

基于 LightGBM-SVM 堆叠算法的强震动记录尖刺波形识别

张越^{1,2}, 周宝峰^{1,2*}, 郭文轩^{1,2}, 温瑞智^{1,2}

(1. 中国地震局工程力学研究所 地震工程与工程振动重点实验室, 黑龙江哈尔滨 150080

2. 地震灾害防治应急管理部重点实验室, 黑龙江哈尔滨 150080)

摘要: 强震动记录中的尖刺是一种常见异常波形, 其产生机理尚不清晰, 需积累大量数据深入研究, 因此尖刺识别具有重要意义。本文提出一种基于波形比例尺自适应预处理方法, 用于提取并强化幅值变化特征, 结合时间尺度判别标准, 降低幅值差异对人工标注的影响。同时提出一种特征表征方法, 将一维数据按采样点幅值的累积分布归一化为特征向量, 以表征强震动记录的空间分布特征。对类别极不平衡数据集, 训练多种机器学习模型, 并对误识别情况进行分析。进一步采用贝叶斯优化的 LightGBM-SVM 堆叠算法实现尖刺波形识别, 测试集马修斯相关系数 (MCC) 超过 86%。结果表明, 所提尖刺判别标准具有稳定性与普适性, 可作为数据质量评估辅助工具, 并为尖刺波形机理研究提供技术支撑。

关键词: 尖刺波形; 强震动记录; 机器学习; LightGBM; SVM; 堆叠算法

中图分类号: P315

收稿日期: 2025-09-17

Spike waveform recognition for strong-motion records based on LightGBM-SVM stacking algorithm

ZHANG Yue^{1,2}, ZHOU Baofeng^{1,2*}, GUO Wenxuan^{1,2}, WEN Ruizhi^{1,2}

(1. Key Laboratory of Earthquake Engineering and Engineering Vibration, Institute of Engineering Mechanics, China Earthquake Administration, Harbin 150080, China

2. Key Laboratory of Earthquake Disaster Mitigation, Ministry of Emergency Mitigation, Harbin 150080, China)

Abstract: Spike in strong-motion record is a common type of abnormal waveform. However, their generation mechanism remains unclear and requires the accumulation of large datasets for further study, making spike identification highly significant. This study proposes a preprocessing method based on adaptive waveform scaling to extract and enhance amplitude variation features, combined with time-scale discrimination criteria, thereby reducing the impact of amplitude differences on manual annotation accuracy. In addition, a novel feature representation approach is introduced, in which one-dimensional data are transformed into feature vectors by normalizing the cumulative distribution of sampling amplitudes, enabling the spatial distribution characteristics of strong-motion records to be represented. Using a highly imbalanced dataset, multiple machine learning models were trained, and cases of misclassification were analyzed. Furthermore, LightGBM-SVM stacking algorithm optimized with Bayesian optimization is adopted to achieve the recognition of spike waveforms, achieving a Matthews correlation coefficient (MCC) exceeding 86% on the test set. The results show that the proposed spike discrimination criterion achieved satisfactory performance, confirming its stability and generalizability. The method can serve as an auxiliary tool for spike waveform screening in data quality assessment and provide technical support for further investigations into the generation mechanism of spike waveforms.

Keywords: Spike waveform; strong-motion record; machine learning; LightGBM; SVM; stacking algorithm

基金项目: 中国地震局工程力学研究所基本科研业务费专项 (2025B02)

作者简介: 张越(1999—), 男, 博士研究生, 主要从事强震观测和地震动特性研究。E-mail: ztfzy741214@163.com, ORCID: <https://orcid.org/0009-0009-6574-9681>

通讯作者: 周宝峰(1978—), 男, 研究员, 博士, 主要从事强震观测和地震动特性研究。E-mail: zbf166@126.com, ORCID: <https://orcid.org/0009-0009-5541-0000>

引言

中国是世界上多地震的国家之一,同时是遭受地震灾害最严重的国家,自2008年中国国家强震动观测网络系统(NSMONS)运行以来,随着我国强震动记录数量与日俱增,具有异常波形的低质量数据的数量也大大增加(Li et al., 2008; Ji et al., 2017),作为地震工程领域的重要研究资料,高质量的强震动记录对震源机制分析、地震动衰减关系建立、烈度速报、地震预警、地震动输入选择等研究意义重大。因此,如何提高强震动数据的质量、提升其利用效率,已成为研究者关注的核心问题(周宝峰等,2017)。其中尖刺作为一种常见的异常波形,其周期短、能量高的特点使其对强震动记录的加速度反应谱有着较大的影响,但受地震差异性的影响,尖刺的识别存在一定的困难,而这类强震动记录存在应用风险或利用率不高的问题,对于应用在像核电站、大坝、海洋平台等地震安全性要求较高的关键工程设施上将会带来一定的安全隐患。针对这类强震动记录研究并给出高效的识别方法,以此给出相应的校正方,提高强震动记录的利用率,提升其科学应用价值,对于我国基于性态地震工程发展以及韧性城乡建设均具有重大意义。

尖刺的形成因素多样,例如环境噪声、仪器故障、电信号传输过程中的干扰等。Douglas (2003)提出尖刺波形可能由模拟强震仪记录数模转化过程中小数点错位引起,在数字强震动仪器广泛应用于强震观测工作的今天,此类问题已基本消除,再难从PGA较大的加速度波形中观察到这类现象。姚鑫鑫(2023)对大量的强震动记录进行分析,在个别强震动记录的噪声段和远场记录中观察到了类似的尖刺波形,并推测这类异常数据可能由故障仪器记录所得,图1所示为一段含有尖刺的仪器响应,其加速度波形中存在多个异常尖峰,放大后如图1(b)所示,在两个采样周期(红色直线之间)的波形中,幅值出现剧烈变化,表现出典型的尖刺特征。周宝峰等(2023)基于振动台试验研究了基墩尺寸对强震动记录的影响,结果表明2m高基墩则可能使PGA放大超过2倍,且傅里叶谱的高频成分显著增加。Arthur (2009)对地震信号复地震道中的瞬时频率尖刺产生的原因及意义进行了研

究,提出尖刺是由破坏性反射干涉、不连续性和噪声引起的,在定量分析中通常会将其作为噪声尖刺进行校正,而在定性分析中却可以作为揭示地层结构信息的关键信号。随着强震观测技术的发展,数模转换导致的尖刺波形已基本消除,但实际强震动记录中仍存在尖刺现象,其成因尚不明确,仍需大量数据深入分析加以研究。

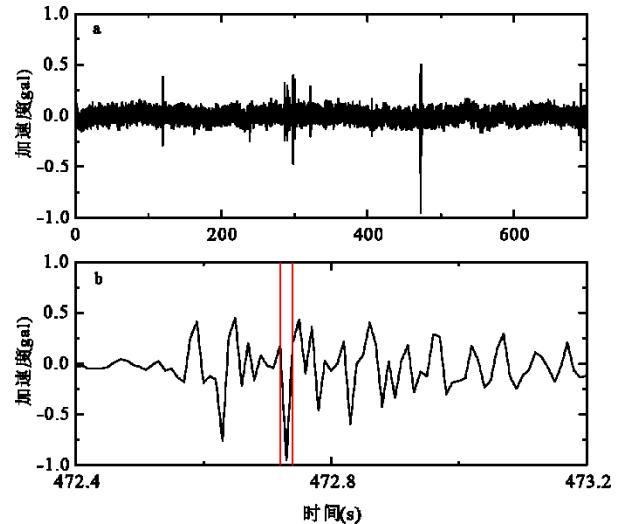


图1含有尖刺波形的仪器响应 (a) 具有尖刺特征的强震动记录加速度时程; (b) 含有尖刺波形的加速度记录时段。

Fig.1 The response of instruments with spike waveforms.

(a)Acceleration time history of strong-motion records with spikes; (b) acceleration time segment containing spike waveform.

幅值比较法是一种定量的尖刺识别方法,该方法将峰值点与周围其他采样点的幅值进行比较,通过比值确定该处是否为尖刺波形。Giuseppe and Claudia (2002)提出一种SPZ算法对每个采样点与其前后区域上计算的平均绝对值进行比较,并在发现两者差异超过某个阈值(适合的阈值范围为平均绝对值的5到20倍)时将采样点幅值归零,以此来抑制尖刺,但此类方法对阈值选取较为敏感。Boore and Bommer (2005)提出“jerk”方法,对加速度迹线求导将尖刺转换为双面脉冲,从而更容易识别尖刺,但这种方法适用于存在较为明显的尖刺的强震动记录,即图1所示类型,但当加速度某些采样点并不作为局部最大值且与之前的采样点幅值差较大时,该采样点易被识别为尖刺。Wang

et al. (2003) 则采用小波分析并提出了一种非线性能量算子 WNEO 算法, 该方法能够有效分离并检测幅值较大的尖刺。

强震动记录中尖刺波形识别与生物、医学领域中的信号尖峰的识别相比有着较大差异 (Yu et al., 2018; Liu and Yue, 2019; Yu et al., 2022), 不同强震动记录的幅值、时程甚至采样周期的差异导致研究人员在将大量的数据放至时间尺度, 幅值大小皆不相同的窗口内比较时, 往往难以得到准确直观的判断 (Yao et al., 2021), 尽管地震工程领域的研究者早已注意到这一与实际不符、与邻近采样点不协调的异常现象, 但在尖刺的识别上始终未能形成统一的标准。为了尽量避免幅值差异较大的多组数据, 在尖刺识别过程中标准不一致的问题, 本文提出一种将采样点幅值按其所对应的持时长短进行调整, 并在原始幅值大小尺度下提取并强化关键特征的方法, 以此辅助研究人员对含有尖刺波形的强震动记录进行标记。

