

基于逻辑回归的巫山县滑坡易发性区划研究^{*}

许嘉慧¹, 张虹¹, 文海家², 孙德亮¹

(1. 重庆师范大学 GIS 应用研究重庆市高校重点实验室, 重庆 401331;

2. 重庆大学 山地城镇建设与新技术教育部重点实验室, 重庆 400045)

摘要:【目的】在已有文献研究基础上,构建基于逻辑回归的滑坡易发性评价模型,并与基于随机森林的模型进行分析比较,探讨出适用于三峡库区巫山县的滑坡易发性评价模型。【方法】选取地质条件、地形地貌、环境条件、人类工程活动等4方面影响下的22个因子作为滑坡易发性影响因子,根据研究区963处历史滑坡点,建立30 m×30 m栅格地理空间数据库;进一步在数据库中,按历史滑坡与非滑坡1:10的比例构建地理空间数据样本集,基于十折交叉验证法选择最佳样本数据后,进行逻辑回归模型训练;最后,利用最佳样本训练后的模型,进行全域滑坡易发性仿真分析,并将结果划分为低、较低、中、较高、高共5种易发性等级。【结果】逻辑回归模型训练集、测试集的受试者工作特征曲线下面积(Area under curve, AUC)值分别为0.803和0.787;有63.24%历史滑坡落入面积比为26.26%的较高滑坡易发性区划中。【结论】基于逻辑回归和随机森林构建的滑坡易发性评价模型均具有良好的稳定性及预测功能,但随机森林模型性能优于逻辑回归模型性能,更适用于三峡库区滑坡易发性区划研究。

关键词:滑坡易发性区划;逻辑回归模型;巫山县;三峡库区;随机森林模型

中图分类号: P694

文献标志码: A

文章编号: 1672-6693(2021)02-0048-09

滑坡是一种极具破坏性的自然灾害,严重威胁人类生命财产安全和影响经济发展^[1]。据世界紧急事件数据库(EM-DAT)记载,2014—2018年期间,滑坡在世界范围内导致4 914人死亡,27 110人无家可归,直接经济损失达21亿美元。因此,找到有效解决方案,减缓、减轻滑坡造成的有关破坏显得尤为重要。

滑坡易发性区划是根据某地地质环境,描述该地区滑坡发生概率的空间分布情况的一种方法,也被认为是减轻滑坡造成影响的通用工具^[2]。基于地理信息系统(GIS)和遥感技术(RS)的各种滑坡易发性评价模型被广泛应用,包括基于物理模型、启发式模型和统计模型^[3]。由于物理模型不适用于面积较大区域,启发式模型在很大程度上依赖于专家经验^[4],因此统计模型受到了广泛关注。国内外学者应用统计模型均取得了较好的滑坡评价效果。如频率比^[5]、层次分析法^[2,6]、证据权^[7-8]等。然而统计模型存在影响因子量化受人为因素干扰、主观性较强等缺陷^[9]。

随着数据挖掘技术的发展,越来越多的研究人员将机器学习方法应用至滑坡易发性评价中。机器学习方法可基于历史滑坡数据,挖掘滑坡发生与影响因子间复杂关系,相比于传统统计方法,具有许多优点^[9]。目前,应用较多的模型有逻辑回归、决策树、支持向量机、人工神经网络等。其中,逻辑回归的方法较为客观、数学计算严谨,且进行统计分析时,数据无须满足正态分布^[10],是滑坡易发性评价中应用最广泛的方法之一。

三峡库区地质构造发育、地形条件复杂,是中国地质灾害发生最多、经济损失最大的地区之一^[11]。自2003年三峡库区蓄水以来,库区内的地质灾害潜在危险明显增多,其中滑坡隐患尤为突出,对滑坡的预测亟待解决^[12]。刘睿等人^[13]以三峡库区巫山县为研究区,基于GIS与随机森林对该区进行滑坡易发性区划,结果表明基于随机森林构建的滑坡易发性评价模型具有较高的准确性及良好的预测功能。本文在该研究的基础上,构建基于逻辑回归的滑坡易发性评价模型,分析比较逻辑回归与随机森林模型性能,总结模型优缺点,探讨出适用于巫山县的滑坡易发性区划模型,为巫山县滑坡灾害风险管理和城市建设规划提供科学参考。

