第56卷第3期 2023年(总229期) 西北地质

NORTHWESTERN GEOLOGY

Vol. 56 No. 3 2023(Sum229)

DOI: 10.12401/j.nwg.2023084

深度随机森林和随机森林算法的滑坡易发性

评价对比——以汉中市略阳县为例

贾俊1,2,毛伊敏3,孟晓捷1,*,高波1,高满新1,武文英1

 (1. 自然资源部黄土地质灾害重点实验室,中国地质调查局西安地质调查中心,陕西西安 710119;2. 地质灾害防治与地质 环境保护国家重点实验室,成都理工大学,四川成都 610059;3. 韶关学院信息工程学院,广东韶关 512005)

摘 要:针对浅层的机器学习模型泛化能力低而导致其滑坡易发性评价模型预测精度不高的问题,笔者围绕陕西省汉中市略阳县城中心为研究区,采用深度随机森林构建区域地灾易发性评价模型来提升预测精度。依据略阳县滑坡成灾机理研究成果,选取坡度、相对高差、坡向、坡型、工程地质岩组、断裂距离、水系距离、公路铁路距离、植被覆盖等9个因子作为易发性评价指标;将研究区栅格单元按5m×5m进行划分并提取评价因子值,输入深度随机森林评价模型,从而获得研究区易发性评价图。依据评价结果略阳县地质灾害可划分为极高易发区、高易发区、中易发区、低易发区4个等级,面积所占比例分别为5.31%、22.97%、42.11%、29.61%,其划分结果与研究区内地质灾害实际发育情况吻合,合理反映研究区地灾分布的总体特征。深度随机森林的地质灾害易发性预测模型在ROC曲线下面积值(AUC)为91.2%,高于随机森林预测模型的86.3%, 表明该模型具有一定的合理性与可行性,可为区域滑坡易发性评价进一步提供新方法。 关键词:滑坡;略阳县;易发性评价;深度随机森林

中图分类号: P694 文献标志码: A 文章编号: 1009-6248(2023)03-0239-11

Comparison of Landslide Susceptibility Evaluation by Deep Random Forest and Random Forest Model: A Case Study of Lueyang County, Hanzhong City

JIA Jun^{1,2}, MAO Yimin³, MENG Xiaojie^{1, *}, GAO Bo¹, GAO Manxin¹, WU Wenying¹

Key Laboratory for Geo-hazards in Loess Area, MLR, Xi'an Center of China Geological Survey, Xi'an 710119, Shaanxi, China;
 State Key Laboratory of Geohazard Prevention and Geoenvironment Protection, Chengdu University of Technology, Chengdu 610059, Sichuan, China;
 School of Information Engineering, Shaoguan University, Shaoguan 512005, Guangdong, China)

Abstract: To address the problem of low prediction accuracy of landslide susceptibility evaluation model due to the difficulty of knowledge reuse and generalization of shallow machine learning model, this paper takes Lueyang County, Hanzhong City, Shaanxi Province as the study area and uses deep random forest to build a regional geological disaster susceptibility evaluation model to improve the prediction accuracy. Firstly, based on the research results of landslide mechanism in Lueyang County, nine factors such as slope, relative height difference, slope direction, slope type, engineering geological rock group, fault distance, river system distance, road

收稿日期: 2023-03-15; 修回日期: 2023-04-23; 责任编辑: 曹佰迪

基金项目:中国地质调查局项目"西北典型地区地质灾害调查"(DD20221739),"黄土高原等典型地区地质灾害精细调查与风险管控"(DD20221739)联合资助。

作者简介:贾俊(1985-),男,高级工程师,主要从事地质灾害调查工作。E-mail: geoj@qq.com。

^{*}通讯作者: 孟晓捷(1986-), 男, 高级工程师, 主要从事地质灾害调查工作。E-mail: 270405820@qq.com。

and railroad distance, and vegetation cover are selected as susceptibility evaluation indexes; secondly, the study area is divided into 5 m \times 5 m raster cells and the values of evaluation factors are extracted and input into the depth random forest evaluation model; finally, the susceptibility evaluation map of the study area is obtained. Based on the evaluation results, geological hazards in Lueyang County can be classified into four levels: very high susceptibility, high susceptibility, medium susceptibility, and low susceptibility, with the proportion of area being 5.31%, 22.97%, 42.11%, and 29.61%. The classification results are consistent with the actual development of geological hazards and reasonably reflect the overall characteristics of geological hazard distribution in the study area. In addition, the area under the ROC curve of the geological hazard susceptibility prediction model of deep random forest is 91.2%, which is higher than 86.3% of the random forest prediction model, indicating that the model is reasonable and feasible, and can provide new ideas for the evaluation of regional landslide susceptibility.