随着强震动记录数据量与日俱增, 机器学习被广泛地应用于强震动记录的校正及应用研究中。宋晋东等 (2023) 以 2022 年青海门源 6.9 级地震为例, 基于机器学习地震预警方法模块, 震级估计与现地阈值报警进行了回溯验证。朱景宝等 (2025) 提出一种结合机器学习与迁移学习的现地地震动峰值预测方法。李春果等 (2025) 提出一种基于谱元法(SEM)的模拟与基于人工神经网络(ANN)的宽频带地震动模拟方法。不同的强震动记录可能具有不同采样点数, 因此难以通过单一的模型直接对一维时序数据进行分类, 故此计算机视觉被广泛应用于波形识别, Zhao et al. (2023) 利用小波分析得到速度时程的小波系数图, 应用卷积神经网络对速度时程中的脉冲进行有效识别, 但在小波分析之前需要人为选取固定时长的记录以适应机器学习模型。仅使用未调整时长的加速度记录, 数据的时程尺度的不一致在绘制二维图像时也会导致前述问题。为此本文提出一种将数据特征转换为一维数据特征向量的处理方法, 该方法可以一定程度上避免上述问题, 并使得机器学习能够被应用于大量的、不同时程长度的强震动记录尖刺识别。经处理后的数据, 即便仅结合简单的集成决策树模型, 也能够实现超过 86% 的马修斯相关系数 (MCC)。同时去时程尺度限制的思路为后续开发可广泛适用于一维数据的机器学习方法提供了有益的借鉴。

1 尖刺波形判别

1.1 数据集选取

考虑到大震能够产生更高的地震动幅值和更丰富的强震动记录, 有利于保证数据的代表性与研究结果的工程应用价值。浅源地震中, 日本 M_{JMA} 与 M_w 在 5–7 级范围内较为一致 (Akio, 1996), 当震级大于 7 时 M_{JMA} 略低于 M_w 。 M_s 在 8 级左右会出现饱和现象, 除此之外 M_s 与 M_w 整体相近, 因此以 7 级以上地震事件作为研究数据集具有可比性与可行性。自 20 世纪 90 年代起, 国际强震动观测网络建设显著推进, 如美国 ANSS、日本 K-NET 及我国“十五”期间台网建设 (李山有等, 2003)。综合考虑, 研究选取了 2000 年后若干 7 级以上的地震事件, 从中选取 PGA 大于 100gal 的强震动记录构建样本数据, 并对其应用了 0.1–40Hz 的巴特沃斯带通滤波 (姚鑫鑫等, 2022), 地震事件信息如表 1 所示。

表 1 数据集地震事件信息

Table 1 Seismic event information of the dataset

| 地区 | 日期 | 震级 | 记录数量 |
|-----------------------|------------|---------------|------|
| Darfield, 新西兰 | 2010-09-04 | M_w 7.0 | 97 |
| El Mayor-Cucapah, 墨西哥 | 2010-08-04 | M_w 7.2 | 78 |
| 十胜, 日本 | 2003-09-26 | M_w 7.1 | 67 |
| 十胜, 日本 | 2003-09-26 | M_w 8.0 | 165 |
| 纪伊半岛, 日本 | 2004-09-05 | M_{JMA} 7.4 | 49 |
| 钏路, 日本 | 2004-11-29 | M_{JMA} 7.1 | 67 |
| 汶川, 中国 | 2008-05-12 | M_s 8.0 | 150 |
| 岩手-宫城, 日本 | 2008-06-14 | M_{JMA} 7.2 | 99 |
| 三陆冲, 日本 | 2011-03-11 | M_{JMA} 7.7 | 51 |
| 三陆冲, 日本 | 2011-03-11 | M_{JMA} 7.4 | 31 |
| 三陆冲, 日本 | 2011-03-11 | M_w 9.0 | 128 |
| 宫城, 日本 | 2011-04-07 | M_{JMA} 7.1 | 212 |
| 三陆冲, 日本 | 2012-12-07 | M_{JMA} 7.3 | 131 |
| 芦山, 中国 | 2013-04-20 | M_s 7.0 | 49 |
| 熊本, 日本 | 2016-04-16 | M_{JMA} 7.3 | 136 |
| 福岛, 日本 | 2016-11-22 | M_{JMA} 7.4 | 60 |
| 福岛, 日本 | 2021-02-13 | M_{JMA} 7.3 | 230 |
| 福岛, 日本 | 2022-03-16 | M_{JMA} 7.4 | 292 |
| Pazarcik, 土耳其 | 2023-02-06 | M_w 7.6 | 43 |
| Elbistan, 土耳其 | 2023-02-06 | M_w 7.8 | 113 |
| 能登半岛, 日本 | 2024-01-16 | M_w 7.5 | 139 |

| | | | |
|---------|------------|---------------|----|
| 十胜, 日本 | 2008-09-11 | M_{JMA} 7.1 | 12 |
| 三陆冲, 日本 | 2011-03-09 | M_{JMA} 7.3 | 23 |
| 福岛, 日本 | 2013-10-26 | M_{JMA} 7.1 | 23 |

九寨沟, 中国 2017-08-08 M_s 7.0 10

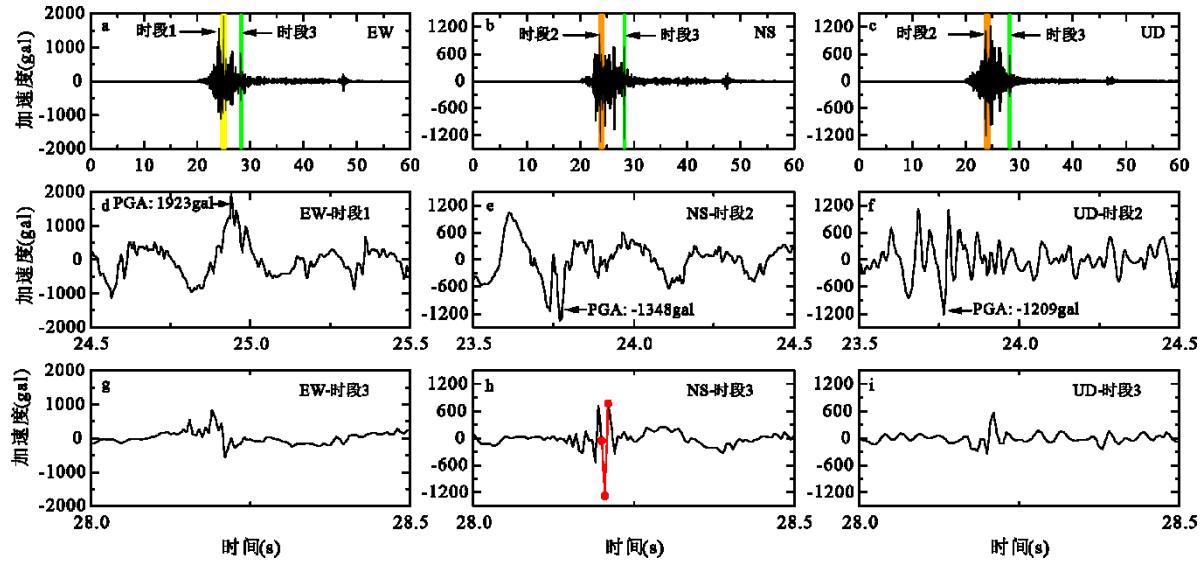


图 2 2017 九寨沟 M_s 7.0 地震 51JZZ 台站东西向强震动记录的加速度波形 (a), 及其时段 1、时段 3 的加速度波形 (d, g); 南北向强震动记录的加速度波 (b), 及其时段 2、时段 3 的加速度波形 (e, h); 竖直向强震动记录的加速度波形 (c), 及其时段 2、时段 3 的加速度波形 (f, i)。

Fig.2 Acceleration waveforms of strong-motion records at station 51JZZ from the 2017 M_s 7.0 Jiuzhaigou Earthquake, acceleration waveform of EW direction strong-motion record (a) and its acceleration waveforms for intervals 1 and 3 (d, g); acceleration waveform for NS direction strong-motion record (b) and its acceleration waveforms for intervals 2 and 3 (e, h); acceleration waveform for UD direction strong-motion record (c) and its acceleration waveforms for intervals 2 and 3 (f, i)

1.2 典型尖刺波形特征分析

2017 年 8 月 8 日九寨沟 M_s 7.0 地震中, 章扎台站(51JZZ)强震动记录东西向 PGA 高达 1.92g, 然而该地区并未出现严重的结构破坏, 因此学界普遍认为该组记录存在异常。图 2 展示了该台站三分向加速度波形: 其中, 东西向记录的时段 1、南北向和竖直向记录的时段 2 分别对应各分向的 PGA 时段。通过对比可以发现, 在 PGA 所在时段内, 加速度波形整体幅值变化较为平缓, 并未出现尖刺波形。相比之下, 南北向记录在时段 3 内的加速度波形(图 2h)中, 标注的三个相邻采样点则形成了典型的尖刺特征; 同时, 在该时刻, 东西向与竖直向的加速度幅值也出现了明显增大。这表明, PGA 并不一定对应于尖刺波形所在的幅值, 且尖刺波形的产生可能受实际的地面运动影响。

