

蒋海昆,王锦红. 2023. 适用于机器学习的地震序列类型判定特征重要性讨论[J]. 地震研究, 46(2): 155–172, doi:10.20015/j.cnki.ISSN1000-0666.2023.0034.

Jiang H K, Wang J H. 2023. Discussion on the importance of the features for the judgement of earthquake sequence types applicable to machine learning [J]. *Journal of Seismological Research*, 46(2): 155–172, doi:10.20015/j.cnki.ISSN1000-0666.2023.0034.

适用于机器学习的地震序列类型判定特征重要性讨论^{*}

蒋海昆¹, 王锦红²

(1. 中国地震台网中心, 北京 100045; 2. 中国地震局地震预测研究所, 北京 100036)

摘要: 基于1970—2021年中国大陆及边邻地区地震目录、地震序列目录和历史地震震源机制资料, 参考以往研究和震后趋势预测实践, 构建基于地震观测数据的机器学习序列类型判定特征样本数据集。基于地震序列分类, 设置多震型、主余型、孤立型3类样本标签, 初步提出44个可用于机器学习地震序列类型判定的备选特征, 包括主震及震源机制相关参数、历史地震序列类型、序列衰减和G-R关系相关参数、震级及频次相关参数。以44个备选特征为基础, 变换震级下限、统计时段等参数, 可以扩充出更多的机器学习备选特征。基于特征与标签之间的关联特性, 评估特征对序列分类的重要性。宏观来看, 震级相关参数、G-R关系和序列衰减相关参数、历史地震序列类型、震源机制相关参数等特征对序列分类有贡献, 其中震级相关参数特征与标签之间的互信息值明显较大且排序稳定。补齐缺失特征不但能够增加可用的训练和检验样本, 还可明显提升特征与序列类型之间的关联性, 这意味着恰当的数据预处理在一定程度上有可能提高特征的序列分类能力。添加原始数据的交互特征是拓展可用特征数量的重要方式之一, 非独立特征经信息交互处理之后显示出与序列标签更强的关联性, 这意味着特征选择应以模型预测效能的综合评价结果为准, 不宜过分强调特征参数的独立性。

关键词: 地震序列类型; 机器学习; 特征; 互信息

中图分类号: P315.7 **文献标识码:** A **文章编号:** 1000-0666(2023)02-0155-18

doi:10.20015/j.cnki.ISSN1000-0666.2023.0034

0 引言

正确的震后趋势预测, 是破坏性地震发生后政府和社会最为关心的问题, 是地震应急、抗震救灾和恢复重建的重要决策依据之一。震后趋势预测的目的是强余震震级估计, 最主要的技术手段是地震序列类型判定。基于历史地震类比和序列参数计算的地震序列类型判定, 是当前广泛采用并且还将长期发挥作用的主要余震预测方法(陈立德等, 1992; 蒋海昆等, 2015)。已有研究显示, 不同参数对序列类型具有或多或少的分辨能力, 但从严格的统计检验结果来看, 单参数序列类型识别正确率并不十分理想(蒋海昆等,

2006a): 一是实际序列参数数值分布较为离散, 即使同一类型的地震序列, 也有较为离散的序列参数数值分布范围; 二是部分参数最优分类标准与主震震级、震后持续时间等因素有关; 三是单参数预测为主, 不同参数对同一个序列的判定结论经常出现矛盾, 因而序列类型判定较多地依赖于“统计+经验”的方法以及研究者的经验和能力。还有非常重要的一点是, 当前的序列类型判定方法对多震型地震和前震几乎没有序列类型甄别能力(蒋海昆, 周少辉, 2020), 而多震型和前震型地震又属于可能导致更严重灾害的地震序列类型。

机器学习(machine learning)是人工智能的重要组成部分, 近年来以机器学习和深度学习为代

^{*} 收稿日期: 2022-09-19.

基金项目: 地震动力学国家重点实验室开放基金(LED2022B05)。

第一作者简介: 蒋海昆(1964-), 研究员, 博士, 主要从事余震统计、余震机理及余震预测研究。

E-mail: jianghaikun@seis.ac.cn.

表的基于大数据的人工智能技术飞速发展,通过数据驱动的技术手段,在样本分类、指标提取、综合决策等方面提供了许多新的工具,在疾病诊断、图像识别、自然语言处理、精准服务、虚拟体验等实际应用场景取得许多突破。在地震预测领域,尤其是针对固定区域、固定时间窗内可能发生地震的震级预测方面,也已有许多探索,具体可参见 Mignan 和 Broccardo (2020)、Al Banna 等 (2020)、Mousavi 和 Beroza (2022) 以及王锦红和蒋海昆 (2023) 的综述。但目前专门针对序列类型判定及余震预测的机器学习研究尚不多见 (王锦红, 蒋海昆, 2023), 因而针对现有主要基于单参数、以及“统计+经验”为主的序列后续趋势预测模式存在的不足,同时考虑更多的影响因素 (特征), 采用人工智能等新技术对序列类型进行多因素的综合判定, 是研究者始终持续探索的一条途径 (韩渭宾等, 1993; 蔡立冬等, 1994; 周翠英等, 1996; 庄昆元等, 2001; 蒋海昆等, 2007a; 李冬梅等, 2013)。从另一个角度来看, 序列类型判定从技术上属于分类问题, 而“分类”正是机器学习的长项。

机器学习是利用计算机基于历史数据 (经验), 建立某种模型 (规律), 并据此预测未来的一种方法或手段 (赵志勇, 2018), 与人类思考和经验总结过程有些类似, 但它能顾及更多的情形, 执行更为复杂的计算。机器学习需要基于历史数据对模型进行训练, 训练的结果被用来对新数据进行预测, 这个结果即模型。训练与预测是机器学习的两个过程, 模型则是过程的中间输出结果。机器学习中的训练与预测可粗略对应于人类思维活动的归纳和推测, 这与震后趋势预测领域当前主流的“统计+经验”的预测模式有些类似。

有监督学习是最为常见的机器学习算法, 它使用标记良好的训练数据进行模型训练, 目的是找到一个映射函数来映射输入变量 (特征) 和输出变量 (标签) 之间的关系。所谓标签, 即样本的分类属性, 在地震序列类型判定中即是常用的多震型、主余型、孤立型等序列类型。所谓特征即用于机器学习的样本的特性, 以往地震序列类

型判定中用到的历史地震活动、序列衰减、 $G-R$ 关系等均属机器学习样本数据集中的特征类数据。监督学习包含样本数据集构建、数据集拆分、模型选择和训练、外推预测等 4 个主要步骤, 其中样本数据集构建是机器学习最重要的基础, 对诸如地震预测这一类机理不明、单项特征与标签之间关系不唯一的分类任务, 如何确定训练数据集的输入特征, 是机器学习数据准备的最重要工作。机器学习应用大多与大数据高度关联, 各种不同模型或算法在输入的数据量达到一定量级后, 都有相近的准确度^①, 这也是机器学习领域推崇“数据为王”的重要原因^②。样本数量、数据质量、对实际场景的覆盖程度以及特征与标签之间的关联性, 都会影响到学习和训练的效果 (王东, 2021)。

综上, 作为机器学习模型训练的基础, 本文将基于 1970—2021 年中国大陆地震目录和历史地震、震源机制等资料, 针对序列类型判定这一目的, 构建样本数据集; 讨论数据集备选特征与标签之间的统计关联特性, 给出推荐的样本数据特征参数集合, 构建适用于不同应用场景、具有较高预测准确性及泛化能力的序列分类模型。

1 地震数据及样本标签

1.1 地震数据

依据中国地震台网统一地震目录^③, 1970—2021 年中国大陆及边邻地区 (国界线外扩 10 km) 共记录到 $M_s \geq 5.0$ 地震 1 336 次, 依据余震破裂范围 (Wells, Coppersmith, 1994) 及余震活动持续时间 (Lolli, Gasperini, 2003) 甄别并删除其中的余震。删除余震后有 $M_s \geq 5.0$ 地震 902 次, 其中 5.0~5.9 级 722 次、6.0~6.9 级 153 次、7.0~7.9 级 25 次、8.0 级以上地震 2 次。以 0.5 级为间隔的地震频次统计结果如表 1 第 II 行所列。地震主要分布于南北地震带、青藏地块、新疆天山地震带和西昆仑地震带, 大陆东部地震主要分布于华北、东北地区。基于中国地震台网统一地震目

① <https://www.cnblogs.com/subconscious/p/4107357.html>.

② <https://hellofuture.orange.com/en/towards-a-less-data-and-energy-intensive-ai/>.

③ <http://data.earthquake.cn/data/>.

录,采用时-空距离方法 (Wells, Coppersmith, 1994; Lolli, Gasperim, 2003) 构建 $M_s \geq 5.0$ 地震

的序列目录,其中204个川滇及附近区域地震的序列目录直接采用赵小艳等人工整理的序列目录^①。

表1 1970—2021年中国大陆及边邻地区 $M_s \geq 5.0$ 地震基本概况

Tab. 1 Statistics of the $M_s \geq 5.0$ earthquakes in Chinese mainland and neighboring areas from 1970 to 2021

样本分类	$M_s 5.0 \sim 5.4$	$M_s 5.5 \sim 5.9$	$M_s 6.0 \sim 6.4$	$M_s 6.5 \sim 6.9$	$M_s 7.0 \sim 7.4$	$M_s 7.5 \sim 7.9$	$M_s \geq 8.0$	合计	比例
I 全部地震	837	266	147	55	25	4	2	1 336	—
II 删除余震后	547	175	107	46	21	4	2	902	—
III 多震型	66 (14)	22 (1)	16 (2)	4	4			112 (17)	15.5%
IV 主余型	234	86	57	33	15	4	2	431	59.8%
V 孤立型	108	36	26	8	—	—	—	178	24.7%
VI 无法确认	139	31	8	1	2 (深震)	—	—	181	—

注:多震型中括号内数值为前震型序列数量;比例计算不含181次无法确认序列类型的地震。

1.2 样本标签

针对地震序列类型判定的机器学习样本数据集,序列类型即是样本标签。国外通常把地震序列类型分为主余型 (mainshock - aftershock)、震群型 (swarm) 或多震型 (multiplets/multievents) (Utsu, 2002; Felzer *et al.*, 2004), 着重序列显著特征的定性描述,但缺乏定量的分类标准。我国由于余震预测的实际需要,从定量的角度以序列中最大地震释放的能量 E_{\max} 占全序列释放能量 E_{total} 之比 $R_E = E_{\max}/E_{\text{total}}$ 进行序列类型划分 (周惠兰等, 1980), 这种分类方式物理含义清晰,但序列活动未结束之前无法计算全序列释放的能量。另一种更为简洁的序列类型划分是基于序列最大与次大地震的震级差 $\Delta M = M_0 - M_1$ 来进行 (吴开统, 1971), 其中 M_0 、 M_1 分别为序列最大与次大地震的震级。上述两种序列分类方法在一定的简化假设条件下等价 (蒋海昆等, 2006b)。在实际序列跟踪及强余震预测中,由于难以明确判断后续类似大小地震究竟是两次 (双震型) 还是两次以上 (震群型), 因而一般把双震型、震群型合称为“多震型” (Felzer *et al.*, 2004)。因而,在我国实际地震序列类型判定工作中一般依据震级差 ΔM 将余震序列分为多震型 ($-0.6 < \Delta M < 0.6$)、主余型 ($0.6 < \Delta M < 2.4$)、孤立型 ($\Delta M > 2.4$) 3类 (蒋海昆等, 2006c, 2015)。考虑到与当前业务工作的衔接,本文服务于机器学习序列类型判定的样本数据集标签 (序列类型) 仍基于序列主震与序列后续最大地震震级差

