产生的特征参数来预测下一次迭代选取样本点时的错误率减小值，从而产生一个可以自适应的质询策略的方法。LAL 算法的核心便是分类阶段特征提取和回归器的训练过程，在 2.1.2.2 节中将详细讲述其原理。

2.1.2.2 LAL 策略

本文中将要介绍的 LAL 算法是 Konyushkova^[69]提出的 LAL-Interactive 算法。

在介绍其原理前，作者需先介绍一下两个基础前置算法：基于蒙特卡洛方法的数据分类特征提取过程，以及 LAL-Independent 算法。

用数据集 L_t 训练出一个分类器模型 f_t ，并对其在测试集 D_t 上的表现做评估，得到分类损失为 l_t ；进而计算分类器状态参数集合 $f_t = \{f_t^1, f_t^2, \dots, f_t^K\}$ ，这些参数的选择取决于具体分类器的类型并且对于训练集的样本发生变化较为敏感，例如，如果 f_t 为随机森林等集成学习算法时，每棵树的最大深度会被认为是有效的参数之一；接下来，随机地从 U_t 中选择一个样本点 x ，这个样本点可以用一个特征参数矩阵来描述，即 $y_x = \{y_x^1, y_x^2, \dots, y_x^R\}$ ，这些特征主要包括了该点与样本中具有标签的数据点的欧式距离等；在加入了新样本点 x 后，新的有标签样本集 $L_x = L_t \cup \{x\}$ 可以用来再次训练分类器从而得到新的映射函数 f_x 及其在测试集 D_t 的分类损失 l_x ，那么将这两次的分类损失作差可以得到新参数 $d_x = l_t - l_x$ ；经过上述过程后得到的两个参数集合可以写成同一个参数向量中，即 $x_t^x = [f_t^1, f_t^2, \dots, f_t^K, y_t^1, y_t^2, \dots, y_t^R]$ ，可以用来描述主动学习过程中每一次迭代的状态以及所选取的数据点的价值；最后经过 M 次迭代后可以得到返回值分别为学习状态矩阵 $X \hat{I} \in M^{(K+R)}$ 与损失改变量矩阵 $D \hat{I} \in M$ 。

在 LAL-Independent 算法中，我们假设质询次数为 Q 次，那么上述的基于蒙特卡洛方法的数据分类特征提取过程也将被重复 Q 次。基于这些得到的数据，可以做出假设并认为分类器的学习状态与所选择的质询样本点对于分类器的表现是有一定关联的，从这个角度来说，可以通过一个回归器函数来预测在某一个分类器的学习状态下质询某一个特定样本点时对于整体分类错误率的降低潜力，从而合理地选择出有价值的、可以使得分类器快速收敛的质询样本点。因此，这里可以寻找一个映射关系 $g: x \rightarrow d$ ，即对于每一个特定的学习状态下选择每一个质询样本点都会对应一个损失减少的预测值。这样，LAL-Independent 算法就可以贪婪地选择具有最大程度上减少损失潜力的样本点，使用数学符号表示为：

$$x^* = \arg \max_{x \in U_t} g(x) \quad (2-3)$$

然而，从上述 LAL 过程中可以看到每次迭代过程中算法都会随机地将数据集划分为 L_t 与 U_t 来进行蒙特卡洛数据特征提取过程而无法真正意义上的根据每次学习状态来更新质询策略。为此，学者提出了 LAL-Interactive 算法来解决这个问题。具体实现的方法如下：

(1) 按照 7:3 的比例初始化训练集 D 与测试集 D_t ，定义一个初始化的随机划分函数 SPLIT；

(2) 在第 t 次迭代时，对数据进行 Q 次蒙特卡洛方法的数据分类特征提取过程，得到第 t 次迭代时的特征数据 X_t 和 D_t ，并使用这些数据训练一个独立的回归

器 $g_t: x_t \rightarrow d_t$; 再根据回归器预测结果可以得到一个新的质询策略 $A(g_t)$ 用来更新上一步迭代的划分函数 **SPLIT**, 即 $\text{SPLIT} \leftarrow A(g_t)$;

(3) 在第 $(t+1)$ 次迭代时, 使用更新的划分函数 **SPLIT** 重复蒙特卡洛方法对数据进行处理与计算, 并重复步骤(2)的过程。

上述过程可以以流程框的形式展现。由上述的过程可以发现, 在每次迭代的数据筛选过程中 **LAL** 算法的质询策略会根据实际样本的分类特性而发生改变, 从而达到自适应的效果。研究表明, 上述介绍的两个 **LAL** 的算法 (**LAL-Independent** 与 **LAL-Interactive**) 在计算步骤上的不同导致了两种方法在运算速度上的差异。一般来说, **LAL-Interactive** 算法需要训练若干个小回归器, 相对来说比较耗费时间。通常的做法是下载比较成熟且已经训练好的回归器模型, 以节约算力。

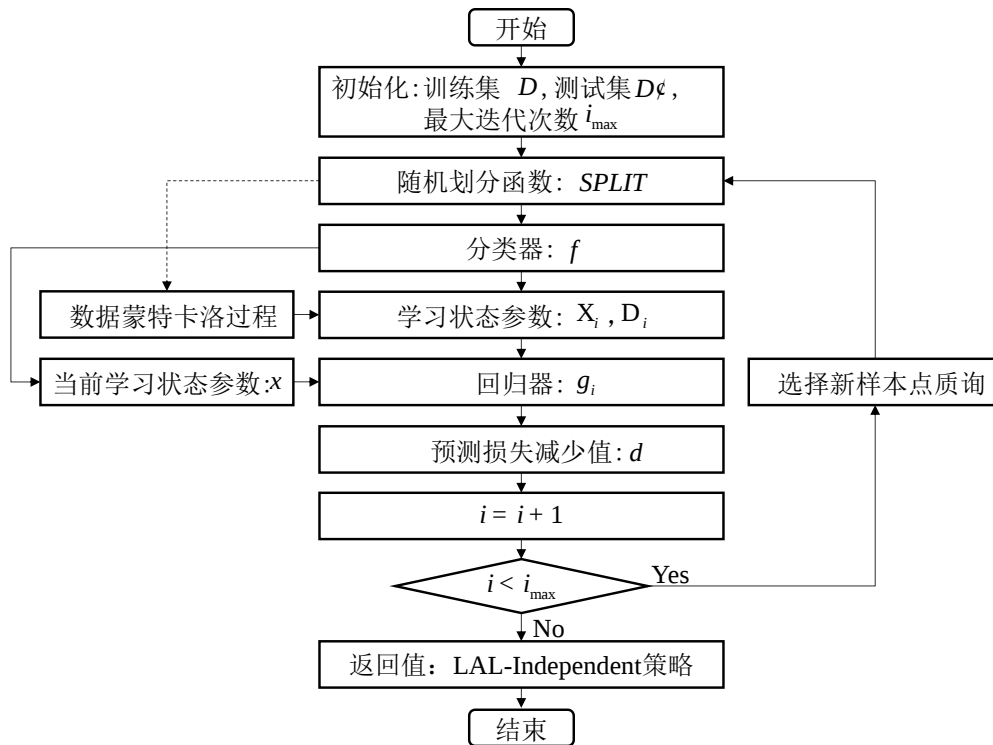


图 2-4 LAL-Independent 算法流程框图

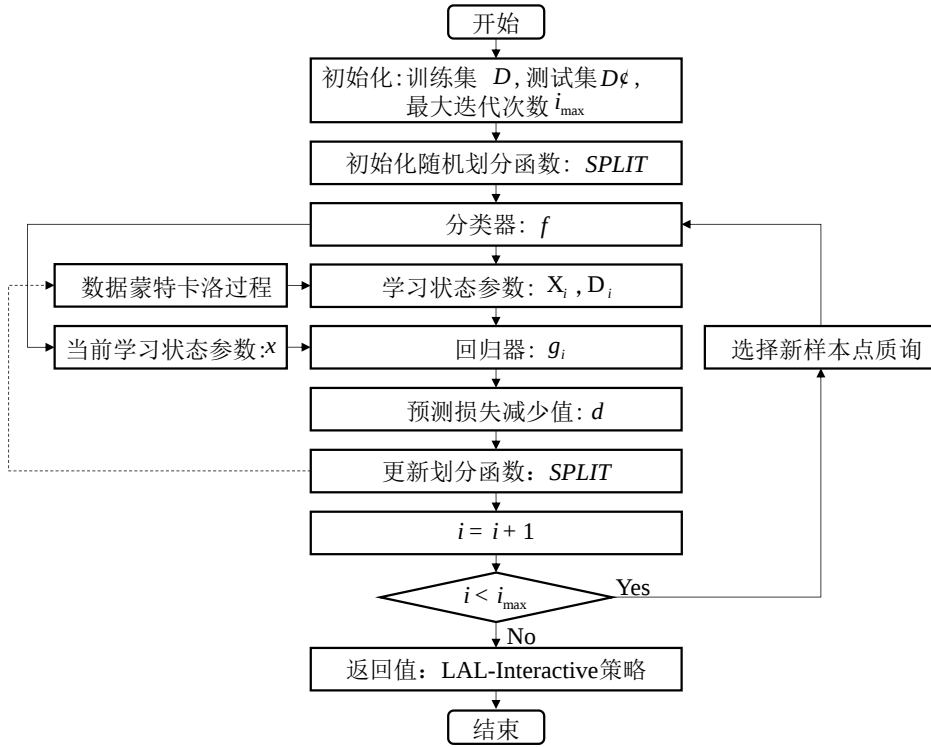


图 2-5 LAL-Interactive 算法流程框图

2.2 AdaBoost 算法

2.2.1 集成学习算法

2.2.1.1 集成学习概述

在训练机器学习模型时,通常认为获得一个能够很好地处理多维特征的完美分类器是几乎不可能实现的,特别是对于包含一定数量的噪声干扰的结构健康监测数据而言,很难权衡不同的样本对于分类效果的准确性。一方面来说,对于从健康监测数据中提取的特征来说,他们对于结构响应的敏感性是不同的。以损伤识别问题为例,某些结构特征对于结构损伤较为敏感,因此对于分类问题的贡献较大;而另外一些特征对于结构损伤较为迟钝,一定程度上对增加模型拟合和收敛的速度和难度。另一方面,许多机器学习模型对于分类结果都有一定的偏向性,一些在分类迭代过程中较为弱势的特征往往无法收到足够的重视,导致分类结果具有偏向性。这对于结构健康监测问题而言往往是较为致命的,以损伤识别为例,一个偏向于忽略损伤的分类器会被认为是一个失败的机器学习模型。目前,为了解决这些问题,集成学习算法被提出并广泛应用于数据挖掘等领域。该算法在处理大型数据集时表现良好,往往显示出了更高的精度和更快的处理速度。

集成学习的基本思路是将多个机器学习算法模型通过一定的策略进行整合，达到提升最终效果的目的。这些弱分类器模型或许无法实现全局的最优分类效果，但是通过整合可以达到优势互补的效果。诸多学者的研究和实验表明，集成多个分类器（相比较于单个分类器）可以显著提高模型的泛化能力和训练精度。集成学习最早来自于 PAC（Probably Approximately Correct，简称 PAC）学习模型，该模型提出了强学习和弱学习的概念：识别准确率比随机猜测高的学习算法定义为弱学习模型，而识别准确率较高且能在多项式时间内完成的学习算法定义为强学习模型。集成学习的任务就是将若干个弱学习模型整合为一个强学习模型的过程。根据不同的集成策略的选择，可以大致将集成学习算法分类为 Bagging、Boosting、Stacking 等，这些策略都是通过不同的迭代方式将单个弱分类器的分类残差作为先验知识来改进分类效果。

2.2.1.2 集成策略

Bagging 算法是最简单、直接的集成策略。Bagging 方法的步骤大致分为以下几部分：1）在总样本量为 N 的数据中进行有放回的抽样 M 次数据（每次抽取数据数量小于等于 N ）；2）每次抽样所得新数据集对应一个弱分类器进行训练（共形成 M 个弱分类器）；3）将 M 个弱分类器的分类结果进行集成，对于分类问题而言，集成方法为投票委员会法，即对 M 个模型结果进行投票决定分类判定值，票数多的选作结果，票数一致则随机决定。根据概率论中 Bootstrap 的思想，由于小样本的不确定性，可以重复计算来提升小样本的精度。随机森林算法（Random Forest）是该方法最普遍的代表性算法，其在构造样本时随机选取特征并为每一个数据集建立一个完全分裂的决策树。

Boosting 算法的主要思路根据每一次模型训练所得结果来调整和指导数据集样本分布，然后再生成下一个模型，最终集成所有的弱分类器。与 Bagging 不同的是，该算法具有不同的集成方式，比较常见的方法是针对每个弱分类器的表现给分类器赋予不同的权重，以增大或减小其在运算最终结果时的“话语权”。AdaBoost 是 Boosting 算法中的代表性算法，该算法认为每个弱学习器应通过权值的方法来改变其对集成结果的影响。Gradient Tree 也是另外一种应用了 Boosting 思想的集成学习算法，其核心思想为将损失函数梯度下降的方向作为优化的目标，常见的算法包括 XGBoost 算法、GBDT 算法等。

总结来说，Boosting 算法通过计算样本权值和分类器权值分别评估了样本对于分类的贡献程度和分类器的表现，从而通过不断迭代来改进分类效果。相比较于 Bagging 算法，虽然二者都采用了重采样，但是由于 Boosting 引入了权重强分类的概念，理论上会取得更优的分类效果。从下图中可以看出 Bagging 算法和

Boosting 算法最核心的不同在于弱分类器模型之间的关系, Bagging 是并行而生, 而 Boosting 则是串行集成。

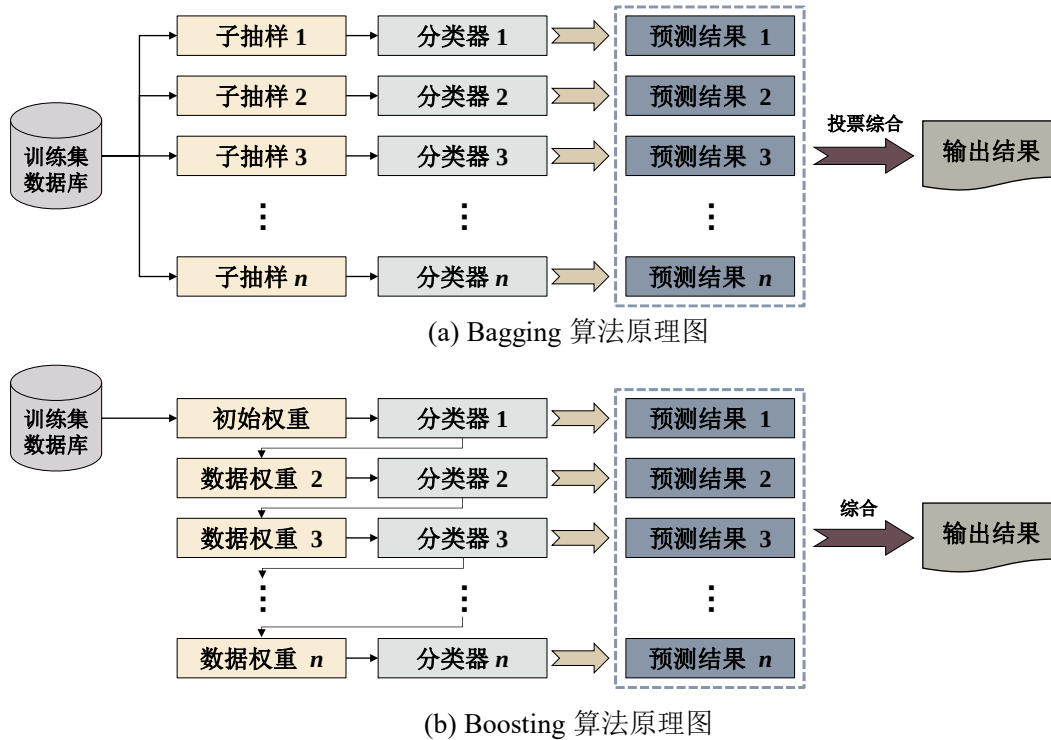


图 2-6 Bagging 与 Boosting 算法原理对比

而 Stacking 的集成方法思想为通过训练集训练多个模型并称之为“一层”, 并使用这一层的模型预测的结果作为输入, 来训练下一层的机器学习模型, 最终通过多层的叠加得到一个较好的预测结果, 并得到一个强学习的分类器。一般来说, 组合器模型通常采用逻辑回归, 也可以使用神经网络等复杂算法, 如 ANN、CNN、DNN 等。

2.2.1.3 基分类器

在集成学习算法中, 基分类器的选取的重要性仅次于集成策略的选择。按照基分类器之间的种类关系可以把集成学习方法划分为异态集成学习和同态集成学习两种: (1) 异态集成学习: 使用各种不同的分类器进行集成, 异态集成学习的两个主要代表是叠加法 (Stack Generalization) 和元学习法 (Meta Learning)。同态集成学习: 集成的基本分类器都是同一种分类器, 只是这些基本分类器之间的参数有所不同; (2) 同态集成的分类器包括有朴素贝叶斯集成、决策树集成、人工神经网络集成、K-近邻集成等。

在本文使用的 AdaBoost 算法使用了决策树分类器 (Decision Tree Classifier) 集成的同态集成方法, 主要控制了以下变量。具体算法原理将在下一节具体阐述。

表 2-1 决策树主要参数

参数	描述	取值范围
max_depth	树的最大深度	整数
min_samples_split	产生分支的最小样本数	整数或浮点数，默认为 2
max_features	寻找分裂点考虑的特征数	小于特征维度的整数

2.2.2 AdaBoost 算法

2.2.2.1 AdaBoost 算法概述

根据 2.2.1.2 小节中关于 Boosting 算法的介绍中可知，基于 Boosting 的机器学习算法主要考虑了两方面问题来提升分类器精度：1）在每一轮选取的数据样本如何分配其权值或概率分布；2）在集成时如何赋予每个弱分类器权值或概率分布来将它们组成一个强分类器。在本节中，作者将介绍 AdaBoost 算法对于这两个步骤的计算原理与方法。

AdaBoost 算法是 Boosting 的最普遍的应用之一，该算法由 Freund 和 Schapire 在 1997 年创造并测试^[70]。“Ada”是“Adaptive”的缩写，意为自适应。下图阐述了 AdaBoost 框架的大致流程：

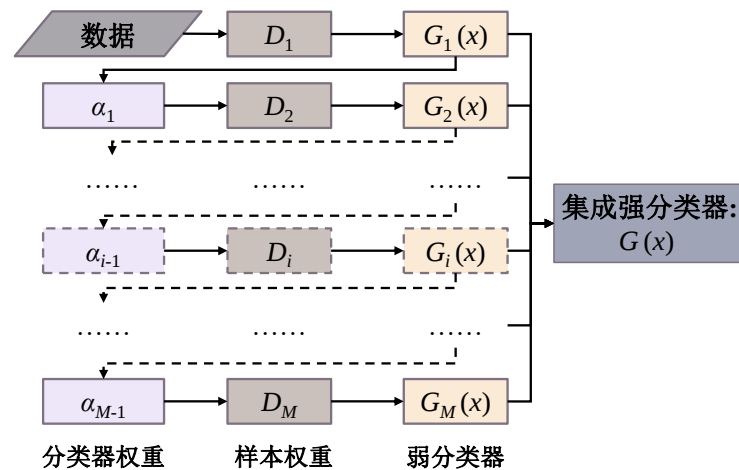


图 2-7 AdaBoost 原理与迭代流程图

首先假设原始数据集包含 N 个样本，其中 $(x_i, y_i), i = 1, 2, \dots, N$ 代表第 i 个样本点的特征量与对应的标签值。那么，由这些样本所构成的数据集 T 可以用下式表达：

$$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\} \quad (2-4)$$

其中，具有 n 个维度特征的实例 $x_i \in X$ ，标签值 $y_i \in Y$ 。在进入迭代之前需要初始化样本权重 D_1 和分类器权重 a_1 ，使得前者中的每一项 $w_{1,i}$ 均为 $1/N$ ，后者的每一项均为 0。在第 m 次迭代时，得到一个弱分类器对于样本集的映射函数，记为 $G_m(X)$ ，其分类表现可用下式表示：

$$G_m(X): X \rightarrow Y \quad (2-5)$$

该弱分类器在训练集样本上的表现可以用错分率 e_m 来表示，即错误分类的样本占总样本数的比例。通过 e_m 可以进一步得到一个评价弱分类器分类效率的指标，即弱分类器的权重 a_m ：

$$a_m = \frac{1}{2} \ln \frac{1 - e_m}{e_m} \quad (2-6)$$

在进行下一步迭代之前，将本次迭代中使用的样本权重 D_m 进行更新，计算过程见下式：

$$D_{m+1} = (w_{m+1,1}, w_{m+1,2}, \dots, w_{m+1,i}, \dots, w_{m+1,N}) \quad (2-7)$$

$$w_{m+1,i} = \frac{w_{m,i}}{Z_m} \exp(-a_m y_i G_m(x_i)) \quad i = 1, 2, \dots, N \quad (2-8)$$

其中， Z_m 为规范化因子，使得 D_{m+1} 成为一个概率分布：

$$Z_m = \sum_{i=1}^N w_{m,i} \exp(-a_m y_i G_m(x_i)) \quad i = 1, 2, \dots, N \quad (2-9)$$

最后，将每次迭代中的弱分类器进行整合集成为一个最终的强分类器：

$$G(x) = \text{sign} \left(\sum_{m=1}^M a_m G_m(x) \right) \quad (2-10)$$

总结来说，AdaBoost 算法主要使用两个关键运算来提高分类器的效率：首先通过提高上一轮分类器错分的样权重本、降低上一轮正确分类的样本权重来强化分类器对于数据的学习；再通过分析每个弱分类器的错分率来调整其权重，表现越好的分类器在最终表决时的话语权就越大，最终集成一个强分类器。文献表明，AdaBoost 分类器算法在诸多数据集中表现出较快的拟合速度以及较高的分类精度。