李渝生等 (2017) 在研究九寨沟 M_s 7.0 地震的发震构造后指出, 其震源断层的破裂过程受一条走向北偏西 331° 、倾角 87° 的断层面控制, 表现为左旋走滑性质。根据公开资料, 九寨沟地震的震源位置为 $E103.812^\circ$ 、 $N32.200^\circ$, 而 51JZZ 台站的经纬度为 $E103.879^\circ$ 、 $N32.290^\circ$, 震源距约 11km。由此计

算得出震源与台站的连线方向基本垂直于断层走向。将东西向与南北向加速度数据绘制成加速度平面轨迹图与加速度差值平面轨迹图(见图 3), 结果显示红色线段分别为两幅轨迹图的最大向量, 其方向分别为 326° 和 332° , 与断层走向高度一致。

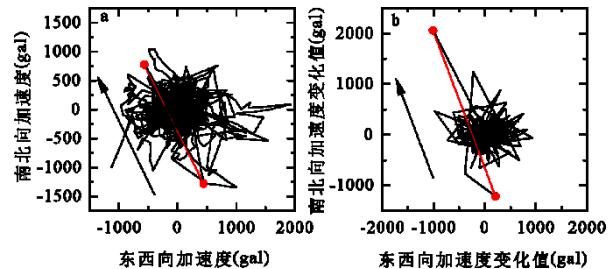


图 3 (a) 加速度平面轨迹图; (b) 加速度差值平面轨迹图

Figure 3. (a) Acceleration planar trajectory; (b). differential acceleration planar trajectory.

结合地震现场调查结果分析, 51JZZ 台站位于九寨沟县章扎镇 301 省道南侧, 建于河沟陡坡之上, 场地类型为 II 类场地, 土质较为松软。对该台站强震仪进行检查与测试表明, 其三个方向极性均正确, 标定记录无误差, 衰减波清晰, 说明仪器工作正常。该台站基墩尺寸为 $0.44m \times 0.44m \times$

1.5m, 与规范 (DB/T17-2018) 规定的 $0.4\text{ m} \times 0.4\text{ m} \times 0.6\text{ m}$ (中国地震局, 2018) 相比存在较大差异。图 4 给出了一组标准截面尺寸、高度 1.6m 基墩的基墩振动台试验数据, 其 PGA 相较于振动台输入数据放大了约 1.7 倍, 在 14–15s 时段的加速度波形中可以明显观察到基墩作用使高频成分增强, 这一现象在傅里叶幅值谱中也有清晰体现, 同时基墩对记录的低频成分亦存在一定影响。由此推测, 基墩尺寸异常导致的高频成分增强可能是造成 51JZZ 台站 PGA 幅值偏高及尖刺波形出现的可能原因。但鉴于该台站场地较软, 且地理位置相对特殊, 因此不能排除场地效应及与断层空间关系等因素对尖刺波形产生的贡献。

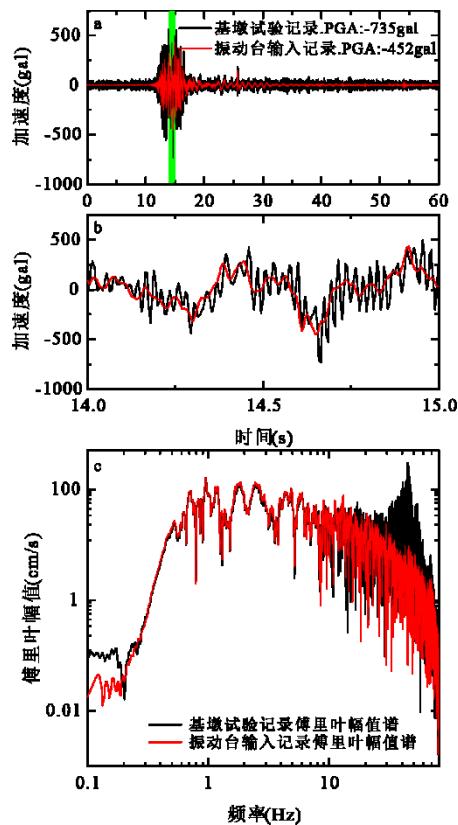


图 4 $0.4\text{ m} \times 0.4\text{ m} \times 1.6\text{ m}$ 尺寸基墩振动台试验加速度波形对比图 (a)、
(b) 及傅里叶幅值谱对比图 (c)

Fig.4 Comparison of acceleration waveforms from the shaking table test of foundation pier with dimensions $0.4\text{ m} \times 0.4\text{ m} \times 1.6\text{ m}$ (a, b),
and comparison of Fourier amplitude spectra (c)

1.3 加速度幅值变化特征提取

尖刺波形的定义尽管有着大致的范围与描述, 但其在强震动记录中的识别一直以来都具有较强的主观因素, 无论是对幅值比较阈值的选取还是利用图像人为筛选, 强震动持续时间与幅值的差异将

会干扰研究者的判断。基于以上问题, 我们希望找到一种数据筛选手段, 将强震动记录的加速度幅值变化特征提取出来, 同时根据其变化的剧烈程度来判断该数据中的疑似尖刺。

对原始数据进行二次微分, 斜率不为 0 且斜率瞬时变化率与该点幅值正负号相反的采样点被定义为峰点, 同理将斜率不为 0 且斜率瞬时变化率与该点幅值正负号相同的采样点定义为谷点, 能够同时作为峰点和谷点的采样点定义为共轭峰谷点, 如图 5 所示。当一个峰点的幅值远远高于周围数据点的幅值时, 将其视为一个尖刺波形, 但对于定量的幅值比较法, 不同研究者对其比值阈值的确定标准并不统一, 因此将明确尖刺波形的概念、减少尖刺波形数据筛选的主观因素影响作为本文提出特征提取方法出发点以此开展后续的工作。

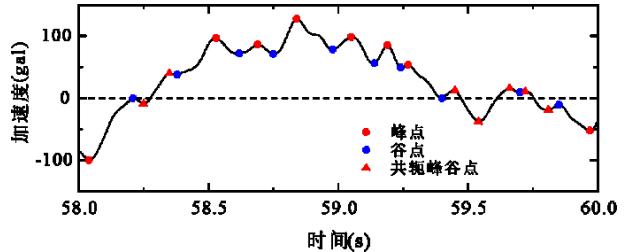


图 5 峰、谷点示意图。

Fig.5 Spike and trough points diagram.

本研究将波峰的形成视为一个完整的过程: 当加速度斜率的绝对值相较于前一采样点增大时, 定义为波峰产生过程的起始; 至峰点时刻, 加速度绝对值达到最大值; 随后, 当加速度斜率的绝对值相较于前一采样点减小时, 定义为该过程的结束。简而言之, 即在峰点两侧的谷点范围内搜索拐点, 并将最靠近峰点的拐点视为波峰产生过程的起点。随后, 计算该峰点与两个拐点的幅值差, 并取其中绝对值最大者作为该峰点的幅值变化值; 同时, 将两个拐点之间的时长与采样周期的比值定义为峰点持时 (见图 6)。需要强调的是, 幅值变化值曲线相较于传统加速度时程曲线具有完全不同的物理意义, 其仅衡量波峰形成过程中加速度的相对变化, 而忽略其余采样点的幅值波动。这种处理方式使数据更为稀疏, 既便于特征观察, 也有利于后续特征提取与分析。

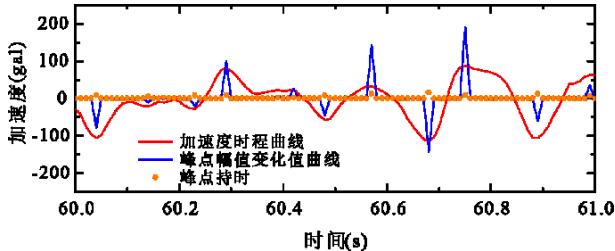


图 6 峰点幅值变化量与峰点持时示意图。

Fig.6 Schematic diagram of spike amplitude variation and spike duration.

在获得峰点幅值变化值后,还需进一步乘以持时系数,以强化可能存在的尖刺特征,便于数据分类。该系数保证峰点持时较长的峰点幅值变化值被削减,而峰点持时较短的峰点幅值变化值基本保持不变。考虑到有些强震动记录存在峰点单侧幅值变化较为剧烈的情况,因此本研究将峰点持时 ≤ 3 的持时系数设定为1;峰点持时 ≥ 5 的峰点持时对应的系数设定为 ν (ν 为持时系数的下限值),这意味着不满足尖刺定义的峰点幅值变化值至多被减少到原本幅值的 ν 倍,很多强震动记录的峰点幅值变化特征散点图,散点分布较为稀疏,特征不够显著,为了让特征向量更具辨识性,只能使 ν 取值尽可能小,因此本研究采用0.1作为持时系数下限值。

$$c = \begin{cases} 1 & , t \leq 3 \\ \frac{1}{2}(1+\nu) & , t=4 \\ \nu & , t \geq 5 \end{cases} \quad (1)$$

1.4 处理结果讨论

基于1.3节提出的方法,对选取的强震动记录进行了处理。图7与图8展示了对强震动记录进行特征提取的主要步骤。图7(a)给出了2008汶川

$M_s8.0$ 地震 062WIX 台站南北向强震动记录的加速度时程曲线,其中突出显示区域为疑似尖刺波形所在区段。将该区域在相同时间尺度下展开,如图7(b)所示,其中红色部分对应于采样周期内加速度变化值较大的一个波峰形成过程。基于此,得到未经持时系数修正的峰点幅值变化特征散点图(图7(c)),可见该峰点的幅值变化特征得以保留。经持时系数作用后(图7(d)),大部分峰点的幅值变化值受到明显调整,而该峰点的幅值变化值基本保持不变,且在负幅值一侧出现了一个较为突出的离群点。因此,该条记录被标记为正样本。