$\Delta M = M_0 - M_1$ 确定。

需要说明的是,部分地震序列主震前短时间内震源区会有震级小于主震的地震发生,即前震活动。全球不同区域的研究结果显示大约 10% ~ 40% 的中强以上地震伴随有前震活动 (陈颢, 1980; Jones, Molnar, 1979; 朱传镇, 王琳瑛, 1989; Reasenber, 1999; 李振, 王辉, 2011; 薛艳等, 2012), 且前震检出率随地震监测能力的提升似有增加的趋势 (Trugman, Ross, 2019)。笼统而言,主震前短时间内震源区震级小于主震的地震均可称为前震 (蒋海昆, 周少辉, 2020), 但为与广泛使用的多震型序列 ($-0.6 < \Delta M < 0.6$) 相区分,可约定符合 $\Delta M \leq -0.6$ 的地震序列为前震序列。由于本文目的是为中强地震震后趋势判定构建机器学习样本数据集,目标地震 (序列主震) 均为 $M_0 \geq 5.0$ 地震,因而即使考虑前震,所涉地震序列中符合主震 $M_0 \geq 5.0$ 且震级差 $\Delta M \leq -0.6$ 的样本集也仅有 17 例。由于前震型、多震型均属后续地震危险性较大的序列类型 (前震型后续存在发生比主震更大地震的危险,多震型后续存在发生与主震同等大小地震的危险),加之本文前震型样本数量较少,因而合并前震序列与多震型为一类,暂且仍称为“多震型”序列。

据此,本文地震序列类型标签确定标准为: ①多震型: $\Delta M < 0.6$ (包含以往提及的双震型、震群型及前震序列); ②主余型: $0.6 \leq \Delta M \leq 2.4$; ③孤立型: $\Delta M > 2.4$ 。

基于地震序列资料计算主震与最大余震震级

① 赵小艳. 2022. 地震预测开放基金 (2021) 结题报告——“基于决策树的西南地区中强地震序列类型判定研究”。

差 $\Delta M = M_0 - M_1$, 依据上述地震序列标签确定标准, 确定表 1 第 II 行所列 902 个地震的样本标签, 其中 181 个样本由于余震区后续无余震记录而无法分辨序列类型, 这些无法分辨序列类型的地震主要分布于西藏、青藏交界、新藏交界、青新交界等地震监测能力较弱的区域, 还包括 1999 年 4 月 8 日、2002 年 6 月 29 日吉林汪清 2 次 7.2 级深源地震以及该区域的几次 6 级深源地震。对 721 个可以确定序列标签 (序列类型) 的样本, 分序列类型及不同震级区间的地震频次统计如表 1 第 III ~ V 行所列。总的来看多震型、主余型、孤立型分别约占 15.5%、59.8% 和 24.7%, 主余型和孤立型合计约占 84.5%, 与前人 78% ~ 87% 的统计结果 (吴开统等, 1990; 蒋海昆等, 2006; 苏有锦等, 2014) 基本一致。后续研究中将无余震记录从而无法确认序列类型的 181 个样本 (表 1 第 VI 行) 统一归并为“孤立型”, 之所以如此处理, 是因为它们都具有后续都无显著余震发生这一共同的特点。最终, 本文数据集样本特征标签分为多震型、主余型和孤立型 3 类, 3 类样本数据集 $G-R$ 关系 b 值分别为 (0.73 ± 0.030) 、 (0.77 ± 0.046) 和 (1.04 ± 0.075) (图 1), 可见多震型、主余型样本中不同震级地震比例基本一致, 孤立型样本中低震级地震明显较多。

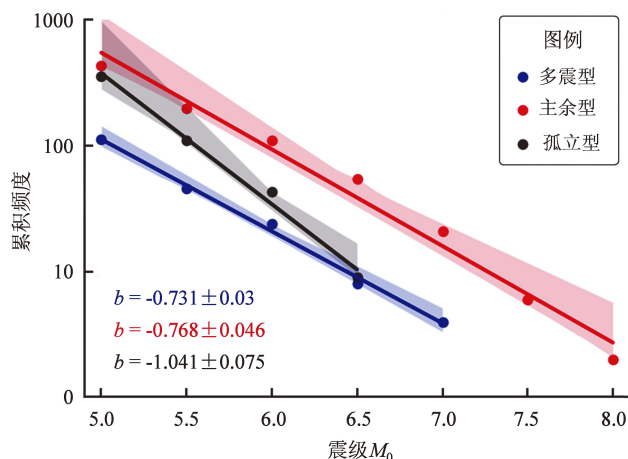


图 1 1970—2021 年中国大陆及边邻地区不同序列类型 $M_s \geq 5.0$ 地震 $G-R$ 关系 (阴影指示 95% 置信区间)

Fig. 1 $G-R$ relationships of the earthquakes with $M_s \geq 5.0$ in different types of sequences in Chinese mainland and the surrounding areas from 1970 to 2021

2 备选特征数据集构建

2.1 现有地震序列类型判定方法

研究显示, 构造和介质不均匀性以及应力水平对地震序列类型有重要影响 (Mogi, 1962; Takayuki, Hirata, 1987; Chen, Knopoff, 1987; Ben-Zion, James, 1993; Ban-Zion, Lyakhovsky *et al*, 2005; Somerville *et al*, 1999; 苏有锦等, 1999; Yamanaka, Kikuchi, 2004; Kanamori, Brodsky, 2004; 蒋海昆等, 2006b; Aochi, Ide, 2009; Marone, Richardson, 2016; 曲均浩等, 2015)。迄今为止尚无完全准确、普适的序列类型判定方法, 当前使用较多的主要有定性类比和定量计算两类。震后早期, 序列数据尚少, 一般只能基于构造及历史地震活动定性类比的方法来判断序列类型; 随着序列地震数据逐渐增多, 基于地震目录的序列参数计算结果被用于序列类型判定, 依据震后不同时段序列参数变化特征, 对序列类型进行定性 (变化趋势) 或定量 (参数统计指标) 的判定, 例如大森公式 p 值、 h 值, $G-R$ 关系 b 值, 归一化能量熵, 地震震级、频次和应变等参数的变化。

基于地震目录的统计地震学方法在当前地震序列类型判定中发挥着不可替代的作用, 而基于数字地震记录的研究结果也得到越来越多的应用。例如震源机制一致性和波形相似性方法就被广泛提及 (陈颢, 1980; Wiemer, Wyss, 2002; 王俊国, 刁桂苓, 2005), 但不同研究者得到的认识有一定差异 (崔子健等, 2012; 黄浩, 付虹, 2014)。由于应力降与震后断层面上的剩余应力水平有关, 因而应力降或视应力也被尝试用于序列类型判定 (陈学忠等, 2003; 钟羽云等, 2004; 秦嘉政等, 2005)、余震区应力水平估计及强余震预测 (杜迎春, 2000; Baltay *et al*, 2011; 王培玲等, 2013; 周少辉, 蒋海昆, 2017)。需要指出的是, 尽管基于地震波的方法因其能够直接获得震源或路径信息而受到重视, 但到目前为止, 要给出地震序列类型的定量判据仍十分困难, 难点首先在于从理论上无法给出不同类型序列“应该”具有的震源特征, 仍然只能采用震例统计的方式进行分析, 而这又由于已研究样本有限使得结果欠缺统计显著性 (Abercrombie, 1995; Ide, Berzoz, 2001; Allman, Shearer, 2009), 加之中小地

震应力降或视应力等震源参数随震级变化 (Dysart *et al*, 1988; Trifu, Radulian, 1989; 吴忠良等, 1999; 赵翠萍等, 2011; 华卫等, 2012; 周少辉, 蒋海昆, 2017), 更是带来了诸多的不确定性。因而, 本文备选特征未包含基于数字地震记录计算

的这些震源或介质参数。

2.2 备选特征

参考现有地震序列类型判定参数和方法, 用于机器学习地震序列类型判定的备选特征见表2, 主要依据参数来源细分为8类44个备选特征。

表2 机器学习序列类型判定备选特征列表

Tab. 2 List of alternative features for judgement of the earthquake sequence types by machine learning

特征序号	参数分类	符号表达	物理描述
1	(1) 主震相关参数	1La	主震纬度/(°)
2		2Lo	主震经度/(°)
3		3Ms	主震震级 M_s
4	(2.1) 主震震源机制 相关参数	4StrA	节面 A 方位角/(°)
5		5DipA	节面 A 倾角/(°)
6		6 RakeA	节面 A 滑动角的绝对值/(°)
7		7StrB	节面 B 方位角/(°)
8		8DipB	节面 B 倾角/(°)
9		9 RakeB	节面 B 滑动角的绝对值/(°)
10		10AziP	P 轴方位角/(°)
11	(2.2) 主震附近区域平均 应力场相关参数	11DipP	P 轴倾角/(°)
12		12MeanAziP	主震附近区域 P 轴方位角平均值/(°)
13		13StdAziP	主震附近区域 P 轴方位角标准差/(°)
14		14MeanDipP	主震附近区域 P 轴倾角平均值/(°)
15	(2.3) 主震应力场相对于 附近区域平均应力场的偏差	15StdDipP	主震附近区域 P 轴倾角标准差/(°)
16		16DiffAziP	主震 P 轴方位角与附近区域 P 轴平均方位角之差的绝对值/(°)
17		17DiffAziP/StdAziP	主震 P 轴方位角与附近区域 P 轴平均方位角之差的绝对值/ P 轴方位角标准差/(°)
18		18DiffDipP	主震 P 轴倾角与附近区域 P 轴平均倾角之差的绝对值/(°)
19	(3) 主震附近区域 $M_s \geq x$ 历史地震序列类型占比	19DiffDipP/StdDipP	主震 P 轴倾角与附近区域平均倾角之差的绝对值/ P 轴倾角标准差/(°)
20		20MxTy1	多震型序列 MMT 占比
21		21MxTy2	主余型序列 MAT 占比
22	(4) 指定时段序列衰减相 关参数	22MxTy3	孤立型序列 IET 占比
23		23pVal	修改的大森公式 p 值
24		24AveDeltaN	实际地震频次与修改的大森公式理论频次之差绝对值的平均值
25		25StdDeltaN	实际与理论频次之差绝对值的标准差
26		26k_Mcon	折合震级线性衰减速率
27		27b_Mcon	折合震级线性衰减纵轴截距 (M_L)
28		28AveDeltaMcon	实际折合震级与线性衰减理论折合震级之差绝对值的平均值
29	(5) 指定时段 $G-R$ 关系 相关参数	29StdDeltaMcon	实际与理论折合震级之差绝对值的标准差
30		30bVal	MLE 方法 b 值
31		31bSD	MLE 方法 b 值标准差
32		32aVal	$G-R$ 关系 a 值
33		33aSD	$G-R$ 关系 a 值标准差
34		34MsPre	$G-R$ 外推震级 (M_s)
35		35 M ₀ - MsPre	主震与 $G-R$ 外推震级之差的绝对值