Keywords: landslide; Lueyang County; susceptibility evaluation; Deep Random Forest Model

陕南秦岭山区由于呈高山峡谷特征的地形地貌, 岩体因强烈的构造活动及风化作用而十分破碎,且区 内软弱变质岩系广泛分布,加上降雨强度大、频率高, 导致滑坡、崩塌、泥石流等地质灾害极为发育,严重 威胁区内生命、财产安全。因此,针对秦岭山区地质 灾害进行大比例尺调查及易发性评价具有重要的实 际意义。

滑坡易发性评价是基于统筹考虑多种诱发滑坡 灾害发生的各项影响因素之间可能的相互组合关系 (向喜琼等, 2005),进一步对滑坡灾害发生的概率进 行的量化等级划分。因此,进行预测精确度较高的易 发性评价对区内避免生命、财产的损失具有决策指导 意义。目前,许多定性和定量方法用于滑坡易发性评 价,前些年常用的定性方法主要有自然历史分析法、 工程地质类比法等(王杰等, 2011)。定性法对滑坡易 发性评价起到一定的推动作用,但仍存在局限性。定 量的方法主要基于机器学习,如决策树(赵建华等, 2004)、支持向量机(黄发明等, 2018)、贝叶斯(Snoek et al., 2012)、神经网络(邱海军等, 2012; 邱维蓉等, 2020)等。另外,近年来所流行的随机森林法以决策 树为基本模型,通过构建不同特征的数据集,经过训 练形成一系列具有差异性的决策树模型,由投票得分 最多决定结果,具有极强的数据挖掘能力、不易过拟、 且对异常值和噪声具有较好容忍度、可获取变量重要 性等优点(李亭等, 2014; 吴孝情等, 2017; 张向营等, 2018; 刘坚等, 2018; 吴润泽等, 2021)。

但是,这些传统的机器学习评价方法无法在完成 当前训练任务的同时保留对上一个任务的记忆,难以 形成知识的复用与泛化,造成易发性评价精度无法极 大提升。深度学习是多任务学习,一方面可增强中间 层的泛化能力,在不断的学习过程中积累、提高有效 的知识表示;另一方面具有知识推理能力,突破了上 述传统机器学习评价方法的局限性。有学者采用深 度神经网络算法(Saro Leeet al., 2004; Dou et al., 2020) 及卷积神经网络算法(Maher et al., 2020; Fang et al., 2020)与传统机器学习算法构建滑坡易发性评价模型 并比较其预测结果,结果验证基于深度学习的模型精 确度相较于传统机器学习算法有着极大的提升。上 述算法需要获得大量的训练样本来构建评价模型,对 人力、物力需求较高,限制其广泛的应用。针对此问 题,有学者进一步提出深度随机森林算法(Zhou et al., 2017),该算法作为决策树的集成模型,具有较少的参 数、深度特征提取和不同数据规模适用性的优势,已 被广泛应用于图像分类、图像识别、语音识别等领域。 因此,笔者采用深度随机森林算法构建易发性评价 模型。

选取合理的评价单元是滑坡易发性评价的基础, 目前相关评价常用的单元划分有栅格和斜坡单元。 由于研究区位于秦岭深处,冲沟及河道密集分布,若 斜坡单元想实现精准划分,难度相当大。另外由于不 同斜坡滑坡数据分布不均,特别是存在多个滑坡点分 布于同一个斜坡上的现象,当采用深度随机森林进行 分析时样本数量偏少且同一斜坡上出现重复计算的 现象,会导致评价结果不太合理。故采取栅格作为评 价单元。

综上所述,充分考虑与发育滑坡的地方具有相似 的环境也同时具有相似的趋势原理,笔者设计能够融 合深度随机森林算法优点与栅格评价单元为一体的 区域滑坡易发性评价方法,并于随机森林算法相对比, 比较滑坡易发性评价预测的精度,为区域滑坡易发性 评价提供一个新的方法和视角。

1 研究区概况与研究方法

1.1 研究区地质概况

研究区位于陕西省汉中市略阳县城的主城区,地 理坐标为 N 106°3′45″~106°15′00″, E 33°20′00″~ 33°25′00″(图 1),面积约为 165.68 km²。地形地貌总 体以高山狭谷为主,可进一步分为剥蚀山地和侵蚀与 堆积河谷地貌。周边多条强烈活动断裂和不同期次 的褶皱发育,以勉县断裂和勉县-洋县断裂带为主。 新构造运动继承第三纪末喜马拉雅运动,以上升运动 为主,河流不断下切,形成高山狭谷。研究区出露的 地层主要有元古界碧口群、震旦系、古生界志留系、 泥盆系、石炭系及各期火成岩和新生界第四系。近年 来,随着山区城镇化和新农村建设的加快,山区城镇 建设规模的扩张,山区城镇建设受到发展空间狭小、 城镇建设用地紧缺的限制。因此,削山造地,开挖坡 脚等现象势必诱发大量的地质灾害(张茂省等,2019a, 2019b)(图 2)。



图 1 研究区位置图 Fig. 1 Location of researching area

西安地质调查中心于 2021 年完成略阳县1:50 000 地质灾害调查评价及重点区域1:10 000 调查评价等 工作,研究区内完成滑坡点调查 137 处(图1),主要分 布于剥蚀中山和剥蚀低山区,其中剥蚀低山区有 91 处滑坡,占总数的 66.5%;剥蚀中山区有 46 处滑坡,占 总数的 33.5%。滑坡以中小型为主,其中大型滑坡 14 处,占比为 10.22%;中型滑坡 69 处,占比为 50.36%; 小型滑坡 54 处,占比为 39.42%,滑坡灾害具有线性分 布特点,主要沿嘉陵江、宝成铁路和 309 省道康勉公 路略阳段两侧分布。



a. 略阳县谢家坪村阴坡里滑坡; b. 略阳县碾子湾村何硖路旁崩塌

图 2 研究区灾害示意图

Fig. 2 Hazard photographs of researching area

1.2 研究方法

1.2.1 深度随机森林算法

深度随机森林算法采用 Bagging 集成学习方法, 对树构成的森林进行集成串联,通过分类器进行特征 学习,从而充分利用深度特征提取来提高分类效果。 其网络模型结构主要包括多粒度扫描模块(图 3a)和 级联森林模块(图 3b)。