2.2.2.2 AdaBoost.SAMME 算法

上一小节介绍的传统的 AdaBoost 算法在进行二分类问题时效果较好，但是 Freund 等人提出的该算法存在以下缺陷：为了使每次迭代的效果进行提升，每次得到的错分率参数 e_m 必须要小于 0.5，否则由式(2.6)计算得出的弱分类器权重 a_m 将为一个负值，这会使得样本权重计算时出现反向迭代（错分的样本权重反而降低）。这个问题在 AdaBoost 进行多分类问题时更加凸显出来，因为当 Y 的维度 K 大于 2 时，分类器的初始迭代错分率往往很难小于 0.5；而相比之下，分类器的分类效果却可以较容易地达到随机猜测的精度，即 $1/K$ 。

基于以上问题，Ji Zhu 等人^[63]提出了一种改进的 AdaBoost 算法——SAMME (Stagewise Additive Modeling using a Multi-class Exponential loss Function)。该算法主要流程如下。

- (1) 初始化样本权重 D_1 和分类器权重 a_1 ，使得前者中的每一项 $w_{1,i}$ 均为 $1/N$ ，后者的每一项均为 0；
- (2) 在第 m 次迭代中，使用样本权重 D_m 得到一个弱分类器对于样本集的映射函数，记为 $G_m(X)$ ；
- (3) 计算出误差率 e_m ：

$$e_m = \frac{\sum_{i=1}^N w_i \tilde{O}(y_i \neq G_m(x_i))}{\sum_{i=1}^N w_i} \quad (2-11)$$

- (4) 计算分类器权重参数 a_m ：

$$a_m = \log \frac{1 - e_m}{e_m} + \log(K - 1) \quad (2-12)$$

- (5) 更新样本权重 D_m 中的每一项 $w_{m,i}$ ：

$$w_{m+1,i} = \frac{w_{m,i}}{Z_m} \exp(a_m \tilde{O}(y_i \neq G_m(x_i))) \quad i = 1, 2, \dots, N \quad (2-13)$$

其中，其中， Z_m 为规范化因子，形式类似于式(2.9)。

- (6) 输出最终的强分类器：

$$G(X) = \arg \max_{k \in \{1, \dots, K\}} \sum_{m=1}^M a_m \tilde{O}(G_m(X) = k) \quad (2-14)$$

可以看到 AdaBoost.SAMME 算法和传统的 AdaBoost 算法相比，最大的不同在于计算样本权重时不再依赖于 e_m 必须要小于 0.5 的先行条件了。对于多分类问

题,为了保证分类器权重 a_m 为正值,仅需满足 $(1 - e_m) > 1/K$,即保证每个弱分类器的预测结果正确率大于随机猜测的猜对概率。即使 e_m 大于0.5,也可以保证 a_m 为正值,从而对提升算法的更新方向做了调整导向。而且,相比较而言,SAMME算法在错分样本上增加的权重实际上大于 AdaBoost 算法。此外,当 $K=2$ 时,SAMME 算法会退化为普通的 AdaBoost 算法。由于 AdaBoost.SAMME 算法本身属于对于 AdaBoost 算法的改进,为了便于书写,在之后的章节中统一简化为 AdaBoost 算法进行二分类与多分类问题。

2.3 LAL-AdaBoost 算法

在本文的研究工作中,作者将 AdaBoost 算法与 LAL 主动学习框架相结合,形成了一个高效的分类器算法 LAL-AdaBoost,该算法不但具有较高的分类精度,而且在处理大量数据时具有较强的优势。在每次的质询环节中通过 LAL 框架学习 AdaBoost 算法的分类学习状态来预测误差减少来产生全新的适应性的质询策略以筛选样本,并重新训练 AdaBoost 算法,以此形成一个良性的循环,在节省标记样本的工作量的同时可以很大程度上提高机器学习模型的效率与精度。

LAL 主动学习框架的回归器算法主要学习的 AdaBoost 学习状态参数大体可以分为两类:集成参数信息与弱分类器参数信息。对于以决策树算法作为弱分类器的 AdaBoost 分类器来说,其主要学习状态参数如下表所示。其中有些参数值不是浮点数,在进行运算时先通过独热编码进行处理后再作为回归器的输入变量来进行运算。

表 2-1 AdaBoost 分类器学习状态参数

类别	参数名称	参数意义
集成参数	n_estimators	弱分类器精度收敛时的数量
	learning_rate	学习率,即在每次提升迭代中的分类权重
	algorithm	计算算法,在本文中选用“SAMME”算法
	train_accuracy	AdaBoost 算法在训练集上的正确率
	sample_weight	AdaBoost 算法在迭代结束后的样本权重
	estimator_weight	AdaBoost 算法在迭代结束后的弱分类器权重
弱分类器参数	criterion	计算决策树划分枝叶的判据,本文中选用基尼不纯度(gini impurity)
	max_depth	决策树精度收敛时的平均最大深度

2.3 本章小结

在本章中，作者主要介绍了主动学习理论以及集成学习算法相关的内容，主要可以总结为以下几点：

1. 首先作者对于主动学习理论进行了简单介绍，并分析了两种主动学习策略 Learning Active Learning(LAL)以及 Uncertainty Sampling(US)。通过一个二分类的分类任务对比了两种算法的特性，展现出了 US 算法在对于均衡数据集与不均衡数据集中的差异性质询表现，并证明了 LAL 算法在对样本不均衡问题时表现更好，可以通过对于样本损失减少的预测来避免不均衡数据集的误分类问题。

2. 介绍了集成学习的相关理论，以及两种常见的集成策略 Bagging 以及 Boosting，并对比了两种算法的异同。二者在进行迭代的过程中，前者中的弱分类器通过又放回的采样过程进行独立的训练，最后通过投票生成最终的强分类器；后者则通过每次迭代中弱分类器的表现来对样本进行样本权重的更新，从而最后整合为一个强分类器。

3. 介绍了基本的 AdaBoost 算法的基本计算迭代流程，并引申出了本课题使用的高效分类器 AdaBoost.SAMME 算法。该算法基于传统 AdaBoost 算法进行了改进，在多分类问题上表现更为出色，在进行样本权重分配时避免了反向迭代的问题；而对于二分类的情形，AdaBoost.SAMME 算法则退化为传统的 AdaBoost 算法。

4. 介绍了 LAL-AdaBoost 算法的基本思想以及其在 SHM 中的应用潜能，并详细列举了 LAL 过程中使用的 AdaBoost 学习状态参数信息。

第 3 章 健康监测数据预处理与异常识别

在本章中，作者将主要介绍对于结构健康检测的加速度数据异常识别的内容。首先作者整理了较为常见的信号处理方法，包括快速傅里叶变换、小波包分解等方法，并针对这些方法进行了对比；基于上述理论，作者提出了一个基于 LAL-AdaBoost 算法的数据异常识别框架；同时为了验证该数据异常识别框架的有效性，作者设计了一个空间网架实验用于采集一批加速度数据，首先通过随机函数等方法生成一批具有异常的加速度数据序列并分析了这些异常值产生的原因以及在时域上的异常形态，包括离群值、基线漂移、矩形波、数据缺失、幅值过小等；分类结果显示，本文所提出的数据异常识别框架效率较高，可以较为准确地识别出可能的异常类型，并且具有较好的鲁棒性与抗干扰性，尤其是当数据样本不均衡时该算法可以依然准确地识别出异常类型，对于实际工程问题来说更具有实用价值；作者同时使用了另外一种更为常见的主动学习算法 US-AdaBoost 作为对比试验，体现出本文提出的 LAL-AdaBoost 方法的质询策略对于不均衡样本而言具有更好的拟合效果。

3.1 信号频域分析方法介绍

3.1.1 快速傅里叶变换

傅里叶变换是工程上常用的时频分析方法。而快速傅里叶变换（Fast Fourier Transform，简称 FFT）的本质是将长序列的信号分解为短序列进行离散傅里叶变换（Discrete Fourier Transform，简称 DFT）计算。DFT 的复指数形式通用表达式可以写为：

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j \frac{2\pi}{N} kn}, k = 0, 1, \dots, N-1 \quad (3-1)$$

其中， $x(n)$ 为原始时域序列， k 为离散谱序列的序号； N 为时间序列长度； j 为虚数单位。下面介绍当 N 为偶数时，如何将长度为 N 的序列转换为 $N/2$ 长度的序列进行 DFT 运算。

首先按照偶数和奇数项将序列划分为两个长度为 $N/2$ 的时间序列，不妨分别记为 $x_1(r)$ 与 $x_2(r)$ ，它们的表达式可以按照下式表示：

$$\begin{cases} x_1(r) = x(2r) \\ x_2(r) = x(2r+1) \end{cases}, r = 0, 1, \dots, \frac{N}{2} - 1 \quad (3-2)$$

那么，为了简便计算过程，可以令 $W_N = e^{-j\frac{2\pi}{N}}$ ， $W_N^{nk} = e^{-j\frac{2\pi}{N}kn}$ ，则由(3-1)可推导出下式：

$$X(k) = \sum_{n=0}^{N-1} x(n)W_N^{nk} = X_1(k) + W_N^{Nk} X_2(k), k = 0, 1, \dots, \frac{N}{2} - 1 \quad (3-3)$$

由上式可知，对于长度为 N 的序列进行 DFT 运算，实际上可以分解为两个 $N/2$ 长度的序列分别进行 DFT 运算。由于系数 W_N^{nk} 具有周期性，可以根据该性质对上述结果继续化简，即：

$$X_1\left(k + \frac{N}{2}\right) = \sum_{r=0}^{N/2-1} x_1(r)W_{N/2}^{r(k+N/2)} = \sum_{r=0}^{N/2-1} x_1(r)W_{N/2}^{rk} = X_1(k) \quad (3-4)$$

同理可得偶数序列的简化表达式，即：

$$X_2\left(k + \frac{N}{2}\right) = X_2(k) \quad (3-5)$$

从公式(3-4)、(3-5)中不难得出结论：以时间序列的第 $(N/2 - 1)$ 个点为分界线，前后两部分序列的 DFT 运算结果其实是完全相等的。按照这个性质，可以推算出对于任意一个长度为 N 的时间序列的 DFT 简化运算公式：

$$\begin{cases} X(k) = X_1(k) + W_N^k X_2(k) \\ X(k + N/2) = X_1(k) - W_N^k X_2(k) \end{cases}, 0 \leq k \leq \frac{N}{2} - 1 \quad (3-6)$$

式(3-6)即为快速傅里叶变换(FFT)的最基本形式。总结来说，FFT 是在 DFT 的基础上进行理论推导及简化步骤，在工程上具有更高的利用价值，因为其运算时由于因子 W_N^{nk} 的存在可以节省非常多的乘法运算。但是进行 FFT 运算时，也需要注意以下几点基本要求：

- (1) 离散信号必须遵循香农奈奎斯特采样定理，即采样频率为奈奎斯特频率；
- (2) 信号长度 N 为 2 的整数次幂；
- (3) 数据长度足够长来拟合稳定信号。

3.1.2 小波包分解

上一节介绍的快速傅里叶变换的基函数为三角函数（或指数函数），而三角函数（或指数函数）是无限长的信号，不具有局部性。对于大多数的平稳信号而言，傅里叶变换可以得到较为精确的频谱。但是对于大多数的结构健康监测数据

而言，信号具有非平稳性，局部可能会发生突变，而且信号的衰减过程往往伴随着诸多环境噪声。此时如果使用 FFT 进行处理，并不能得到各个频率的信号在时域上的分布情况。实际上，当两段在时域上相差很大的信号在总体上包含相同几个频率的成分时，通过 FFT 方法得到的频谱图可能会高度一致。因此，我们在对一段非平稳信号进行时频分析过程中，我们不仅希望知道这段信号包含哪些频率成分，而且还想知道信号频率随时间变化的情况。

而小波变换则可以很大程度上从原理上解决这个问题。相比于傅里叶变换使用的三角函数基，小波变换使用的是小波函数基，其最主要的特点是正交性、且在时域迅速衰减。这种小波函数基包含两个变量：尺度变量 a 和平移变量 t 。尺度变量通过控制伸缩来控制小波的频率，而平移变量反应的是小波在时域中出现的位置，对于一个小波变换函数：

$$WT(a, t) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \psi\left(\frac{t - t_0}{a}\right) \frac{dt}{a} \quad (3-7)$$

从上式可以看出，傅里叶变换的变量只有频率 ω ，小波变换的变量则包括尺度变量 a 和平移变量 t ，前者对应于某段时域数据的频率，后者对应于该频段在时域上发生的时间。由此计算可以同时得到该信号存在什么样的频率成分，同时也可以知道它在时域上的具体位置。

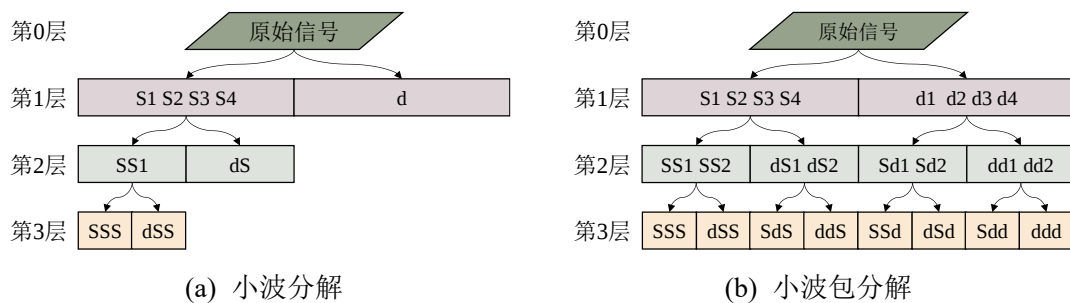


图 3-1 小波分解与小波包分解示意图（以 3 层运算为例）

而上述小波变换的弊端在于其只能在低频段的信号中有较高的分辨率，而在高频段上分辨率较差。小波包分解则可以解决这一问题，它可以认为是小波变换的精细化方法，可以同时低频段与高频段进行时频分析，其原理图如上图所示。基于小波包分解的方法可以对每一层的节点提取出相对应的信号能量，从而实现对于多尺度空间能量特征的提取过程，将这些能量值按照尺度函数（即频率值的大小）进行排序，可以作为特征向量进行分析。对于第 j 层的第 i 个节点，其能量值的计算如下：

$$E(i, j) = \sum_{k=1}^N \left| \hat{p}_s(n, j, k) \right|^2 \quad (3-8)$$

其中, p_s 为小波包变换系数, n 为每个小波包能量值排序 ($n = 0, 1, L, 2^{j-1}$)。然而在实际的基于小波包能量值的特征提取过程中, 我们只关注不同节点能量的占比, 而不关心每个节点具体的能量值大小, 因此往往会做归一化处理而求得每个节点的小波包能量占比用于特征向量的构造。

为了对比本小节提出的几种信号频域分析处理的方法, 作者在下表中进行了总结。总体来说, 当采集的信号为平稳信号时, 可以使用快速傅里叶变换 (FFT) 进行处理; 当采集信号为非平稳信号, 即在特定时间段内出现了较为明显的突变时, 小波包变换 (WPT) 表现更为出色, 因为它不仅可以确定组成不同频率参数, 同时可以进行定位运算。

表 3-1 本节介绍的信号处理方法对比

方法	描述	适用范围
离散傅里叶变换 (DFT)	对于离散时域数据进行的全局变换	平稳信号
快速傅里叶变换 (FFT)	对于 DFT 的改进变换, 减少乘法运算工作量	平稳信号
小波分解 (WD)	时频局部分分析方法, 对于高频段分辨率较差, 在低频段分辨率较高	非平稳信号
小波包变换 (WPT)	在小波分解的基础上对于频带多层次划分, 提高了全频段上的视频分辨率	非平稳信号

3.2 数据异常检测方法

3.2.1 常见时域加速度数据异常类型

对于加速度信号而言, 最常见的几种数据异常包括: 离群值过多、数据幅值偏小、矩形波、基线漂移、数据缺失等。在本节, 作者主要对于这几种数据异常的波形特征进行汇总, 并根据相关资料总结了出现该异常的原因。

(1) 离群值: 数据串中出现较多的数据点的幅值远高于正常数据的平均幅值。一般认为如果数据中超过 5% 的点为离群值点, 那么这些离群值点将给这一时域序列的分析过程带来较大偏差。离群值的产生有很多原因, 较为常见的有传感器底座与结构未固定牢靠、传输电缆损坏、设备电路故障等。

(2) 基线漂移: 数据点的基准线呈现某种趋势, 一般认为趋势为特定斜率的一次函数或毫无规律的抖动。这种数据异常情况通常是由于数据采集正式开始之

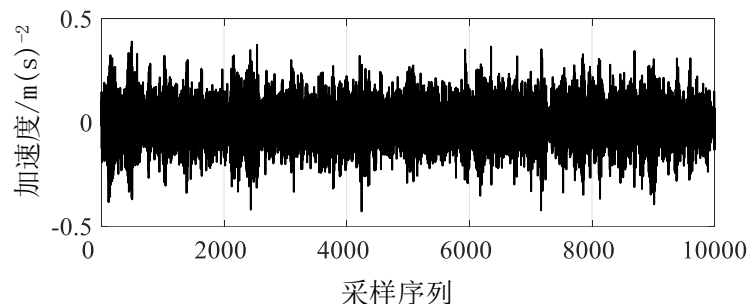
前未将采集设备校准所引起的，示波过程中电信号无法在原点上下波动，而在采集过程中，由于加速度幅值的逐渐增大，基线可能会进一步偏移。这种现象也有可能是加速度传感器自身出现了故障所导致的，亦或者当加速度传感器底座与结构固定不稳固的时候，结构振动幅度较大，传感器与结构接触面之前出现了相对滑动，在时域上往往会表现为一个逐渐增大或衰减的动态加速度值，使得结构的基线出现漂移。

(3) 矩形波：加速度数据在时域上呈现矩形波状。矩形波一般认为是由两个二进制（2 级）是从逻辑电路中产生的。由于逻辑电路是同步操作的，严格规定的时间间隔，使方波快速转换和定时参考信号适当“时钟”被使用。在加速度传感器中产生此类波形一般认为是接线或者传感器与采集仪之间的接线出现了故障，误将闭合电路的电磁脉冲电流显示成了加速度信号进行采集。若示波采样环节出现此类错误，应当检查传感器的连接是否正常以及采集设备是否出现了短路。

(4) 部分数据缺失：数据缺失主要分为两种情形，第一种是个别数据点的缺失，这一般是由于采集设备设置的传感器灵敏度错误导致采样频率相对于传感器的参数而言过高而导致部分数据无法采样；另一种是某一段的数据点均为 0 或固定值或在采集仪中显示为 NaN 值（Not a Number）。

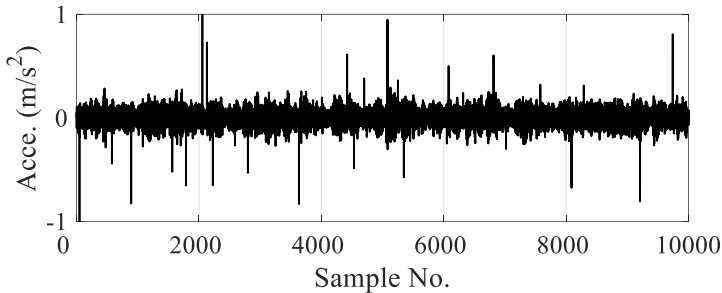
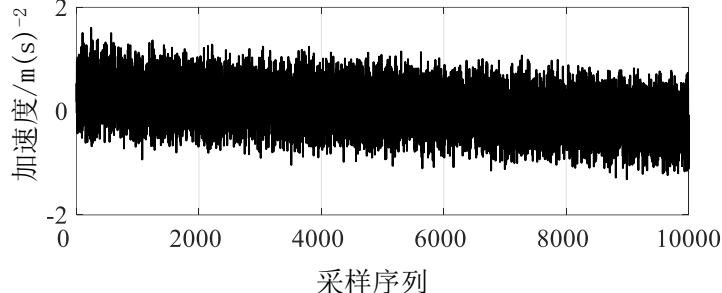
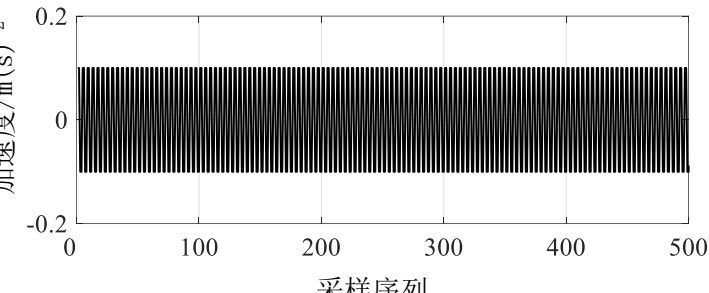
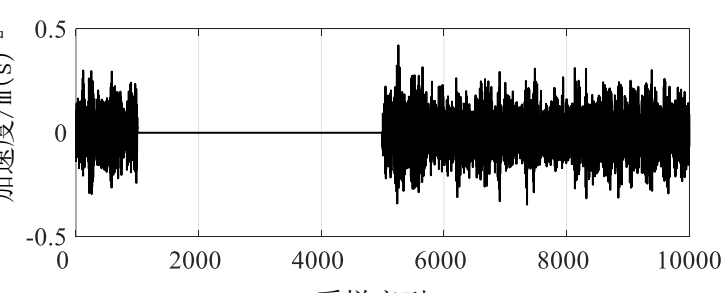
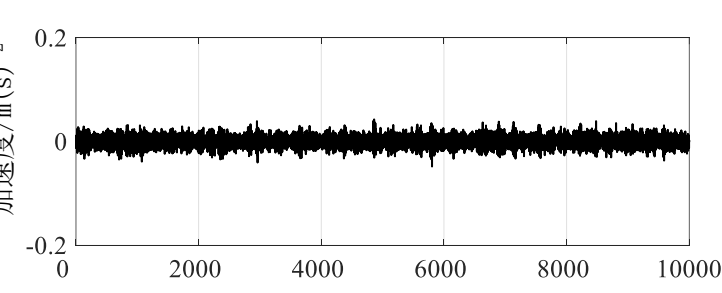
(5) 幅值过小：所采集的数据点的幅值相对于正常的加速度值偏小，导致时域分析时各个统计指标（均值、峰峰值、均方差等）出现异常，影响特征提取。这种数据异常情形一般是由于采集设备在初始化时参数设置错误导致电信号转换为加速度值时出现错误。也存在加速度传感器与结构接触不良，或者安装位置处有缓冲结构或材料，加速度值变小。