图8(a)展示了2024能登半岛 $M_w7.5$ 地震 ISK005 台站南北向强震动记录的加速度时程曲线。在突出显示区域内可观察到一个较为明显的异常尖峰。将其展开后(图8(b)),可以发现该峰点的持时较长,不符合尖刺波形的定义。图8(c)给出了该记录未经持时系数修正的峰点幅值变化特征散点图,其中存在两个显著的离群点,然而经持时系数作用后(图8(d)),这些离群点在原幅值变化特征的尺度下已基本消失,因此该条记录被标记为负样本。

“jerk”方法通过对加速度进行微分将单侧脉冲转换为双侧脉冲以更好地实现尖刺波形识别。对已标记的强震动记录应用“jerk”方法,许多被判定为含有尖刺波形的强震动记录其加速度微分曲线在 jerk 图像中呈现出较为明显的双侧脉冲,如图9(c)所示。但仍有相当一部分记录仅表现为单侧脉冲(图9(a)),或两侧脉冲幅值差异较大(图9(b))。这是因为本文研究聚焦于波峰产生过程而非单个峰点与周围采样点之间的幅值差异。

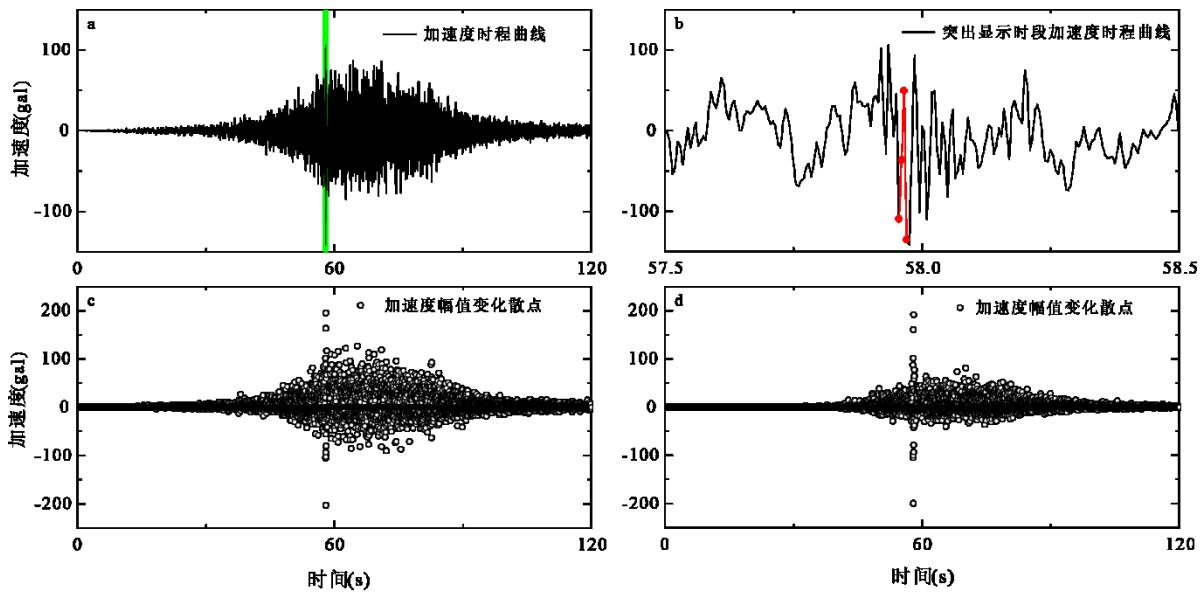


图 7 2008 汶川 $M_s8.0$ 地震 062WIX 台站南北向强震动记录的峰值幅值变化特征提取。(a) 强震动记录的加速度时程曲线; (b) 加速度时程曲线中绿色区域的波形细节展示图; (c) 前处理过程初步得到的峰值幅值变化特征散点图; (d) 经持时系数作用后的峰值幅值变化特征散点图

Fig.7 Spike amplitude variation of 062WIX station NS-direction strong-motion record from the 2008 Wenchuan $M_s8.0$ Earthquake. (a) Acceleration time history; (b) detailed waveform of highlighted segment;(c) scatter plot of spike amplitude variation before spike duration correction;(d) scatter plot of spike amplitude variation after correction.

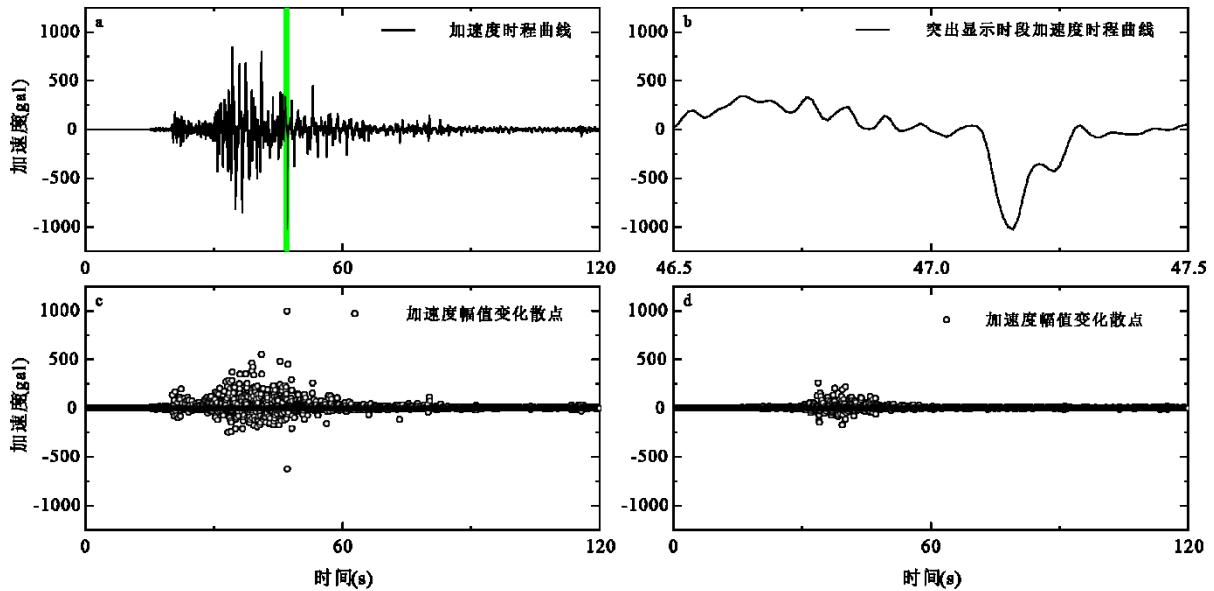


图 8 2024 能登半岛 $M_w7.5$ 地震 ISK005 台站南北向强震动记录的峰值幅值变化特征提取。(a) 强震动记录的加速度时程曲线; (b) 加速度时程曲线中突出显示时段波形图; (c) 前处理过程初步得到的峰值幅值变化特征散点图; (d) 经持时系数作用后的峰值幅值变化特征散点图

Fig.8 Spike amplitude variation of ISK005 station NS-direction strong-motion record from the 2024 Noto Peninsula $M_w7.5$ Earthquake. (a) Acceleration time history; (b) detailed waveform of highlighted segment;(c) scatter plot of spike amplitude variation before spike duration correction;(d) scatter plot of spike amplitude variation after correction.

当某一峰点仅在一侧表现出较大的幅值变化时, 也会被视为尖刺波形, 从而导致 jerk 图像中仅在一侧出现显著脉冲。此外, 部分 jerk 图像未能呈现明显脉冲, 这是由于非峰点处的微分值过大, 掩盖了峰点对应的 jerk 脉冲特征。另一方面, 对于标记为不含尖刺波形的强震动记录, 其 jerk 图

像中也可能出现类似的双侧或单侧脉冲, 如图 10 所示为 2008 年汶川 $M_s8.0$ 地震 051AXT 台站东西向记录的 jerk 图像, 该记录被标记为不含尖刺波形的负样本, 其加速度微分曲线呈现出明显的双侧脉冲, 但与之对应的加速度时程曲线在双侧脉冲出现的时段却整体较为平滑, 仅包含一处幅值较小的

高频波动。这说明当加速度中出现高频成分时,即使加速度幅值较低、幅值变化较小, jerk 图像中仍可能产生明显的脉冲。进一步分析表明,图中两侧脉冲分别出现在 67.322s 与 67.328s, 相隔 6 个采样周期(采样周期为 0.002s), 不满足 Boore 对尖刺波形的严格定义。

综上所述,单纯依靠研究者对强震动记录加速度时程曲线的直观判断,并不足以作为识别尖刺波形存在与否的可靠依据。这是因为尖刺波形在部分记录中表现并不显著,容易受到高频噪声或幅值波动的干扰,从而造成误判。除本文所展示的两条典型记录外,还有相当数量的强震动记录在时程图像中未能直观呈现出异常特征,但通过第 1.3 节提出的特征提取方法,可以有效捕捉并量化峰点的幅值变化,并进一步利用持时系数对该特征进行强化。