(1)多粒度扫描模块

设滑坡影响因子数目为n(称为n维),滑动窗口 数目为m(称为m维),把滑坡影响因子的n维数据划 分为(n-m+1)个数据实例,将它们输入到2个随机森 林进行训练,生成k个(k为滑坡易发性等级数目,文 中设置4个易发性等级)概率向量,合并结果,获得 (n-m+1)×k 维概率向量。

(2)级联森林模块

首先将通过粒度扫描模块得到的概率向量输入, 在 2 个随机森林(图 3b 中的黑色部分)和 2 个完全随 机森林(图 3b 中的蓝色部分)进行训练。随机森林和 完全随机森林使用基尼指数(Gini index, 简称 Gini)进 行树的节点分裂, 其公式为:

$$Gini(D) = 1 - \sum_{i}^{k} p_i^2 \tag{1}$$

式中: D 是空间数据库的样本数据集, p_i 是该样本 集中记录属于 k_i(表示极高易发、高易发、中易发、低 易发和极低易发)种类的概率, 并用|k_i, p|/|D|进行计算。

$$Gini_{A}(D) = \frac{|D_{1}|}{|D|}Gini(D_{1}) + \frac{|D_{2}|}{|D|}Gini(D_{2}) \quad (2)$$



图 3 深度随机森林模型流程结构(据 Zhou, 2017) Fig. 3 Process structure illustration of Deep Random Forest Model

式中:A 是空间数据库中滑坡影响因子属性,按 照滑坡影响因子属性A 将样本集进一步分为D₁、D₂。 对于离散属性A,选择该属性最小基尼指数的子集作 为它的分裂指数。

每个森林生成一个 k 维的概率向量与原始的概 率向量 [2×(n-m+1)×k]合并作为下一级联层的输入 并采用 K 折交叉验证方法进行训练, 避免出现过拟合。 其计算公式如下:

$$\beta = \frac{\sum_{i=1}^{n} \mathcal{Q}_i}{p} \tag{3}$$

式中:p为数据集划分子集的个数,Q_i为第i次划 分的结果。在训练过程中级联层依次迭代,直到分类 精度相较于未扩展前(称最大级联层数)没有显著提 高,训练过程终止,从而构建基于深度森林的滑坡易 发性评价模型。

1.2.2 数据来源

笔者所采用的滑坡详细调查数据来源于秦岭南 部地区灾害地质调查项目 2020 年开展的略阳县幅 1:50 000 及1:10 000 地质灾害调查成果。针对该区 域进行易发性评价所采用的基础数据来源于陕西省 测绘局(包括1:50 000 和1:10 000 地形图及 DEM 数据),用于提取高程、坡度、相对高差、坡向、坡型、 水系、公路铁路等信息;地层岩性及断裂分布等图件 来源于中国地质调查局西安地质调查中心,植被归一 化指数的数据来源于中国科学院资源环境科学数据 中心。

2 评价因子选取及分级

2.1 评价因子选取

滑坡受多种因素影响,在野外调查及滑坡灾害信息编录的基础上,依据地质灾害形成的控制因素、影响因素、诱发因素、变形迹象、已有地质灾害等5大 类多源数据,主要利用已有地质灾害发生前和发生后 对比数据,设计建立训练数据集(张茂省等,2019a, 2019b),结合研究区地质环境条件,调研前人对与研 究区类似的地质条件进行滑坡发生机理及影响因素 的相关文献(张春山等,2008;张永双等,2011;孟庆华, 2011;邱海军,2012;王涛等,2013;周样样,2013;周静 静等,2019;吴常润等,2020),选取9类评价因子,分 别为坡度、相对高差、坡向、坡型、工程地质岩组、断 裂距离、河流水系距离、公路铁路距离、植被归一化 指数等。各因子的等级划分见图 4。

2.2 评价因子分级

基于选择的9类评价因子及分区进行信息量计 算,结合研究区总面积(S)为165.68km²,滑坡点总数 (N)为137个,得到各评价因子分区面积(Si)、各评价 因子灾害点数(Ni)及各评价因子(Ii)等(表1)。

统计结果显示(表 1),在坡度评价因子中, 10°~40°区间为滑坡多发区,滑坡数量为105处,占全 部滑坡的 76.64%; 在相对高差评价因子中, 高差 300 m以内为滑坡多发区,滑坡数量为128处,占全部滑 坡的93.43%;在坡向评价因子中,滑坡分布较为平均; 在曲率评价因子中,直线型坡滑坡为119处,占全部 滑坡的 86.86%; 在工程地质岩组评价因子中, 坚硬岩 组、软硬相间岩组、松散岩组发生的滑坡为125处,占 全部滑坡的 91.24%, 这里值得一提的是, 坚硬岩组发 生40处滑坡,主要是由于表层堆积层滑坡沿节理或 基覆界面所发生;在断裂距离评价因子中,距离1000 m以内发生滑坡共128处,占全部滑坡的93.43%;在 水系距离评价因子中,距离 400 m 以内发生滑坡共 117处,占全部滑坡的85.40%;在公路铁路距离评价 因子中,距离1000m以内发生滑坡共113处,占全部 滑坡的 82.48%; 在 NDVI 植被归一化指数中, < 0.68 为发生滑坡的主要区域,共发生滑坡132处,占全部 滑坡的 96.35%。