^{*} 收稿日期:2020-05-22 修回日期:2020-11-11 网络出版时间:2021-04-14 15:12

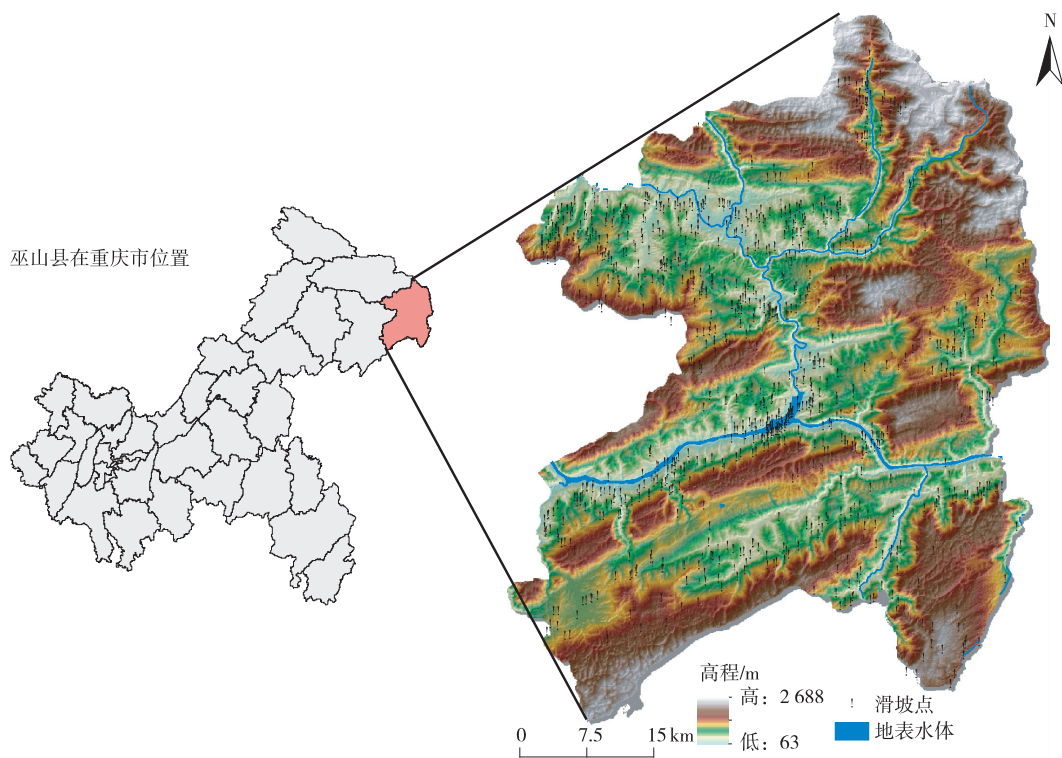
资助项目:国家重点研发计划(No. 2018YFC1505501);国家自然科学基金(No. 41807498);教育部人文社科规划基金(No. 20XJAZH002)

第一作者简介:许嘉慧,女,研究方向为大数据与地质灾害,E-mail: jiahuiXu@aliyun.com;通信作者:张虹,女,副教授,博士,E-mail: zhangh@cqnu.edu.cn

网络出版地址:https://kns.cnki.net/kcms/detail/50.1165.n.20210414.1426.005.html

1 研究区概况

研究区巫山县地处北纬 $30^{\circ}46' \sim 31^{\circ}28'$, 东经 $109^{\circ}33' \sim 110^{\circ}11'$, 位于重庆市东部, 是三峡库区重庆市区内首淹首迁县^[14-15]。巫山县位于川鄂湘黔隆起褶皱带、大巴山弧和川东褶皱带的交界地带, 构造应力场复杂。当地地貌以山地丘陵为主, 境内地层除白垩系和第三系缺失外, 自寒武系至第四系地层均有出露。气候为亚热带季风气候, 年平均气温约为 18°C , 年平均降雨量约为 $1\,049\text{ mm}$, 且降雨集中发生在 5 月至 9 月, 占全年总降雨量的 69%。受复杂的山地地貌、多变的岩性褶皱及城镇规划建设、库区移民迁镇等人类工程活动及气候水文影响, 研究区地质灾害发生频繁。图 1 显示了巫山县地理位置和历史滑坡点分布。