续表 2

特征序号	参数分类	符号表达	物理描述
36	(6) 指定时段归一化能量熵	36Entropy	指定时段序列归一化能量熵
37	(7) 指定时段最大余震	37MsASK	最大余震震级 (M_s)
38	震级相关参数	38 DeltaMsASK	主震与最大余震震级差的绝对值
39	(8) 指定时段 $M_L \geq x$ 小震	39CumN_MLx	累积频次
40		40AveN_MLx	日平均频次
41		41NorN_MLx	相对震后第 1 天的归一化频次
42		42MsAveMLx	平均震级 (M_L)
43	频次及震级相关参数	43MsStdMLx	平均震级标准差 (M_L)
44		44MconMs3.0	$M_s \geq 3.0$ 地震折合震级 (M_L)

(1) 主震相关参数 (表 2 特征 1~3)

余震活动强度与主震大小定性正相关, 因而余震活动水平与主震大小直接关联 (Båth, 1965; Helmstetter, Sornette, 2003; Žalohar, 2014)。序列类型具有一定的区域特征 (刁守中等, 1995; 王华林等, 1997; 郭大庆等, 1998; 蒋海昆等, 2006b), 结合区域及发震构造特征的历史地震活动类比, 可在一定程度提供序列类型判定的参考依据。因而, 开展序列类型判定备选特征选择应考虑主震空间位置即经、纬度参数的影响。

(2) 主震震源机制相关参数及相对于附近区域平均应力场的偏差 (表 2 特征 4~19)

地震序列类型与构造运动或主震破裂形式有一定关系 (陈颙, 1980; 秦保燕, 刘武英, 1992; Reasenber, 1999; 苏有锦等, 1999; 蒋海昆等, 2006b; 张国民等, 2010), 因而主震震源机制相关参数 (表 2 特征 4~11) 被纳入作为机器学习的备选特征。考虑导致地震破裂的应力场特征, 主震附近区域平均 P 轴方位角、倾角及其标准差 (表 2 特征 12~15), 主震 P 轴方位角和倾角相对于区域平均结果的偏差及离散程度 (表 2 特征 16~19) 也一并纳入作为机器学习的备选特征。此处的“平均结果”来源于主震附近区域以往地震相关参数的统计, “附近区域”是以主震为圆心、以 R 为半径的圆域, R 与震级 M_w 相关, 可粗略地认为等同于主震破裂尺度 (Wells, Coppersmith, 1994):

$$R = 10^{\frac{M_w - 5.08}{1.16}} + 10 \quad (1)$$

式 (1) 右侧加 10 为考虑定位误差而人为增加的一个常量 (单位: km)。震级标度 M_s 与 M_w 之间

采用下式进行粗略转换 (Giacomo *et al.*, 2015):

$$M_w = e^{(-0.222 + 0.233M_s)} + 2.863 \quad (2)$$

(3) 主震附近区域历史地震序列类型相关参数 (表 2 特征 20~22)

基于与历史地震序列类型的定性类比, 是当前最主要的序列类型判定手段之一 (蒋海昆等, 2015), 表 2 第 20~22 行分别为主震附近半径为 R 的圆域范围内、震级 $M_s \geq x$ 地震序列中, 多震型、主余型及孤立型所占的比例。与主震震级相关的 R 由式 (1) (2) 计算得到, 历史地震序列类型基础数据来源于当前正在实时运行并准实时更新的 CAAFs 系统基础数据 (刘珠妹等, 2019)。

(4) 序列衰减相关参数 (表 2 特征 23~29)

序列衰减是余震活动的最主要特征, 修改的大森公式 $n(t) = K(t+c)^{-p}$ 是对余震衰减的最经典描述, 衰减系数 p 分布在 0.6~2.5 之间, 均值为 1.1 (Utsu *et al.*, 1995; Freed, Lin, 2001; Scholz, 2002; Lyakhovsky *et al.*, 2005; 贾若, 蒋海昆, 2014)。 p 值与区域地壳介质状况有关 (Creamer, Kisslinger, 1993; Rabinowitz, Steinberg, 1998; Jones, Craven, 1990; 曲均浩等, 2015), 余震衰减还受控于应力状态 (Narteau, 2009)。在修改的大森公式基础上, 刘正荣等 (1979)、刘正荣和孔绍麟 (1986) 提出的 h 值方法, 在余震跟踪预测中发挥着重要作用。因而, 本文为机器学习序列类型判定而构建的备选特征数据集中包含了大森公式相关参数的计算结果 (表 2 特征 23~29)。其中特征 23 为大森公式衰减系数 p 值; 特征 24、25 分别为指定时段基于修改的大森公式计算的完备震级以上的理论地震频次与实际地震频次之差的

绝对值及标准差,两者共同表征了修改的大森公式与实际地震频次变化的贴合程度;特征26为指定时段单位时间折合震级的线性衰减速率(假定为线性衰减),特征27为折合震级线性衰减纵轴截距(M_L),特征28、29分别为实际折合震级与线性衰减理论折合震级之差绝对值的平均值及标准差,三者间接表征了序列应变能释放时间衰减特征,以及理论预期相对于实际的偏差。

(5) G-R关系相关参数(表2特征30~35)

完整地震序列的震级-频度关系遵循G-R关系 $\log N = a - bM$,比例系数 b 值介于0.6~1.1之间,与构造及背景应力状态有关(Utsu, 2002)。G-R关系相关参数是机器学习地震预测研究中使用最为广泛的一类参数(王锦红,蒋海昆, 2023)。表2特征30、31为MLE方法计算的G-R关系比例系数 b 值及标准差;特征32、33为G-R关系比例系数 a 值及标准差(Gulia, Wiemer, 2019);特征34为基于G-R关系外推的最大余震震级,即依据 b 值截距方法(国家地震局科技监测司, 1990;刘正荣, 1995; Shcherbakov *et al*, 2013)估计的最大余震震级,这在当前最大余震震级预测中发挥着重要作用。已有研究显示,震后足够长时间之后,序列G-R关系外推最大余震震级逐渐趋近于真实的最大余震震级(苏有锦等, 2014);特征35为序列主震震级与G-R关系外推最大余震震级之差的绝对值。

(6) 归一化能量熵(表2特征36)

归一化能量熵是描述地震序列能量分配均匀程度的一个统计量,反映了序列地震能量分配均匀程度随时间的变化(朱传镇,王琳瑛, 1989;王琳瑛,舒曦, 1997)。余震序列性质判定单参数判据的统计研究显示,归一化能量熵具有相对较强的序列分类能力(蒋海昆等, 2006a)。

(7) 序列地震震级相关参数(表2特征37、38)

余震活动强度与主震大小定性正相关,早期一般认为主震与最大余震震级差 ΔM 是一个与主震震级无关的常数(平均约为1.2;Bath, 1965)。进一步的研究显示,Bath定律并不严格适合所有余震序列,实际 ΔM 介于0~3之间,受震源机制、震源深度、序列类型、区域构造特征、断层之间相互作用等多种因素的影响(Kisslinger, Jones, 1991; Helmstetter, Sornette, 2003; 蒋海昆等,

2006c, 2015; 苏有锦等, 2014; Žalohar, 2014; Rodríguez - Pérez, Zúñiga, 2016; Apostol, 2021),这意味着同等强度主震的余震活动水平可以明显不同。在实际预测应用方面,Shcherbakov等(2013)、Shcherbakov和Turcotte(2004)提出改进的Bath定律并引入推定最大余震震级对最大余震震级进行估算;也有研究利用主震震级、余震分布尺度等参数之间的统计关系(蒋海昆等, 2007c; 吕晓健等, 2010)、震级差(刘蒲雄等, 1996)等来对最大余震震级进行估计。基于以上研究,表2特征37、38分别为指定时段最大余震震级、序列主震与该最大余震震级差的绝对值。随震后时间的推移,特征37将趋近于序列最大余震,特征38会趋近于序列主震与序列最大余震震级差 ΔM ,而 ΔM 是序列标签的确定依据。

(8) 序列地震频次相关参数(表2特征39~44)

已有研究显示,震后不同时段小震频次(王志东等, 1982; 陈荣华等, 1994)及应变释放(戴英华等, 1990)特征一定程度上含有序列的分类信息。据此,表2特征39~41分别为震后指定时段 $M_L \geq x$ 小震的累积频次、日均频次及相对于震后首日的归一化频次;特征42~44分别为 $M_L \geq x$ 小震的平均震级、平均震级标准差和折合震级,两者分别从小震频次和震级/应变释放的角度,反映指定时段余震活动的平均水平以及震级分布的离散程度。

3 基于互信息的特征重要性评估

3.1 特征数据概况

根据前人提出的序列类型判定方法或参数,本文分8类列出44个备选特征,自特征23之后所有基于地震序列目录计算的参数中,指定时段可包含一系列震后不同的统计时段,特征20~22中 x 亦可取不同的震级下限,如此可得到多个组合备选特征,分别对应震后不同时段震中附近区域 $M_s \geq x$ 地震样本中,多震型、主余型及孤立型所占的比例。同样,特征39~43中涉及的震级下限 x ,在保证数据基本完备的情况下也可取如 $M_L 3.0$ 、 $M_L 3.5$ 等多个数值。因而,以上述44个备选特征为基础,可以扩充出更多的机器学习备选特征。具体到本文,针对表1中所列902次

$M_s \geq 5.0$ 地震样本, 以震后 3 天资料为例, 计算表 2 所列 44 个参数, 构建包含 902 个样本的机器学习样本集, 每个样本包含 44 个备选特征。为简单起见, 特征 20~22 仅考虑 $x = M_s 5.0$ 的情形, 即仅使用震中附近区域 $M_s \geq 5.0$ 地震多震型、主余型及孤立型所占的比例; 特征 39~43 中仅考虑 $x = M_L 3.0$ 的情形。

44 个备选特征中, 特征 23 之后的序列参数计算要求一定的样本量, 加之由于地震监测能力的差

异, 部分早期地震序列以及发生在青藏高原监测能力较低区域的地震, 有些特征参数无法给出计算结果。由 44 个备选特征的数据缺失情况统计可见 (图 2), 主震参数及主震附近区域历史地震相关参数的完备性相对较高, 达 98%; 涉及主震震源机制的相关参数, 特征完备性接近 70%; 涉及统计时段震级差或震级相关参数的特征, 特征完备性接近 60%; 修改的大森公式及 G-R 关系相关参数, 特征完备性仅及 32%。图 2 中 108Lab2 为序列标签。