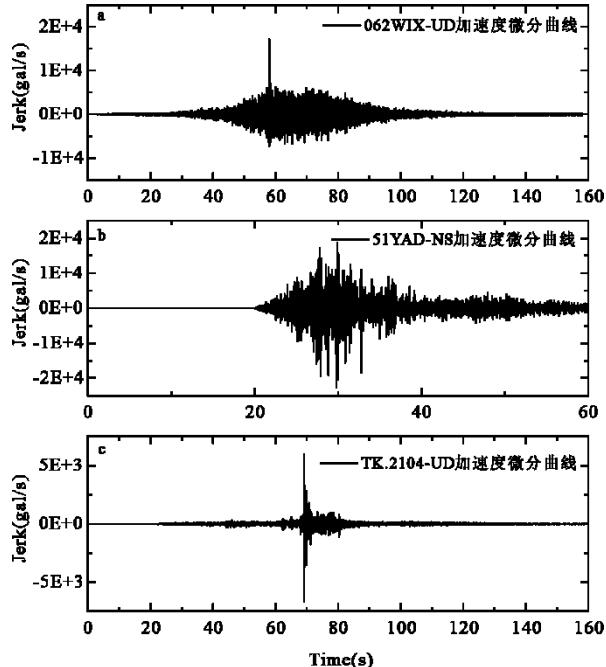


图 9 含尖刺波形强震动记录 jerk 图像, 从上至下分别来自 2008 汶川 M_8 .0 地震, 2013 芦山 M_s 7.0 地震, 2023 土耳其-叙利亚 M_w 7.8 地震。

Fig.9 Jerk images of strong-motion records containing spike waveforms. From top to bottom, correspond to the 2008 Wenchuan M_8 .0 Earthquake, the 2013 Lushan M_s 7.0 Earthquake, and the 2023 Turkey - Syria M_w 7.8 Earthquake.

该方法不仅降低了人为主观判断带来的不确定性,而且为尖刺波形的识别提供了更加客观的依据,辅助研究者更准确地对含有尖刺波形的强震动记录进行标记。

基于现有的数据考虑到仪器参数、地区特性以及发布数据的完整性等复杂因素,无法对单一因素进行研究分析,因此本文仅聚焦于现有数据的尖刺

波形识别,不就尖刺波形的成因过多讨论。

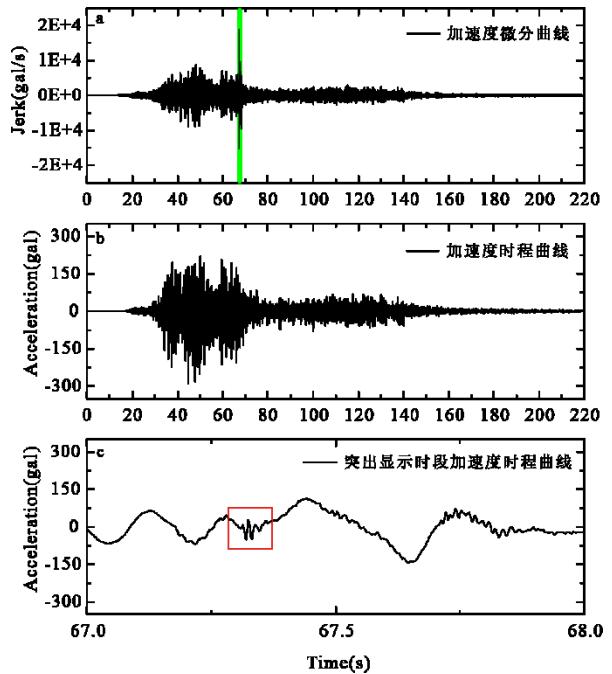


图 10 2008 汶川 M_8 .0 地震 051AXT 台站东西向强震动记录 jerk 图像 (a)、加速度波形 (b) 及 jerk 图像突出显示时段加速度波形 (c)。

Fig.10 Jerk image (a), acceleration waveform (b) and acceleration waveform at the highlighted interval in jerk image. (c) of 051AXT station EW-direction strong-motion record from the 2008 Wenchuan M_8 .0 Earthquake.

2 特征提取

强震动记录时长不一致,给一维时序数据的直接利用及其在机器学习模型中的应用带来了困难。研究者通常采用两种方式来解决数据长度差异:其一是对原始数据进行截取,其二是通过补零将较短的记录扩展至统一的数据点数(Wang et al., 2020)。其中,补零方法能够有效保证数据维度的一致性,且不会影响第 1.3 节所提出方法的处理结果。然而,当采样点数超过统一设定值时,截取方式却难以确保异常波形等关键特征得以完整保留(人工截取可以在一定程度上保留有价值的信息,但自动化程序难以实现这一点)。鉴于此,本文提出了一种新的特征提取方法,将数据的转化为维度一致的特征向量,从而避免截取造成的特征损失。

该方法首先计算峰点幅值变化值的双侧累积分布,并以未经峰点持时作用的峰点幅值变化值最大值为边界,将其划分为 200 个区间,统计处理后峰点幅值变化值在各区间范围内的点数。随后,将双侧累积分布取倒数以强化尖刺特征,并对数据进行归

一化处理,从而得到最终的特征向量。图 11 展示了第 1.3 节中两条强震动记录的双侧累积分布图像及其特征向量可视化结果。其中,图 11 (a) 和图 11 (b) 分别对应图 7 (d) 与图 8 (d) 的双侧累积分布。可以看出,不同峰点数的强震动记录,其累积分布数值存在较大差异,但峰点数量本身并非判别尖刺波形的有效特征。若直接将其归一化作为模型输入,反而会削弱模型的判别能力并降低可解释性。尖刺波形的特征应体现为小数值占主导,因此通过取倒数操作可强化小数值的作用,同时抑制大数值的影响。图 11 (c) 图 11 (d) 分别为典型的尖刺、非尖刺记录特征向量,其特征向量可视化图像出现了明显的差异,且特征向量值被约束于[0, 1]范围,达到了归一化的目的。

以上述特征提取方法作为辅助手段对数据集行分类,最终样本集含有尖刺波形的正样本 67 条,不含尖刺波形的负样本 2388 条。

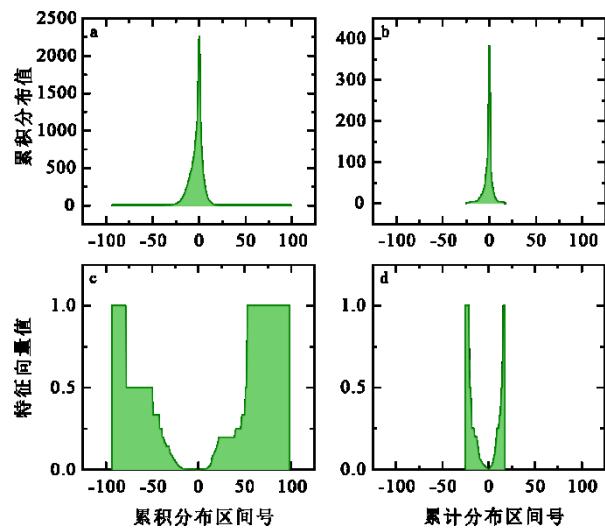


图 11 图 7 的双侧累积分布 (a) 及其特征向量可视化图像 (c); 图 8 的双侧累积分布 (b) 及其特征向量可视化图像 (d)。

Fig.11 Bilateral cumulative distribution (a) and its feature vector visualization image (c) for Fig.7; bilateral cumulative distribution (b) and its feature vector visualization image (d) for Fig.8.

3 机器学习在强震动记录“尖刺”识别中的应用

3.1 LightGBM-SVM 的决策树堆叠模型 Stacking

本研究采用第 2 节所述的数据集训练了 7 种机器学习模型。结果表明,经过贝叶斯优化的 SVM 模型在分类准确率、F1 分数、召回率及马修斯相关系数 (Matthews Correlation Coefficient, MCC)

方面均表现最佳; LightGBM 模型次之。在此基础上,进一步采用决策树算法对 LightGBM 与 SVM 进行堆叠 (Stacking),结果显示该集成模型对正类样本的识别准确率上显著提升,且仅在某一类型的正类数据上存在稳定的误判现象。最后,将未参与训练和测试的独立数据集作为验证集,对堆叠模型的性能进行了进一步分析与验证。

LightGBM (Light Gradient Boosting Machine) 是一种基于决策树的分布式梯度提升框架,具有更高的效率和准确性 (Ke et al., 2017; Chen et al., 2023)。它采用了 Gradient-based One-Side Sampling (GOSS)、Exclusive Feature Bundling (EFB) 技术通过优先保留梯度较大的样本,减少梯度较小的样本,从而加速训练过程并保持准确性同时将不互斥的特征进行捆绑,减少特征维度,提升训练速度并降低内存开销,此外 LightGBM 采用 Leaf-Wis 树构建方式,进一步提升了模型的准确性和收敛速度,并利用直方图方法来加速分割过程,特别适用于高维数据集的特征提取,对异常值和缺失值较为敏感导致其具有良好的鲁棒性,在处理大规模数据集、稀疏数据和高维度特征的工作中表现尤为出色。

支持向量机 (Support Vector Machine, 简称 SVM) 是一种常用于分类和回归分析的监督学习模型 (Kim et al., 2020; Pakniat et al., 2025), SVM 的核心思想是搜索一个最优的超平面,将不同类别的样本尽可能地分离并最大化超平面两侧的间隔,以此构建出稳定的决策边界,使得 SVM 具有良好地处理高维数据和小样本数据的能力,并通过最大化类间间隔来保证分类的稳定以及对样本小扰动的鲁棒性,因此该模型适用于二分类及多分类问题。