在下表中，展示了五种常见的数据信号的异常情形的举例。相对于正常数据而言，这几种数据异常情形在时域上均有较为明显的图形特征。

表 3-2 正常与异常加速度时域数据示例	
类型	时域加速度数据波形示例
正常数据	

（下页续表）

(接上页续表)

类型	时域加速度数据波形示例
离群值过多	
基线漂移	
矩形波状	
部分数据缺失	
幅值过小	

3.2.2 基于 LAL-AdaBoost 算法的数据异常识别框架

作者基于第 2 章所提出的主动学习框架以及集成学习算法,结合本章所述信号处理方法以及常见的时域数据异常类型,在本节中提出了一个数据异常识别框架。该框架主要分为三个部分:数据预处理与特征提取、LAL-AdaBoost 质询与训练、结果验证。

第一部分:预处理与特征提取。由于 NaN 值无法进行小波包分解的运算,因此首先剔除包含缺失值的样本段;对于时域加速度数据进行小波包变换,得到从低频到高频的各段数据;这里小波包的层数可以依据具体的数据实例而定,一般来说,层数越多分析的越细致,所提取的特征更全面,但是对于机器学习算法的精度要求越高;之后提取出每一频段数据的小波包能量,并作归一化处理,得到小波包占比作为特征向量,用于后续的分析。

第二部分:LAL 训练过程。这里首先将上一步得到的特征向量按照 7:3 的比例划分为训练集和测试集,其中测试集将直接用于第三部分的验证阶段作为输入;训练集随机地挑选出 5%作为初始标记样本用于训练 AdaBoost 分类器,剩余的 95%用作后续的筛选样本集;每次迭代都会训练一个回归器来根据学习状态参数对样本进行预测从而作为样本筛选的依据,这里用的 LAL 算法可以是第 2 章介绍的 LAL-Independent 算法或 LAL-Interactive 算法,可以依照具体数据而定,一般来说,为了节省计算时间建议优先选用 LAL-Independent 算法;在经过预先设定好的迭代次数后,可以得到一个最终的 AdaBoost 算法。

第三部分:验证。训练集作为输入来验证第二部分中得到的 AdaBoost 算法的精度,最终根据 AdaBoost 算法的分类结果得出报告,来验证整个框架的有效性;这里一般用到的指标包括混淆矩阵、 F_1 值等。

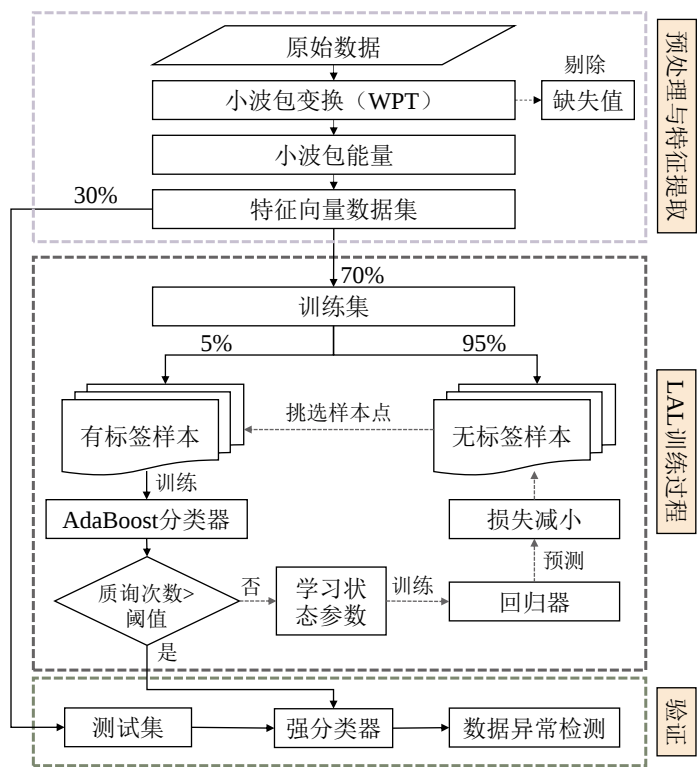


图 3-2 数据异常识别框架

3.3 基于 LAL-AdaBoost 算法的空间网架数据异常识别研究

为了验证 3.2.2 节中提出的数据异常识别框架，在本节中介绍了一个空间网架实验的数据采集方案与过程，用于分析数据异常识别的主要过程以及验证该框架的可行性。

3.3.1 实验设计与数据采集

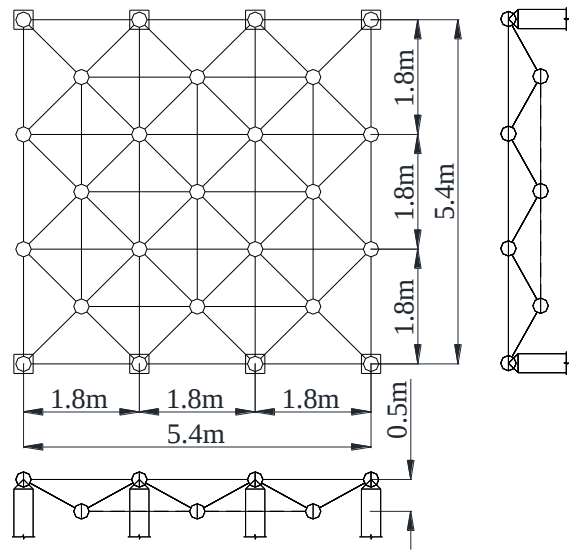


图 3-3 空间网架实验设计平面图与立面图（单位：mm）

该实验对象为一 $5.4\text{m} \times 5.4\text{m}$ 正放四角锥空间网架，共包含 9 个单元，每个单元尺寸为 $1.8 \times 1.8\text{m}$ ，高 0.5m 。八根工字钢柱子使用 C30 混凝土固定在地面，网架下弦节点距离地面 1m 。网架整体结构采用 Q235 钢材，节点采用螺栓球节点，支撑采用 $250 \times 125 \times 5\text{mm}$ 的工字钢柱子。主要构件的尺寸见下表：

表 3-3 主要构件明细表

序号	类型	规格	数量
1	螺栓球	$\phi 100 \times 4$	25
2	螺栓	M10	144
3	圆钢管	$\phi 48 \times 3.5$	72
4	套筒	21 $\times$ 25	144
5	封板	48 $\times$ 12	144

具体设计方案介绍如下：

（1）节点尺寸：螺栓球节点采用 $\phi 100$ ，切削量为 4mm ，螺栓型号为 M12，根据所在位置的不同共设计了 6 种螺栓球节点（仅开孔位置不一样），具体分布位置如下图所示，分别编号为 A1、A2、A3、A4、A5、A6。

（2）杆件尺寸：上、下弦杆和斜腹杆均采用 $\phi 48 \times 3.5$ 圆钢管，弦杆与腹杆理论长度和下料长度有所差异，数量上分别都为 36 根（共 72 根）。杆件端部由

螺栓套筒和销钉组成，可通过扳手旋转固定在节点上。

(3) 支座设计：支座整体主要由底板、螺栓垫板、双面肋板、单面肋板组成。其中地板尺寸为 $250 \times 250\text{mm}$ ，支座高 300mm 。螺栓球通过点焊方式固定在支座上，焊接之前确保 8 个支座位于同一高度上，尽量减少误差。

(4) 实验场地：实验场地位于天津市津南区北闸口镇前进村某一院落内，实验场地四面有 2 米高挡板维护，网架上方无任何遮挡。其中室内放置采集仪、计算机等相关设备，通过线缆与布置在网架上的传感器进行连接。



图 3-4 实验场地实拍图

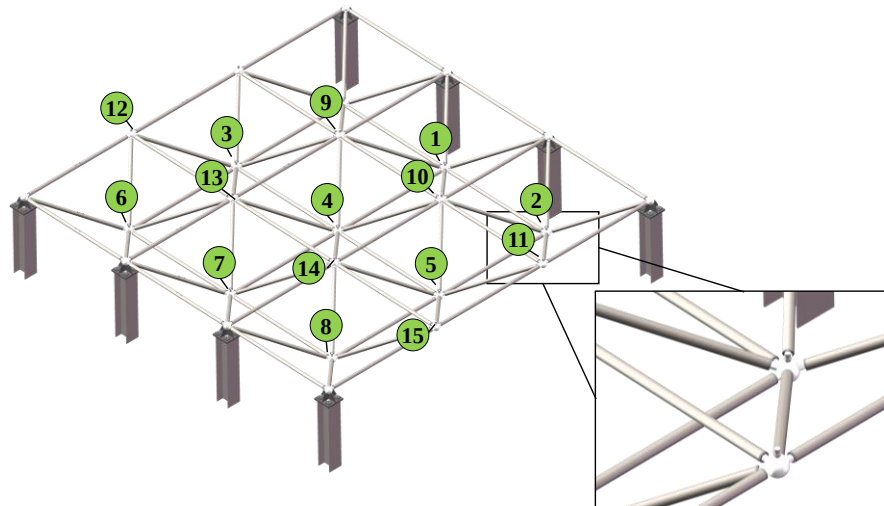
(5) 数据采集：采集数据所使用的设备型号以及规格如下表所示。

表 3-4 主要硬件设备明细表

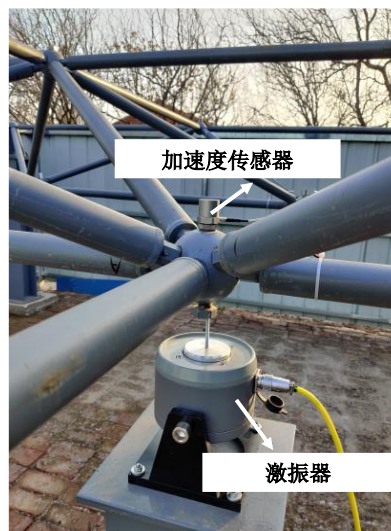
类别	型号	规格	数量
激励设备	激振器	YE-5	最大激振力 50N
	信号发生器	YE5871A	2~20kHz
	功率放大器	YE1311	阈值 28V 8.4A
采集设备	加速度传感器	众晟 ZS-AI202C	灵敏度:2V/g,量程:±2.5g, 频响:0.2Hz-1kHz
	采集仪	东方所动态采集仪	16 通道
	计算机	惠普工作站	cpu: i7-6x

激励方式采用激振器对正中央单元的下弦节点进行激励。信号发生器可以产生高斯白噪声(频带 $2\text{Hz} \sim 20\text{kHz}$)，通过功率放大器将放大的信号源输入激振器，

产生稳定的激励。激振器作用点位置坐标位于途中红箭头指向处。共 15 个 z 向加速度传感器布置在网架上弦、下弦节点上，支座处不布置加速度传感器。所有加速度传感器均通过线缆与室内的采集仪连接。具体布置方案如下图所示。



(a) 加速度传感器布置图



(b) 加速度传感器布置实拍图

图 3-5 加速度传感器

本次实验采集方式为每隔一个小时采集 3 次数据，每次采集时间为 2 分钟，采样频率统一为 500Hz，每次采集到的数据长度为 30000。本次实验共采集了两周共计 14 天的时间，为了便于后续的存储与读取，将数据存储结构设计为“日期-时间-采集次数-传感器”的格式，如下图所示。

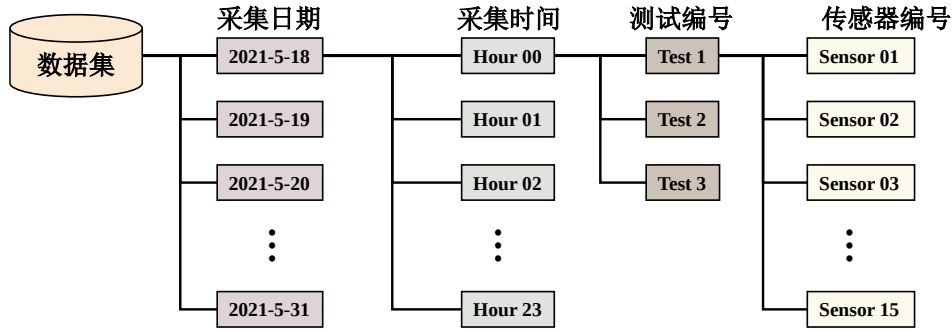


图 3-6 数据结构图

3.3.2 构造异常数据类型与方法

在本节中，主要介绍人造异常数据的构造方法，主要运用到了 python 自带 random 等函数包。具体的函数介绍如下表所示。

表 3-5 构造异常数据所用函数介绍

函数名	实现功能
<code>randint(x_1, x_2)</code>	在(x_1, x_2)的范围内随机抽取 1 个随机整数。
<code>uniform(x_1, x_2)</code>	在(x_1, x_2)的范围内随机抽取 1 个 float64 编码的浮点数。
<code>append(x)</code>	在 list 格式的数据中加入新元素 x 。

3.3.2.1 离群值

当离群值达到总样本数据点的 5%时，可以认为该样本集中出现了较多的离群值。具体应用随机函数实现数据异常构造伪代码如下：

```
>> Begin
>> 新建离群值数据点的指针列表 rand
>> for 每个指针 in rand :
>>     每个指针 = randint(0,数据总长度)    # 数据中随机抽取样本点
>>     End
>> for 每个离群点 in 原数据串:
>>     放大每个数据为 uniform(3,8)倍数    # 离群值点幅值随机扩大 3~8 倍
>>     End
>> End
```

3.3.2.2 基线漂移

为了模拟真实工程状况中出现的基线漂移情形，在整成数据中加入一个随机斜率的一次函数的趋势，来模拟因为设备故障而出现的异常数据情况。具体实现的伪代码如下：

```
>> Begin
>> # 首先构造一次函数
>> 斜率 $k$  = uniform(-2,2) / 数据总长度
>> 截距 $b$  = uniform(-2,2)
>> 一次离散函数表达式  $y_0 = kx + b$ 
>> # 生成一个带有高斯白噪声的一次函数序列
>> for 每个数据点 in  $y_0$ :
>>     加入噪声  $y_{\epsilon} = y_0 + \text{uniform}(-0.1, 0.1)$ 
>>     End
>> # 最后将原始数据和包含噪声的一次函数进行叠加
>> for 每个数据点 in 原数据:
>>     漂移数据  $y_{\epsilon} = \text{原数据 } y + y_{\epsilon}$ 
>>     End
>> End
```

3.3.2.3 矩形波状

因未知设备电路的相关信息，故在本文中生成随机步长的矩形波，循环周期数由数据总长度与步长的比值确定，数据上下阈值定为 $\pm 0.1\text{m/s}^2$ 。具体地，可以通过在一个二维数组中循环添加周期数据来实现矩形波的构造，伪代码如下：

```
>> Begin
>> 新建一个空白的序列 $S$ 
>> 定义上下阈值  $y_0, y_1$ ，以及周期长度  $x_{unit}$  与周期占比  $ratio$ 
>> for 每个矩形波单元 in 总数据长度:
>>      $S$ 中添加初始 $x$ 值和 $y$ 值( $x_0, y_0$ )
>>     定义中间值  $m = \text{初始值 } x_0 + \text{周期长度 } x_{unit} * \text{周期占比 } ratio$ 
>>      $S$ 中继续添加带上限阈值的新样本点( $m, y_0$ )和( $m, y_1$ )
>>     定义  $ed = \text{初始值 } x_0 + \text{周期长度 } x_{unit}$ 
>>      $S$ 中最后添加带上限阈值的新样本点( $ed, y_1$ )
```

```
>> End
>> End
```

3.3.2.4 部分数据缺失

假设当超过样本数据总长度 30% 的数据点缺失时，视为异常情况。随机在总数据长度中选取两个间隔超过 30% 数据总长的时间节点进行截取，将这一范围内的数据点替换为 NaN 值。具体实现的伪代码如下所示：

```
>> Begin
>> # 设置缺失数据段的开始以及结束节点
>> 定义缺失长度范围  $range = 30\% \times length$ ,  $length$  为数据总长度
>> 定义缺失开始节点  $time_1 = randint(0, range)$ 
>> 定义缺失结束节点  $time_2 = randint(range + 0.3length, length)$ 
>> for 数据点 in 范围( $time_1, time_2$ ):
>>     del(数据点)
>>     数据点位置 append(NaN)
>> End
>> End
```

3.3.2.5 幅值过小

对于幅值过小的异常情形，只需将单位制等比例缩小即可，这里采用随机函数对每个数据点的加速度值进行 2 至 5 倍的缩小。具体实现过程如下：

```
>> Begin
>> # 设置缺失数据段的开始以及结束节点
>> for 数据点 in 范围( $time_1, time_2$ ):
>>     数据点 / uniform(2,5)
>> End
>> End
```

3.3.3 分类结果与讨论

在本小节中，生成的异常数据将和正常的数据一起形成两组数据集作为对照试验。第一种数据集包含正常数据以及五种异常数据，且每一类的数据量均为 500；

第二类数据同样包含正常数据以及物种异常数据类型，但其中正常数据占 50%，其余五种异常数据类型各占 10%。一般来说，第二种不均衡的数据集类型更加符合真实的 SHM 数据库类型，正常数据占据主体，出现异常数据属于小概率事件。然而不均衡的数据集会给分类器带来挑战，往往会造成训练出一个带有偏向性的分类器。在本小节的实验中，每次进行试验时会将数据集随机地按照 7：3 的比例划分为训练集和测试集，并且将讨论三种分类器分别对于均衡数据集和不均衡数据集的情况进行讨论分类效果，其主要参数信息如下表所示：

表 3-6 三种分类器参数信息

名称	主动学习	质询策略	分类器	标记样本比例
LAL-AdaBoost	是	LAL	AdaBoost	10%
US-AdaBoost	是	US	AdaBoost	10%
AdaBoost	否	无	AdaBoost	100%

前两种算法分别为使用主动学习框架下的 AdaBoost 算法，根据质询策略的不同记为 LAL-AdaBoost 与 US-AdaBoost；第三种分类器使用基础的 AdaBoost 算法，不涉及任何主动学习框架，用于对照实验。对于样本标签的数量，初始训练集含标签的样本占总数据 2%，通过质询策略进行迭代后每次质询一个样本点进贴标签的工作，直到含标签样本占总训练集比例达到 10%就停止质询并结束训练过程。AdaBoost 算法的具体参数如下表所示：

表 3-7 AdaBoost 分类器参数设置

集成方式参数		弱分类器（决策树算法）参数设置		
分类器数量	学习率	最大层数	分裂样本数	分裂准则
50	0.5	10	2	Gini Impurity

特别地，对于 LAL 算法中使用到的回归器的算法进行介绍。随机森林回归器算法（Random Forest Regressor，简称 RFR）由多个决策树算法（Decision Tree，简称 DT）组成，通过投票委员会进行整合。该算法对于多维度特征的拟合具有较好的表现。在本算例中，采用 50 个最大深度为 40 层的 DT 组成 RFR 算法进行对于学习参数的回归过程。

最终的分类效果通过总体正确率 *accuracy* 和 F_1 值进行评判。这两个指标的具体推导过程如下：

假设一类样本的标签值为 y ，那么非该类样本的标签值记为 $y \neq$ 且满足

$y \in Y$ 。对于任意一个分类器在测试集上对于标签为 y 的样本的分类结果可以分为以下四种可能性：（1）真阳性 tp ：对于真实值为 y 的标签预测值为 y ；（2）真阴性 tn ：对于真实值为 $y \notin Y$ 的标签预测值为 $y \notin Y$ ；（3）假阳性 fp ：对于真实值为 $y \notin Y$ 的标签预测值为 y ；（4）假阴性 fn ：对于真实值为 y 的标签预测值为 $y \notin Y$ 。由此可以引出精确度(*precision*)和召回率(*recall*)的概念，其表达式如下：

$$precision = \frac{tp}{tp + fp} \quad (3-9)$$

$$recall = \frac{tp}{tp + fn} \quad (3-10)$$

由此可以给出正确率 *accuracy* 和 F_1 值的计算公式：

$$accuracy = \frac{tp + tn}{tp + tn + fp + fn} \quad (3-11)$$

$$F_1 = 2 \frac{precision \times recall}{precision + recall} \quad (3-12)$$

其中，对于数据异常识别问题而言，我们比较关注的指标为召回率、正确率以及 F_1 值；由于相较于误判正常样本为异常值，遗漏数据异常的代价是更大的，因此我们不期望任何包含异常的样本数据被分类器错分类为正常样本。

3.3.3.1 均衡数据集分类表现

为了避免实验的偶然性对实验结果的干扰，在每组实验时对样本集进行 10 次随机划分训练集和测试集分别进行实验。在均衡数据集上，三个分类器的表现非常接近。理论上来说，AdaBoost 算法由于充分考虑了整个训练集的分布情况，在测试集上的表现应该最佳。对于主动学习框架下的机器学习算法而言，其质询策略决定了对于样本偏向性的捕捉，如果一个质询算法无法准确质询到易被错分类的样本，那么加入训练集的样本往往会扩大分类器对于某一类标签样本的错分率。

下图展示了 LAL-AdaBoost 与 US-AdaBoost 算法的分类曲线。起始点对应的分类状态是 AdaBoost 通过最初给予的标签训练集数据学习模型在测试集上的精度，可以看到两种方法的起始精度均偏低。但是随着质询次数逐渐增加，分类器的精度也随之增大。黑色虚线代表的是 AdaBoost 的分类正确率，在经过 80 次质询后，LAL-AdaBoost 与 US-AdaBoost 的精度均与之接近。这说明两种主动学习的方法都可以准确地筛选出具有价值的质询样本点，使得分类器的精度最大程度