3 滑坡易发性评价

笔者将研究区栅格文件按照 5m×5m进行划分, 共分割成 6 627 144 个栅格。每个栅格看作一个点,分 别从研究区的 9 个评价因子中提取数据。样本数据 中滑坡与非滑坡的比例将影响模型的特征学习,一个 平衡的比例使模型精确度达到最优值。笔者获取 137 处滑坡点为滑坡样本数据集,依据研究人员的经 验,选取 142 处非滑坡点作为非滑坡样本数据集,随 后随机从 137 处滑坡点和 142 处非滑坡点构成的样本 集中抽取 70% 作为训练样本集, 30% 作为测试样本集。 为避免非滑坡区样本选取时易将潜在滑坡视为非滑 坡而造成的误差(刘坚等, 2018),通过对各因子分级 的信息量数据进行分析,人工从低易发区选取非易发 点。在这过程当中,提取所有栅格单元的评价因子值 并进行归一化处理。



图 4 滑坡与单评价因子分类图 Fig. 4 The landslides division and evaluation factors classification

通过 Arcpy 接口, 传输 ArcGIS 各单因子评价数据 于 Python 语言环境内并使其实现核完成随机森林算 法和深度随机森林算法预测,再将算法结果数据从 Python语言环境中传输回ArcGis10.5软件进行易发性

| 因子 | <u> </u> | S1(km ²) | N1(个) | li |
|--------------|----------------------------------|----------------------|----------|------------|
| | < 10° | 10.535 625 | 8 | -0.085 244 |
| | 10° \sim 20° | 13.396 1 | 21 | 0.639 636 |
| 坡度 | $20^\circ \sim 30^\circ$ | 35.779 55 | 46 | 0.441 342 |
| 72.12 | $30^\circ \sim 40^\circ$ | 54.226 925 | 38 | -0.165 514 |
| | $40^\circ \sim 50^\circ$ | 38.621 9 | 12 | -0.978 836 |
| | > 50° | 13.119 575 | 12 | 0.100 879 |
| 相对高差 | $16\sim 175~m$ | 24.930 925 | 53 | 0.944 260 |
| | 175~238 m | 55.388 15 | 46 | 0.004 353 |
| | 238~300 m | 49.224 425 | 29 | -0.339 017 |
| | 300~379 m | 26.436 65 | 8 | -1.005232 |
| | 379~605 m | 9.698 6 | 1 | -2.081 904 |
| | 0~45° | 23.746 850 | 21 | 0.067 150 |
| | 45° \sim 90° | 21.137 625 | 23 | 0.274 517 |
| | 90°~135° | 17.924 900 | 10 | -0.393 528 |
| | $135^\circ \sim 180^\circ$ | 18.869 250 | 14 | -0.108 399 |
| 坡同 | $180^\circ \sim 225^\circ$ | 20.641 625 | 26 | 0.420 864 |
| | 225°~275° | 22.018 400 | 18 | -0.011 429 |
| | 275°~315° | 20.152 475 | 13 | -0.248 300 |
| | 315°~360° | 21.185 550 | 12 | -0.378 335 |
| | <-0.5(凹形坡) | 11.109 9 | 12 | 0.267 147 |
| 曲率 | -0.5~0.5 (直线形坡) | 143.597 | 119 | 0.002 190 |
| | >0.5(凸形坡) | 10.969 775 | 6 | -0.413 307 |
| 工程地质岩组 | 坚硬岩组 | 65.882 425 | 40 | -0.308 915 |
| | 半坚硬岩组 | 29.980 725 | 13 | -0.645 528 |
| | 软硬相间岩组 | 58.766 425 | 46 | -0.054 852 |
| | 松散岩组 | 11.050 425 | 39 | 1.451 170 |
| | < 100 m | 31.79 | 26 | -0.010 915 |
| | $100 \sim 200 \text{ m}$ | 25.87 | 26 | 0.195 167 |
| 距断裂距离 | $200\sim500~{ m m}$ | 49.79 | 53 | 0.252 595 |
| | $500 \sim 1000{ m m}$ | 34.60 | 23 | -0.218 224 |
| | > 1000 m | 23.64 | 9 | -0.775 554 |
| | < 200 m | 69.69 | 86 | 0.400 425 |
| 距河流水系 距离 | $200 \sim 400 \text{ m}$ | 52.38 | 31 | -0.334 384 |
| | $400\sim 600$ m | 27.98 | 17 | -0.308.056 |
| | $600 \sim 800 \text{ m}$ | 11.37 | 1 | -2 240 549 |
| | > 800 m | 4 28 | 2 | -0 569 794 |
| 距公路、铁路 距离 | < 100 m | 18 99 | 58 | 1 306 584 |
| | $100 \sim 500 \text{ m}$ | 46.12 | 41 | 0 072 488 |
| | $500 \sim 1000\mathrm{m}$ | 38.06 | 14 | -0 809 915 |
| | $1.000 \sim 1.500 \text{ m}$ | 26.29 | 11 | -0 681 226 |
| | > 1500 m | 36.23 | 12 | -0 834 704 |
| | $-0.41 \sim 0.07$ | 1.99 | | 0.034 /94 |
| NDVI | 0.41 - 0.07 | 1.00 | 0 | 1 415 100 |
| | $0.07 \approx 0.52$ | /.04 | 20 | 0.047.005 |
| | $0.52^{\circ} = 0.52$ | 20.23 | 30 50 | 0.94/905 |
| | 0.52 - 0.08 | 91.74 | 50 | -0.416 840 |
| | $0.68 \sim 0.84$ | 38.18 | 5 | -1.842 838 |