注: 底图数据来源于重庆标准地图服务网, 审图号: 渝 S(2019)055 号, 下同

图 1 研究区位置及滑坡点分布

Fig. 1 Location and landslide distribution of Wushan county

2 地理空间数据库

2.1 影响因子

选择合适的滑坡影响因子作为输入变量是易发性评价的一个关键步骤。根据 Ayalew 和 Yamagishi 的观点^[16], 基于 GIS 的滑坡易发性影响因子应是可测量的、可操作的、完整的和非冗余的。为消除影响因子的不同而对构建的模型产生的影响, 本文选取的因子与以往研究保持一致, 即选取了地形地貌(高程、起伏度、坡向、坡度、坡位、曲率、剖面曲率、平面曲率、微地貌、地形湿度指数^[17]、粗糙度指数^[18]、输沙指数^[19]、水流动力指数^[20])、地质条件(岩性、距断层距离、倾坡类型^[21])、环境条件(距河流距离、土地利用类型、归一化植被指数、多年平均降雨)、人类活动(距道路距离^[22-24]、POI 核密度^[25])共 4 方面影响下的 22 个影响因子, 构建巫山县滑坡易发性评价指标体系^[13]。

2.2 数据处理

本研究所有数据及数据源与以往研究保持一致^[13]。起伏度、坡向、坡度、坡位、微地貌、曲率、剖面曲率、平面曲率、粗糙度指数、地形湿度指数、输沙指数和水流动力指数均通过对 30 m 分辨率的 ASTER GDEM 数字高程模型处理后获得; 归一化植被指数通过计算 Landsat 8 OLI 卫星影像后获得; 岩性和断层通过矢量化来获得 1:200 000 的地质图提取; 依次对断层、河流和道路进行换成, 得到距断层、距河流和距道路距离图层; 坡向和倾

向相减并重分类来提取倾坡^[21];POI 由百度地图服务提供,本文从研究区提取了与人口有关约 12 688 条 POI 记录,包括商业场所、居民社区、教育设施(例如幼儿园、小学、中学)、临床设施和风景名胜等,POI 核密度基于该数据通过 ArcGIS 10.5 软件的核密度工具生成。

对高程等连续因子进行重分类,以便进一步分析。根据现场调查、专家经验以及以往文献,对每个连续因子建立了重分类方案^[13]。同时,为减少数据离散性,将重新分类后的 22 个因子进行归一化及线性变换,将值减小到 $[0,1]$ 区间。

以 $30\text{ m} \times 30\text{ m}$ 分辨率的栅格单元作为滑坡易发性评价基本单元,建立滑坡影响因子地理空间数据库可视化专题图层(部分),如图 2 所示。