上与 AdaBoost 靠拢。在拟合速度上, LAL-AdaBoost 表现要略微优于 US-AdaBoost。最终三种分类器总体正确率都可以达到 96% 左右, 基本认为可以较为准确地识别出正常与异常数据。

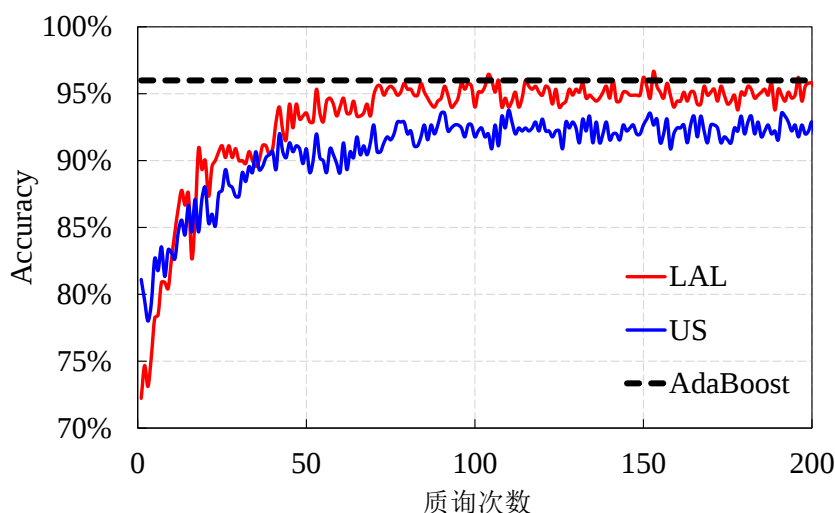


图 3-7 三种分类器对于均衡数据集的分类曲线

为了分析三种分类模型对于各类样本的偏向性, 计算出三组实验模型对于每一类标签数据的 F_1 值进行分析。在此图中各数据异常类型为便于展示, 用英文简写表示 (normal: 正常数据、trend: 基线漂移数据、outlier: 离群值过多数据、minor: 幅值偏小数据、square: 矩形波状数据、missing: 部分缺失数据)。可以从图中看出三组分类器表现十分接近, 对于每一类标签的数据没有体现出较为明显的趋向性, 分类效果较好。对于标签为 normal 和 outlier 的样本, 三组分类器的 F_1 值均达到了 0.95 左右; 而对于其他四种样本, 三组分类器的 F_1 值均接近于 1.00, 说明分类效果趋近于完美。

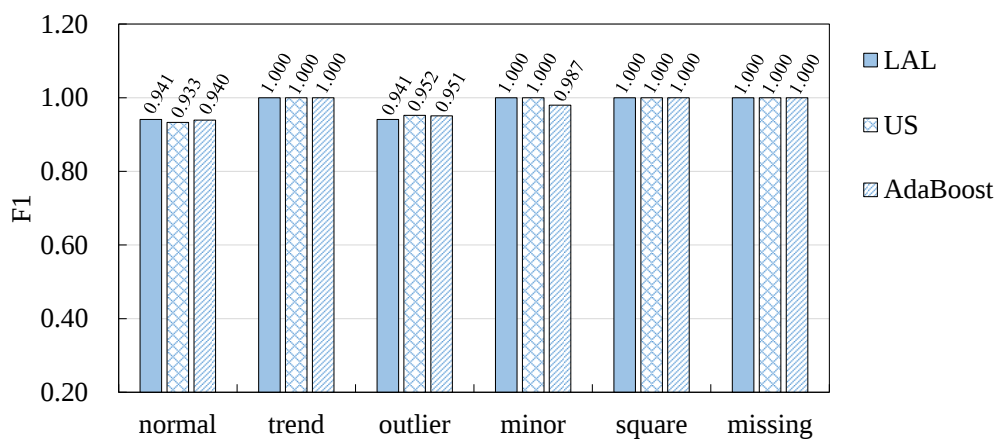


图 3-8 三种分类器对于每一类样本的 F_1 值

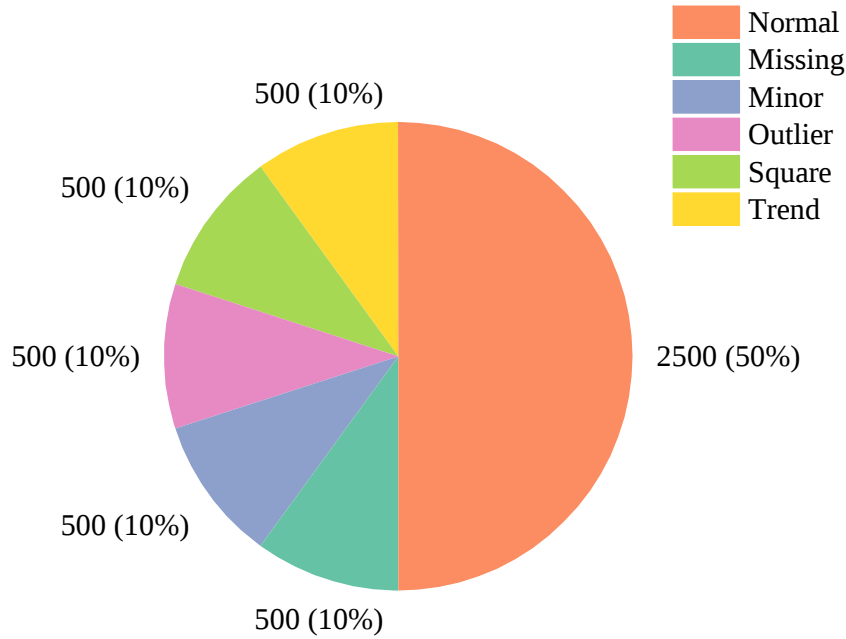


图 3-9 不均衡数据集各标签的数据量对比

3.3.3.2 不均衡数据集分类表现

对于不均衡数据集，三种分类器的分类结果表现出了较大的差距。不均衡数据集中各标签的数据量比例如图 3-9 所示。在实际工程中，此种数据不均衡情况较为常见，而且往往会对分类效果产生不利影响，因为分类器为了满足分类总体的精度值而牺牲掉少数类别标签的样本。一般的做法为对于数据本身进行改造与处理，例如过采样和降采样的方法，然而这些方法的局限性体现在无法充分利用数据和其中的特征进行训练模型。LAL 主动学习的算法在这方面的做法是通过训练出一个回归器来预测可能出现的分类结果，从而筛选出可以尽量减少偏差的样本点。然而，主动学习的质询策略如果无法合理地判断所要质询的样本点，将会对分类器的偏差进行进一步地放大从而对不同标签的数据产生更加悬殊的分类效果。

从分类结果来看，LAL-AdaBoost 表现最为稳定，表现出了较高的总体分类正确率（约 93%，与 AdaBoost 类似）。LAL-AdaBoost 在拟合速度上也明显优于 US-AdaBoost 算法，前者在质询 50 次时已经趋于稳定，后者则在质询达到 100 次时才开始收敛。质询停止时，LAL-AdaBoost 的分类正确率明显高于 US-AdaBoost，具体来说，前者在少数标签的样本上表现更为出色，这一点可以在 F_1 值上看出（图 3-11）。在对于标签为 outlier 的样本集上，LAL-AdaBoost 达到了

0.89 的 F_1 值，遥遥领先了 US-AdaBoost 的 0.73，甚至比 AdaBoost 的 0.86 还要略高。这说明在质询过程中 LAL-AdaBoost 可以准确预测出可能的样本分布情况，使得分类器效果往克服偏向性的分类结果靠拢。此外，在其他标签的分类对比中可以发现 AdaBoost 对于 minor 标签的数据分类效果欠佳，原因是 minor 数据很容易和 outlier 数据混淆，因为在小波包能量分布上二者有一定的相同之处。而基于 LAL 主动学习框架的分类器则可以克服这一缺点，避免一些干扰分类的数据混入训练集中，从而提升分类精度。

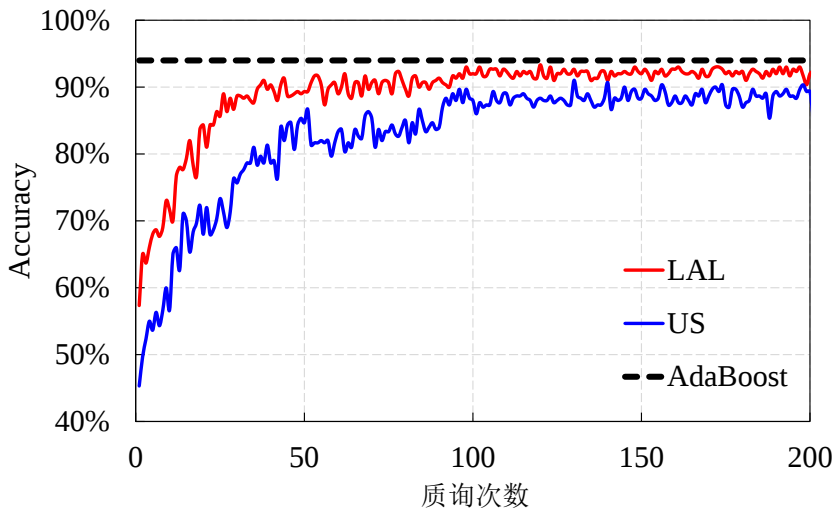


图 3-10 三种分类器对于不平衡数据集的分类曲线

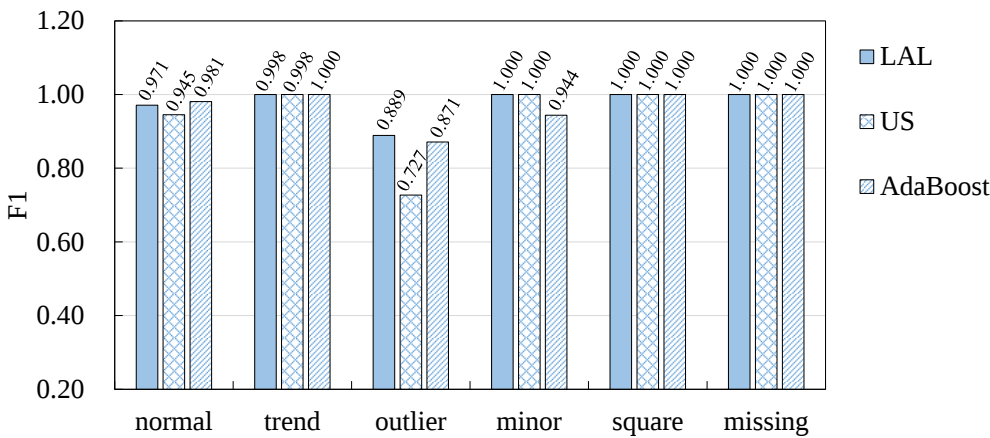


图 3-11 三种分类器对于每一类样本的 F_1 值

为了更直观地分析出实验数据样本对于分类结果的影响，在下图中以混淆矩阵的形式绘制出了分类结果展示，并在特定的分类框内挑选了具有代表性的一段

或几段数据进行展示。矩阵的纵列代表 LAL-AdaBoost 的预测值，矩阵的横行代表数据样本的真实标签。可以从该矩阵中进行以下分析来解释分类结果：

(1) 对于真实标签为 Normal 类的样本而言，整体召回率较高，仅有少部分样本被错误分类为了 Trend、Outlier 与 Minor 类。

(2) 对于真实标签为 Trend 类的样本而言，整体来说分类效果较好，少数样本由于趋势不明显容易被忽略，但在实际工程中，微小的基线漂移情况不会很大程度上影响数据的分析过程。

(3) 对于真实标签为 Outlier 类的样本而言，召回率相较于其他类样本偏低。主要的原因是某些样本中离群值的比例较低，或部分离群点的幅值较低，导致识别不出 Outlier 样本。大多数被错误分类的该类样本的预测值为 Normal，二者在时域与频域上的特征较为接近，其小波包能量占比分布规律类似，故易出错。

(4) 对于真实标签为 Minor 类的样本而言，少数样本的幅值缩小比例不足以与 Normal 类样本体现出差距，因此被错分。在频域上，其小波包能量占比与 Normal 类数据也比较类似。

(5) 对于真实标签为 Square 与 Missing 类的样本而言，总体来说错分的标签较少，除了少数 Missing 类样本由于缺失段长度较短，去除 NaN 值后数据依然具有完整性，因此被错误分类为 Normal。

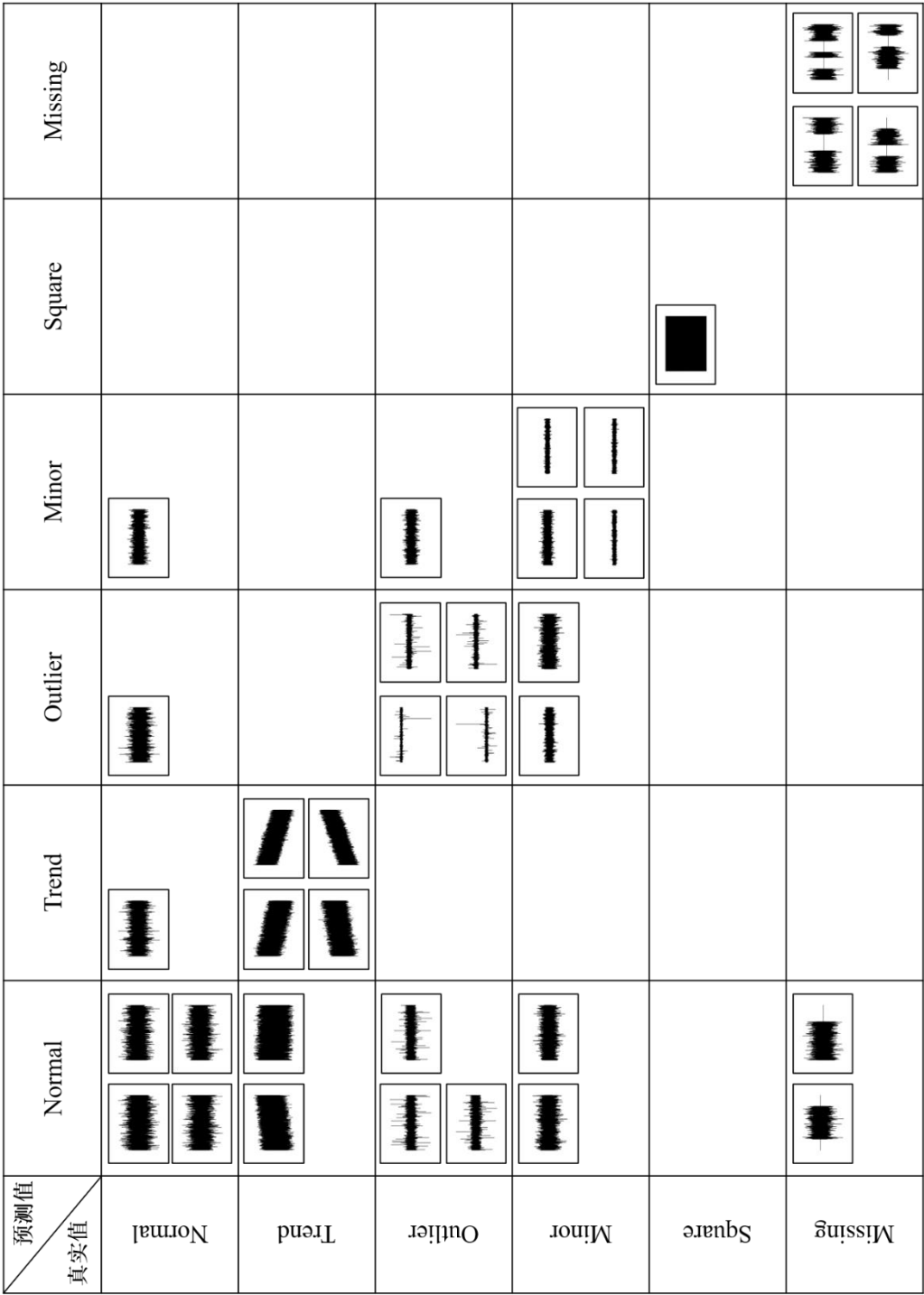


图 3-12 分类结果以混淆矩阵形式展示

3.4 本章小结

在本章中，作者主要介绍了关于结构健康检测数据预处理与异常识别的相关研究，并将 LAL-AdaBoost 算法引入到了空间网架检测数据分析中。具体工作内容如下：

1. 介绍了两种常见的数据异常处理的方法：快速傅里叶变换（FFT）和小波

包分解 (WPT)。二者都可以对时域信号进行频域分析, 但是前者侧重于对于数据总体的频域分析, 而后者则适用于突变信号的频域分析。

2. 提出了基于 LAL-AdaBoost 算法的数据异常识别框架, 包含了包括特征提取、LAL 训练过程、验证结果三个主要部分。通过对于时域数据的小波包变换得到小波包能量占比进行归一化处理的特征向量作为 LAL-AdaBoost 的输入, 用于训练 LAL-AdaBoost 算法来进行质询过程以及训练过程, 最后通过输出 F_1 值以及分类曲线来对结果进行验证以及分析。

3. 设计了一个空间网架实验并构造出了数据集: 首先介绍了空间网架结构设计方案以及数据采集过程; 通过随机函数等方法构造出了一批包含五种异常类型的加速度时域数据集用于后续研究, 并将数据集分为两组, 第一组为均衡数据, 而第二组为不均衡的数据集, 即正常数据占据大部分。

4. 分别使用均衡数据集与不均衡数据集进行验证 LAL-AdaBoost 数据异常识别框架, 在经过 10% 样本的质询后输出最终分类结果, 从分类曲线与 F_1 指标的计算结果中可以看出 LAL-AdaBoost 不仅在均衡数据集上表现较好, 而且在数据集样本不均衡时也可以避免干扰而获得较好的分类结果。

第 4 章 损伤指标与损伤识别方法研究

在本章中，作者提出了一个全新的损伤识别方法，通过基于动力测试的损伤指标并通过 AdaBoost 算法来研究指标对于损伤的敏感性进行判定；具体的损伤识别框架在网架结构实验上进行了验证。首先，通过有限元模拟，可以得到网架结构的各损伤工况的加速度时域数据集，利用这些数据来训练一个高精度的 LAL-AdaBoost 分类器模型用于捕捉结构损伤的信息；具体地，作者评估了损伤识别框架的鲁棒性以及泛化性分别被加以验证；同时，为了体现出 LAL-AdaBoost 算法的优越性，US 算法同时被用于对比实验。

4.1 基于动力测试的损伤指标

对于损伤识别流程来说，能否从原始数据中提取出关键敏感的损伤指标决定了后续数据分析工作能否顺利进行。基于动力测试的健康检测中，较为常见的损伤指标一般为时域或频域的相关信息，常见的指标有峰峰值、峰值因子、脉冲因子、裕度因子、基本频率、重心频率、频率均方根、频率标准差等。这些指标在统计学上具有一定的数学意义，可以反映出一段加速度信号在时域上的波形以及其在频域上的频段分布情况。研究表明，对于长期动力监测的数据分析而言，这些指标是有效的，而且原理较为简单，提取指标的过程也比较容易。然而，上述基于统计学的时域与频域指标只能反映单个传感器的信号特征，各个传感器采集数据之间存在独立性；而对于复杂的空间结构而言，多个传感器形成的传感器网络（Sensor Network）内部存在极高的关联度，对于分析损伤具有重要的意义，不应该被忽略。

在本文中，提出了一个全新的考虑传感器网络互相关性的损伤指标，用于作为损伤定位研究的输入变量。该指标的核心为计算两段同时采集的序列的互相关系数，计算过程如下：

(1) 首先选择某一个位置的传感器位置的传为参考点，该传感器在某段时间内采集的长度为 l 的数据可以写作 $X_R = [x_{R1}, x_{R2}, x_{R3}, L, x_{Rl}]$ ；同时另一处传感器采集的长度为 l 的数据可以写作 $X = [x_1, x_2, x_3, L, x_l]$ 。假设两个加速度传感器为同向采集的传感器。

(2) 按照下式分别计算两个序列 X_R 与 X 的均差值（Deviation Mean，简称 DM），其中 m 为序列的均值。

$$DM(X) = [x_1 - m, x_2 - m, L, x_l - m] \quad (4-1)$$

(3) 求出相关性系数。首先定义函数 $\text{dot}(X_1, X_2)$ 为两个序列每一项乘积的和，即：

$$\text{dot}(X_1, X_2) = \sum_{i=1}^l (x_{1i} x_{2i}) \quad (4-2)$$

按照下式求出相关性系数：

$$\text{cov}(X_R, X) = \frac{\text{dot}(X_R, X)}{l} \quad (4-3)$$

根据上述计算理论，首先考虑两个布置在同一个结构上不同位置的传感器 i 与传感器 j ，其在同一时间段内采集到的时域数据流分别命名为 d_i 与 d_j 。由于在本实验中，为了模拟自然状态下的结构响应主要采用的是白噪声激励方式，同时考虑到两端序列在时间上会存在延迟，因此在这里主要关注数据的频域特性。通过快速傅里叶变换（3.1.1 节），可以分别得到 d_i 与 d_j 的频谱，下面计算他们的互相关系数：

$$f_{ij} = \text{cov}(\text{FFT}(d_i), \text{FFT}(d_j)) \quad (4-4)$$

将上式推广到包含 n 个传感器的网络中，充分考虑每一个传感器与其他 $n-1$ 个传感器之间的相互关系，该指标可以用一个 $n' \times n$ 的矩阵的形式来表达（ n 表示传感器总数量）：

$$F = \begin{bmatrix} f_{11} & f_{12} & f_{13} & \dots & f_{1n} \\ 0 & f_{22} & f_{23} & \dots & f_{2n} \\ 0 & 0 & f_{33} & \dots & f_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & f_{nn} \end{bmatrix} \quad (4-5)$$