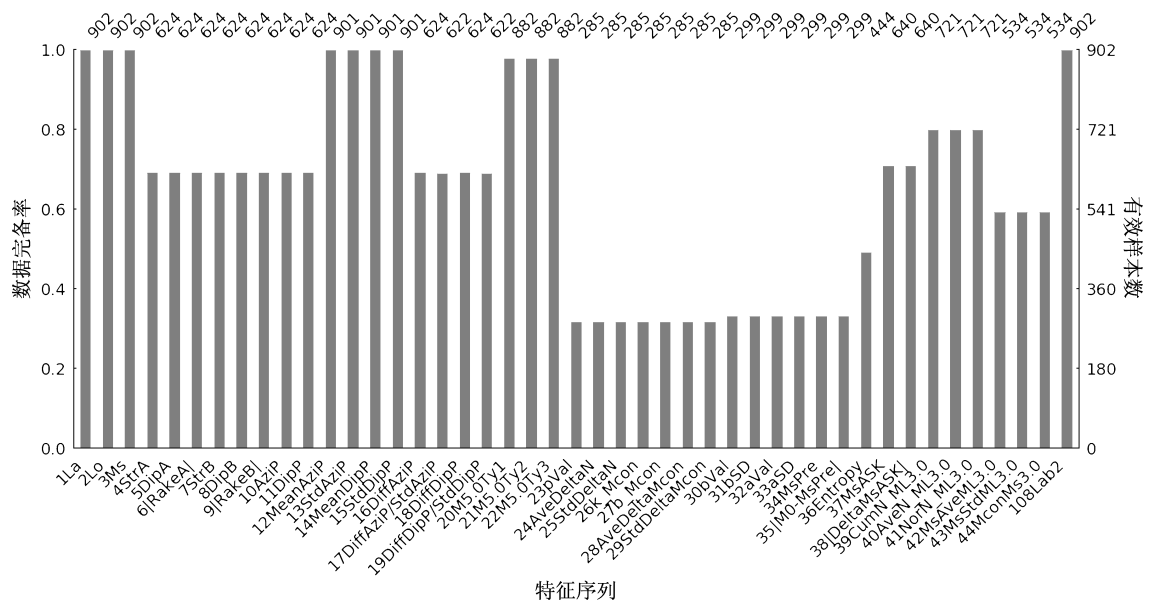


图 2 1970—2021 年中国大陆及近邻 $M_s \geq 5.0$ 地震样本集特征缺失情况统计

Fig. 2 Statistics of missing features in sample sets ($M_s \geq 5.0$ earthquakes in Chinese mainland and the surrounding areas from 1970 to 2021)

机器学习算法一方面期望有尽可能多的特征用于序列分类, 另一方面过高的特征维度又可能导致维度灾难的发生。对高维向量的降维处理, 不仅能显著降低运算和存储开销, 还有利于提高模型训练效率、提高分类效果 (杨秋良等, 2021)。特征降维的主要依据是特征重要性, 主要有特征选择和特征抽取两大类算法, 其中特征选择由于计算复杂度较低而被广泛采用 (刘健, 张维明, 2008)。尽管诸如决策树、随机森林等机器学习算法本身就包含有相关的特征重要性评估 (Breiman, 2001; 赵志勇, 2018), 但有研究表明, 机器学习模型默认的特征重要性更“偏好”连续型的类型变量, 因为连续型的类型变量在树节点上更容易找到切分点, 换言之之更容易过拟合从而得到“好”的训练模型 (Strobl et al., 2007)。

本文不涉及具体的机器学习模型, 仅从特征与分类标签之间相关性的角度, 通过互信息方法展示和讨论上述备选特征对地震序列分类的重要性, 为后续实际模型选择或训练提供参考。

3.2 互信息计算

在概率论和信息论中, 互信息可用于表征随机变量之间的相互依赖或相关性程度, 其基础是信息熵。信息熵是系统不确定性或复杂性的一种度量, 早在 20 世纪 40 年代即已提出。假设随机变量 $X = \{x_1, x_2, \dots, x_n\}$ 的概率分布为 $p(x)$, 则 X 的信息熵定义为:

$$H(X) = - \sum_{x \in X} p(x) \log p(x) \quad (3)$$

信息熵 $H(X)$ 越大, 信息量越大、系统越复杂。

若随机变量 X 和 $Y = \{y_1, y_2, \dots, y_n\}$ 服从联合分布 $P(X, Y)$, Y 的概率分布为 $p(y)$, 则 X 和 Y 的联合熵表示为 (Cai *et al*, 2018):

$$H(X, Y) = - \sum_{x \in X} \sum_{y \in Y} p(x, y) \log p(x, y) \quad (4)$$

式中: $H(X, Y)$ 表示多个随机变量信息量的总和。

给定 X 变量的条件下, Y 的不确定程度由条件熵 $H(X|Y)$ 表示^①:

$$H(Y|X) = - \sum_{x \in X} p(x) \sum_{y \in Y} p(x, y) \log p(x, y) \quad (5)$$

对于两个随机变量 X 和 Y , 互信息熵 (简称互信息 MI , Mutual Information) 定义为 (李欣倩等, 2022)^②:

$$I(X; Y) = H(Y) - H(Y|X) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \quad (6)$$

互信息值 $I(X; Y)$ 用于衡量变量之间的依存关系, 具体描述两组变量之间的信息共享及相互依赖程度 (徐洪峰, 孙振强, 2019)。换言之, MI 可用于确定 X 中所含 Y 的信息量, 或在 X 已知时确定 Y 所减少的不确定性 (李欣倩等, 2022)。在机器学习特征工程中, 常用 MI 来度量特征与标签之间的相关性强弱, 即讨论特征对标签分类 (categorical) 的重要性。 X 和 Y 完全无关或相互独立时 $I(X; Y) = 0$; 反之 $I(X; Y)$ 值越大, 表示该特征与样本标签之间的相关性越强 (周志华, 2016)^③。利用 scikit-learn 库^④基于 k 最近邻居距离方法计算互信息值 (Kraskov *et al*, 2004; Ross, 2014), 最近邻居数取缺省值 $k = 3$ 。

3.3 基于互信息的特征重要性评价及相关问题讨论

3.3.1 基于互信息的特征重要性

作为地震序列类型判定的依据, 特征与序列类型 (标签) 之间的相关性强弱, 是衡量特征重要性的重要指标。图3给出特征与类型之间的互信息计算结果, 从下往上为互信息值从大到小的排序, 表

征相关性由高到低, 亦可理解为各特征对序列分类重要性由大到小的排序。由于序列资料的限制, 许多样本有不同的特征缺失, 图3删除了所有有特征缺失的样本, 因而仅有285个样本参加计算。

图3a为未进行任何更进一步的数据预处理, 是针对最原始数据的互信息计算结果。从图3a总的来看, 各特征与序列类型之间相关性均不十分显著, 其中特征38 (主震与统计时段最大余震震级差的绝对值) 与序列类型之间具有最大的关联性, 互信息约为0.33; 震级相关参数 (特征38、37、44)、能量熵 (特征36)、历史地震序列类型 (特征20)、主震震级及 $G-R$ 关系外推震级相关参数 (特征35) 等特征互信息相对较高 ($MI > 0.1$)。除此之外, 其它诸如小震频次和震级相关参数 (特征43、42、28、41、25、40、39、24)、主震纬度 (特征1)、 $G-R$ 关系 b 值及相关参数 (特征27、34、31、33)、震源机制相关参数 (特征18、11、19、14) 等特征也有较弱的分类贡献率, 互信息大于0。

图3b为对不平衡样本数据进行了处理。不平衡样本是指样本数据集各类标签不均衡, 例如主余型样本就远多于多震型样本。通常多数类与少数类样本比例接近100:1时可认为属于不平衡样本^⑤, 不平衡样本对模型训练结果有一定影响。在样本总量受限的情况下, 对不平衡数据集的处理主要从数据本身和算法两方面入手。由于本文不涉及具体的机器学习算法, 因而主要从数据的角度进行不平衡数据处理。针对图3a数据, 采用随机重采样算法^⑥, 从少数类样本中对特征进行随机采样, 以组合构建新的 (伪) 样本, 从而达到各分类样本数均衡的目的。不平衡数据处理之后的互信息计算结果如图3b所示。从不平衡数据处理前、后的结果对比来看 (图3a、b之间的连线), 除个别特征 (如特征11、22、40、26、39) 重要性排序有较大差异外, 大多数特征的重要性排序及互信息值基本一致。这意味着, 对删除特征缺失样本后的数据集, 是否进行不平衡数据处理对特征分类重要性影响不大。

① https://blog.csdn.net/qq_40765537/article/details/114838485.

② <https://zhuanlan.zhihu.com/p/128091167>.

③ https://blog.csdn.net/weixin_46072771/article/details/106188753.

④ https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.mutual_info_classif.html.

⑤ <https://www.cnblogs.com/kamekin/p/9824294.html>.

⑥ <https://github.com/scikit-learn-contrib/imbalanced-learn>.

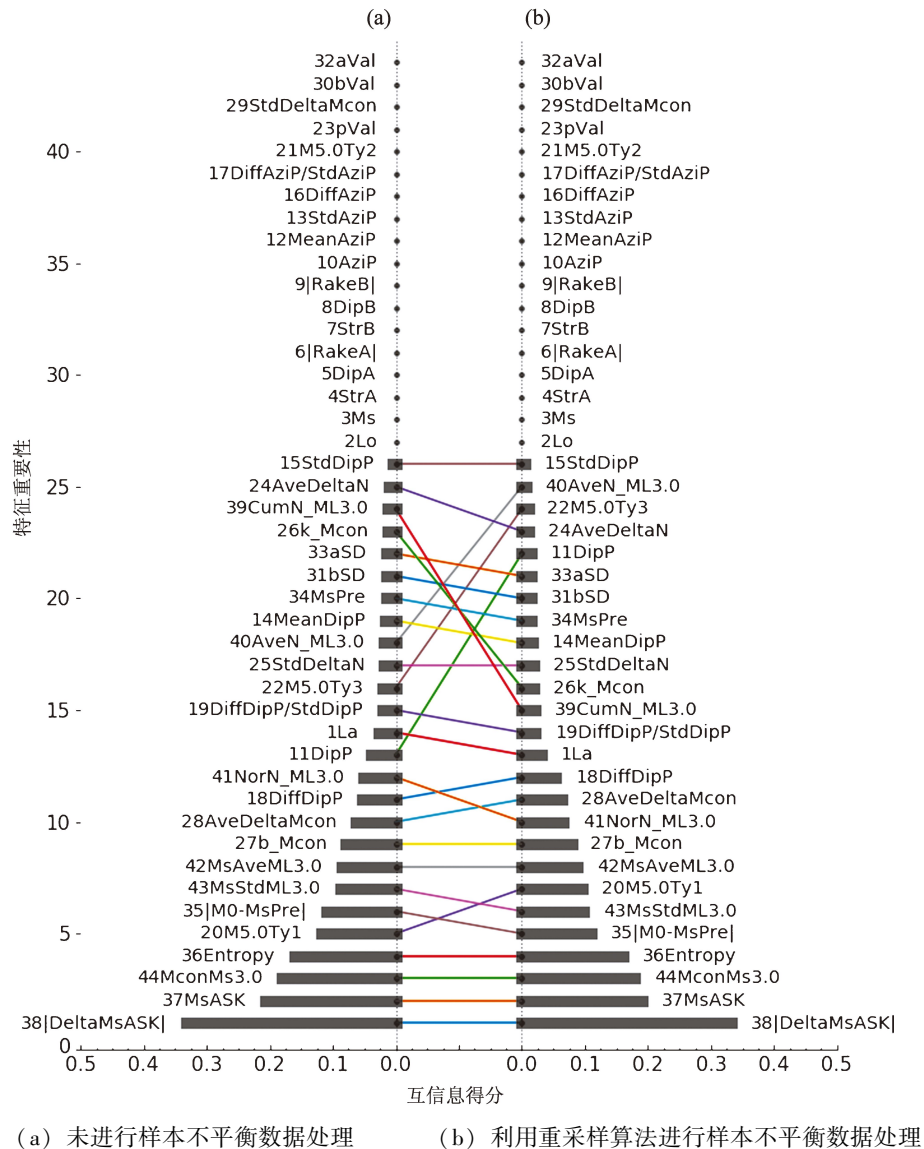


图3 基于互信息的机器学习特征重要性 (删除特征缺失样本)