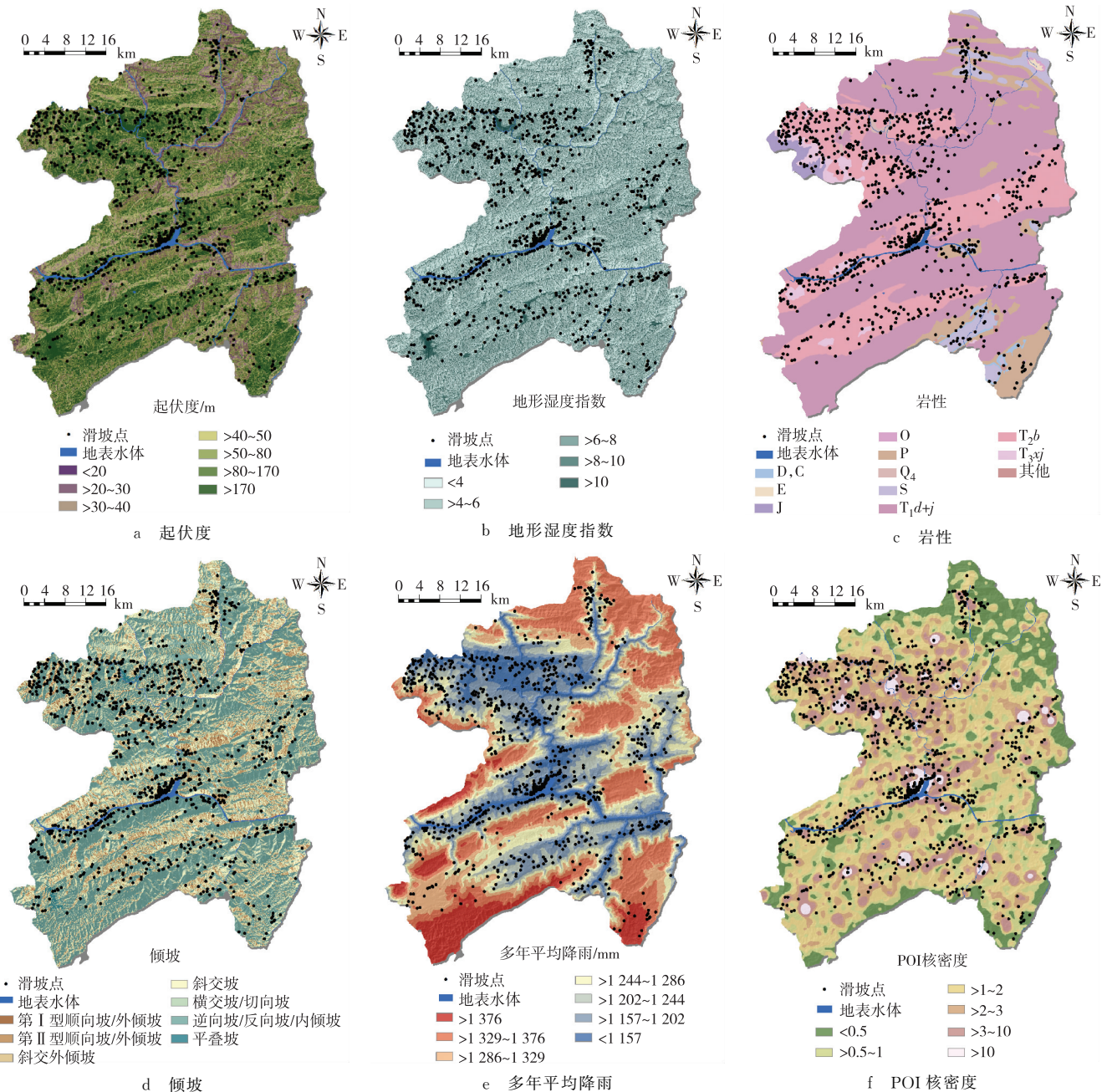


图 2 部分滑坡影响因子专题图层

Fig. 2 Partial conditioning parameters of landslide

2.3 共线性诊断

多重共线性是线性逻辑回归模型中棘手的问题,降低变量间的相互联系强度是解决共线性的关键,若不处理则会导致观测值的变化异常敏感,估计值偏差严重。容忍度(TOL)与方差膨胀因子(VIF)乘积为 1,是诊断多

重共线性常用指标。通常当 VIF 大于等于 10 或 TOL 小于 0.1, 可认为自变量间共线性严重^[2]。本研究中, 影响因子的多重共线性基于 SPSS 统计软件进行, 结果显示起伏度、坡度、曲率 3 个因子未通过共线性检验。因此, 将通过检验的 19 个影响因子进入模型训练。

3 滑坡易发性区划

3.1 模型概述

逻辑回归模型由多变量控制, 该算法优势在于自变量无需满足正态分布^[26]。基于该模型, 滑坡易发性能力和自变量间关系可表述为:

$$f(z) = \frac{1}{1 + e^{-z}},$$

其中: $z = w_1x_1 + w_2x_2 + \dots + w_Mx_M + b$ 是一个加权线性组合模型, b 为函数截距, w_M ($M=1, 2, 3, \dots, 19$) 为相关性系数, 自变量 x_M ($M=1, 2, 3, \dots, 19$) 代表 19 个影响因子。 $f(z)$ 代表滑坡发生概率, 位于 $[0, 1]$ 之间。当函数值接近 1 时, 表明该点易发生滑坡灾害; 当函数值接近 0 时, 表明该点滑坡发生较难^[4]。