式(4.5)中的矩阵是一个对角元素均为 1 的上三角矩阵，然而其中包含了有效信息的项的数量为 $n(n-1)/2$ 。因此该特征矩阵可以展开写作一维形式作为简化处理，用下式表达：

$$F = [f_{12}, f_{13}, \dots, f_{1n}, f_{23}, f_{24}, \dots, f_{2n}, f_{34}, \dots, f_{3n}, \dots, f_{(n-1)n}] \quad (4-6)$$

4.2 基于 LAL-AdaBoost 算法的损伤识别框架

在本节中，作者主要总结了本课题中进行损伤识别研究的技术路线图。本课题中的损伤识别框架旨在通过 LAL-AdaBoost 算法去拟合所提出的损伤指标与对应损伤发生事件的深层映射关系，并为后续的空间结构的长期监测打下理论与模

型基础。本算例中仅考虑 6 种损伤工况，然而对于实际结构而言，往往不止于此；因此其需要更多的实验数据进行拟合，分类样本的类别较多，这对于机器学习算法而言无疑是加重了其训练成本，并且训练结果也不无得到保障。而本研究框架所使用的 LAL-AdaBoost 则可以显著减少训练样本的数量，通过筛选样本减少训练所需时间，从而提高机器学习算法的训练效率。同时，由于主动学习框架带来的训练样本量的减少，机器学习训练的时间也得到了缩减，可以更快地拟合出高精度的分类曲线。

在本课题提出的损伤识别框架中，训练样本所使用的数据库源自于有限元模拟，由此建立各工况下结构的响应信息；损伤识别任务被分为了两部分，分别是损伤检测与损伤定位，其中前者是对于结构中是否存在损伤进行判断，属于一个二分类问题；而后者则是基于前者的判断后进一步确定损伤发生所在的单元号。通过 LAL-AdaBoost 算法，可以捕捉到结构自身对于损伤发生后产生的响应变化，从而对损伤进行识别。

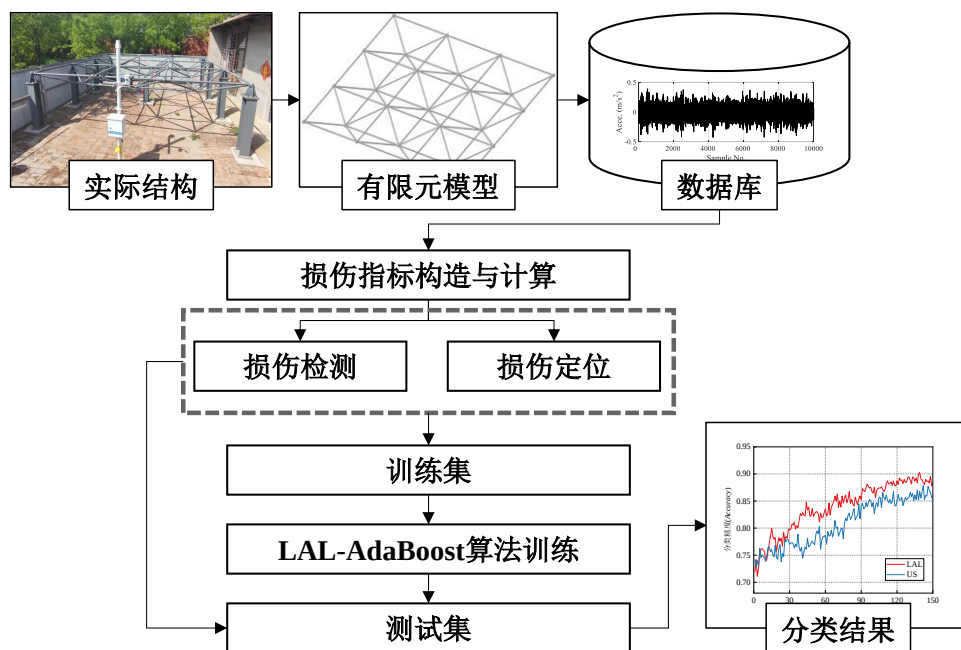


图 4-1 损伤识别框架

为了保证 LAL 主动学习算法不被划分标签样本集与无标签样本集的偶然性影响，在每次进行实验时，将数据集重复随机划分十次，分别进行分类实验。这十次实验所得到的分类结果后，通过求均值而得到可以评判主动学习算法以及分类器性能的混淆矩阵、分类曲线、 F_1 值等有效指标。

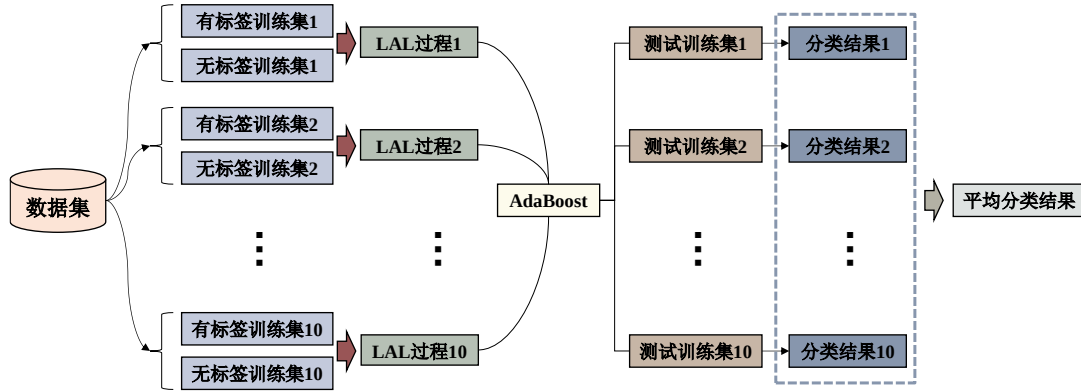


图 4-2 主动学习实验流程

4.3 基于 LAL-AdaBoost 算法的空间网架结构损伤识别研究

4.3.1 健康与损伤数据库建立

在空间网架结构中，杆件较多而且往往具有对称性，为了实现广义上的损伤识别研究，不能局限于个别杆件的损伤工况作为研究对象；而且在实际结构中，受到监测环境与监测成本的限制，往往不能均布传感器，因此复杂的空间结构的动力数据不可能完整获取。基于上述原因，作者在本课题中以每个正放四角锥单元为一个损伤单元进行定位研究，通过建立有限元模型并折减单元内的杆件的刚度来模拟实际工况中可能出现的损伤情况。具体做法为：若某正方四角锥单元内出现了某种程度的损伤，那么可以从组成该单元的四根上弦杆与四根腹杆中随机地选取两根杆件的截面刚度进行一定比例的折减，从而构造出相对应工况的损伤，模拟出类似于现实中结构可能出现的损伤类型。

由于本课题的空间网架实验结构具有双轴对称的特性，在本论文中主要讨论其中四个独立单元的损伤工况（共计九个单元）。由于实际结构的损伤监测数据一般难以获得，而且也往往无法得到全部工况的数据进行对比分析，因此针对损伤识别的研究一般会基于有限元模型进行训练数据的获取与特征提取过程。

通过 ANSYS Mechanical APDL 软件进行有限元模型的建立，基本建模信息如下表所示。由于支座处球节点为完全焊接节点，而且工字钢柱子的截面刚度较结构本身刚度而言较大，因此可以假设节点为完全刚性节点；具体构件建模信息如表 4-1 所示；输入荷载使用 MATLAB 生成伪随机高斯白噪声作用在中心下弦节点模拟激振器激励，激振频率为 500Hz；最后使用 post26 进行后处理，并输出网架结构各节点 z 向加速度时程数据。

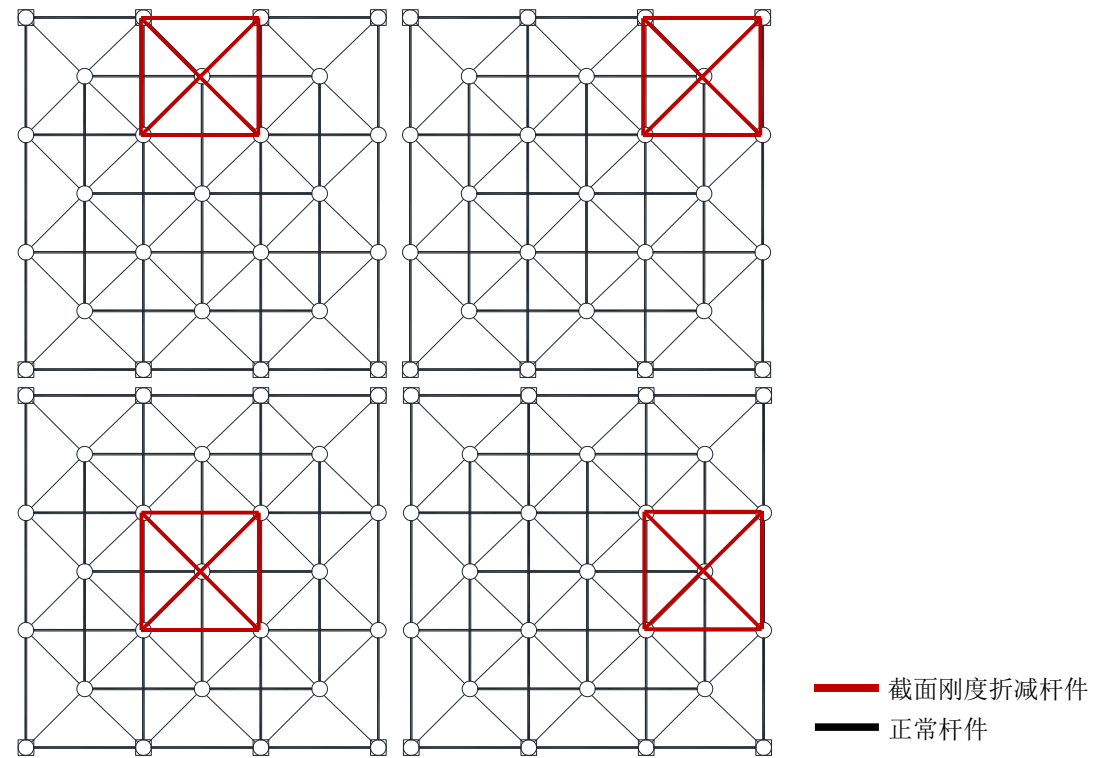


图 4-3 损伤模拟示意图

表 4-1 有限元建模信息

构件	单元	参数信息
杆件	BEAM188	泊松比 0.3，弹性模量 206Gpa，CTUBE 子类型，并保持截面尺寸与实际结构一致
损伤杆件	BEAM188	将 CTUBE 定义截面尺寸进行一定程度的缩减
支座	COMBIN14	x, y, z 三个方向加刚度极大的弹簧约束

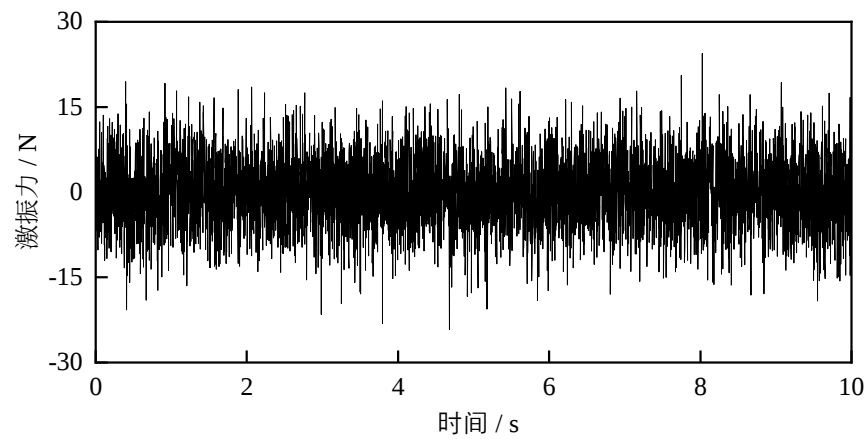


图 4-4 有限元模拟输入荷载（10s）

表 4-2 有限元模型模拟数据与实测数据前三阶频率对比

阶数	实测数据	有限元模拟	相对误差
1	25.4Hz	32.7Hz	28.7%
2	42.5Hz	50.2Hz	18.1%
3	72.2Hz	64.8Hz	11.4%

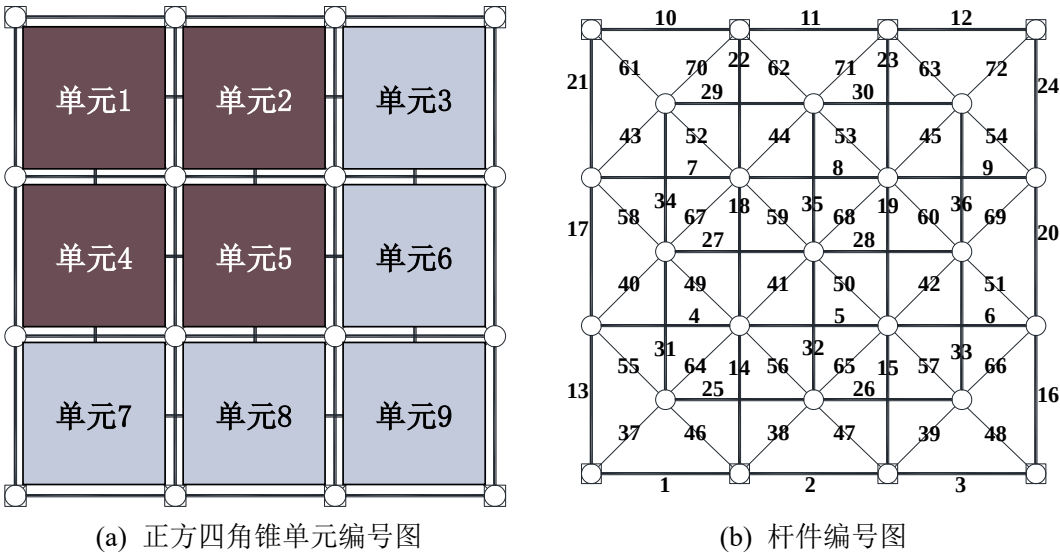


图 4-5 构件及单元编号图

表 4-3 损伤工况设计

工况	损伤杆件号	截面刚度折减比例	样本数量	研究内容
0	无损伤	0%	每组数量均	对照组
1	5,8,18,19,41,50,59,68	10%	为 90 组，每	研究损伤程
2	5,8,18,19,41,50,59,68	30%	组包含 10s	度识别
3	5,8,18,19,41,50,59,68	50%	的动力数	
4	8,11,22,23,4,53,62,71	50%	据，采样频	研究损伤位
5	9,12,23,24,45,54,63,72	50%	率 500Hz	置识别
6	6,9,19,20,42,51,60,69	50%		

通过对于有限元模拟与实测数据所得的结构前三阶频率的对比（表 4-2）可以基本认为有限元模型的动力响应与实测结构较为接近。在本次模拟实验中，共设计了一种基准健康状态工况和六种损伤工况，具体设计情况如上表所示。由于有限元模型按照圆形管状单元进行建模，因此通过缩减圆管截面尺寸可以模拟出截面刚度的缩减。对于损伤程度的研究工作，选取 5 号正方四角锥单元为例进行

研究,分别计算了以下几种损伤程度的结构加速度响应:0%(基准工况)、10%、30%、50%。一般来说,真实结构可能存在的损伤一般不会出现刚度折减超过50%的情形,因为这往往意味着杆件失稳;对于复杂的空间结构而言,每个节点的超静定次数较高,因此在本实验中仅考虑最多50%的截面刚度折减比例。对于损伤位置识别的研究,考虑了结构具有对称性,因此工况3、4、5、6分别针对于正方四角锥单元5、6、8、9的损伤情形,目标为通过算法来识别定位损伤的单元位置,这里仅考虑杆件单元截面刚度折减50%的情形。以上的每种工况均分别进行90次的采样,每次采样频率为500Hz,采样时间为10s,激励条件为带宽为0.2Hz~2000Hz的高斯白噪声。本算例所拟定的几种损伤程度,其研究重点在于定性地判别不同程度的杆件刚度折减在进行损伤检测与损伤定位过程中对算法分类效果的干扰,进而验证算法的鲁棒性。

4.3.2 损伤诊断研究

在本小节中,作者主要研究了LAL-AdaBoost算法对于结构损伤的判定,主要分为以下几个研究内容:(1)单一程度的损伤检测。当结构存在单一损伤程度(50%)的损伤时,进行损伤检测的研究,输出为存在或不存在损伤;(2)泛化性研究。本数据库中仅包含几个特定损伤程度的监测数据,为了验证方法的泛化性,可以先使用50%损伤程度的数据训练模型,使用10%和30%损伤程度的数据作为输入验证,研究模型是否可以发现较为轻微的损伤;(3)鲁棒性研究。当数据样本中存在多个损伤程度的样本时,探究模型是否可以将其区分。为了体现出LAL-AdaBoost算法的有效性,同时使用了US-AdaBoost算法进行了对比研究,分别比较二者的分类曲线以及分类结果。

4.2.3.1 单一程度损伤检测

对于一个损伤识别算法,最基本的功能为检测到结构是否出现了损伤,即损伤检测。对于不同位置的损伤,通过计算4.1节中提出的损伤指标并借助机器学习模型来分析该指标的改变,可以学习到损伤引起的损伤指标变化规律,从而用于判定结构是否发生损伤。

(1)数据集组成:基准工况0+损伤工况3~6。其中,将标签设计为二分类问题进行分类算法的训练,即:

$$y_i = \begin{cases} 0, & \text{if } i = \text{健康工况} \\ 1, & \text{if } i = \text{损伤工况} \end{cases} \quad (4-7)$$

(2) LAL-AdaBoost 训练过程: 首先按照4.2.2节提出的损伤识别框架进行数

数据集的划分工作。为了节约训练成本，使用 LAL 算法进行数据的筛选过程。为了防止随机划分样本的不确定性，将训练集随机地按照 *initial_rate* 的比例重复划分 10 次，即参数 *split_count*=10，并且在每次划分时都确保数据集 L_0 与 U_0 均包含标签为 0 或 1 的样本。具体地 LAL 算法的参数如下表所示。

表 4-4 LAL 参数设置（单一程度损伤检测）

参数名称	参数意义	参数取值
<i>initial_rate</i>	初始迭代时 L_0 占总样本的比例	0.05
<i>test_ratio</i>	测试集占总样本的比例	0.3
<i>split_count</i>	重复划分数据集的次数	10
<i>stop_criterion</i>	停止迭代条件	‘number of queries’=150
<i>cls_est</i>	用于运算分类器状态的随机森林回归器的决策树的数量	Default = 50
<i>cls_depth</i>	决策树的最大深度	Default = 40
<i>batch_size</i>	每次质询挑选的数据点的个数	1
<i>mode</i>	选择 LAL 算法（见 2.1 节），可选 LAL-Independent 或 LAL-Interactive	Default = ‘LAL-Interactive’

(3) 分类结果讨论：对比 LAL 算法与 US 算法对于该数据集的分类表现，分别绘制分类曲线（横坐标为质询次数，纵坐标为分类正确率 *Accuracy*，并保持两个算法的 *batch_size* 均为 1）。在质询次数达到 40 次之前，LAL 算法保持更高的正确率；在 40~70 段二者精度类似，并同时处于较为平稳的阶段；而在 70 次之后的质询过程中，LAL 算法明显优于 US 算法，后者在这段训练过程中精度维持在 87% 左右，并出现上下起伏；最后在质询次数达到 130 次左右时，二者几乎同时收敛在 93% 左右的精度上。综上所述，LAL-AdaBoost 算法在单一程度（50%）的损伤检测中表现良好，不仅可以最终通过质询 30% 左右的样本达到较为理想的精度，而且在分类曲线上表现出了较快的拟合速度。

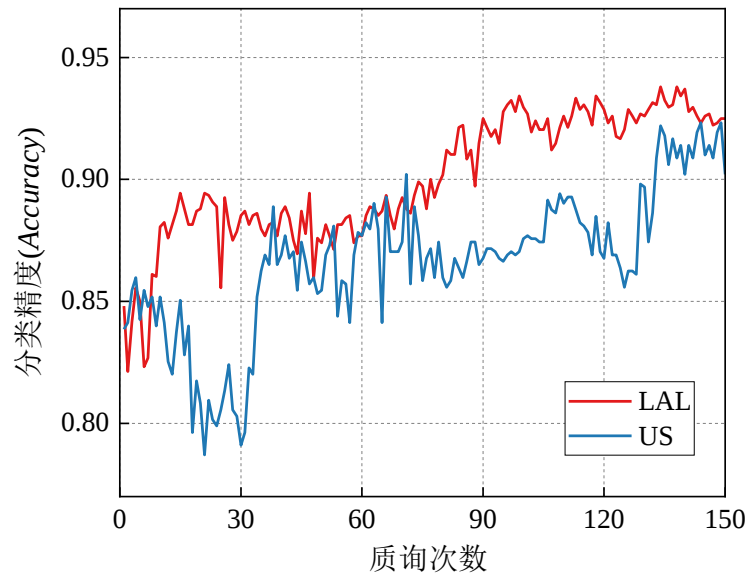


图 4-6 单一程度损伤检测算法分类曲线

从 LAL-AdaBoost 的混淆矩阵的结果来看，分类效果总体效果较好。0 类样本的召回率偏低，仅有 78.3%，这说明一部分健康工况的数据被误判为了损伤工况，这说明该算法偏于保守；0 类样本的精确度较高，达到了 90.0%，说明该算法可以有效地捕捉到结构内部出现的各种程度的损伤，可以做的有效的预警。

预测值 真实值	0	1	召回率
0	18 17.0%	5 4.7%	78.3% 21.7%
1	2 1.9%	81 76.4%	97.6% 2.4%
精确度	90.0% 10.0%	94.2% 5.8%	93.4% 6.4%