Fig. 3 Sorts of feature importance based on mutual information
(for data set deleting the samples with missing feature)

3.3.2 缺失特征补齐对互信息计算结果影响

类似图3对含有缺失特征样本进行删除的数据处理方式,对大数据问题不会有明显影响,但对地震预测这一类小样本问题,会进一步加剧样本不足的矛盾。机器学习数据预处理算法中有许多缺失特征处理方式,图4为采用同类样本中位值对缺失特征进行补齐的结果。

具体做法是:对表2中每一个特征,首先分多震型、主余型、孤立型3类,计算该特征中位值,之后对该类样本中缺失该特征的样本,以该中位值进行补齐。例如对多震型样本的G-R关系 b 值(30 bVal),首先基于多震型样本

中无 b 值缺失的样本,计算 b 值的中位值,进而对有 b 值缺失的多震型样本,用该中位值进行补齐,对主余型、孤立型样本做类似处理。如此处理的好处是,所有样本都可参与互信息计算。并且从初步结果来看,缺失特征补齐与否,对特征与类型之间的相关性有较大影响,主要体现在4个方面:

(1) 缺失特征补齐之后绝大多数特征均显示与序列类型之间具有或多或少的相关性(图4),且各特征互信息值较图3所示的未进行缺失特征处理的结果有明显增大,例如特征38的互信息值就增大到0.49。

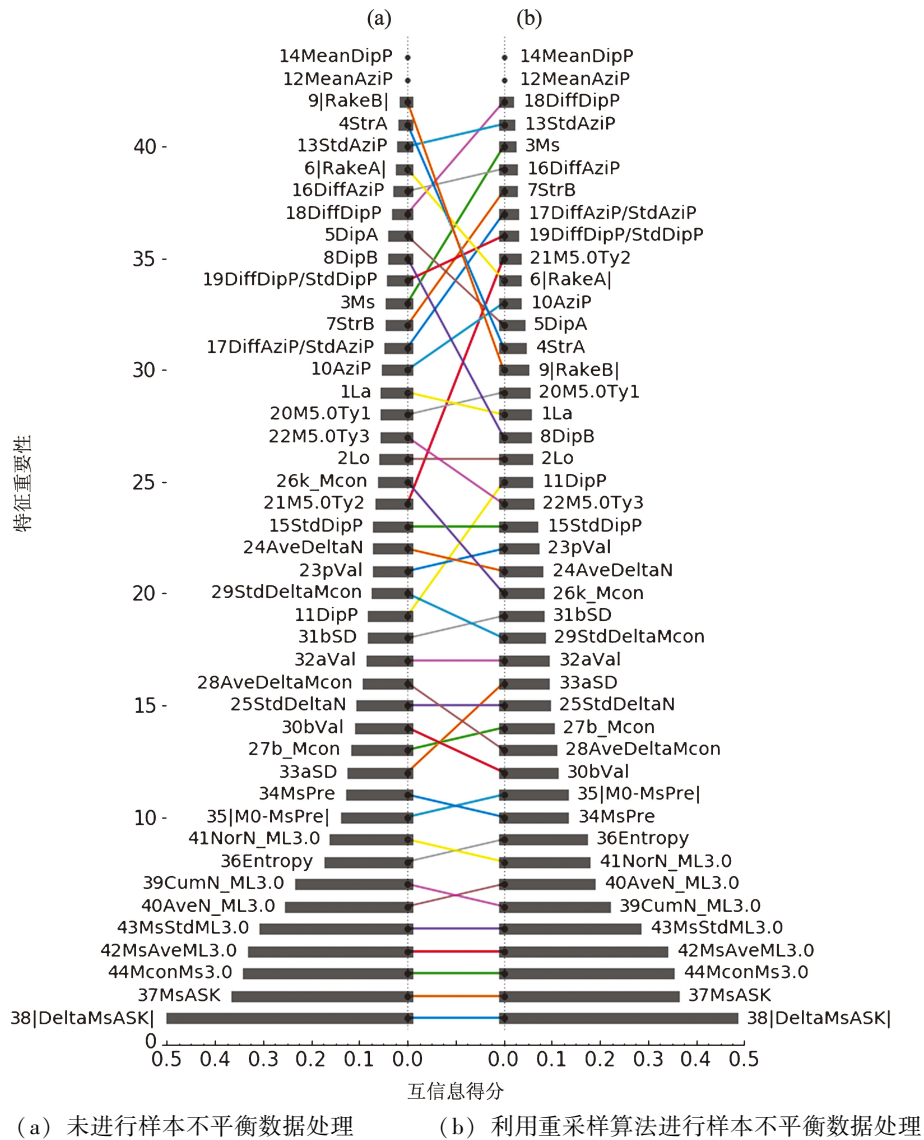


图4 基于互信息的机器学习特征重要性 (以中位值补齐缺失特征)

Fig. 4 Sorts of feature importance based on mutual information
(for data set filling the missing feature by the median value)

(2) 针对缺失特征补齐之后的样本数据, 是否进行不平衡数据处理, 对重要性较大特征的排序仍然影响不大。从图4具体来看, 对序列分类具有较大贡献的特征的重要性排序始终比较稳定, 关联性最好的特征 ($MI > 0.2$) 在不平衡数据处理前、后完全一致, 这些特征仍然以震级、频次相关参数为主 (特征 38、37、44、42、43、40、39); $MI > 0.1$ 的特征相关性排序也基本一致, 这些特征包括能量熵 (36)、主震及 G-R 关系外推震级相关参数 (35)、G-R 关系和序列衰减相关参数 (特征 34、33、27、30); 此外主震破裂形式和 P 轴方位相关参数 (特征 11、15、10)、历史地

震序列类型 (特征 21、22、23) 等也有一定的序列分类能力, 其 $MI > 0.05$ 。这种相对稳定的特征重要性排序, 对后续序列类型判定机器学习模型训练非常重要。

(3) 缺失特征补齐前后的特征重要性排序并不完全一致 (图5), 这意味着是否进行缺失特征补齐对特征重要性排序有影响。所幸的是, 对序列分类重要性最大的前几个特征排序几乎完全一致, 例如特征 38、37、44、43、42、36、35 等, 而对序列类型判定真正起“驾辕”作用的, 可能就是这些与序列类型相关性最强的、重要性排序比较稳定的特征。

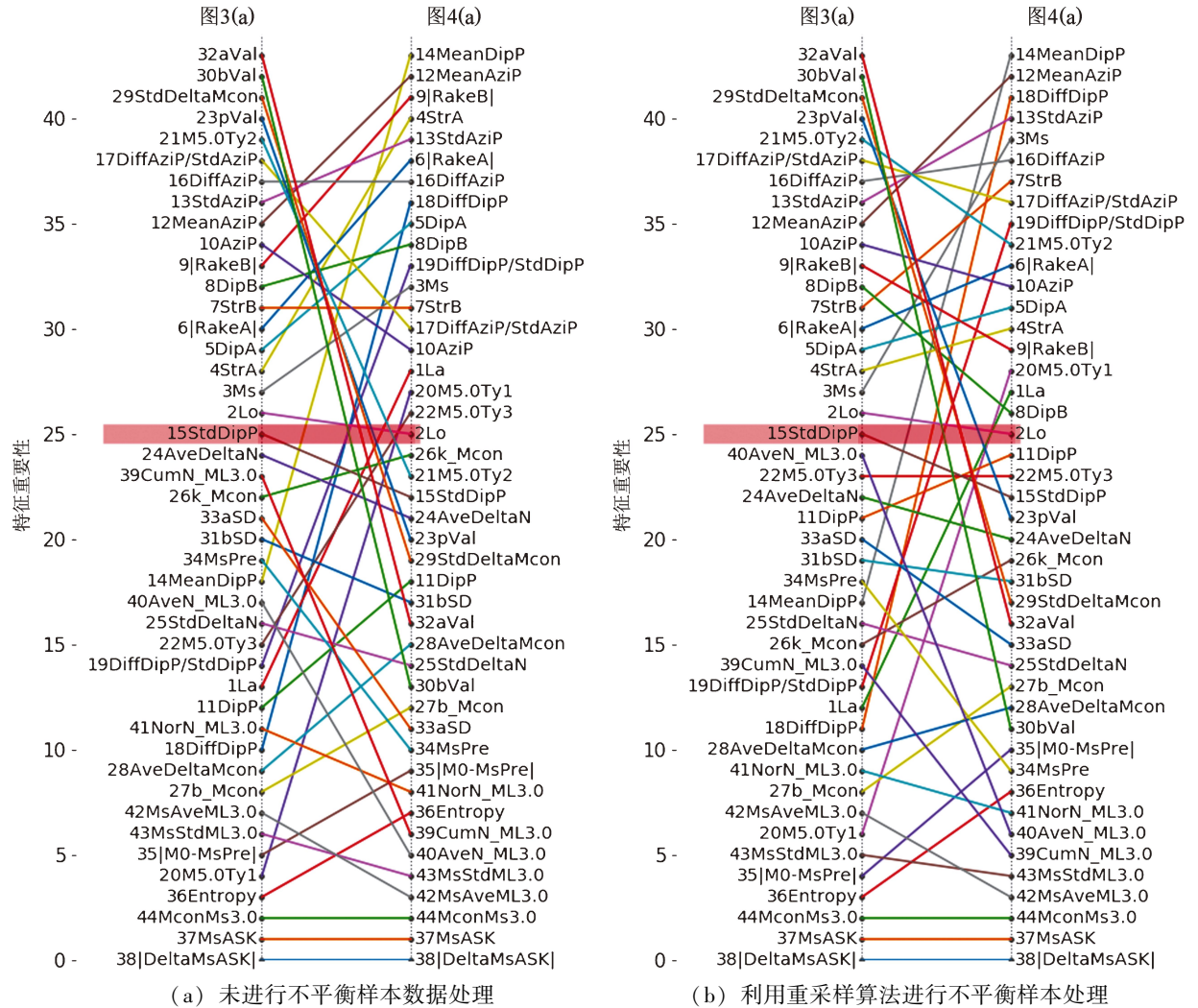


图5 缺失特征处理方式对特征重要性排序的影响

Fig. 5 Influence of the missing feature processing on feature importance sort

(4) 图5红色水平条上侧与图3上部互信息等于或接近于0的区域相对应, 这些特征未进行任何数据预处理之前对序列分类没有贡献。但由图5对比可见, 这些先前对序列分类没有贡献的特征, 在进行缺失值补齐处理之后, 部分特征、尤其是序列衰减相关参数(特征23、29)、序列G-R关系相关参数(特征30、32)的互信息值排序明显提前, 序列分类的贡献度明显提升; 而缺失值补齐处理前有一定序列分类能力的历史地震序列类型(特征20)、震源机制相关参数(特征18、19、14)等特征, 其重要性排序在缺失值补齐之后明显降低。这意味着, 恰当的数据处理方式在一定程度上会提高大多数特征的序列分类能力, 但同