3.2 模型训练

在本研究中, 正样本(滑坡点)和负样本(非滑坡点)构成了全体数据集。正样本由 963 个历史滑坡点组成, 每个滑坡点都视为一个网格单元。将剔除历史滑坡点 500 m 缓冲区、研究区河流所在区域作为非滑坡区域, 并按 1:10 的比例在非滑坡区域随机提取非滑坡点(9 630 个)作为负样本^[9]。为降低随机取样方式对结果产生的干扰, 采用十折交叉验证法进行样本选取。十折交叉验证将全体数据集(963 个正样本和 9 630 个负样本)随机平均划分成 10 个互不相交的子集, 每次用一个子集进行模型精度验证, 其余子集进行模型训练^[13]。表 1 为逻辑回归模型十折交叉验证准确性结果。可以看到, 逻辑回归模型测试样本平均精度为 0.908, 其中, 样本 2 测试集精度(0.922)相对更高。因此, 基于样本 2 的训练数据集, 构建逻辑回归模型并进行全域仿真分析。

3.3 模型评价

受试者工作特征(Receiver operating characteristic, ROC)曲线及曲线下面积(Area under curve, AUC)值, 是综合检验和评价滑坡易发性评价模型精度的常用指标^[4]。ROC 曲线以灵敏度和特异性为水平轴和垂直轴, AUC 值可定量表示模型预测的准确性, 值为 1 代表理想模型, 值为 0 代表模型无意义值, 大于 0.5 且越靠近 1, 表明模型诊断效果良好, 模型预测精度较高^[9]。图 3 为逻辑回归模型 ROC 曲线, 可以发现, 逻辑回归模型训练集 AUC 值为 0.803, 表明构建的模型具有一定稳定性; 测试集 AUC 值为 0.787, 表明该模型具有一定预测能力。

表 1 十折交叉验证准确性

Tab. 1 The accuracy of 10-fold cross-validation

样本号	精度		样本号	精度	
	训练	测试		训练	测试
1	0.909	0.905	6	0.909	0.907
2	0.908	0.922	7	0.909	0.912
3	0.908	0.919	8	0.909	0.908
4	0.908	0.919	9	0.910	0.895
5	0.910	0.897	10	0.909	0.900

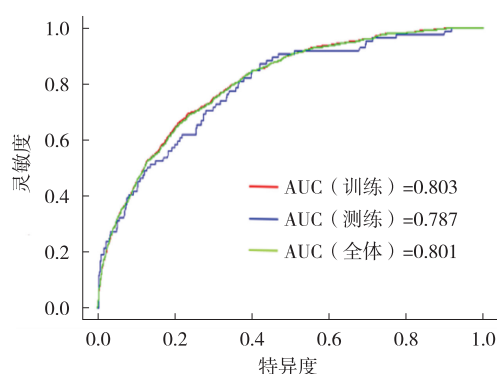


图 3 逻辑回归模型的 ROC 及 AUC 值

Fig. 3 ROC and AUC values of logistic regression

3.4 区划结果

应用训练好的逻辑回归模型至整个研究区, 模拟出研究区范围每个栅格滑坡发生概率 P , 并根据专家经验^[2], 将它划分为 5 个等级, 于是得到巫山县滑坡易发性区划图(图 4)。可以发现, 巫山县大部分地区均处于滑坡低易发区, 且集中分布于研究区东部及东北部; 高易发区集中分布于长江及支流两岸, 以巫山西北部为主, 与历史滑坡点分布几乎一致。

同时, 利用统计分析来定量评估该模型的有效性(表 2)。可以发现, 极低和低易发区、高和极高易发区面积

分别占研究区总面积的 62.16%和 26.26%,而两者的历史滑坡数分别占滑坡总数的 24.61%和 63.24%;随着易发性等级的升高,滑坡在各等级占比逐渐增大,滑坡密度与易发性等级具有正相关关系,模型预测结果具有一定可靠性。