贝叶斯优化 (Bayesian Optimization) 是一种用于优化目标函数的全局优化方法 (Ghavamzadeh and Engel, 2007), 它是一种序列决策过程, 通过构建目标函数的代理模型逐步推测并选择下一个最有可能带来最好结果的试验点, 从而逐步完成对模型的超参数优化, 各机器学习模型均采用贝叶斯优化的方式进行超参数优化, 参与优化的主要参数、取值范围以及最优参数如表 2 所示。

由于正负样本数据集类别极不平衡,在训练过程中考虑添加类别权重并将 MCC 作为贝叶斯优化指标 (Chicco and Jurman, 2020)。MCC 是一种常用于评估二分类问题中分类器性能的指标,能够在类别不平衡的情况下提供更为稳定和全面的评估,它考虑了四个指标: 真正例 (TP)、真反例 (TN)、

假正例 (FP) 和假反例 (FN)，其计算公式为：

(2)

表 2 各机器学习模型主要参数、优化范围及最优参数

Table 2 The main parameters, optimization range and optimal parameters of each machine learning model

| 模型 | 参数 | 超参数搜索范围 | 最优参数 |
|----------|------------|------------------------------------|---------|
| LightGBM | 决策树最大叶数 | [5, 100] | 12 |
| | 决策树数量 | [10, 1000] | 947 |
| | 学习率 | [0.01, 0.5] | 0.3 |
| SVM | 正则化参数 | [1, 1000] | 72 |
| | RBF 核函数的参数 | [-5, -1] | -1.09 |
| DNN | 全连接层神经元个数 | [64, 128] | 116 |
| | 神经元丢弃率 | [0.1, 0.5] | 0.124 |
| KNN | 邻居数 | [1, 19] | 1 |
| | 权重 | {Uniform, Distance} | Uniform |
| CNN | 卷积核数量 | [16, 64] | 17 |
| | 卷积核尺寸 | [3, 7] | 6 |
| | 训练批量大小 | [16, 64] | 40 |
| LR | 学习率 | [0.001, 0.01] | 0.009 |
| | 正则化强度 | [0.001, 1] | 0.945 |
| | 求解器类型 | {Liblinear, Sag, Newton-Cg, Lbfgs} | Sag |
| NB | 惩罚项类型 | {L1, L2} | L2 |
| | 平滑参数 | [1E-9, 1E-6] | 6E-9 |

对 LightGBM 与 SVM 的误识别案例进行分析发现，两种模型在正类样本上的错误识别仅有少量交集。因此本文采用决策树模型作为元学习器，将 LightGBM 与 SVM 作为基模型构建堆叠 (Stacking) 模型。堆叠是一种典型的集成学习方法，通过融合多个模型的预测结果来提升整体预测性能 (Long et al., 2025)。其具体流程为：以 LightGBM 与 SVM 的预测结果作为输入，训练一个新的决策树模型，并由该模型重新输出最终预测结果。

3.2 多模型识别结果分析

为验证所提方法在不同机器学习模型中的适用性，本文选取了 7 种传统机器学习模型及 1 种集成模型进行训练。将强震动记录数量大于 30 的地震事件作为验证集与测试集。为增加样本数量，对数据集中的正样本特征向量进行顺序翻转，从而构造出原始数据量两倍的正样本集。训练与测试集按照 4:1 的比例划分，并在每次随机划分数据后重复训练 10 次，统计结果如表 3 所示。结果表明，集

成模型 Stacking 在 MCC 的统计学描述指标上整体优于其他模型，仅在标准差方面略低于 SVM 与 LogisticRegression (LR)；而 Naive Bayes (NB) 的各项指标表现最差。

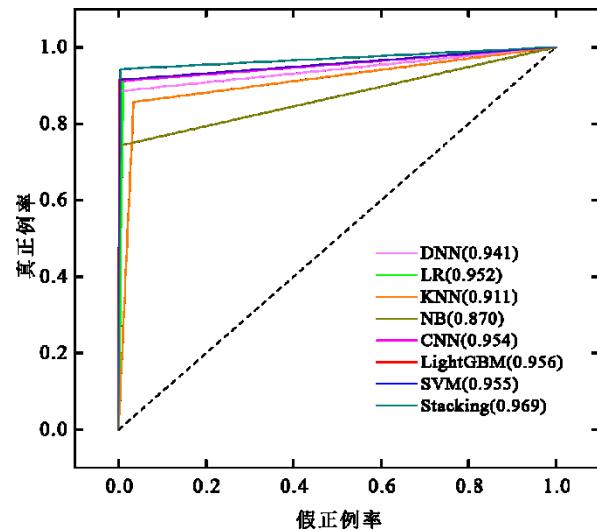


图 12 ROC 曲线

Fig.12 ROC curve.

由图 12 的 ROC 曲线可知, 集成模型 Stacking 的 AUC (曲线下面积) 值最高, 为 0.969; 其次为 LightGBM 和 SVM, AUC 值分别为 0.956 和 0.955; Naive Bayes (NB) 的 AUC 值最低, 仅为 0.870。这一结果与表 4 中的统计性描述基本一致。需要注意的是, Stacking 模型在负样本 (即不含尖刺波形的数据) 上的误识别率相对较高, 但本研究更关注正样本的识别准确性。从表 3 的混淆矩阵可以看出, Stacking 模型对正样本的误识别率显著降低, 其在负样本上的误识别率仍处于可接受范围内。相比之下, 尽管 NB 在正类样本上的误识别率低于 KNN, 但其对负类样本的误识别率远高于其他模型, 这一结论与表 4 及 ROC 曲线的结果保持一致。

表 3 机器学习模型对测试集混淆矩阵

Table 3 Confusion matrices of machine learning models for the test set.

| 模型 | TP | TN | FP | FN |
|----------|----|-----|----|----|
| LightGBM | 32 | 577 | 3 | 1 |
| SVM | 32 | 576 | 3 | 2 |
| DNN | 31 | 576 | 4 | 2 |
| CNN | 32 | 576 | 3 | 2 |
| LR | 32 | 572 | 3 | 6 |
| NB | 30 | 558 | 5 | 20 |
| KNN | 26 | 576 | 9 | 2 |
| Stacking | 33 | 575 | 2 | 3 |

对误识别案例的深入分析进一步验证了 Stacking 集成模型的优越性与局限性。由图 13 可见, 单一模型在面对不同类型的正类样本时存在明显短板, SVM 对于图 13 (a)、(c) 所示类型的正类样本识别能力较弱, 这说明在处理某些尖刺波形特征向量时, SVM 对小尺度特征差异不够敏感; LightGBM 则在图 13 (b)、(c) 所示类型的正类样本识别准确率较低, 说明树模型在处理特征分布高度不均衡时, 容易受离群点影响, 导致分类边界偏移容易将其与负样本混淆。当两者进行堆叠后, Stacking 模型能有效弥补 SVM 与 LightGBM 的局部不足。在图 13 (a)、(b) 所示类型的正类样本中, Stacking 的识别准确率显著提升, 说明其在融合特征空间信息后对典型尖刺波形表现更加稳健。然而, 对于图 13 (c) 这类样本, Stacking 的表现仍不理想。造成这一现象的主要原因在于部分强震动记录即使经过持时因子修正, 其峰点幅值变化特征仍保留异常大的离群点。例如 2008 汶川 $M_{s}8.0$ 地震 051DFB 台站垂直向强震动记录 (图 14 所示)

中, 未经持时因子校正的峰点幅值变化最大值高达 804gal。校正后虽被削弱, 但仍保留了一个 290gal 的离群点, 而其余数据点均低于 100gal。这种情况下, 样本在统计意义上属于正类, 但其特征分布与部分负类样本过于接近, 导致模型在训练与预测过程中倾向于将其误判为负类。整体而言集成模型 Stacking 误识别的正样本特征更为稳定。

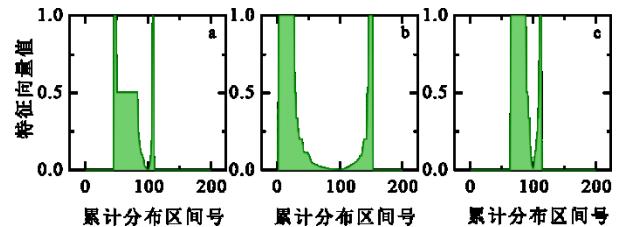


图 13 误识别正样本特征向量可视化图像。(a)、(b) 分别为 SVM、LightGBM 误识别类型, (c) 为两者共同误识别类型。

Fig.13 Visualization of the misidentified positive sample feature vectors. (a) SVM misidentification type, (b) LightGBM misidentification type, (c) common misidentification type of both.