图 4-7 单一程度损伤检测算法混淆矩阵

4.2.3.2 微小程度的损伤检测

在本小节中，作者进一步探究 LAL-AdaBoost 算法的泛化性能。假设以下损伤识别情形：对于损伤程度较为严重的工况，往往模型可以较好的提取到敏感性强的特征从而进行损伤诊断；然而实际中经常发生的是损伤程度较轻的微小损伤，截面刚度折减比例相对来说较小，这就对于模型的泛化性提出了更高的要求。在本小节中，作者使用损伤程度为 50%与 30%的模拟数据进行训练 LAL-AdaBoost 算法，使用 10%损伤程度的数据进行验证（该批数据不参与模型的训练过程），

以验证机器学习模型的泛化性能。

(1) 数据集组成：基准工况 0+损伤工况 2~3（训练集）+损伤工况 1（测试集）。其中，将标签设计为二分类问题进行分类算法的训练，与上一节一致（公式 4.7）。

(2) LAL 训练过程参数设定：如下表所示。

表 4-5 LAL 参数设置（预测微小程度的损伤）

参数名称	参数意义	参数取值
<i>initial_rate</i>	初始迭代时 L_0 占总样本的比例	0.05
<i>split_count</i>	重复划分数据集的次数	10
<i>stop_criterion</i>	停止迭代条件	‘number of queries’=50
<i>cls_est</i>	用于运算分类器状态的随机森林回归器的决策树的数量	40
<i>cls_depth</i>	决策树的最大深度	20
<i>batch_size</i>	每次质询挑选的数据点的个数	1
<i>mode</i>	选择 LAL 算法（见 2.1 节），可选 LAL-Independent 或 LAL-Interactive	Default = ‘LAL-Interactive’

由于此研究数据集数量相对较少，因此适当地简化了用于处理分类器状态参数的随机森林回归算法：每个决策树的最大深度限制为 20 层，同时减少随机森林中决策树的数量为 40 个，以防止过拟合现象发生，同时节约算力。此外，由于数据集数量的相对减少，将停止迭代条件修改为“质询次数=50 次”，在确保 LAL 具有足够的性能的同时，提高学习的效率，并降低学习的成本。

(3) 分类结果讨论：从分类曲线上看，两种算法在训练集上的拟合速度几乎一致，拟合精度上 LAL-AdaBoost 算法有微弱优势。同样地，由于二者使用了相同的基分类器算法 AdaBoost，二者在测试集上的表现也较为接近：LAL-AdaBoost 精度为 91.1%，US-AdaBoost 精度为 88.9%。这说明在结果出现微小损伤时，分类器算法依然可以较为准确地识别出结构是否处于健康状态。

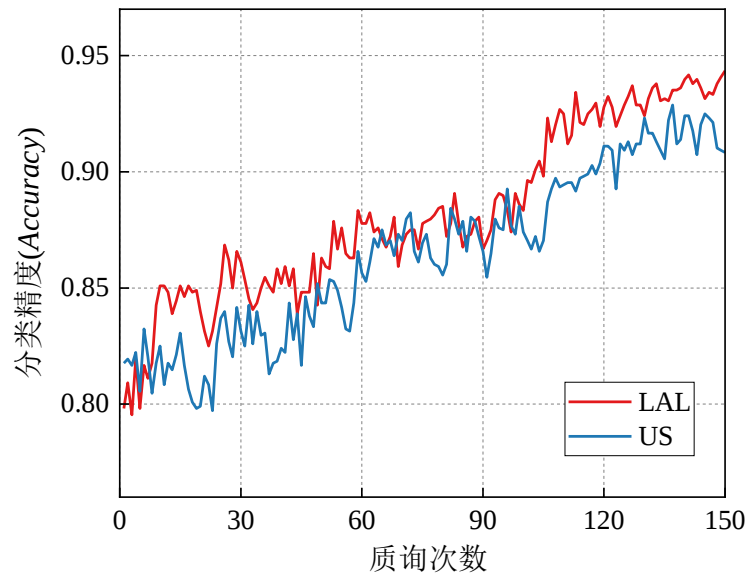


图 4-8 微小损伤检测算法分类曲线

4.2.3.3 多种损伤程度的损伤检测及分类

在本小节中，作者主要讨论 LAL-AdaBoost 对于损伤程度的回归问题研究。针对于本节研究所使用的数据库，共有一种健康工况与三种损伤工况进行识别。本研究的意义在于：将结构损伤粗略归一化到若干程度的损伤比例中，为结构的安全评估做参考。因此，对于损伤程度估计的回归问题其实可以简化为多分类问题就行研究。在进行多分类器研究中，能否将不同损伤对应的损伤指标之间的边界划分是该研究的重点。

(1) 数据集组成：基准工况 0+损伤工况 1~3。其中，将标签设计为多分类问题进行分类算法的训练，即：

$$y_i = \begin{cases} 0, & \text{if } i = \text{健康工况}0 \\ 1, & \text{if } i = \text{损伤工况}1 \\ 2, & \text{if } i = \text{损伤工况}2 \\ 3, & \text{if } i = \text{损伤工况}3 \end{cases} \quad (4-8)$$

(2) LAL-AdaBoost 训练过程：数据集的划分工作与 4.2.3.2 节类似，这里不做赘述。具体地 LAL 算法的参数如下表所示：

表 4-6 LAL 参数设置（多损伤程度研究）		
参数名称	参数意义	参数取值
<i>initial_rate</i>	初始迭代时 L_0 占训练集的比例	0.05
<i>test_ratio</i>	测试集占总样本的比例	0.3

（下页续表）

(接上页续表)

参数名称	参数意义	参数取值
<i>split_count</i>	重复划分数据集的次数	10
<i>stop_criterion</i>	停止迭代条件	‘number of queries’=150
<i>cls_est</i>	用于运算分类器状态的随机森林回归器的决策树的数量	Default = 50
<i>cls_depth</i>	决策树的最大深度	Default = 40
<i>batch_size</i>	每次质询挑选的数据点的个数	1
<i>mode</i>	选择 LAL 算法（见 2.1 节），可选 LAL-Independent 或 LAL-Interactive	Default = ‘LAL-Interactive’

(3) 分类结果讨论：分别使用 LAL-AdaBoost 算法与 US-AdaBoost 两种方法进行对比实验。在拟合速度和精度上，二者表现类似，但是 LAL-AdaBoost 要略优于 US-AdaBoost；相比较而言，LAL-AdaBoost 算法的分类曲线更加平稳，在质询次数到达 60 次左右时，精度就已经达到了 85%左右，相比较于 US-AdaBoost 的 77%要领先较多；最终达到质询上限时，LAL-AdaBoost 精度仍然领先了 3%左右，说明其效果较好，可以在复杂的多分类问题中有效地将结构损伤程度进行分类。

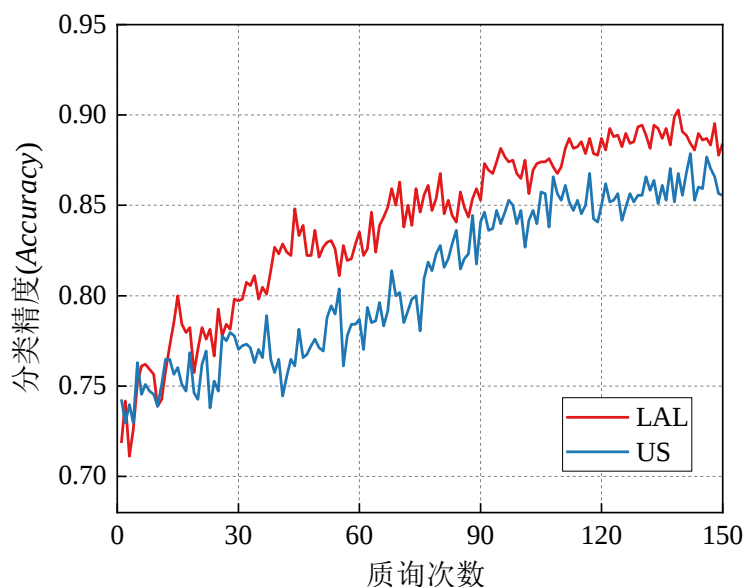


图 4-9 多种损伤程度分类算法分类曲线

从混淆矩阵来分析，总体上各类样本的召回率与精确度都较高，说明该算法的分类效果较好，在测试集中仅有少部分样本被错分。根据分类结果可以分别求

出 LAL-AdaBoost 算法与 US-AdaBoost 算法对于各标签样本的 F_1 值，从该指标上来看，前者的总体 F_1 值要大于后者，尤其是样本标签为 0（即健康工况）时，LAL 算法表现出了较为明显的优越性，领先了 8%。

真实值 \ 预测值	0	1	2	3	召回率
0	24 22.2%	0 4.7%	0 0	1 0.9%	96.0% 4.0%
1	0 0	25 23.1%	1 0.9%	0 0	96.2% 3.8%
2	0 0	2 1.9%	25 23.1%	2 1.9%	86.2% 13.8%
3	4 3.7%	0 0	0 0	24 22.2%	85.7% 14.3%
精确度	85.7% 14.3%	92.6% 7.4%	96.2% 3.8%	88.9% 11.1%	90.7% 9.3%

图 4-10 多种损伤程度分类算法混淆矩阵

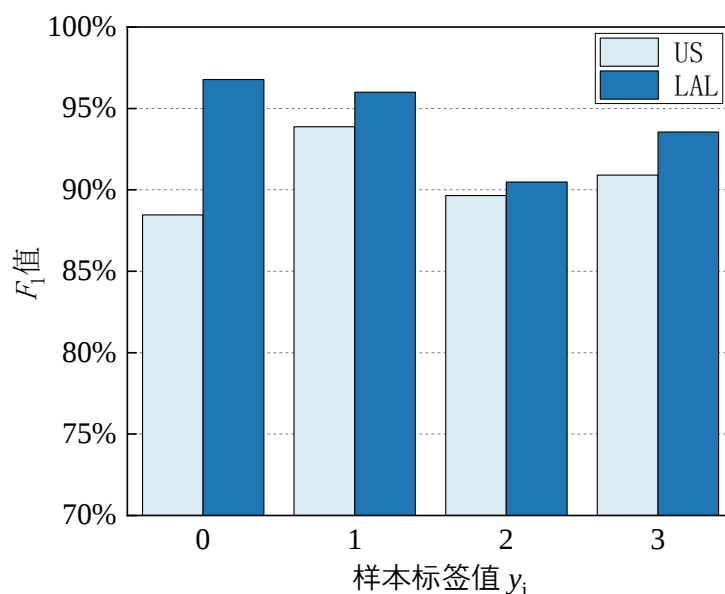


图 4-11 各标签样本的 F_1 值

4.3.3 损伤定位研究

4.3.3.1 不同位置处的损伤分类

在本小节中，作者主要探究关于损伤定位的模型训练及预测过程。

(1) 数据集组成：基准工况 0+损伤工况 3~6。其中，将标签设计为多分类问题进行分类算法的训练，即：

$$y_i = \begin{cases} 0, & \text{if } i = \text{健康工况} \\ 1, & \text{if } i = \text{5号单元损伤} \\ 2, & \text{if } i = \text{6号单元损伤} \\ 3, & \text{if } i = \text{8号单元损伤} \\ 4, & \text{if } i = \text{9号单元损伤} \end{cases} \quad (4-9)$$

如此设计后，标签 y 可以有效地反映出当前结构损伤出现的位置信息，保证机器学习算法输出的值即为定位损伤的单元号。

(2) LAL-AdaBoost 训练过程：数据集的初始划分保证随机性，重复十次分别进行实验。具体地 LAL 算法的参数如下表所示：

表 4-7 LAL 参数设置（损伤定位研究）

参数名称	参数意义	参数取值
<i>initial_rate</i>	初始迭代时 L_0 占训练集的比例	0.1
<i>test_ratio</i>	测试集占总样本的比例	0.3
<i>split_count</i>	重复划分数据集的次数	10
<i>stop_criterion</i>	停止迭代条件	‘number of queries’=150
<i>cls_est</i>	用于运算分类器状态的随机森林回归器的决策树的数量	Default = 50
<i>cls_depth</i>	决策树的最大深度	Default = 40
<i>batch_size</i>	每次质询挑选的数据点的个数	1
<i>mode</i>	选择 LAL 算法（见 2.1 节），可选 LAL-Independent 或 LAL-Interactive	Default = ‘LAL-Interactive’

(3) 分类结果：分类结果从混淆矩阵中可以看出，各标签样本的召回率(Recall)以及精确度(Precision)均较高，最终总体正确率达到了 91.9%，结果较为理想。对于预测值为 0 的样本，仅有少部分损伤位置的数据被误判，这说明该模型的鲁棒性较强，在多个分类位置的损伤同时出现时，AdaBoost 算法依然可以有效地提

取出分类边界。各个损伤位置样本在测试集上的召回率(Recall)均在 90%左右，说明模型整体性能较强，可以充分学习到损伤指标在不同位置的损伤发生时所产生的变化。

预测值 真实值	0	1	2	3	4	召回率
0	21 15.6%	3 2.2%	0 0	0 0	0 0	87.5% 12.5%
1	3 2.2%	27 20.0%	0 0	0 0	0 0	90.0% 10.0%
2	0 0	0 0	30 22.2%	0 0	0 0	100.0% 0
3	0 0	0 0	3 2.2%	23 17.0%	0 0	88.5% 11.5%
4	1 0.7%	1 0.7%	0 0	0 0	23 17.0%	92.0% 8.0%
精确度	84.0% 16.0%	87.1% 12.9%	90.9% 9.1%	100.0% 0	100.0% 0	91.9% 8.1%

图 4-12 损伤定位分类算法混淆矩阵

4.3.3.2 损伤定位抗噪性分析

为了进一步对该算法的鲁棒性进行验证，在本小节中作者对模拟的加速度时域数据增加一定程度的高斯白噪声来模拟真实情境下出现的结构响应。在本次实验中，作者设计了三组不同程度的噪声影响下 LAL-AdaBoost 算法对于损伤定位分类的效果。

为了定量控制加入到时域信号中的高斯白噪声，在这里引入信噪比（Signal Noise Ratio，用 b 表示）的概念，其表达式为：

$$b = \frac{P_{\text{signal}}}{P_{\text{noise}}} \quad (4-10)$$

其中 P_{signal} 与 P_{noise} 分别代表了信号与噪声的功率。所谓信号的功率即为信号的能量或强度，对于常见的连续（或离散）信号通过对序列的求平方后求得积分（或求和）可以得到该指标。由公式（4-10）看出，信噪比 b 值越大，其包含的噪声越小。当信噪比 b 的单位为分贝（dB）时，上式可以改写为以下形式：

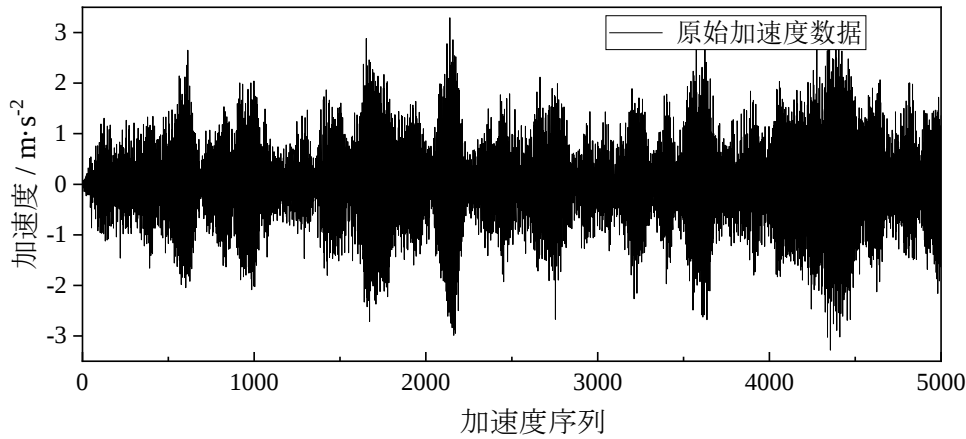
$$b(\text{dB}) = 10 \log_{10} \frac{P_{\text{signal}}}{P_{\text{noise}}} = 10 \log_{10} \frac{\sum_{n=0}^{N-1} x^2(n)}{\sum_{n=0}^{N-1} s^2(n) - \sum_{n=0}^{N-1} x^2(n)} \quad (4-11)$$

其中 s 为包含噪声的序列, x 为原始序列。由于高斯白噪声均值为 0 且方差为 1, 可以得到最终的加噪声信号表达式:

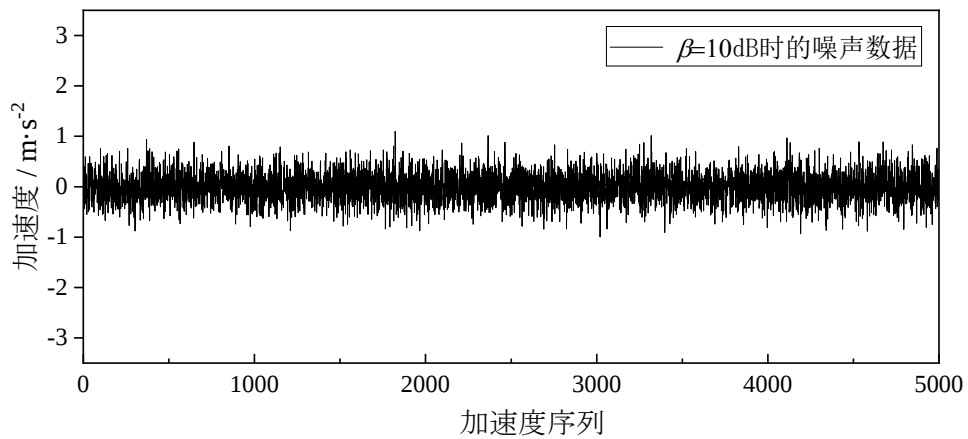
$$s = x + 10^{-\frac{b}{10}} \text{gN}(0,1) \text{rms}(x) \quad (4-12)$$

在上式中, $N(0,1)$ 为服从标准正态分布的随机数, rms 函数为任意离散序列的均方根计算过程。

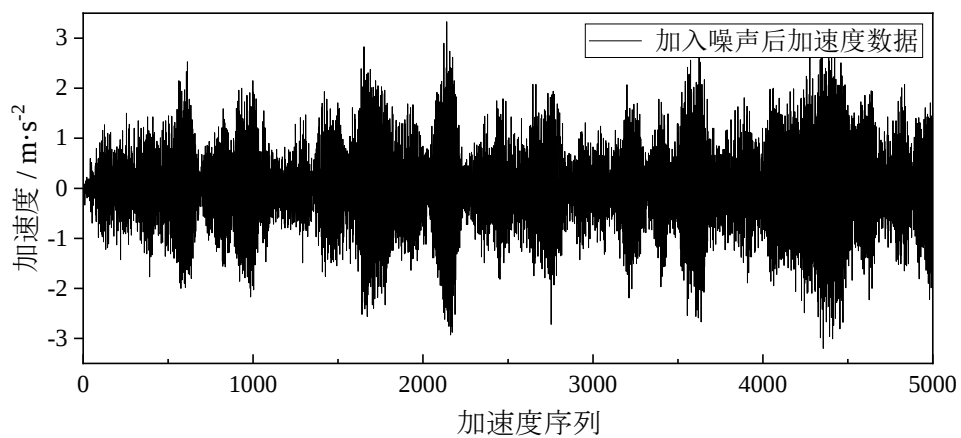
下图展示了在加入不同程度的噪声后加速度数据在时域和频域上的区别。可以看到噪声的加入对于加速度序列的时频域转换处理上会产生影响, 加入了高斯白噪声后可以有效地模拟实际工程中采集信号可能会发生的情形。随着噪声等级的增加, 加速度数据时域上波动变大, 在功率谱上表现出较大的差异。下图中展示了在加入控制信噪比为 10dB 的高斯白噪声前后加速度时域与频域的变化。可以看出在频域上, 加入噪声后的功率谱密度函数明显出现了较大的随机波动, 峰值出现了轻微偏移, 属于实测条件下常见的情形(灰色部分为两功率谱重叠部分)。



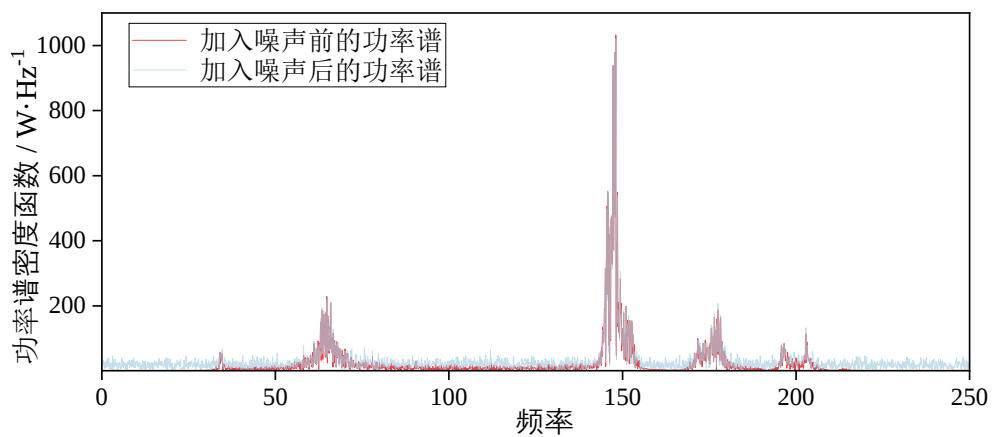
(a) 加速度原始数据时域图



(b) $b = 10\text{dB}$ 的高斯白噪声数据



(c) 加入 $b = 10\text{dB}$ 高斯白噪声后的加速度时域数据



(d) 加入噪声前后加速度数据频域对比图

图 4-13 $b = 10\text{dB}$ 时信号与原始信号在时频域对比图

为了充分验证本课题提出的分类算法在研究空间结构损伤定位问题中的抗噪声性能，在本小节中设计了三种噪声情况： $b = 40\text{dB}$ （轻微噪声）、 $b = 20\text{dB}$ （中等噪声）、 $b = 10\text{dB}$ （较强噪声）。分别针对以上几种情况进行训练与测试，得出