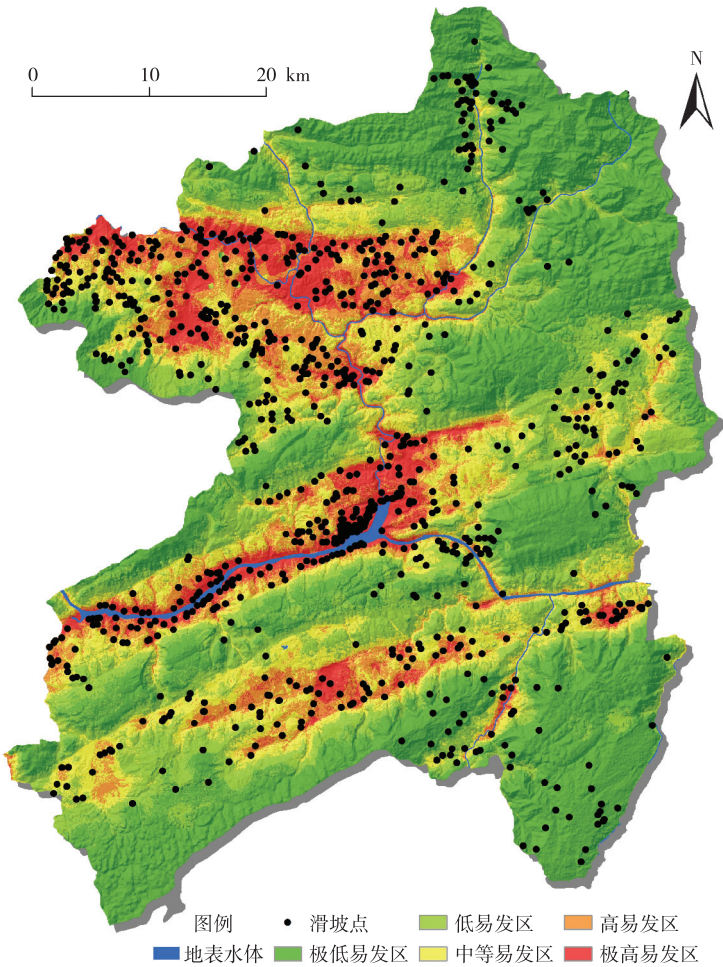


图 4 巫山县滑坡易发性区划图

Fig. 4 Landslide susceptibility mapping of Wushan county

表 2 RF 滑坡易发性区划分级统计

Tab. 2 Statistic result of landslide susceptibility in different grades

滑坡概率	易发性级别	栅格数/个	面积比/%	滑坡数/个	滑坡占比/%	密度/(个·km ⁻²)
<0.034	极低易发区	1 251 562	38.35	104	10.80	0.092
>0.034~0.10	低易发区	777 077	23.81	133	13.81	0.190
>0.10~0.18	中等易发区	378 176	11.59	117	12.15	0.344
>0.18~0.32	高易发区	569 839	17.46	282	29.28	0.550
>0.32	极高易发区	287 103	8.80	327	33.96	1.266

4 讨论与结论

不同的影响因子对滑坡发生影响程度各不相同,进行因子影响规律及重要性分析,对滑坡防治等具有指导意义。比值比(Odds ratio, OR)通过估计逻辑回归模型中每个预测变量“有意义”增量,来评估响应变量与预测变量间的关系^[27]。在滑坡易发性评估中,OR 值等于 1 表示该影响因子对滑坡发生没有影响;OR 值大于 1 表明该因子促进滑坡发生;OR 值小于 1 表明该因子对滑坡发生具有保护作用。由表 3 可以看出,POI 核密度、水流

动力指数、距断层距离、归一化植被指数、平面曲率、坡位和距河流距离 OR 值均大于 1,表明这些因子对滑坡发生具有促进作用,即这些因子值越大,发生滑坡的可能性越大;距道路距离、微地貌、坡向、输沙指数、剖面曲率、倾坡类型、粗糙度指数、土地利用、多年平均降雨、起伏度、地形湿度指数、岩性和高程 OR 值小于 1,表明这些因子对滑坡发生具有保护作用,即这些因子值越大,发生滑坡的可能性越小。

同时,基于逻辑回归模型的逐步回归方法得到的相关系数,对影响因子具有一定的解释能力,它对因子重要性的排序具有一定参考价值。本文通过逐步回归筛选因子,共剩下 12 个影响因子,各个相关系数绝对值排序如图 5 所示。可以发现,高程、多年平均降雨、POI 核密度重要性位于前 3,对滑坡发生贡献较大。而这 3 个因子在随机森林构建的模型中,重要性也位于前 3^[13],表明这 3 个因子是导致巫山县滑坡发生最为重要的因素。