时也可能导致部分特征的序列分类能力降低, 具体如何取舍需视最终模型需求而定。

3.3.3 信息交互对特征重要性的影响

理论上总是希望特征之间彼此独立以拓展尽可能宽泛的信息来源, 但实际上彼此完全独立的信息来源总是很少, 特征的独立性假设往往难以达成(李欣倩等, 2022)。但另一方面, 机器学习有一个重要的概念即特征间的信息交互(feature interaction)。已有研究显示, 很多特征单独来看可能与标签无关, 一旦组合在一起却与标签集明显相关(Jakulin, Bratko, 2003)。信息交互在诸如信息推送等人工智能应用中已被广泛采用^①。丰富特征表达、拓展可用特征数量的重要

① <https://zhuanlan.zhihu.com/p/384271606>.

方式之一即添加原始数据的交互特征 (interaction feature)，例如针对线性模型不仅可以学习偏移，还可以学习斜率^①。由表2各特征参数物理含义可见，特征之间或多或少具有物理上的关联性，这从特征参数之间的树状关系 (图6) 亦可看出。由图6中虚线可将特征粗略划分为具有一定相关性的4个部分，自左至右第I部分是序列G-R关系及序列衰减关系相关参数，第II部分是主震及震中附近震源机制相关参数，第III部分是主震及震中附近历史地震序列类型相关参数，第IV部分是序列地震震级及频次相关参数。实

际上，上述所有用于序列类型判定的备选特征，都来源于地震目录及震源机制，特征之间自然具有或多或少的相关性，但在特征组合之后仍显示出与序列标签更强的关联性。类似的特征组合方式，在机器学习地震预测中已有许多应用 (Reyes *et al*, 2013; Asencio - Cortes *et al*, 2016, 2018; Shodiq *et al*, 2017, 2018)。实际上，著名的G-R关系、修改的大森公式等都来源于地震目录，但它们又都提供了对地震序列震级-频度关系、地震序列衰减特征等的全新认识，某种程度上说，这就是信息交互的典型做法。

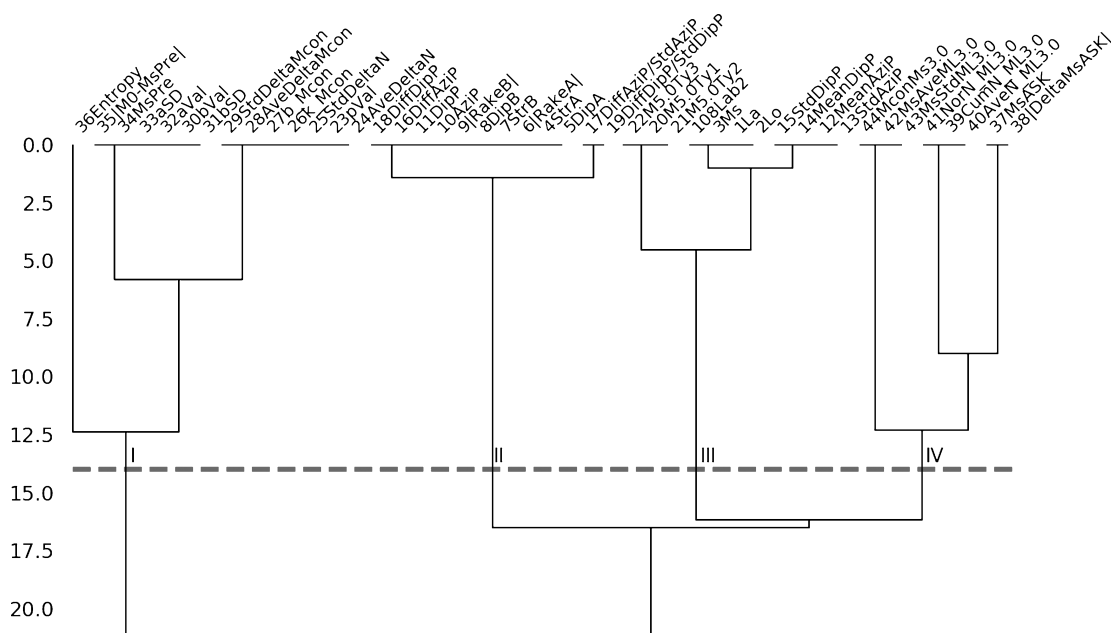


图6 特征参数关联性树状图^②

Fig. 6 Tree diagram of the dependency of the feature parameters

4 结论及讨论

基于1970—2021年中国大陆及边邻地区地震数据，参考以往研究和余震预测实践，以服务于震后序列类型判定为目的，构建基于地震观测数据的地震序列类型判定机器学习特征样本数据集。基于互信息方法，通过样本特征数据与标签之间的相关性，评估特征对序列分类的重要性，以利于构建适用于不同应用场景、具有较高预测准确

性及较强泛化能力的序列分类模型，得到以下主要认识：

(1) 初步提出可用于机器学习地震序列类型判定的备选特征。参考现有地震序列类型判定参数和方法，初步提出44个可用于机器学习地震序列类型判定的备选特征，这些特征包括主震三要素相关参数、主震及附近区域震源机制相关参数、主震附近区域历史地震序列类型、序列衰减相关参数、序列G-R关系相关参数、序列能量相关参数、序列地震震级及频次相关参数等8类。以震后

^① <https://blog.csdn.net/snail9610/article/details/105925305/>.

^② <https://github.com/ResidentMario/missingno>.

3 天观测数据为例, 构建了包含 902 个样本、每个样本包含 44 个备选特征的机器学习样本集。根据地震序列类型设置样本“标签”, 分为多震型、主余型、孤立型 3 类。基于对后续危险性的考虑, 将前震型序列合并至多震型。902 个样本中多震型、主余型、孤立型分别约占 15.5%、59.8% 和 24.7%, 其中主余型和孤立型合计为 84.5%, 与前人研究结果基本一致。

由于序列参数的计算要求一定的样本量, 加之部分监测能力较低区域地震记录不完备, 因而部分样本的部分特征参数 (如 $G-R$ 关系 b 值、序列衰减系数 p 值等) 无法计算。总的来看, 主震及主震附近区域历史地震相关参数完备性较高, 约为 98%, 主震震源机制相关参数完备性接近 70%, 序列地震震级相关参数完备性接近 60%, 修改的大森公式及 $G-R$ 关系等相关参数的完备性仅及 32%。

(2) 近 60% 的备选特征显示与序列类型之间具有一定的关联性。基于互信息方法分析特征与标签 (序列分类) 之间的相关性强弱, 讨论各特征对序列分类的重要性。在删除缺失特征样本的情况下, 有近 60% 的特征 (26 项) 与序列类型之间互信息大于 0, 但相关性均不太强, 最大互信息值仅为 0.33。这些特征中, 震级相关参数、能量相关参数、历史地震序列类型、主震震级及序列 $G-R$ 关系外推震级相关参数等特征的互信息相对较高 ($MI > 0.1$), 其它如小震频次相关参数、主震纬度、 $G-R$ 关系 b 值及相关参数、震源机制相关参数等特征也有相对弱的序列分类能力。

(3) 补齐缺失特征可明显提升特征的序列分类能力。缺失特征补齐的数据预处理方式, 不但可显著增加可用的模型训练和检验样本量, 更可以明显提升特征与序列分类之间的关联性。本文以中位值补齐样本缺失特征的情况下, 大多数特征 (42 项, 约占 95%) 显示出与序列类型之间具有或多或少的关联特性, 且各特征的互信息值明显增大, 最大达 0.49。

(4) 特征间之间的信息交互有利于提升特征的序列分类能力。信息交互是机器学习特征构建的重要手段。本文结果显示, 即使实际资料中特征的独立性假设难以达成, 但依据这些彼此并不完全独立的参数构建的特征, 仍显示一定的特征

重要性, 即具有一定的序列分类能力。换言之, 在机器学习数据处理过程中, 即使已明确知道彼此或多或少相关的特征, 也不宜一开始即完全删除。特征选择是一个复杂的过程, 尽管为节省内存和运算开支, 特征选取应在保留尽可能多分类信息的前提下仅保留尽可能少的特征 (姜文煊等, 2021), 但特征之间的信息交互对特征与标签之间关联性的提升作用亦须重视, 在当前机器学习仍属“黑箱”操作的前提下, 最终的特征选择应以模型预测效能的综合评价结果为准。

(5) 震级相关参数对序列类型判定的特征重要性相对较大。宏观来看, 震级相关参数、能量相关参数、 $G-R$ 关系和序列衰减相关参数、历史地震序列类型、震源机制相关参数等的特征重要性排序相对靠前, 对序列分类有贡献, 其中震级相关参数特征与标签之间的互信息明显较大且排序稳定。本文结果还显示一些有趣的现象, 例如主震空间位置尤其是主震纬度, 以及主震震源机制相对于区域应力场的偏差等参数, 对序列分类似乎也具有一定的贡献率, 前者实际上与此前序列类型具有一定区域特征这一认识 (王华林等, 1997; 蒋海昆等, 2006b) 相契合, 后者此前研究未见提及, 但这也提示我们, 通过大数据或人工智能所揭示的现象或特征, 有可能为地震序列分类物理本质的探寻提供思路或指引。

(6) 需重点关注机器学习算法的小尺度样本数据适用性问题。需要指出的是, 机器学习算法着重于大样本数据共性特征的提取和学习, 其能够发挥作用的基础是大数据。事实上已有研究关注到中尺度数据集 ($\sim 10K$ 尺度样本数) 的算法适用性问题 (Grinsztajn *et al*, 2022), 但与此相比, 用于地震预测的数据或学习样本可谓十分匮乏, 加之如本文所展示的那样, 不同的数据处理方式对结果还有复杂的影响, 因而尽管基于机器学习的地震预测已经展现出一定的应用前景 (Al Banna *et al*, 2020; Mignan, Broccardo, 2020; 王锦红, 蒋海昆, 2023), 但就余震预测而言, 机器学习是否比现有“经验+统计”的传统预测方法有更好的预测效率, 有待进一步的深入探讨, 各种机器学习算法对小尺度样本数据 (K 尺度样本数) 的模型适应性问题即是当前面临的重要问题之一。