了分类曲线与 F_1 值，见图 4-13 与图 4-14。从分类曲线上来看，随着噪声增强，算法拟合的爬升速度减慢，最后的精度相差 5% 左右，各工况均在 85% 以上；在轻微噪声与中等噪声的影响下，分类精度提高的速率基本相同，最终完成迭代后的精度大致均落在了 90% 左右；而当信噪比为 10dB 时，分类精度明显下降，拟合速度减慢，最终经过 150 次质询后精度达到约 97%。从 F_1 值的计算结果上来看，算法对于各类样本的分类精度均维持在较高的水平；当信噪比为 10dB 时，LAL-AdaBoost 算法对于健康工况的样本分类效果略差，但对于损伤工况的位置识别精度普遍较高。

为了对比 LAL-AdaBoost 算法的抗噪性，将 US-AdaBoost 算法重复上述的实验过程，在不同程度的噪声作用下得到其分类精度的对比如下表所示。通过对比两种算法在噪声影响下的受干扰程度可以得出，本文提出的算法具有更高的抗噪声性能。

表 4-8 LAL-AdaBoost 与 US-AdaBoost 不同噪声下分类正确率对比

噪声等级	LAL-AdaBoost	US-AdaBoost
无噪声	91.9%	91.2%
40dB	89.9% (-2.0%)	90.2% (-1.0%)
20dB	88.5% (-3.4%)	87.2% (-4.0%)
10dB	86.8% (-5.1%)	85.4% (-5.8%)

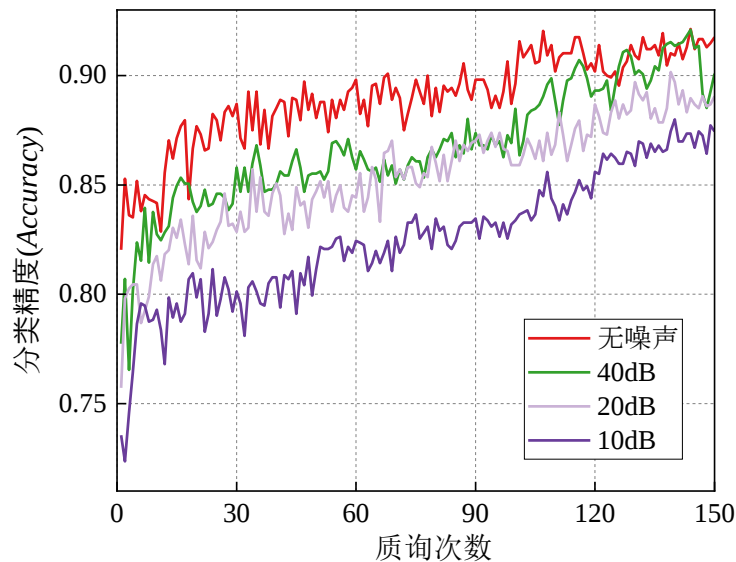


图 4-14 不同信噪比的噪声下 LAL-AdaBoost 算法分类曲线

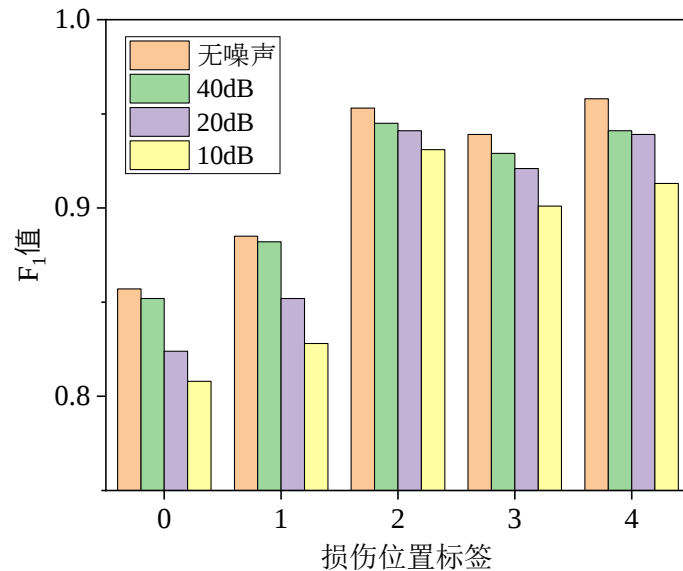


图 4-15 不同信噪比的噪声下 LAL-AdaBoost 算法 F_1 指标

4.4 泛化性研究：天津一中网架数值模型验证

本文提出的损伤指标与损伤识别方法在本章 4.3 节中介绍的空间网架结构上进行了有效性的充分验证。然而，对于实际的网架监测项目而言，往往结构的体型会更为复杂与庞大，而且传感器的布置通常不可能遍历到每个单元。为了进一步验证本文提出方法的有效性与泛化性，在本节中作者将引入天津一中网架结构健康监测项目的数值模型，通过设计四种不同位置的损伤工况分别进行损伤定位的实验研究。

4.4.1 工程概况



图 4-16 天津一中预应力网架

本节算例为天津一中预应力网架结构健康监测项目,该项目主要对于天津一中体育馆二层的羽毛球馆(图 4-16)进行长期动力监测,以确保其在运营期间内所产生的网架损伤及时被诊断与修复^[3]。该项目所研究网架结构由正方四角锥单元组成,上弦杆件以混凝土肋梁代替,下弦杆件纵向设置了预应力拉索;结构纵向总跨度为 18.6m,横向跨度为 13.8m,共包含 $14 \times 10 = 140$ 个单元;该网架结构四周与钢筋混凝土结构进行耦合,可以视为上弦节点与结构主体设置为铰接支座条件。

该健康监测项目动力监测部分主要设计了 9 个加速度传感器测点(图 4-17),在每个测点布置一个 z 向的加速度传感器进行长期采集,数据储存在采集仪中,并通过无线网络上传至云端进行储存。传感器通过特质的夹具固定在结构的下弦节点上(图 4-18),线缆延下弦杆件连接至二层地面的采集仪。

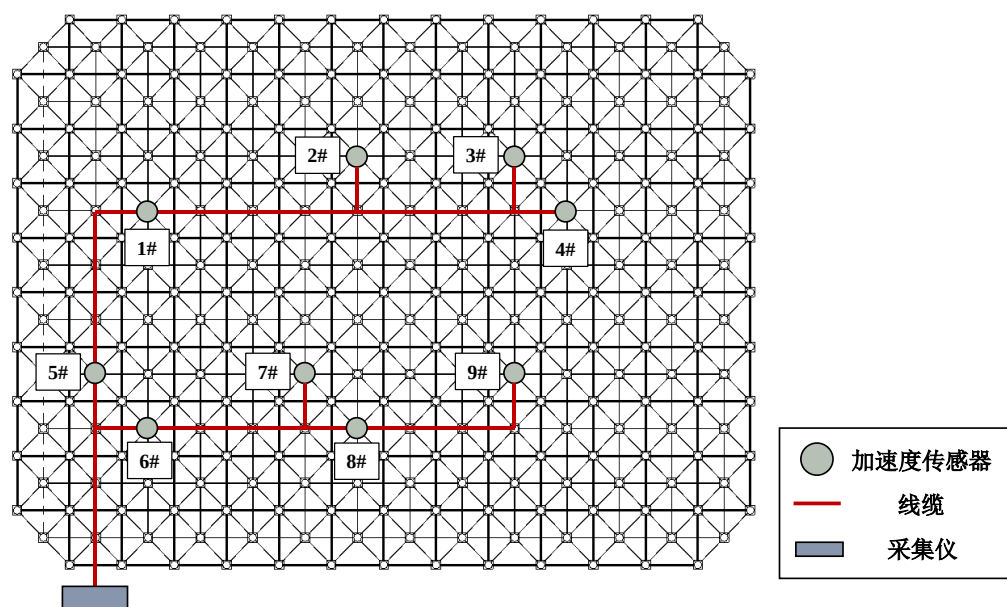


图 4-17 传感器布置图



图 4-18 传感器安装图

4.4.2 数值模型与工况模拟

由于该网架在进行监测过程中还未出现损伤的情况，因此在本节中可以合理假定目前监测条件作为基准工况使用 ANSYS 软件进行有限元建模。具体建模信息可以参考文献^[3]。

由于此网架结构的跨度较大，在考虑损伤定位研究时主要选择接近跨中位置的 4 个单元进行损伤的设置，如图 4-18 所示。在进行损伤的模拟时，通过与 4.3.1 节类似的方法，在某一特定单元内随机选取 2 根腹杆对其弹性模量缩减 30%来模拟单元内可能会出现的情况。具体的工况设计情况如表 4-9 所示，分别对于不同位置单元损伤进行模拟，以探究本文提出的 LAL-AdaBoost 算法对于损伤检测与损伤定位判别的精确性。模拟的采样频率为 50Hz，每组工况采集时间为 1000s，在进行特征提取之前将数据划分为以 10s 为一组的加速度数据。

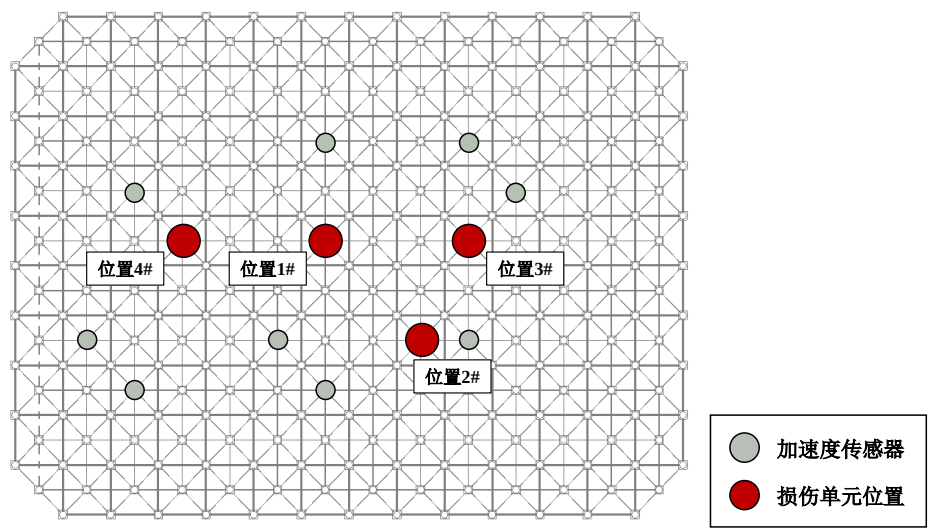


图 4-18 损伤杆件位置选取

表 4-9 数值模拟工况设计

工况号	损伤单元位置	研究内容
0	无损伤	基准工况设定
1	1#位置单元	不同损伤位置的判定
2	2#位置单元	
3	3#位置单元	
4	4#位置单元	

4.4.3 特征提取

根据本章 4.1 节所提出的损伤指标计算方法可以得出概算例所布设的 9 个加速度传感器之间的相关性指标，每个特征向量为 36 个浮点数组成的 list 结构。

对于本算例所提取的损伤指标而言，不同传感器之间具有较强的相互关系。为了减轻机器学习模型的训练负担并降低数据样本的冗杂度，这里采用了主成分分析（principal component analysis，简称 PCA）的方法进行数据降维的工作，该算法在现有的研究中被证明可以在结构动力分析的过程中有效降低数据冗杂度对于损伤识别的影响^[71]。下面简要介绍一下 PCA 的主要原理。

PCA 是一种多元统计分析方法，其基本思想是降维，即在保证原始数据信息损失较少的前提下，将高维相关变量转换为低维无关变量。新变量保留了原始变量的大部分信息，即原始变量的主成分^[72]。通过对高维空间中分布的样本点对低维空间计算映射关系，可以实现数据的降维过程。以一个二维特征 (x_1, x_2) 样本为例，样本点 A 、 B 、 C 在二维坐标系的分布如图 4-19 所示。这里引入一个新坐标轴 y_1 ，设置为新的变量，并求得三个样本点在该坐标轴上的投影 A' 、 B' 、 C' 。坐标值的平方和 $OA'^2 + OB'^2 + OC'^2$ 表示样本在该变量 y_1 上的方差和。主成分分析的目标即为求得在此坐标变换过程中得到的个样本方差和最大值，作为第一主成分。从图 4-19 中不难推导得，主成分分析等价于在旋转变量坐标时选取距离样本点距离平方和最小的轴作为第一主成分，而第二主成分则取与第一主成分正交的轴选取，以此类推。

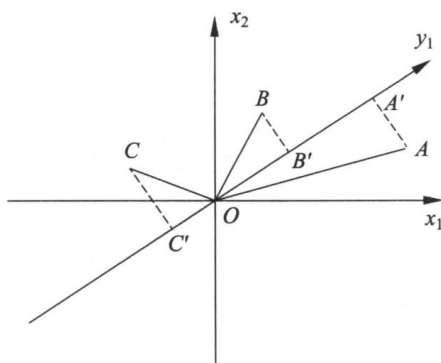


图 4-19 主成分分析几何解释^[71]

在进行主成分分析的过程中，应注意确定主成分的个数的选取，过多的主成分会导致数据降维效果差，而过少的主成分则会导致新特征的数据近似误差过大，无法有效保留原数据所包含的信息。因此，一般研究中会采用 PCA 过程前后数据的方差百分比 p 来衡量主成分个数 k 值的合理性，其表达式如下所示：

$$p_k = \frac{\sum_{j=1}^k \lambda_j}{\sum_{j=1}^n \lambda_j} \quad (4-13)$$

上式中, λ_j 为对应于 PCA 过程中第 j 个特征向量的特征值, n 为特征向量的总数。在本次算例中, 将 p_k 值定为 0.90 进行降维过程, 理论上最大程度上保留原始样本的方差值。

4.4.4 损伤识别研究

在本节的损伤识别研究中, 相关参数设定如下表所示。在此算例中样本数量较为庞大, 因此适当增加了回归器模型的复杂程度以高效拟合。

表 4-10 LAL 参数设置

参数名称	参数意义	参数取值
<i>initial_rate</i>	初始迭代时 L_0 占训练集的比例	0.1
<i>test_ratio</i>	测试集占总样本的比例	0.3
<i>split_count</i>	重复划分数据集的次数	10
<i>stop_criterion</i>	停止迭代条件	‘number of queries’=150
<i>cls_est</i>	用于运算分类器状态的随机森林回归器的决策树的数量	Default = 60
<i>cls_depth</i>	决策树的最大深度	Default = 50
<i>batch_size</i>	每次质询挑选的数据点的个数	1
<i>mode</i>	选择 LAL 算法 (见 2.1 节), 可选 LAL-Independent 或 LAL-Interactive	Default = ‘LAL-Interactive’

在本节中主要进行两部分的研究, 分别是损伤检测与损伤定位。前者属于较为简单的二分类问题, 后者属于多分类的判定问题。

4.4.4.1 损伤检测

在本小节中作者将探究本文提出的 LAL-AdaBoost 算法对于损伤检测的灵敏度与精确度。在此分类问题中, 将输出设置为 $y=0$ 或 1, 即对于结构是否出现了损伤进行判定工作, 使用 LAL-AdaBoost 算法与 US-AdaBoost 算法进行对比实验, 分别通过分类曲线与混淆矩阵的形式进行对比 (图 4-20、表 4-11)。可以得

出结论，两种算法的正确率比较接近，在进行损伤识别过程中均取得了较为理想的结果，LAL-AdaBoost 算法在 F_1 值的对比中有微弱的优势。

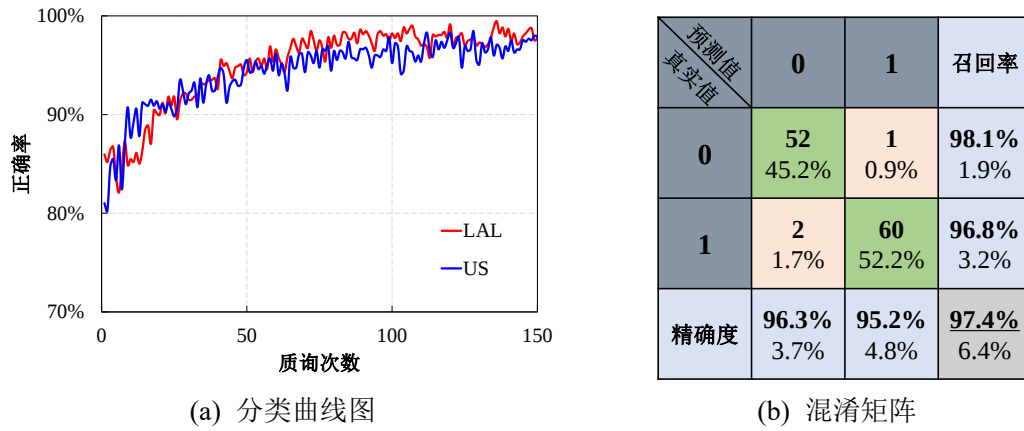


图 4-20 损伤检测分类结果

表 4-11 F_1 值对比

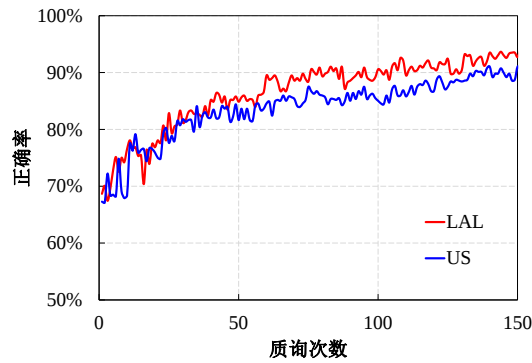
算法	健康标签	损伤标签
LAL-AdaBoost	0.967	0.987
US-AdaBoost	0.955	0.979

4.4.4.2 损伤定位

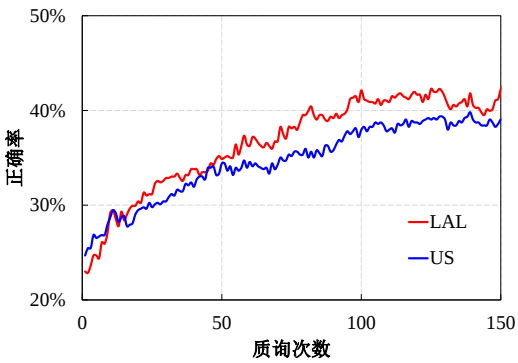
在本小节中作者将探究本文提出的 LAL-AdaBoost 算法对于损伤定位判定的正确率，属于多分类问题。在本算例中，为了验证 4.4.3 节中所提出的基于 PCA 技术的特征提取过程的有效性，将经过 PCA 处理后的数据集（记为 1#数据集）和计算出的原始损伤指标数据集（记为 2#数据集）进行对比，分别给出分类曲线和混淆矩阵，对比二者差异。

从分类曲线的结果中可以看出，LAL-AdaBoost 算法在 1#数据集上的表现要明显优于 2#数据集。经过 150 次的迭代过程，算法在 1#数据集上的精度达到了 93%左右，而算法在 2#数据集上的精度仅为 40%左右，说明未经过 PCA 过程的数据集所包含的结构响应信息被干扰。从混淆矩阵的结果可以得出，1#数据集的分类结果在各标签值的分类召回率与精确度上效果都要更好。同样对比 LAL-AdaBoost 与 US-AdaBoost 的分类曲线可得，前者的拟合速度与精度相较而言更为理想。从训练时间上来看，1#数据集的拟合速度明显更为迅速，效率更高。综合来看，在更复杂的结构上运用本文提出的损伤识别算法时通过合理运用 PCA 的数据处理方法可以有效地提高损伤识别的精确度与分类效率。

Fout! Gebruik het tabblad Start om 标题 1,章标题 toe te passen op de tekst die u hier wilt weergeven.