表 3 比值比计算结果
Tab. 3 Calculation results of odds ratio

影响因子	OR 值	影响因子	OR 值	影响因子	OR 值
POI 核密度	3. 829	距道路距离	0. 993	土地利用	0. 639
水流动力指数	2. 286	微地貌	0. 869	多年平均降雨	0. 508
距断层距离	1. 902	坡向	0. 863	起伏度	0. 419
归一化植被指数	1. 606	输沙指数	0. 826	地形湿度指数	0. 345
平面曲率	1. 139	剖面曲率	0. 784	岩性	0. 110
坡位	1. 063	倾坡类型	0. 756	高程	0. 022
距河流距离	1. 014	粗糙度指数	0. 651		

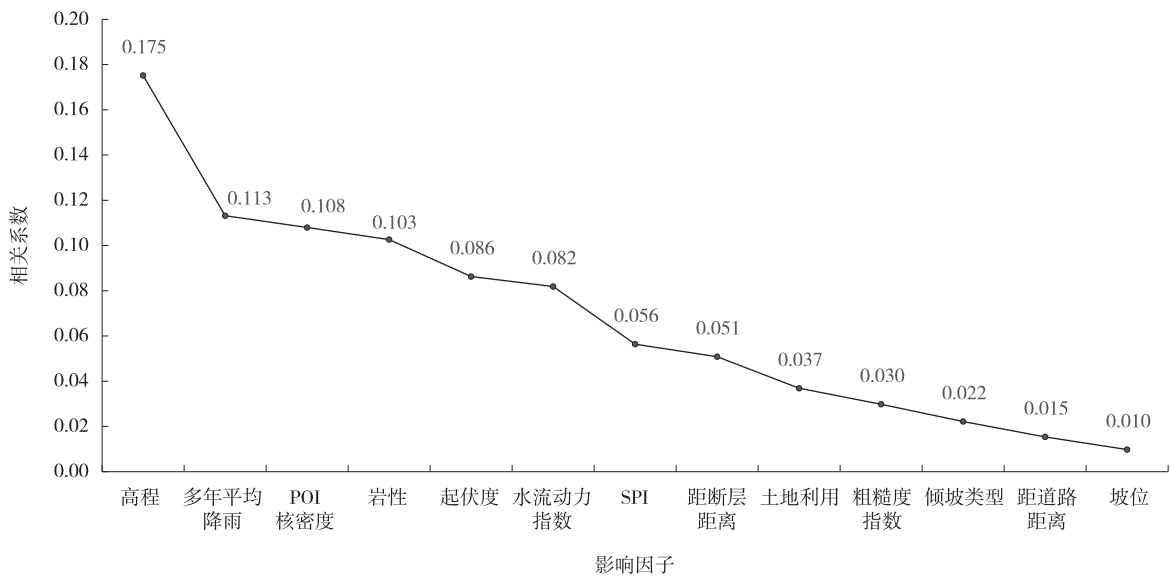


图 5 影响因子重要性排序
Fig. 5 The sorting of conditioning factors

将本研究构建的逻辑回归模型与之前研究构建的随机森林模型^[13]进行分析比较,可以发现,逻辑回归模型训练集 AUC 值(0. 803)低于随机森林模型 AUC 值(1. 00),逻辑回归测试集 AUC 值(0. 787)也同样低于随机森林测试集 AUC 值(1. 00);同时,逻辑回归模型在较低易发区滑坡密度(0. 092 个·km⁻²)高于随机森林模型(0. 036 个·km⁻²),在较高易发区滑坡密度(1. 266 个·km⁻²)低于随机森林模型(2. 01 个·km⁻²)。所有指标均表明,基于逻辑回归和随机森林构建的滑坡易发性评价模型均具有一定的参考价值,但随机森林模型性能优于逻辑回归模型性能,更适用于巫山县的滑坡易发性区划。然而,Hong 等人^[28]与本文得到了相反的结果,他们的研究表明,逻辑回归模型性能优于随机森林。尽管本文与文献^[28]所选的研究区均为山地城市,且滑坡类型相同(中/小土质滑坡),但二者的研究设计完全不同。首先,对于滑坡影响因子,本文除了考虑到地形地貌和环境条件外,还考虑了人类工程活动的影响,而人类工程活动对滑坡发生产生的影响是无法忽视的;其次,对