表 1 各评价因子信息量表

 Tab. 1
 Weighted information values of individual evaluation factors

评价。在地质灾害相关评价中,针对栅格赋值并归一 化后的离散数据常用的分区方法有自然断点法(邱维 蓉等,2020)、基于数理统计的自然拐点法(王佳运等, 2020)、百分位数法(吴孝情等, 2017)、相等间隔法(田 春山等,2016)等。在空间统计分析中,自然断点法以 自然(或非人为设定)的转折点和断点为分区界线,实 现将研究对象分成性质相似的群组,实现每一组内部 数据的相似性最大。基于数理统计的自然拐点法其 思想和自然断点法类似,都是寻找数值中间的自然转 折点,但该方法在针对归一化的海量数据求导后,存 在拐点阈值难以界定的问题,虽然可以通过 Matlab或 Pyhton 的帮助而找到拐点,但这个过程是将求 导后的数据结果进行曲线拟合而完成,在拟合过程中 存在一定的数据偏离。百分位数法和相等间隔法是 通过人为确定数值的间隔来完成分区,掺杂一定的主 观性,难以客观体现出不同区域边坡的易发性。因此, 笔者选择自然断点法来完成易发性分区(图 5、图 6), 通过栅格统计工具,完成易发性评价等级的相关统计 分析。



图 5 略阳县随机森林滑坡易发性评价图

Fig. 5 Landslide susceptibility assessment map of Lueyang County base on Random Forest Model

通过上述步骤分别得到随机森林易发性评价图 (图 5)和深度随机森林易发性评价图(图 6),按自然 资源部中国地质调查局地质调查技术标准(DD 2019-08)——地质灾害调查技术要求(1:50000),将 研究区分为极高易发区、高易发区、中易发区及低易 发区。

据 ArcGIS 分区统计显示,随机森林易发性评价 图内,极高易发区的面积为 5.32 km²,占比为 3.21%,发 生滑坡数量为 22 个;高易发区面积为 18.63 km²,占比 为 11.24%,发生滑坡数量为 40 个;中易发区面积为 114.65 km²,占比为 69.20%,发生滑坡数量为 74 个;低 易发区面积为 27.08 km²,占比为 16.34%,发生滑坡数





量为1个。

深度随机森林易发性评价图内,极高易发区面积为 8.8 km²,占比为 5.31%,发生滑坡数量为 27 个;高易发区面积为 38.05 km²,占比为 22.97%,发生滑坡数量为 62 个;中易发区面积为 69.77 km²,占比为 42.11%,发生滑坡数量为 41 个;低易发区面积为 49.06 km²,占比为 29.61%,发生滑坡数量为 7 个。

将随机森林和深度随机森林法所得到的易发性 分区结果与滑坡分布的实际情况做对比(表 2), 是一 种常用的对评价结果初步检验(薛强, 2015)。随易发 性等级的提高, 各等级中所包含的滑坡数量同步递增, 同时滑坡实际发生的比率(c/a)也在增大, 说明这 2 种 方法所得出的易发等级与滑坡发生的实际情况基本 吻合, 划分结果较为理想。

极高易发区和高易发区主要分布于中低山区与 河谷接壤地带,地层由松散岩组的坡积层碎石土和软 硬相间岩组如片岩、千枚岩、板岩等构成,工程地质 条件属于强度较低,稳定性较差。其中坡积层较为松 散,利于地表水入渗,下伏基岩又起到阻水作用。因 此,在该层中储存的孔隙水易导致坡体失稳、破坏而 形成滑动或滑塌。软硬相间的岩组更多受强烈活动 的断裂和多期褶皱的影响,造成地表破碎,节理裂隙 发育,为孕灾提供了物质基础。加上人类工程活动在 河谷两岸,削坡修建四通八达的公路网、铁路及房屋, 特别是嘉陵江两岸、宝成铁路与省道公路沿线均容易 由于边坡失稳而导致滑坡形成。