对于负样本而言, 三种模型的误识别类型整体上更为随机, 没有表现出稳定或系统性的模式。这表明负样本整体特征分布相对复杂, 且模型之间在对负样本的判别过程中并未形成一致性的偏差。

3.3 独立验证集分析

为了检验训练得到的机器学习模型对未参与训练的数据集的识别效果, 本研究选取强震动记录数量小于 30 的地震事件组成独立验证集, 并加入

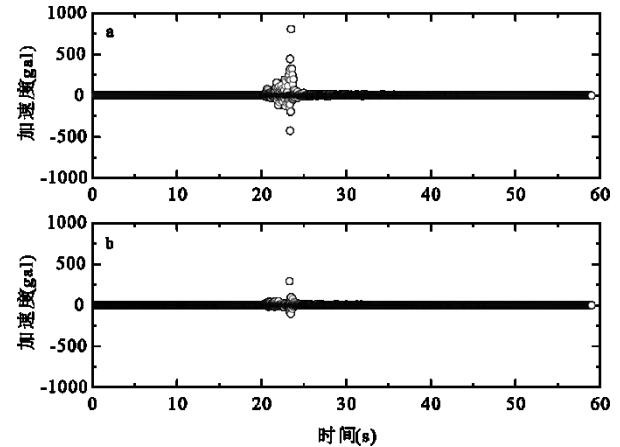


图 14 图 13(c)所对应的持时因子作用前峰点幅值变化特征散点图(a)及作用后的峰点幅值变化特征散点图(b)。

Fig.14 Scatter plots of spike amplitude variation for Fig.14(c). (a) Scatter plots of spike amplitude variation before spike duration correction, (b) after correction.

少量故障仪器的响应作为补充, 同上对地震事件里

的正样本特征向量做顺序翻转。

各机器学习模型对验证集的混淆矩阵如表 5 所示, 可以发现 LR 与 NB 的识别效果不佳, 其余机器学习模型对正样本以及负样本的识别准确率均较高, 这证明本文提出的特征提取方法在应用机器学习识别尖刺波形的工作中具有普适性与稳定性。观察验证集混淆矩阵发现, DNN 对负样本的识别准确率相较于 SVM、LightGBM 基本相同, 在小样本条件下 DNN 对负样本的识别准确率高于

SVM、LightGBM 是可以接受的。同时由于测试集的正样本较少, 为较标准的正样本类型数据, 因此三者对正样本的识别准确率较为一致, 对误识别样本进行分析, 除 DNN 外, 其他机器学习模型均对同一个负样本进行了误判, 该样本的幅值变化特征散点图与特征向量可视化图像如图 15 (a)、(c) 所示, 其并未含有明显的尖刺波形特征, 特征向量可

表 4 各机器学习模型对含尖刺记录识别性能 MCC 统计学描述

Table 4 Statistical description of the recognition performance of each machine learning model for records with spike (MCC)

| 模型 | 均值 | 标准差 | 最小值 | 最大值 | 中位值 |
|----------|-------|-------|-------|-------|-------|
| DNN | 0.882 | 0.050 | 0.800 | 0.946 | 0.879 |
| LR | 0.856 | 0.032 | 0.808 | 0.921 | 0.857 |
| SVM | 0.901 | 0.035 | 0.846 | 0.946 | 0.899 |
| LightGBM | 0.901 | 0.059 | 0.779 | 0.951 | 0.925 |
| KNN | 0.829 | 0.041 | 0.756 | 0.889 | 0.835 |
| CNN | 0.838 | 0.058 | 0.707 | 0.898 | 0.851 |
| NB | 0.648 | 0.034 | 0.592 | 0.713 | 0.638 |
| Stacking | 0.925 | 0.036 | 0.866 | 0.965 | 0.931 |

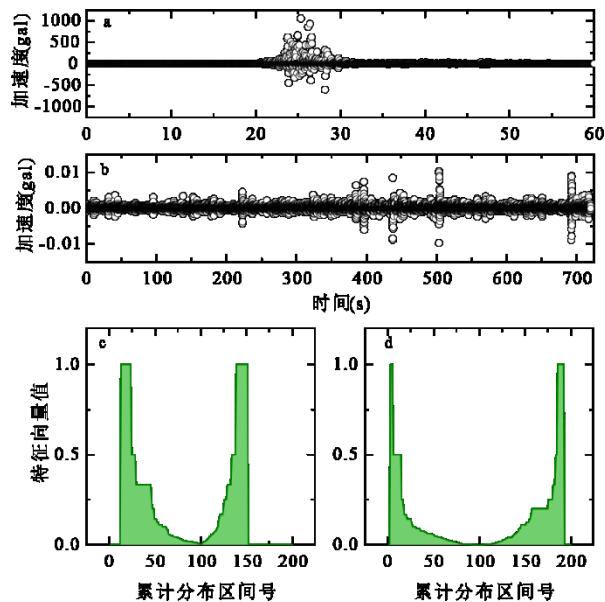


图 15 验证集识别负样本峰点幅值变化特征散点图 (a) 及其特征向量可视化图像 (c); 误识别正样本峰点幅值变化特征散点图 (b) 及其特征向量可视化图像 (d)

Fig.15 Scatter plot of amplitude variation characteristics of spike points of negative samples in the validation set (a) and its feature vector visualization image (c), as well as scatter plot of amplitude variation characteristics of spike points of positive samples (b) and its feature vector visualization image (d).

可视化图像在底部呈现出较宽的平台段, 这可能是机

器学习模型将其识别为正样本的原因; 对于误识别的正样本, 均属于故障仪器的响应, 如图 15 (b)、(d) 所示, 该强震动记录由故障仪器所记录, 经检查其加速度时程曲线存在大量尖刺波形, 但由于该记录时程过长, 同时含有多个幅值相近的离群值点, 其特征向量可视化图像与某类负样本的可视化图像较为接近, 因此被机器学习模型识别为负样本。

表 5 机器学习模型对测试集混淆矩阵

Table 5 Confusion matrices of machine learning models for the test set.

| 模型 | TP | TN | FP | FN |
|----------|----|----|----|----|
| LightGBM | 12 | 65 | 2 | 1 |
| SVM | 12 | 65 | 2 | 1 |
| DNN | 12 | 66 | 2 | 0 |
| CNN | 12 | 65 | 2 | 1 |
| LR | 12 | 65 | 2 | 3 |
| NB | 11 | 64 | 3 | 2 |
| KNN | 12 | 65 | 2 | 1 |
| Stacking | 12 | 65 | 2 | 1 |

由此可见, 对于多波包、多尖刺的强震动数据, 本文提出的方法难以对其进行精准的分类, 需要对多波包强震动记录进行分段或截取, 对于故障仪器的响应可以采用异常模态检测的方法判断其中是否存在尖刺波形。

4 结论

本文首先基于实际强震动观测经验说明了尖刺波形的客观存在,以及尖刺波形产生机理研究的重要性,进而针对强震动记录尖刺波形自动识别问题,提出了一种基于峰点幅值变化特征与持时系数加权机制的特征提取方法,并结合去时程化处理策略,构建了适用于机器学习模型输入的标准特征向量表示形式。在此基础上,系统对比了包括逻辑回归(LR)、朴素贝叶斯(NB)、支持向量机(SVM)、K最近邻(KNN)、深度神经网络(DNN)、卷积神经网络(CNN)、LightGBM、以及LightGBM-SVM决策树堆叠模型(Stacking)在内的八种机器学习方法的识别效果,重点评估了各模型在精度、稳定性、泛化能力以及误识别特征上的表现。研究得到以下主要结论:

(1) 本文提出的峰点幅值变化与峰点持时联合特征提取方法,能够在不依赖固定窗口长度的情况下有效提取尖刺波形的异常特征,具备良好的适应性与通用性,适用于不同震级、采样频率和记录时长的强震动数据,显著提升了模型的输入统一性与分类准确性。

(2) 在所有模型中,Stacking模型表现最为优异,其在训练集与独立验证集上的MCC平均值分别达到0.925和0.901,优于其他单一模型。相比之下,朴素贝叶斯和KNN模型表现不稳定,准确率和抗噪性能较低,不适合用于高噪声或特征分布不均的数据环境。

(3) 误识别样本分析表明,SVM与LightGBM在处理某些特定类型的尖刺记录上存在盲区,而Stacking模型通过整合两者优势,有效提升了对边界样本的识别能力,降低了正类数据误判概率,并在特征相近负样本上保持了良好的稳定性,体现了其在样本不平衡情况下的鲁棒性。

(4) 对于包含多个尖刺或尖刺密集型的记录,当前方法存在识别能力下降的问题,主要原因在于遍历整条记录特征提取难以准确捕捉局部尖刺信息。后续研究应引入时序分段策略、动态窗口机制等局部强化方法,以进一步提升复杂记录的识别精度。

(5) 尖刺波形产生的因素多种多样,难以对单一因素进行研究分析,需要对尖刺波形数据进行大量的积累并结合现场调查、仪器检测等工作才能对尖刺波形的产生提出假设、探明机理并进一步发

展修正方法。

综上所述,本文构建的基于特征提取与集成学习的尖刺波形识别体系,在强震动数据异常检测中展现出较高的准确性与实用性。该方法可作为强震动记录自动化处理流程中的关键一环,为后续地震动参数提取、地震危险性分析及抗震设计等提供更为可靠的数据基础。