(a) 1#数据集分类曲线图



(b) 2#数据集分类曲线图

预测值 真实值	0	1	2	3	4	召回率
0	42 28.8%	1 0.7%	0 0	0 0	0 0	97.7% 2.3%
1	0 0	23 15.8%	0 0	0 0	1 0.7%	95.8% 4.2%
2	2 1.4%	0 0	22 15.1%	1 0.7%	1 0.7%	84.6% 15.4%
3	1 0.7%	0 0	2 1.4%	24 16.4%	0 0	88.9% 11.1%
4	0 0	0 0	1 0.7%	1 0.7%	24 16.4%	92.3% 7.7%
精确度	93.3% 6.7%	95.8% 4.2%	88.0% 12.0%	92.3% 7.7%	92.3% 7.7%	92.5% 7.5%

(c) 1#数据集混淆矩阵(LAL-AdaBoost)

预测值 真实值	0	1	2	3	4	召回率
0	12 8.2%	6 4.1%	3 2.1%	2 1.4%	3 2.1%	46.2% 53.8%
1	6 4.1%	12 8.2%	6 4.1%	5 3.4%	4 2.7%	36.4% 63.6%
2	7 4.8%	7 4.8%	13 8.9%	3 2.1%	2 1.4%	40.6% 59.4%
3	8 5.5%	8 5.5%	2 1.4%	16 11.0%	6 4.1%	40.0% 60.0%
4	4 2.7%	1 0.7%	1 0.7%	2 1.4%	7 4.8%	46.7% 53.3%
精确度	32.4% 77.6%	35.3% 74.7%	52.0% 48.0%	57.1% 42.9%	31.8% 68.2%	41.2% 58.8%

(d) 2#数据集混淆矩阵(LAL-AdaBoost)

图 4-21 损伤定位分类结果

表 4-12 1#数据集与 2#数据集测试集结果对比

数据集	正确率		训练时间	
	LAL-AdaBoost	US-AdaBoost	LAL-AdaBoost	US-AdaBoost
1#	92.5%	90.2%	1256 s	984 s
2#	41.2%	39.0%	3547 s	2897 s

4.5 本章小结

在本章中，作者主要探究了关于 LAL-AdaBoost 算法在进行空间结构损伤识别研究中的应用，通过四个分类场景分别进行结果的分析：单一程度损伤检测、微小程度的损伤检测、多种程度的损伤检测以及分类、损伤定位分类。主要分为以下几部分工作内容：

1. 结合传感器分布情况以及加速度信号频域分析的方法，提出了一个全新的损伤指标，该指标利用相关性函数以及 FFT 频谱变换的技术描述了空间结构中不同传感器之间的相互关系，在损伤发生时会产生相应的变换，从而捕捉到相应的损伤敏感信息。

2. 提出了一个基于 LAL-AdaBoost 算法的损伤识别框架，该框架通过结合对实际结构的分析以及有限元模拟的数据进行损伤识别分析，从而对实际结构的动力检测进行指导。

3. 通过对一个空间网架结构进行有限元模拟建立了结构健康以及损伤工况的数据库，并验证了所提出的损伤识别框架；主要进行了四组不同情景的实验：单一程度的损伤检测、微小程度的损伤检测、多种损伤程度的损伤检测、损伤定位检测等；结果表明 LAL-AdaBoost 算法的精度较高，而且具有较强的泛化能力，可以完成较为复杂的分类任务。

4. 为了进一步验证所提出的损伤识别框架对于噪声干扰的鲁棒性，设计了一组不同信噪比的加速度数据用于对比试验，在信噪比为 40dB、20dB、10dB 的高斯白噪声干扰下分析其分类曲线与 F_1 值的大小，证明了所提出的方法在较强噪声的干扰下依然可以较为准确地定位损伤。

5. 通过天津一中预应力网架结构健康监测项目验证了本文提出方法的有效性，首先将特征提取后的样本进行主成分分析，并分别进行损伤检测与损伤定位的验证。结果显示，经过 PCA 处理过的数据样本具有更好的分类效果。

第 5 章 总结与展望

5.1 全文工作总结

本文提出了一种基于 LAL-AdaBoost 智能算法的空间网架损伤识别框架，包含数据预处理与异常检测、损伤指标提取与损伤识别三部分主要内容，并通过一个空间网架实验结构来验证了本文提出的方法的有效性。主要完成的工作内容总结如下：

1. 在数据预处理方法研究中，提出了一种基于 LAL-AdaBoost 算法框架的数据异常识别方法，针对结构健康监测中常见的五种数据异常类型（离群值、基线漂移、数据缺失、数据幅值过小、矩形波），采用了一种基于主动学习理论的分类器模型用于在训练集中筛选出具有代表性的数据异常类型进行质询标签，并用筛选出的数据训练 AdaBoost 分类器用于数据异常的分类任务。

2. 为了验证提出的数据异常识别框架，设计了一个空间网架实验终于采集加速度响应数据，并构造了包含五种异常类型的数据集，通过 LAL-AdaBoost 与 US-AdaBoost 进行了对比实验；结果显示，对于均衡数据集二者拟合速度与分类精度类似；但当数据集不均衡时，LAL-AdaBoost 算法在拟合速度上明显优于 US-AdaBoost 算法，而且最终得到了 93% 的总体精度，领先后者 4% 左右的正确率；在 F_1 值的对比中，LAL-AdaBoost 算法对于少数样本的分类结果更理想，达到了 0.9 左右，说明该算法避免了有偏向性的分类结果，达到较高精度的同时避免了欠拟合与过拟合问题的发生。

3. 提出了一个基于加速度信号频域分析并结合传感器布置的损伤指标，从而适用于空间结构的 SHM。该指标充分考虑了空间结构的自身特性与加速度信号特征提取技术，可作为机器学习算法的输入使用。

4. 对空间网架实验结构使用 ANSYS 进行有限元建模，并模拟出了六种不同位置、不同程度的损伤，构造了健康与损伤数据库。使用上述的损伤指标作为 LAL-AdaBoost 算法的输入，分别进行了四种监测场景的验证：单一损伤程度检测、微小损伤程度检测、多损伤程度分类、多损伤位置分类。通过与 US 主动学习算法的对比试验的结果表明，从分类曲线、混淆矩阵、 F_1 值等分析中可以得出，LAL 算法精度高且拟合速度较快，具有更优秀的分类效果，而且该算法对于数据量的要求较低，因此在实际工程中更具有使用价值。通过抗噪性实验得出该 LAL-AdaBoost 损伤识别框架具有较高的抗噪性与鲁棒性。最后，通过天津一中预应力网架监测项目进行了模拟验证本文提出的损伤识别算法的有效性。

5.2 后续工作展望

本文主要研究了于 LAL-AdaBoost 智能算法在数据驱动的网架损伤识别监测中的应用。在未来的工作中，可以在以下几个方面展开更深入的研究工作：

1. 本文研究的数据异常识别方法本身具有较强的泛化性，在未来的研究中可以针对其他结构上采集的实测数据进行相应的验证工作，而不止局限于本文使用的空间网架结构。同时出现多个一场模式的多标签分类问题。

2. 本文提出的损伤指标在一个 9 单元的正方四角锥网架中效果较好，但对于实际的大跨度空间结构而言，则需要充分考虑传感器的优化布置问题，确保可以通过传感器之间的相关性找到对应损伤位置。

3. 本文使用的有限元模型由于与实际结构不可避免地存在一定偏差，因此动力响应可能会与实际结构存在部分不一致。在未来的研究中，可以讲有限元模型更新的方法引入到损伤识别框架中，进行更加精确的研究。

参考文献

- [1] 董石麟, 邢栋, 赵阳. 现代大跨空间结构在中国的应用与发展[J]. 空间结构, 2012, 18(1): 3-16.
- [2] Zhai C, Liu W, Xie L. Progress of Research on City Seismic Resilience Evaluation [J]. Journal of Building Structures, 2018, 39(9): 1-9.
- [3] 韩庆华, 马乾, 刘名, 徐杰. 温度变化下基于固有频率聚类分析的空间网格结构损伤诊断[J]. 华东交通大学学报, 2021, 38(4): 8-17.
- [4] Brownjohn J M W, De Stefano A, Xu Y L, et al. Vibration-Based Monitoring of Civil Infrastructure: Challenges and Successes [J]. Journal of Civil Structural Health Monitoring, 2011, 1(3-4): 79-95.
- [5] Ou J, Li H. Structural Health Monitoring in Mainland China: Review and Future Trends [J]. Structural Health Monitoring, 2010, 9(3): 219-231.
- [6] 孙利民, 尚志华, 夏烨. 大数据背景下的桥梁结构健康监测研究现状与展望[J]. 中国公路学报, 2019, 32(11): 1-20.
- [7] Bao Y, Chen Z, Wei S, et al. The State of the Art of Data Science and Engineering in Structural Health Monitoring [J/OL]. Engineering, 2019, 5(2): 234-242.
- [8] Bakhary N, Hao H, Deeks A J. Structure Damage Detection Using Neural Network With Multi-Stage Substructuring [J]. Advances in Structural Engineering, 2010, 13(1): 95-110.
- [9] An Y, Li B, Ou J. An Algorithm for Damage Localization in Steel Truss Structures: Numerical Simulation and Experimental Validation [J]. Journal of Intelligent Material Systems and Structures, 2013, 24(14): 1683-1698.
- [10] Talebinejad I, Fischer C, Ansari F. Numerical Evaluation of Vibration-Based Methods for Damage Assessment of Cable-Stayed Bridges [J]. Computer-Aided Civil and Infrastructure Engineering, 2011, 26(3): 239-251.
- [11] Jaishi B, Ren W-X. Structural Finite Element Model Updating Using Ambient Vibration Test Results [J]. Journal of Structural Engineering, 2005, 131(4): 617-628.
- [12] Morassi A, Tonon S. Dynamic Testing for Structural Identification of a Bridge [J]. Journal of Bridge Engineering, 2008, 13(6): 573-585.
- [13] 邱飞力, 张立民, 张卫华. 基于模态柔度矩阵的结构损伤识别[J]. 噪声与振动控制, 2015, 35(4): 101-106.
- [14] Cheung S H, Beck J L. Bayesian Model Updating Using Hybrid Monte Carlo Simulation with Application to Structural Dynamic Models with Many

- Uncertain Parameters [J]. *Journal of Engineering Mechanics*, 2009, 135(4): 243-255.
- [15] 方圣恩. 基于有限元模型修正的结构损伤识别方法研究[D]. 中南大学, 2010.
- [16] 张欣, 刘洋, 高丹盈. 基于第2代小波的有限元模型更新方法[J]. *振动、测试与诊断*, 2015, 35(4): 660-665.
- [17] Pandey A, Biswas M. Damage Detection in Structures Using Changes in Flexibility [J]. *Journal of Sound and Vibration*, 1994, 169(1): 3-17.
- [18] Ko J M, Sun Z G, Ni Y Q. Multi-Stage Identification Scheme for Detecting Damage in Cable-Stayed Kap Shui Mun Bridge [J]. *Engineering Structures*, 2002, 24(7): 857-868.
- [19] Cao H, Friswell M I. Nondestructive Damage Evaluation Indicator Based on Modal Flexibility Curvature [J]. *Engineering Mechanics*, 2006, 23(4): 33-38.
- [20] 曹晖, 张新亮, 李英民. 利用模态柔度曲率差识别框架的损伤[J]. *振动与冲击*, 2007, 07(6): 116-124.
- [21] 张军, 刘建勋, 徐进, 等. 结构损伤诊断的模态柔度差曲率法[J]. *工程力学*, 2011, 28(12): 112-118.
- [22] 黄江, 梁亚斌, 冯谦. 环境温度影响下基于振动模态柔度曲率的结构损伤监测方法[J]. *大地测量与地球动力学*, 2016, 36(12): 1121-1125.
- [23] Ding K, Chen T. Study on Damage Detection of Bridge Based on Wavelet Multi-Scale Analysis [J]. *Advanced Materials Research*, 2013, 639-640(1): 1010-1014.
- [24] Hester D, González A. A Wavelet-Based Damage Detection Algorithm Based on Bridge Acceleration Response to a Vehicle [J]. *Mechanical Systems and Signal Processing*, 2012, 28: 145-166.
- [25] Dervilis N, Worden K, Cross E J. On Robust Regression Analysis as a Means of Exploring Environmental and Operational Conditions for SHM Data [J/OL]. *Journal of Sound and Vibration*, 2015, 347: 279-296.
- [26] Reynders E, Roeck G De. Vibration-Based Damage Identification: The Z24 Bridge Benchmark [J]. *Encyclopedia of Earthquake Engineering*, 2014, 72(1): 1-8.
- [27] Spiridonakos M D, Chatzi E N, Sudret B. Polynomial Chaos Expansion Models for the Monitoring of Structures under Operational Variability [J]. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, 2016, 2(3): 1-13.
- [28] Cury A, Cremona C, Dumoulin J. Long-Term Monitoring of a PSC Box Girder Bridge: Operational Modal Analysis, Data Normalization and Structural Modification Assessment [J/OL]. *Mechanical Systems and Signal Processing*, 2012, 33: 13-37.

- [29] Shokrani Y, Dertimanis VK, Chatzi EN, et al. Structural Damage Localization under Varying Environmental Conditions [C]. Proceedings of 11th HSTAM International Congress on Mechanics, 2016, Athens, Greece.
- [30] Joan R, Moughty JJ. Bridge Damage Detection Based on Vibration Data: Past and New Developments [J]. Frontiers in Built Enviroment, 2017, 3, 4.
- [31] Duan Z, Yan G, Ou J, et al. Damage Detection in Ambient Vibration Using Proportional Flexibility Matrix with Incomplete Measured DOFs [J]. Structural Control and Health Monitoring, 2007, 14(2): 186-196.
- [32] K N. Damage Detection of a Steel Truss Bridge Using Frequency Response Function Curvature Method [D]. KTH Royal Institute of Technology, 2013.
- [33] Yeum C M, Dyke S J. Vision-Based Automated Crack Detection for Bridge Inspection [J]. Computer-Aided Civil and Infrastructure Engineering, 2015, 30(10): 759-770.
- [34] 刘涛, 李爱群, 缪长青, 等. 基于数据融合的结构损伤识别方法研究[J]. 工程力学, 2008, 25(1): 24-29.
- [35] Guo H Y. Structural Damage Detection Using Information Fusion Technique [J]. Mechanical Systems and Signal Processing, 2006, 20(5): 1173-1188.
- [36] Li H, Bao Y, Ou J. Structural Damage Identification Based on Integration of Information Fusion and Shannon Entropy [J]. Mechanical Systems and Signal Processing, 2008, 22(6): 1427-1440.
- [37] Bao Y, Li H, An Y, et al. Dempster-Shafer Evidence Theory Approach to Structural Damage Detection [J]. Structural Health Monitoring, 2012, 11(1): 13-26.
- [38] Bao Y, Li H, Ou J. Applications of Compressive Sensing Technique in Structural Health Monitoring [J]. Key Engineering Materials, 2013, 558: 561-566.
- [39] Bao Y, Tang Z, Li H. Compressive-Sensing Data Reconstruction for Structural Health Monitoring: A Machine-Learning Approach [J]. Structural Health Monitoring, 2020, 19(1): 293-304.
- [40] Loh C H, Hung T Y, Chen S F, et al. Damage Detection in Bridge Structure Using Vibration Data under Random Travelling Vehicle Loads [J]. Journal of Physics: Conference Series, 2015, 628, 012044.
- [41] Bao Y, Tang Z, Li H, et al. Computer Vision and Deep Learning-Based Data Anomaly Detection Method for Structural Health Monitoring [J]. Structural Health Monitoring, 2019, 18(2): 401-421.
- [42] Tang Z, Chen Z, Bao Y, et al. Convolutional Neural Network-Based Data Anomaly Detection Method Using Multiple Information for Structural Health Monitoring [J]. Structural Control and Health Monitoring, 2019, 26(1): e2296.
- [43] Ni F T, Zhang J, Noori M N. Deep Learning for Data Anomaly Detection and

- Data Compression of a Long-Span Suspension Bridge [J]. *Computer-Aided Civil and Infrastructure Engineering*, 2020, 35(7): 685-700.
- [44] Kim N, Hong M. Measurement of Axial Stress Using Mode-Converted Ultrasound [J]. *NDT and E International*, 2009, 42(3): 164-169.
- [45] Raghavan A, Cesnik C E S. Review of Guided-Wave Structural Health Monitoring [J]. *Shock and Vibration Digest*, 2007, 39(2): 91-114.
- [46] Zhou L, Chen S X, Ni Y Q, et al. A Hybrid Model for Real-Time Monitoring of Multiple Bolt Looseness Using Electromechanical Impedance and Graph Convolutional Networks [J]. *Smart Materials and Structures*, 2021, 30(3): 1-20.
- [47] Chen S X, Zhou L, Ni Y Q, et al. An Acoustic-Homologous Transfer Learning Approach for Acoustic Emission-Based Rail Condition Evaluation [J]. *Structural Health Monitoring*, 2021, 20(4): 2161-2181.
- [48] Huo L, Chen D, Kong Q, et al. Smart Washer - A Piezoceramic-Based Transducer to Monitor Looseness of Bolted Connection [J]. *Smart Materials and Structures*, 2017, 26(2): 1-18.
- [49] Cole P T, Watson J R. Acoustic Emission for Corrosion Detection [J]. *Advanced Materials Research*, 2006, 13-14: 231-236.
- [50] Li S, Wei S, Bao Y, et al. Condition Assessment of Cables by Pattern Recognition of Vehicle-Induced Cable Tension Ratio [J]. *Engineering Structures*, 2018, 155: 1-15.
- [51] Zaurin R, Khuc T, Catbas FN. Hybrid Sensor-Camera Monitoring for Damage Detection: Case Study of a Real Bridge [J]. *Journal of Bridge Engineering*, 2016, 21(6): 05016002.
- [52] Valença J, Puente I, Júlio E, et al. Assessment of Cracks on Concrete Bridges Using Image Processing Supported by Laser Scanning Survey [J]. *Construction and Building Materials*, 2017, 146: 668-678.
- [53] Lin Y Z, Nie Z H, Ma H W. Structural Damage Detection with Automatic Feature-Extraction through Deep Learning [J]. *Computer-Aided Civil and Infrastructure Engineering*, 2017, 32(12): 1025-1046.
- [54] Han Q, Zhao N, Xu J. Recognition and Location of Steel Structure Surface Corrosion Based on Unmanned Aerial Vehicle Images [J]. *Journal of Civil Structural Health Monitoring*, 2021, 11: 1375-1392.
- [55] Seung H S, Oppor M, Sompolinsky H. Query by Committee [C]. *Proceedings of The ACM Workshop on Computational Learning Theory*, 1992: 287-294.
- [56] S. E, M. F, B S. RALF: A Reinforced Active Learning Formulation for Object Class Recognition [C]. *Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012: 3626-3633.
- [57] Zhao Y, Shi Z, Zhang J, et al. A Novel Active Learning Framework for

- Classification: Using Weighted Rank Aggregation to Achieve Multiple Query Criteria [J/OL]. Pattern Recognition, 2019, 93: 581-602.
- [58] Bull L, Worden K, Manson G, et al. Active Learning for Semi-Supervised Structural Health Monitoring [J/OL]. Journal of Sound and Vibration, 2018, 437: 373-388.
- [59] Bull L A, Rogers T J, Wickramarachchi C, et al. Probabilistic Active Learning: An Online Framework for Structural Health Monitoring [J/OL]. Mechanical Systems and Signal Processing, 2019, 134: 106294.
- [60] Hughes A J, Barthorpe R J, Dervilis N, et al. A Probabilistic Risk-Based Decision Framework for Structural Health Monitoring [J/OL]. Mechanical Systems and Signal Processing, 2021, 150: 107339.
- [61] Hughes A J, Bull L A, Gardner P, et al. On Risk-Based Active Learning for Structural Health Monitoring [J/OL]. 2021: 1-28.
- [62] Zhao J, Sun S, Wang H, et al. Promoting Active Learning with Mixtures of Gaussian Processes [J/OL]. Knowledge-Based Systems, 2020, 188: 105044.
- [63] Hastie T, Rosset S, Zhu J, et al. Multi-Class AdaBoost [J]. Statistics and Its Interface, 2009, 2(3): 349-360.
- [64] Wen X, Shao L, Xue Y, et al. A Rapid Learning Algorithm for Vehicle Classification [J]. Information Sciences, 2015, 295: 395-406.
- [65] Wu Y, Ke Y, Chen Z, et al. Application of Alternating Decision Tree with AdaBoost and Bagging Ensembles for Landslide Susceptibility Mapping [J]. Catena, 2020, 187: 104396.
- [66] Furuta H, Hattori H. Health Monitoring System Using AdaBoost Technique [C]. Proceedings of Structural Health Monitoring And Intelligent Infrastructure, 2006: 1255-1261.
- [67] Hattori H, Gul M, Catbas F, et al. Structural Health Monitoring and Damage Detection Using AdaBoost Technique[C]. Proceedings of the 6th International Conference on Bridge Maintenance, Safety and Management (IABMAS), 2012: 384-391.
- [68] Yang Y, Ma Z, Nie F, et al. Multi-Class Active Learning by Uncertainty Sampling with Diversity Maximization [J]. International Journal of Computer Vision, 2015, 113(2): 113-127.
- [69] Konyushkova K, Sznitman R, Fua P. Learning Active Learning from Data [C]. Proceedings of Conference and Workshop on Neural Information Processing Systems, 2017: 1-10.
- [70] Freund Y, Schapire R E. A Decision Theoretic Generalization of On-Line Learning and an Application to Boosting [J]. Journal of Computer and System Sciences, 1997, 55: 119-139.

- [71] Yan AM, Kerschen G, De Boe P et al. Structural damage diagnosis under changing environmental conditions—part 2: local PCA for nonlinear cases [J]. *Mechanical Systems and Signal Processing*, 2005, 19:865–880.
- [72] Han QH, Ma Q, Xu J, et al. Structural health monitoring research under varying temperature condition: a review [J]. *Journal of Civil Structural Health Monitoring*, 2021, 11:149–173.

发表论文和参加科研情况说明

1. Dang DZ, Xu J. An ensemble learning-based structural damage identification method using acceleration data [C]. In proceedings of the 8th International Conference of Vibration Engineering, Program of Structural Dynamics and Control, Shanghai, China. 2021.7. DOI: 10.26914/c.cnkihy.2021.016464 （国际会议论文，第一作者，已发表）

2. Xu J, Dang DZ, Ma Q, Liu X, Han QH. A Novel and Robust Data Anomaly Detection Framework using LAL-AdaBoost for Structural Health Monitoring [J]. Journal of Civil Structural Health Monitoring, 2021. （SCI 三区，第二作者，已接收）

致 谢

在天津大学建筑工程学院历时近两年半的硕士学习已经接近尾声，回想这段宝贵的学习经历，太多的感慨难以言表。经过这段时间的学习，我能够顺利完成学业，离不开课题组教师和同学们以及家人们对我的悉心指点与帮助。

首先要衷心感激导师徐杰副教授、课题组带头人韩庆华教授，两位教师为我创造了良好的学习环境和浓厚的学习氛围。徐老师的学识渊博、平易近人，韩老师的严谨的治学态度和踏实的工作作风，都给我留下了深刻的印象，他们是我今后在事业和生活上的楷模。两位导师的教诲将不断激励我在以后的工作学习中不断努力、顽强拼搏，做一个对社会和国家有用的人。同时也要衷心感激课题组所有的老师们对我的教导、关心和帮忙。

其次要感谢同门的马乾师兄、刘璇师兄、孙健婕师姐以及所有同门师兄师姐在我的科研与生活上的帮助，他们在百忙之中对我的指导令我受益匪浅；同时要感谢跟我同级的孙彤、牛先俐、刘名、张翼鹏、鹿子鸣以及师弟师妹们的帮助与陪伴。

同时我要感激我的父母，他们默默地在生活、学习上为我提供了一切可能的帮忙，使我在繁忙的科研之余能享受家庭的温馨。

最后我想再次感谢帮助和关心过我的所有人，祝好！