4 评价结果验证分析

不论是随机森林法还是深度随机森林法,所得到 的易发性分区结果都基本与研究区的地貌、构造、工

| Tab. 2 Comparison of susceptibility classification and actual landslide occurrence rate | | | | | | | | |
|---|-----|-------------|--------|----------|--------|--|--|--|
| 评价 | 易发性 | а | b | с | c/a | | | |
| 方法 | 等级 | 易发分区面积占比(%) | 分区滑坡数量 | 滑坡百分比(%) | 滑坡发生比率 | | | |
| | 低 | 16.34 | 1 | 0.73 | 0.04 | | | |
| 随机森林 | 中 | 69.21 | 74 | 54.01 | 0.78 | | | |
| | 高 | 11.24 | 40 | 29.20 | 2.60 | | | |
| | 极高 | 3.21 | 22 | 16.06 | 5.00 | | | |
| 深度 随机森林 | 低 | 29.61 | 7 | 5.11 | 0.17 | | | |
| | 中 | 42.11 | 41 | 29.93 | 0.71 | | | |
| | 高 | 22.97 | 62 | 45.26 | 1.97 | | | |
| | 极高 | 5.31 | 27 | 19.71 | 3.71 | | | |

表 2 易发性等级划分与滑坡实际发生比率对比表

程地质岩组、人类活动等有着强烈的联系,它们的滑 坡分布趋势是相一致的。在此引进一个衡量易发性 分区评价的检验指标,来对这2种方法的优劣性进行 分析。

在地质灾害易发性评价中,前人常使用 ROC(Receiver Operating Characteristic Curve)曲线对分区结果 进行验证,通过计算 AUC(Area Under Curve)值来评 价区划结果的准确性(Chung et al., 2003; 孟晓捷等, 2022)。笔者在采用 ROC 曲线进行对比时, y 轴表示 把实际为真值(滑坡)的预测为真值(滑坡)的概率;x 轴表示把实际为假值(非滑坡)的预测为真值(滑坡) 的概率,采用构图法描绘危险性和特异性的相互关系。 随后得到 ROC 曲线内的面积,即 AUC 值。当 AUC≤ 0.5, 表明评价模型的预测能力无效; 当 0.5< AUC≤0.7, 表明评价模型的预测能力具有一定的准确 性;当0.7<AUC≤0.9,表明评价模型的预测能力具有 较好的准确性;当AUC≥0.9,表明评价模型的预测能 力较高。

在模型验证过程中,按照上节的训练集与测试集 的比例,从样本集中随机抽5组训练集与测试集;接 着依据5组测试集中真实值与预测值,计算每组RF 与DRF的AUC值,求这5组的平均值,即获得RF和 DRF的 AUC 值分别为 86.3% 和 91.2%, 从 而 绘 制 ROC 曲线(图 7)。

基于随机森林和深度随机森林的滑坡预测模型 的 AUC 值都高于 0.7, 说明这 2 个模型的预测能力具 有一定的准确性。深度随机森林评价模型的 AUC 值 比随机森林高出 4.9%, 因此深度随机森林的模型预测 精确度略高于随机森林模型。其产生的原因是:虽然 随机森林和深度随机森林都采用集成学习方法,但是



图 7 深度随机森林与随机森林易发性 评价 ROC 曲线对比图



随机森林只是同时训练多个决策树,使用类似投票方 式获取最终结果来提高预测精确度;而深度随机森林 通过多粒度扫描技术,其特征学习能力得到进一步的 提升。另外深度随机森林采用K折交叉验证方法避 免过拟合,从而提高了滑坡危险性评价易发性评价模 型的精确度。

结论 5

(1)笔者以秦岭山区腹地汉中市略阳县城主区域 为研究对象,在1:5万及1:1万调查基础上,选取坡 度、相对高差、坡向、坡型、工程地质岩组、断裂距离、 河流水系距离、公路铁路距离及 NDVI 植被归一化指 数等9个要素作为易发性评价指标。基于 ArcGIS 平 台和随机森林及深度随机森林的融合,将评价结果进 一步分为极高易发区、高易发区、中易发区和低易发

区。经 ROC 曲线验证, 深度随机森林评价模型预测 精度达到 91.2%。其评价结果具有一定的可行性及科 学合理性。

(2)研究区内滑坡点共有 137 处,主要分布于剥 蚀中山、低山区。剥蚀低山区有 91 处滑坡,占总数的 66.5%;剥蚀中山区有 46 处滑坡,占总数的 33.5%。滑 坡以中小型为主,其中大型滑坡 14 处,占比为 10.22 %; 中型滑坡 69 处,占比为 50.36%;小型滑坡 54 处,占比 为 39.42%。极高易发区、高易发区主要分布于嘉陵 江两岸、宝成铁路与省道公路沿线,地貌类型为秦岭 中低山区,工程地质岩组主要为松散岩组及软硬相间 岩组,距断裂 500 m 以内,距河流水系 400 m 以内,距 公路铁路等工程活动 500 m 以内的地区。

参考文献(References):