参考文献：

- 蔡立冬, 宫家文, 甘俊人, 等. 1994. 地震序列类型预测的人工神经网络方法[J]. 地震研究, 17(1): 40–45.
- 陈立德, 蔡静观, 孙志民. 1992. 震后趋势早期判定的初步研究[J]. 地震研究, 15(4): 355–364.
- 陈荣华, 吴开统, 刘杰, 等. 1994. 不同地震序列类型的早期特征[J]. 地震, 14(1): 44–47.
- 陈学忠, 王小平, 王林瑛, 等. 2003. 地震视应力用于震后趋势快速判定的可能性[J]. 国际地震动态, (7): 1–4.
- 陈颢. 1980. 用震源机制一致性作为描述地震活动性的新参数[J]. 地球物理学报, 2(3): 39–47.
- 崔子健, 李志雄, 陈章立, 等. 2012. 判别小震群序列类型的新方法研究——谱振幅相关分析法[J]. 地球物理学报, 55(5): 1718–1724.
- 戴英华, 李钦祖, 王泽皋, 等. 1990. 地震现场综合地震学预报方法[J]. 地震, 10(1): 1–13.
- 刁守中, 王红卫, 华爱军. 1995. 中国大陆地区地震序列显著地震的时间分布特征[J]. 中国地震, 11(4): 315–326.
- 杜迎春. 2000. 1998年张北地震及其较大余震的应力降[J]. 华北地震科学, 18(2): 66–69.
- 郭大庆, 刘蒲雄, 袁一凡, 等. 1998. 地震现场工作大纲和技术指南[M]. 北京: 地震出版社.
- 国家地震局科技监测司. 1990. 地震学分析预报方法程序指南[M]. 北京: 地震出版社.
- 韩渭宾, 王虹, 曾健, 等. 1993. 中强以上地震的震后趋势早期综合判断方法的研究[J]. 地震学报, 15(1): 15–21.
- 华卫, 陈章立, 郑斯华, 等. 2012. 水库诱发地震与构造地震震源参数特征差异性研究——以龙滩水库为例[J]. 地球物理学进展, 27(3): 924–935.
- 黄浩, 付虹. 2014. 2008年以来滇西地区地震序列的谱振幅相关系数变化特征[J]. 地震学报, 36(4): 631–639.
- 贾若, 蒋海昆. 1994. 基于同震库仑应力变化的汶川余震序列频次研究[J]. 中国地震, 31(1): 74–90.
- 姜文煊, 段友祥, 孙歧峰. 2021. 基于交互信息的混合特征选择算法[J]. 应用科学学报, 39(4): 545–558.
- 蒋海昆, 代磊, 侯海峰, 等. 2006a. 余震序列性质判定单参数判据的统计研究[J]. 地震, 26(3): 17–25.
- 蒋海昆, 李永莉, 曲延军, 等. 2006b. 中国大陆中强地震序列类型的空间分布特征[J]. 地震学报, 28(4): 389–398.
- 蒋海昆, 曲延军, 李永莉, 等. 2006c. 中国大陆中强地震余震序列的部分统计特征[J]. 地球物理学报, 49(4): 1110–1117.
- 蒋海昆, 杨马陵, 付虹, 等. 2015. 震后趋势判定参考指南[M]. 北京: 地震出版社.
- 蒋海昆, 郑建常, 代磊, 等. 2007a. 中国大陆余震序列类型的综合判定[J]. 地震, 27(1): 17–25.
- 蒋海昆, 郑建常, 吴琼, 等. 2007b. 中国大陆中强以上地震余震分布尺度的统计特征[J]. 地震学报, 29(2): 151–164.
- 蒋海昆, 周少辉. 2020. 前震: 预测意义及识别方法[J]. 地震地磁观测与研究, 41(5): 223–225.
- 李冬梅, 周翠英, 朱成林, 等. 2013. 基于SVM的地震序列类型早期预测研究[J]. 地震研究, 36(1): 69–73.
- 李欣倩, 杨哲, 任佳. 2022. 基于互信息与层次聚类双重特征选择的改进朴素贝叶斯算法[J]. 模式识别与人工智能, 41(2): 36–69.
- 李振, 王辉. 2011. 前震序列时间空间统计特点[C]//中国地球物理学会. 中国地球物理学会第二十七届年会论文集. 北京: 中国地球物理学会, 353.
- 刘健, 张维明. 2008. 基于互信息的文本特征选择方法研究与改进[J]. 计算机工程与应用, 44(10): 135–137.
- 刘蒲雄, 陈修启, 吕晓健, 等. 1996. 地震序列的后续显著地震的预测研究[J]. 地震学报, 18(1): 27–33.
- 刘正荣, 孔绍麟. 1986. 地震频度衰减与地震预报[J]. 地震研究, 9(1): 6–8.
- 刘正荣, 钱兆霞, 王维清, 等. 1979. 前震的一个标志——地震频度的衰减[J]. 地震研究, 2(4): 1–9.
- 刘正荣. 1995. *b*值特征的研究[J]. 地震研究, 18(2): 168–173.
- 刘珠妹, 蒋海昆, 李盛乐, 等. 2019. 基于震例类比的震后趋势快速判定技术系统建设[J]. 中国地震, 34(4): 602–615.
- 吕晓健, 高孟潭, 陈丹. 2010. 全球大陆7级浅源大地震强余震震级和空间分布特征[J]. 地震, 30(3): 108–122.
- 秦保燕, 刘武英. 1992. 发震构造类型与震型预测[J]. 西北地震学报, 14(1): 29–36.
- 秦嘉政, 钱晓东, 叶建庆, 等. 2005. 2001年施甸 $M_s 5.9$ 地震序列的震源参数研究[J]. 地震学报, 27(3): 250–259.
- 曲均浩, 蒋海昆, 宋金, 等. 2015. 介质黏滞性质对余震活动影响的数值模拟[J]. 地震地质, 37(1): 53–67.
- 苏有锦, 李忠华, 赵小艳, 等. 2014. 全球7级以上地震序列研究[M]. 昆明: 云南大学出版社.
- 苏有锦, 刘祖荫, 蔡民军, 等. 1999. 云南地区强震分布的深部地球介质背景[J]. 地震学报, 21(3): 313–332.
- 王东. 2021. 机器学习导论[M]. 北京: 清华大学出版社.
- 王华林, 周翠英, 耿杰. 1997. 中国大陆及邻区地震序列类型的分区特征和震源环境讨论[J]. 地震, 17(1): 34–42.
- 王锦红, 蒋海昆. 2023. 基于地震数据的机器学习地震预测研究进展综述[J]. 地震研究, 46(2), doi: 173–187.
- 王俊国, 刁桂苓. 2005. 千岛岛弧大地震前哈佛大学矩心矩张量(CMT)解一致性的预测意义[J]. 地震学报, 27(2): 178–183.
- 王林瑛, 舒曦. 1997. 利用熵值及多分形方法研究地震序列类型的早期判定[C]//地震短临预报的理论与方法——“八五”攻关三级课题论文集. 北京: 地震出版社, 136–142.
- 王培玲, 姚家骏, 刘文邦, 等. 2013. 玉树地区两次强震序列应力降对比研究[J]. 内陆地震, 27(4): 295–302.
- 王志东, 焦远碧, 吴开统. 1982. 地震序列的持续时间与震级的关系[J]. 地震, 2(5): 34–39.
- 吴开统, 焦远碧, 吕培苓, 等. 1990. 地震序列概论[M]. 北京: 北京大学出版社.
- 吴开统. 1971. 地震序列的基本类型及其在地震预报中的应用[J]. 地震战线, 7(11): 45–51.
- 吴忠良, 陈运泰, Mozaffari P. 1999. 应力降的标度性质与震源谱高频衰减常数[J]. 地震学报, 21(5): 460–468.