References

- Akio K., 1996. Comparison of magnitudes estimated by the Japan Meteorological Agency with moment magnitudes for intermediate and deep earthquakes. *Bulletin of the Seismological Society of America*, 86(3): 832–842. doi: <https://doi.org/10.1785/BSSA0860030832>
- Arthur, E.B., 2009. The origin and significance of spikes in complex seismic trace attributes. *SEG Technical Program Expanded Abstracts*, 28(1): 1048-1052. doi: 10.1190/1.3255029
- Boore D, M., Bommer, J.J., 2005. Processing of strong-motion accelerograms: needs, options and consequences. *Soil Dynamics and Earthquake Engineering*, 25(2): 93-115. doi: 10.1016/j.soildyn.2004.10.007
- Chen, H.Y., Li, X.Y., Feng, Z.B., et.al., 2023. Shield attitude prediction based on Bayesian-LGBM machine learning. *Information Sciences*, 632: 105-129, doi: 10.1016/j.ins.2023.03.004
- Chicco, D., Jurman, G., 2020. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC genomics*, 21: 1-13, doi: 10.1186/s12864-019-6413-7
- China Earthquake Administration, 2018. Specification for the Construction of Seismic Station-Strong Motion Station. *Beijing: Seismological Press* (in Chinese).
- Douglas, J., What is a Poor Quality Strong-Motion Record? 2003. *Bulletin of Earthquake Engineering*, 1: 141–156. doi: 10.1023/A:1024861528201
- Ghavamzadeh, M., Engel, Y., 2007. Bayesian policy gradient algorithms. *Advances in neural information processing systems 19: Proceedings of the 2006 Conference*, Vol.19: 457-464. doi: 10.7551/mitpress/7503.003.0062
- Giuseppe, S., Claudia, B., 2002. Local-trace zeroing and spike zeroing: Two short automated noise-rejection routines to remove noise and spikes on seismic traces. *Geophysics*,

- 67 (1): 188–196. doi: 10.1190/1.1451510
- Ji, K., Wen, R.Z., Ren, Y.F., 2017. Site classification for National Strong Motion Observation Network System (NSMONS) stations in China using an empirical H/V spectral ratio method. *Journal of Asian Earth Sciences*, 147(1): 79-94. doi: 10.1016/j.jseaes.2017.07.032
- Ke, G.L., Meng, Q., Finley, T., et.al., 2017. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Advances in Neural Information Processing Systems* 30 (NIPS 2017).
- Kim, S., Lee, K., You, K., 2020. Seismic discrimination between earthquakes and explosions using support vector machine. *Sensors*, 20(7): 1879,10.3390/s20071879
- Li, C.G., Wang, H.W., Wen, R.Z., et.al., 2025. Simulation of Broadband Ground Motion in Yangbi Earthquake by Integrating Spectral Element Method and Artificial Neural Networks. *Earth Science*, 1-14 (in Chinese with English abstract).
- Li, S.Y., Jin, X., Liu, Q.F., et.al., 2003. Prospect of strong motion observation in China. *Earthquake Engineering and Engineering Vibration*, (02):1-7 (in Chinese with English abstract).
- Li, X.J., Zhou, Z.H., Yu, H.Y., et.al., 2008. Strong motion observations and recordings from the great Wenchuan Earthquake. *Earthquake Engineering and Engineering Vibration*, 7: 235–246. doi: 10.1007/s11803-008-0892-x
- Li, Y.S., Huang, C., Yi, S.J., et.al., 2017. Study on Seismic Fault and Sorce Rupture Tectonic Dynamic Mechanism of Jiuzhaigou M_s 7.0 Earthquake. *Journal of Engineering Geology*, 25(04): 1141-1150 (in Chinese with English abstract).
- Liu, D.Q., Yue, S.G., 2019. Event-Driven Continuous STDP Learning with Deep Structure for Visual Pattern Recognition, *IEEE Transactions on Cybernetics*. 49(4): 1377-1390. doi: 10.1109/TCYB.2018.2801476
- Long, T., Akbari, M., Fakharian, P., 2025. Prediction of soil liquefaction using a multi-algorithm technique: Stacking ensemble techniques and Bayesian optimization. *Journal of Soft Computing in Civil Engineering*, 9(2): 32-55. doi: 10.22115/scce.2024.453006.1860
- Pakniat, S., Najafizadeh, J., Kadkhodaavval, M., 2025. Machine learning for earthquake engineering analysis: Comparing regression models to predict peak ground acceleration. *World Journal of Advanced Research and Reviews*, 26(02): 856-867. doi: 10.30574/wjarr.2025.26.2.1714
- Song, J.D., Zhu, J.B., Wei, Y.X., et.al., 2023. Backtracking Verification of Machine Learning Earthquake Early Warning Magnitude Estimation and On-Site Threshold Alarm for Menyuan M_6 . 9 Earthquake in Qinghai on January 8, 2022. *Chinese J. Geophys.*,66(7):2903-2919 (in Chinese with English abstract).
- Wang, H.J., Jin, P., Liu, G.Z., 2003. Automatic spikes detection in seismogram. *Acta Seismologica Sinica*, 16: 348–355. doi: 10.1007/s11589-003-0039-0
- Wang, W.Y., Ji, K., Wen, R.Z., et.al., 2020. Impact of Strong Ground Motion's Process Procedure on the Structural Nonlinear Time-History Analysis. *Engineering Mechanics*, 2020, 37(S): 42-50, 62. doi: 10.6052/j.issn.1000-4750.2019.04.S003
- Yao, X.X., Ren, Y.F., Kishida, T., et.al., 2022. The Procedure of Filtering the Strong Motion Record: Denoising and Filtering. *Engineering Mechanics*, 39 (S1): 320-329 (in Chinese with English abstract).
- Yao, X.Y., 2023. Study on the Elaborate Data Processing of Strong Motion Record and the Flatfile Parameters in China (Dissertation). Institute of Engineering Mechanics, China Earthquake Administration, Harbin (in Chinese with English abstract).
- Yao, X.Y., Zhou, Q.Z., Wang, C., et.al., 2021. An adaptive seismic signal denoising method based on variational mode decomposition. *Measurement*, 177:109277. doi: 10.1016/j.measurement.2021.109277
- Yu, Q., Li, S.L., Tang, H.J., et.al., 2022. Toward Efficient Processing and Learning With Spikes: New Approaches for Multispike Learning. *IEEE Transactions on Cybernetics*, 52(3): 1364-1376. doi: 10.1109/TCYB.2020.2984888
- Yu, Q., Wang, L.B., Dang, J.W., 2018. Efficient Multi-spike Learning with Tempotron-Like LTP and PSD-Like LTD. In: Cheng L, Leung A, Ozawa S ed. *Neural Information Processing*. ICONIP 2018. Lecture Notes in Computer Science, vol 11301, 545 – 554
- Zhao, G.C., Xu, L.J., Lin, S.B., et.al., 2023. An Easy-To-Update Pulse-Like Ground Motion Identification Method Based on Siamese Convolutional Neural Networks. *Journal of Earthquake Engineering*,

28(1): 1-19. doi: 10.1080/13632469.2023.2199433

Zhou, B.F., Song, Q., Ren, Y.F., et.al., 2023. The Study on the Influence and Application of Foundation Pier Height on Strong Motion Records. *Earth Science*, 49(02): 414-424 (in Chinese with English abstract).

Zhou, B.F., Yu, H.Y., Wen, R.Z., et.al., 2017. Preliminary study on data quality in strong motion records. *Seismological and Geomagnetic Observation and Research*, 38(1): 69-75 (in Chinese with English abstract).

Zhu, J.B., Liu, H.Y., Luan, S.C., et.al., 2025. Prediction of On-Site Peak Ground Motion Based on Machine Learning and Transfer Learning. *Earth Science*, 50(5): 1842-1860 (in Chinese with English abstract).

附中文参考文献

李山有, 金星, 刘启方, 等, 2003. 中国强震动观测展望. 地震工程与工程振动, (02):1-7.
李渝生, 黄超, 易树健, 等, 2017. 九寨沟 7.0 级地震的地震断裂及震源破裂的构造动力学机理研究. 工程地质学报, 25(04): 1141-1150.
李春果, 王宏伟, 温瑞智, 等, 2025. 融合谱元法与人工神

经网络的漾濞地震宽频带地震动模拟. 地球科学, 1-1

4.

宋晋东, 朱景宝, 韦永祥, 等, 2023. 2022 年 1 月 8 日青海门源 6.9 级地震机器学习地震预警震级估计与现地阈值报警的回溯验证. 地球物理学报, 66(07): 2903-2919.

姚鑫鑫, 2023. 我国强震动记录精细化数据处理及 Flatfile 参数研究(博士学位论文). 哈尔滨: 中国地震局工程力学研究所.

姚鑫鑫, 任叶飞, 岸田忠大, 等, 2022. 强震动记录的数据处理流程: 去噪滤波. 工程力学, 39 (S1): 320-329.

中国地震局, 2018. 地震台站建设规范 强震动台站(DB/T 17-2018). 北京: 地震出版社.

周宝峰, 宋泉, 任叶飞, 等, 2023. 基墩高度对强震动记录的影响及应用研究. 地球科学, 49(02): 414-424.

周宝峰, 于海英, 温瑞智, 等, 2017. 强震动记录数据质量初探. 地震地磁观测与研究, 38(1): 69-75.

朱景宝, 刘赫奕, 栾世成, 等, 2025. 基于机器学习和迁移学习的现地地震动峰值预测. 地球科学, 50(5): 1842-1860.

通讯作者: **周宝峰(1978—)**, 男, 博士, 研究员。主要从事强震观测和地震动特性研究(E-mail: zbf166@126.com, ORCID: <https://orcid.org/0009-0009-5541-0000>);

作者简介: **张越(1999—)**, 男, 博士研究生。主要从事强震观测和地震动特性研究(E-mail: ztfzy741214@163.com, ORCID: <https://orcid.org/0009-0009-6574-9681>);

郭文轩(1997—), 男, 博士研究生。主要从事强震观测和地震动特性研究(E-mail: gwx9781@163.com);

温瑞智(1968—), 男, 博士, 研究员。主要从事地震工程与防灾减灾研究(E-mail: ruzhi@iem.ac.cn)。