- 黄发明,殷坤龙,蒋水华,等.基于聚类分析和支持向量机的滑 坡易发性评价[J].岩石力学与工程地质学报,2018,37(1): 156-167.
- HUANG Faming, YIN Kunlong, JIANG Shuihua, et al. Landslide susceptibility assessment based on clustering analysis and support vector machine[J]. Chinese Journal of Rock Mechanics and Engineering, 2018, 37(1): 156–167.
- 李亭,田原,邬伦,等.基于随机森林方法的滑坡灾害危险性区 划[J].地理与地理信息科学,2014,30(6):25-30.
- LI Ting, TIAN Yuan, WU Lun, et al. Landslide susceptibility Mapping Using Random Forest[J]. Geography and Geo-Information Science, 2014, 30(6): 25–30.
- 刘坚,李树林,陈涛,等.基于优化随机森林模型的滑坡易发性 评价[J]. 武汉大学学报·信息科学版,2018,43(7): 1085-1091.
- LIU Jian, LI Shulin, CHEN Tao, et al. Landslide Susceptibility Assesment Based on Optimized Random Forest Model[J]. Geomatics and Information Science of Wuhan University, 2018, 43(7): 1085–1091.
- 孟庆华.秦岭山区地质灾害风险评估方法研究—以陕西凤县为 例[D].北京:中国地质科学院,2011.
- MENG Qinghua. Study on the Methods of Geo-hazards Risk Assessment in Qinling Mountain: A case studyof Feng County, Baoji City, Shaanxi Province[D]. Beijing: Chinese Academy of Geological Sciences, 2011.
- 孟晓捷,张新社,曾庆铭,等.基于加权信息量法的黄土滑坡易 发性评价——以1:5万天水市麦积幅为例[J].西北地质, 2022,55(2):249-259.
- MENG Xiaojie, ZHANG Xinshe, ZENG Qingming, et al. The Susceptibility Evaluation of Loess Landslide Based on Weighted Information Value Method: Taking 1: 50 000 Map of Maiji Dis-

trict of Tianshui City As an Example[J]. Northwestern Geology, 2022, 55(2); 249–259.

- 邱海军.区域滑坡崩塌地质灾害特征分析及其易发性和危险性 评价研究——以宁强县为例[D].西安:西北大学,2012:83.
- QIU Haijun. Study on the Regional Landslide Characteristic Analysis and Hazard Assessment: A case study of Ningqiang County[D]. Xi'an: Northwest University, 2012: 83.
- 邱维蓉,吴帮玉,潘学树,等.几种聚类优化的机器学习方法在 灵台县滑坡易发性评价中的应用[J].西北地质,2020, 53(1):222-233.
- QIU Weirong, WU Bangyu, PAN Xueshu, et al. Application of Several Cluster-optimization-based Machine Learning Methods in Evaluation of Landslide Susceptibility in Lingtai County[J]. Northwestern Geology, 2020, 53(1): 222–233.
- 田春山,刘希林,汪佳,等.基于 CF 和 Logistic 回归模型的广东 省地质灾害易发性评价[J].水文地质工程地质,2016, 43(6):154-170.
- TIAN Chunshan, LIU Xilin, WANG Jia. et al. Geohazard susceptibility assessmentbased on CF model and Logistic Regression models in Guangdong[J]. Hydrogeology & Engineering Geology, 2016, 43(6): 154–170.
- 王佳运,毕俊擘,杨旭东,等.山西吕梁山区城镇边坡风险分级 与优化[J].地质通报,2020,39(12):2004-2012.
- WANG Jiayun, BI Junbo, YANG Xudong, et al. Risk classification and optimization of to-wn side slope in Lüliang Mountain[J]. Geological Bulletin of China, 2020, 39(12): 2004–2012.
- 王杰,马凤山,郭捷,等.一种改进的区域滑坡危险性评价模型 及其应用[J].中国地质灾害与防治学报,2011,22(2): 14-19.
- WANG Jie, MA Fengshan, GUO Jie, et al. An improved regional landslide hazard assessment, model and its application[J]. The Chinese Journal of Geological Hazard and Control, 2011, 22(2): 14–19.
- 王涛,吴树仁,石菊松,等.秦岭中部太白县地质灾害发育特征 及危险性评估[J].地质通报,2013,32(12):1976-1983.
- WANG Tao, WU Shuren, SHI Jusong, et al. Case Study Of Landslide Characteristics And Hazard Assessment in Taibai County, Central Qinling Mountains[J]. Geological Bulletin of China, 2013, 32(12): 1976–1983.
- 吴常润,赵冬梅,刘澄静,等.基于 GIS 和信息量模型的陇川县 滑坡易发性评价[J].西北地质,2020,53(2):308-320.
- WU Changrun, ZHAO Dongmei, LIU Chengjing, et al. Landslide Susceptibility Assessment of Longchuan County Based on GIS and Information Value Model[J]. Northwestern Geology, 2020, 53(2): 308–320.
- 吴润泽,胡旭东,梅红波,等.基于随机森林的滑坡空间易发性 评价:以三峡库区湖北段为例[J].地球科学,2021,46(1): 321-330.
- WU Runze, HU Xudong, MEI Hongbo, et. al. Spatial Susceptibility Assessment of Landslides Based on Random Forest: A Case

Study from Hubei Section in the Three Gorges Reservoir Area[J]. Earth Science, 2021, 46(1): 321–330.