- 徐洪峰,孙振强. 2019. 多标签学习中基于互信息的快速特征选择方法[J]. 计算机应用,39(10):2815–2821.
- 薛艳,刘杰,余怀忠,等. 2012. 2011 年日本本州东海岸附近 9.0 级地震活动特征[J]. 科学通报,57(8):634–640.
- 杨秋良,王钰,杨杏丽,等. 2021. 基于互信息 F 统计量特征选择技术的地基气象云图分类[J]. 计算机与现代化,(2):18–23.
- 张国民,钮凤林,邵志刚,等. 2010. 中国大陆 $M_s \geq 7.8$ 大震余震活动差异性特征及其成因研究[J]. 地震,30(4):1–12.
- 赵翠萍,陈章立,华卫,等. 2011. 中国大陆主要地震活动区中小地震震源参数研究[J]. 地球物理学报,54(6):1478–1489.
- 赵志勇. 2018. Python 机器学习算法[M]. 北京:电子工业出版社.
- 钟羽云,张帆,张震峰,等. 2004. 应用强震应力降和视应力进行震后趋势快速判定的可能性[J]. 防灾减灾工程学报,24(1):8–14.
- 周翠英,张宇霞,王红卫. 1996. 以模式识别方法提取地震序列早期判断的综合指标[J]. 地震学报,18(1):118–124.
- 周惠兰,房桂荣,章爱娣,等. 1980. 地震震型判断方法探讨[J]. 西北地震学报,2(2):45–59.
- 周少辉,蒋海昆. 2017. 景谷 6.6 级、鲁甸 6.5 级地震序列应力降变化对比研究[J]. 中国地震,33(1):23–37.
- 周志华. 2016. 机器学习[M]. 北京:清华大学出版社.
- 朱传镇,王琳璞. 1989. 震群信息熵异常与地震预报[C]//地震预报方法实用化研究文集(地震学专辑). 北京:学术书刊出版社,229–242.
- 庄昆元,王炜,章纯,等. 2001. 震后趋势决策支持系统 PTSS[J]. 西北地震学报,23(4):21–28.
- Abercrombie R E. 1995. Earthquake source scaling relationships from -1 to 5 using seismograms recorded at 2.5 km depth[J]. J Geophys Res,100(B12):24015–24036.
- Al Banna M H, Taher K A, Kaiser M S, *et al.* 2020. Application of artificial intelligence in predicting earthquakes: state-of-the-art and future challenges[J]. IEEE Access,8:192880–192923.
- Allman B P, Shearer P M. 2009. Global variations of stress drop for moderate to large earthquakes[J]. J Geophys Res,114(B1):B01310.
- Aochi H, Ide S. 2009. Complexity in earthquake sequences controlled by multiscale heterogeneity in fault fracture energy[J]. J Geophys Res,114(B3):B03305.
- Apostol B F. 2021. Correlations and Bath's law[J]. Results in Geophysical Sciences,doi:org/10.1016/j.rings.2021.100011.
- Asencio-Cortés G, Morales-Esteban A, Shang X, *et al.* 2018. Earthquake prediction in California using regression algorithms and cloud-based big data infrastructure[J]. Computers & Geosciences,115(5):198–210.
- Baltay A, Ide S, Prieto G, *et al.* 2011. Variability in earthquake stress drop and apparent stress[J]. Geophys Res Lett,38(6):L06303.
- Ben-Zion Y, James R R. 1993. Earthquake failure sequences along a cellular fault zone in a three-dimensional elastic solid containing asperity and nonasperity regions[J]. J Geophys Res, B8:14109–14131.
- Ben-Zion Y, Lyakhovsky V. 2006. Analysis of aftershocks in a lithospheric model with seismogenic zone governed by damage rheology[J]. Geophys J Int,165(1):197–210.
- Breiman L. 2001. random forests[J]. Machine Learning,45:5–32.
- Båth M. 1965. Lateral inhomogeneities in the upper mantle[J]. Tectonophysics,2(6):438–514.
- Cai J, Luo J W, Wang S L, *et al.* 2018. Feature selection in machine learning: a new perspective[J]. Neurocomputing,300(26):70–79.
- Chen Y T, Knopoff L. 1987. Simulation of earthquake sequences[J]. J Geophys Res,91(3):693–703.
- Creamer F H, Kisslinger C. 1993. The relation between temperature and the Omori decay parameter for aftershock sequences near Japan[J]. EOS74,43(S):417.
- Dysart P S, Snoke J A, Sacks I S. 1988. Source parameters and scaling relations for small earthquakes in the Matsushiro region, southwest Honshu, Japan[J]. Bull Seism Soc Am,78(2):571–589.
- Felzer K R, Abercrombie R E, Ekstrom G. 2004. A common origin for aftershocks, foreshocks, and multiplets. Bulletin of the Seismological Society of America[J]. 94(1):88–98.
- Freed A M, Lin J. 2001. Delayed triggering of the 1999 Hector Mine earthquake by viscoelastic stress transfer[J]. Nature,411(6834):180–183.
- Giacomo D D, Bondár I, Storchak D A, *et al.* 2015. ISC – GEM: Global instrumental earthquake catalogue (1900–2009), III. Re-computed M_s and m_b , proxy M_w , final magnitude composition and completeness assessment[J]. Physics of the Earth and Planetary Interiors,239(1B):33–47.
- Grinstajn L, Oyallon E, Varoquaux G. 2022. Why do tree-based models still outperform deep learning on tabular data[J]. NeurIPS arXiv:2207.08815v1[cs.LG]18 Jul 2022.
- Gulia L, Wiemer S. 2019. Real-time discrimination of earthquake foreshocks and aftershocks[J]. Nature,574(7777):193–199.
- Helmstetter A, Sornette D. 2003. Bath's law derived from the Gutenberg-Richter law and from aftershock properties[J]. Geophys Res Lett,30(20):1–4.
- Ide S, Beroz G C. 2001. Does apparent stress vary with earthquake size?[J]. Geophys Res Lett,28(17):3349–3352.
- Jakulin A, Bratko I. 2003. Analyzing attribute dependencies[J]. Lect Notes Artif Intell,2838:229–240.
- Jones A G, Craven J A. 1990. The North American central plains conductivity anomaly and its correlation with gravity, magnetism, seismic, and heat flow data in the province of Saskatchewan[J]. Phys Earth planet Inter,60(1–2):169–194.
- Jones L M, Molnar P. 1979. Some characteristics of foreshocks and their possible relationship to earthquake prediction and premonitory slip on fault[J]. J Geophys Res,84(B7):3596–3608.
- Kanamori H, Brodsky E E. 2004. The physics of earthquakes[J]. Rep Prog Phys,67(8):1429–1496.
- Kisslinger C, Jones L M. 1991. Properties of aftershock sequences in southern California[J]. J Geophys Res,96(B7):11947–11958.
- Kraskov A, Stogbauer H, Grassberger P. 2004. Estimating mutual information[J]. Phys Rev E,69:66138–66154.

- Lolli B, Gasperini P. 2003. Aftershocks hazard in Italy Part I: Estimation of time – magnitude distribution model parameters and computation of probabilities of occurrence [J]. *Journal of Seismology*, 7 (2): 235 – 257.
- Lyakhovsky V, Ben – Zion Y, Agnon A. 2005. A visco – elastic damage rheology and rate-and state-dependent friction [J]. *Geophys J Int*, 161 (1): 179 – 190.
- Marone C, Richardson E. 2016. Connections between fault roughness, dynamic weakening, and fault zone structure [J]. *Geology*, 44 (1): 79 – 80.
- Mignan A, Broccardo M. 2020. Neural network applications in earthquake prediction (1994 – 2019): Meta-analytic and statistical insights on their limitations [J]. *Seismological Research Letters*, 91 (4): 2330 – 2342.
- Mogi K. 1962. Magnitude – frequency relation for elastic shocks accompanying fractures of various materials and some related problems in earthquakes [J]. *Bull Earthq Res Inst*, 40: 831 – 853.
- Mousavi S M, Beroza G C. 2020. A machine – learning approach for earthquake magnitude estimation [J]. *Geophysical Research Letters*, 47 (1): e2019GL085976.
- Narteau C. 2009. Common dependence on stress for the two fundamental laws of statistical seismology [J]. *Nature*, 462 (3): 642 – 645.
- Rabinowitz N, Steinberg D M. 1998. Aftershock decay of three recent strong earthquakes in the Levant [J]. *BSSA*, 88 (6): 1580 – 1587.
- Reasenber P A. 1999. Foreshock occurrence before large earthquakes [J]. *J Geophys Res*, 104 (B3): 4755 – 4768.
- Reyes J, Morales – Esteban A, Martínez – Álvarez F. 2013. Neural networks to predict earthquakes in Chile [J]. *Applied Soft Computing*, 13 (2): 1314 – 1328.
- Rodríguez – Pérez Q, Zúñiga F R. 2016. Bath's law and its relation to the tectonic environment: A case study for earthquakes in Mexico [J]. *Tectonophysics*, 687: 66 – 77.
- Ross B C. 2014. Mutual information between discrete and continuous data sets [J]. *PLoS ONE*, 9 (2): e87357.
- Scholz C H. 2002. *The mechanics of earthquakes and faulting* [M]. New York: Cambridge Univ Press.
- Shcherbakov R, Goda K, Ivanian A. *et al.* 2013. Aftershock statistics of major subduction earthquakes [J]. *Bull Seism Soc Am*, 103 (6): 3222 – 3234.
- Shcherbakov R, Turcotte D L. 2004. A modified form of Bath's law [J]. *Bull Seism Soc Am*, 94 (5): 1968 – 1975.
- Shodiq M N, Kusuma D H, Rifqi M G. *et al.* 2017. Spatial analysis of magnitude distribution for earthquake prediction using neural network based on automatic clustering in Indonesia [C]//International Electronics Symposium on Knowledge Creation and Intelligent Computing (IESKCIC). IEEE, 246 – 251.
- Shodiq M N, Kusuma D H, Rifqi M G. *et al.* 2018. Neural network for earthquake prediction based on automatic clustering in Indonesia [J]. *International Journal on Informatics Visualization (JOIV)*, 2 (1): 37 – 43.
- Somerville P, Irikura K, Graves R. *et al.* 1999. Characterizing crustal earthquake slip models for the prediction of strong ground motion [J]. *Seismological Research Letters*, 70 (1): 59 – 80.
- Strobl C, Boulesteix A, Zeileis A. *et al.* 2007. Bias in random forest variable importance measures: Illustrations, sources and a solution [J]. *BMC Bioinformatics*, 8 (1): 25.
- Takayuki W, Hirata G A. 1987. Omori's Power law aftershock sequences of microfracturing in rock fracture experiment [J]. *J Geophys Res*, 92 (B7): 6215 – 6221.
- Trifu C I, Radulian M. 1989. Asperity distribution and percolation as fundamentals of an earthquake cycle [J]. *Phys Earth Planet Inter*, 58 (4): 277 – 288.
- Trugman D T, Ross Z. 2019. Pervasive foreshock activity across southern California [J]. *Geophysical Research Letters*, 46 (15): 8782 – 8781.
- Utsu T, Ogata Y, Matsuura R S. 1995. The Centenary of the Omori formula for a decay law of aftershock activity [J]. *J Phys Earth*, 43 (1): 1 – 33.
- Utsu T. 2002. Statistical features of seismicity [J]. *International Geophysics*, 81 (A): 719 – 731.
- Wells D L, Coppersmith K J. 1994. New empirical relationships among magnitude, rupture Length, rupture width, rupture area, and surface displacement [J]. *Bulletin of the Seismological Society of America*, 84 (4): 974 – 1002.
- Wiemer S, Wyss M. 2002. Mapping spatial variability of the frequency – magnitude distribution of earthquakes [J]. *Adv Geophys*, 45: 259 – 302.
- Yamanaka Y, Kikuchi M. 2004. Asperity map along the subduction zone in northeastern Japan inferred from regional seismic data [J]. *J Geophys Res*, 109: B07307.
- Žalohar J. 2014. Explaining the physical origin of Bath's law [J]. *Journal of Structural Geology*, 60 (B2): 30 – 45.

Discussion on the Importance of the Features for the Judgement of Earthquake Sequence Types Applicable to Machine Learning

JIANG Haikun¹, WANG Jinhong²

(1. *China Earthquake Networks Center, Beijing 100045, China*)

(2. *Institute of Earthquake Forecasting, China Earthquake Administration, Beijing 100036, China*)

Abstract

Based on the catalog and focal mechanism of earthquakes in Chinese mainland since 1970 and referring to the previous research and practice on estimation of aftershock activity tendency, a feature sample dataset for judgement of earthquake sequence types by machine learning has been constructed. Three labels—multiplet mainshocks type, mainshock-aftershock type, as well as isolated earthquake type—have been set up according to the earthquake sequences. Forty-four alternative features that can be used for machine learning for earthquake sequence type judgement have been proposed preliminarily, including mainshock and focal-mechanism-related parameters, historical earthquake sequence types, sequence decay and G-R relationship-related parameters, magnitude- and frequency-related parameters. Based on the 44 alternative features, more features can be expanded by different threshold magnitude or statistical period. Based on the mutual information between features and labels, the feature importance or contribution rate of feature parameters to sequence classification has been evaluated. In summary, the magnitude-related parameters, G-R relationship, sequence-decay-related parameters, historical earthquake sequence type, focal mechanism related parameters are contributory for sequence classification. Especially, the mutual information between magnitude-related parameters and labels are obviously large and the ranking is stable. Our results show that the complementing of missing features can not only increase the available samples for model training and testing, but also significantly improve the correlation between features and labels, which means that appropriate data pre-processing on features may improve the ability of sequence classification to a certain extent. Adding interactive features of original data is one of the important ways to expand the number of available features, the independent features show a stronger correlation with sequence labels after information interaction processing in this paper, reminding us that the feature selection should be based on the results of efficiency estimation of the final model, and the feature independence should not be overemphasized.

Keywords: earthquake sequence type; machine learning; feature; mutual information