- 吴孝情,赖成光,陈晓宏,等.基于随机森林权重的滑坡危险性 评价:以东江流域为例[J].自然灾害学报,2017,26(5): 119-129.
- WU Xiaoqing, LAI Chengguang, CHEN Xiaohong, et al. A landslide hazard assessment based on random forestweight: a case study in the Dongjiang River Basin[J]. Journal of Natural Disarsters, 2017, 26(5): 119–129.
- 向喜琼.区域滑坡地质灾害危险性评价与风险管理[D].成都: 成都理工大学,2005:22.
- XIANG Xiqiong. Regional Landslide Hazard Assessment and Risk Management[D]. Chengdu: Chengdu University of Technology, 2005: 22.
- 薛强,张茂省,李林,等.基于斜坡单元与信息量法结合的宝塔 区黄土滑坡易发性评价[J].地质通报,2015,34(11): 2108-2115.
- XUE Qiang, ZHANG Maosheng, LI Lin. Loess landslide susceptibility evaluation based on slope unit and information value method in Baota District, Yan'an[J]. Geological Bulletin of China, 2015, 34(11): 2108–2115.
- 张春山,韩金良,孙炜锋,等.陕西陇县地质灾害危险性分区评价[J].地质通报,2008,27(11):1795-1801.
- ZHANG Chunshan, HAN Jinliang, SUN Weifeng, et al. Assessments of geohazard danger zoning in Longxian County, Shaanxi, China[J]. Geological Bulletin of China, 2008, 27(11): 1795–1801.
- 张茂省, 贾俊, 王毅, 等. 基于人工智能 (AI) 的地质灾害防控体 系建设[J]. 西北地质, 2019a, 52(2): 103-116.
- ZHANG Maosheng, JIA Jun, WANG Yi, et al. Construction of Geological Disarster Prevention and Control System Based on AI[J]. Northwestern Geology, 2019a, 52(2): 103–116.
- 张茂省,薛强,贾俊,等.山区城镇地质灾害调查与风险评价方 法及实践[J].西北地质,2019b,52(2):125-135.
- ZHANG Maosheng, XUE Qiang, JIA Jun, et al. Methods and Practice for the Investigationand Risk Assessment of Geo-hazards in Mountains Towns[J]. Northwestern Geology, 2019b, 52(2): 125–135.
- 张向营,张春山,孟华君,等.基于 Random Forest 和 AHP 的贵德 县北部山区滑坡危险性评价[J].水文地质工程地质,2018, 45(4):142-149.
- ZHANG Xiangying, ZHANG Chunshan, MENG Huajun, et al. Landslide hazard evaluation in the northern mountainous area of Guide County based on Random Forest and AHP[J]. Hydrogeology & Engineering Geology, 2018, 45(4): 142–149.
- 张永双,苏生瑞,吴树仁,等.强震区断裂活动与大型滑坡关系 研究[J].岩石力学与工程学报,2011,30(Z2):3503-3513.

- ZHANG Yongshuang, SU Shengrui, WU Shuren, et al. Research On Relationship Between Fault Movement And Large-Scale Landslide In Intensive Earthquake Region[J]. Chinese Journalof Rock Mechanics and Engineering, 2011, 30(Z2): 3503 – 3513.
- 赵建华,陈汉林,杨树峰,等.基于决策树算法的滑坡危险性区 划评价[J].浙江大学学报(理学版),2004,31(4):465-470.
- ZHAO Jianhua, CHEN Hanlin, YANG Shufeng, et al. Landslide risk assessment based on decision tree arithmetic[J]. Journal of Zhejiang University (Science Edition), 2004, 31(4): 465–470.
- 周静静,赵法锁,李辉,等.陕西省地质灾害与影响因素相关性研究[J].灾害学,2019,34(2):228-234.
- ZHOU Jingjing, ZHAO Fasuo, LI Hui, et al. Correlational Research Between Geological Hazards and impact Factors in Shaanxi Province[J]. Journal of Catastrophology, 2019, 34(2): 228–234.
- 周样样. 陕南地区强降雨条件下突发型地质灾害成因机制研 究[D]. 西安: 长安大学, 2013.
- ZHOU Yangyang. Study on Formation Mechanism of Abrupt Geological Hazards of Southern Shaanxi Region in Condition of Strong Rainstorm[D]. Xi'an: Chang'an University, 2013.
- Chung C J F, Fabbri A G. Validation of spatial prediction modelsfor landslidehazard mapping[J]. Natural Hazards, 2003, 30(3): 451–472.
- Dou J , Yunus A P , Merghadi A , et al. Different sampling strategies for predicting landslide susceptibilities are deemed less consequential with deep learning [J]. Science of the Total Environment, 2020, 720: 137320. https://doi.org/10.1016/j.scitotenv. 2020.137320.
- Maher Ibrahim Samee, Biswajeet Pradhana, Saro Lee, et al. Application of convolutional neural networks featuring Bayesian optimization for landslide susceptibility assessment[J]. Catena, 2020, 186: 10424. https://doi.org/10.1016/j.catena.2019.104249.
- Saro Lee, Joo Hyung Ryu, Joong Sun Won, et al. Determination and application of the weights for landslide susceptibility mapping using an artificial neural network[J]. Engineering Geology, 2004, 71(3-4): 289–302.
- Snoek J, Larochelle H, Adams R P. et al. Practical Bayesian Optimization of Machine Learning Algorithms[J]. Advances in Neural Information Processing Systems, 2012, 4: 1–12.
- Fang Zhice, Wang Yi, Ling Peng, et al. Integration of convolutional neural network and conventional machine learning classifiers for landslide susceptibility mapping[J]. Computer & Geosciencees, 2020, 139: 104470. https://doi.org/10.1016/j. cageo.2020.104470.
- Zhou Zhihua, Ji Feng. Deep Forest: Towards An Alternative to Deep Neural Networks [J]. Statistics, 2017, 71(3–4): 289–302.