于训练样本,相关文献表明仅使用几十个或几百个样本数据可能无法证明机器学习的全部性能,尤其是随机森林模型,随机森林只有在大量样本训练时才具有强大的性能^[29]。本研究共使用了 10 593 个样本(963 个正样本和 9 630 个负样本)构建模型,而文献[28]仅使用 163 个正样本来构建模型。因此本文得出结论,可能是由于选择的影响因子及样本数量的不同,致使本研究与文献[28]得出了完全相反的结论。孙德亮等人^[10]选择了 16 个影响因子,且考虑人类工程活动产生的影响,并使用了大量样本(1 520 个正样本和 15 200 个负样本)来分别构建随机森林和逻辑回归模型,得出结论与本文结论保持一致,即随机森林模型性能优于逻辑回归。因此,在三峡库区地区,全面、综合考虑各种影响因素和大量样本数量情况下,基于随机森林构建的滑坡易发性评价模型性能优于基于逻辑回归构建的滑坡易发性评价模型。

综上所述,就预测能力而言,逻辑回归和随机森林模型是完全可比的。尽管本研究结果显示随机森林模型预测性能相对更好,但逻辑回归在实践中更容易应用于滑坡易发性评价,同样具有应用价值。逻辑回归模型计算简单,训练速度快,内存资源占用小,但它的线性特征使得在实际计算过程中应用该模型更容易产生多重共线性等问题。随机森林可处理复杂非线性问题,且无需进行因子筛选,可处理样本不均衡问题,不易产生过拟合,泛化能力强,但作为一个黑盒算法,它的输出结果有时无法解释,且运算速度相对较慢,建模时间长。对于不同的研究区,二者的优劣暂无确切定论,在具体研究时,需根据研究区特定的地形地质、水文环境等条件及样本数据量等多方面综合考虑,分析比较不同模型,构建最适用于具体研究区的滑坡易发性评价模型。

参考文献:

- [1] WANG Y M, FENG L W, LI S J, et al. A hybrid model considering spatial heterogeneity for landslide susceptibility mapping in Zhejiang province, China[J]. *Catena*, 2020, 188: 104425.
- [2] 许嘉慧, 孙德亮, 王月, 等. 基于 GIS 与改进层次分析法的奉节县滑坡易发性区划[J]. *重庆师范大学学报(自然科学版)*, 2020, 37(2): 36-44.
XU J H, SUN D L, WANG Y, et al. Landslide susceptibility mapping of Fengjie county based on GIS and improved analytic hierarchy process[J]. *Journal of Chongqing Normal University (Natural Science)*, 2020, 37(2): 36-44.
- [3] LUO W, LIU C C. Innovative landslide susceptibility mapping supported by geomorphon and geographical detector methods[J]. *Landslides*, 2017, 15: 465-474.
- [4] 孙德亮. 基于机器学习的滑坡易发性区划与降雨诱发滑坡预报预警研究[D]. 上海: 华东师范大学, 2019.
SUN D L. Mapping landslide susceptibility based on machine learning and forecast warning of landslide induced by rainfall[D]. Shanghai: East China Normal University, 2019.
- [5] 郭子正, 殷坤龙, 黄发明, 等. 基于滑坡分类和加权频率比模型的滑坡易发性评价[J]. *岩石力学与工程学报*, 2019, 38(2): 287-300.
GUO Z Z, YIN K L, HUANG F M, et al. Evaluation of landslide susceptibility based on landslide classification and weighted frequency ratio model[J]. *Chinese Journal of Rock Mechanics and Engineering*, 2019, 38(2): 287-300.
- [6] 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 X Y, ZHANG C S, FENG H J, 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.
- [7] 郭长宝, 唐杰, 吴瑞安, 等. 基于证据权模型的川藏铁路加查—朗县段滑坡易发性评价[J]. *山地学报*, 2019, 37(2): 240-251.
GUO C B, TANG J, WU R A, et al. Landslide susceptibility assessment based on WOE model along Jiacha-Langxian county section of Sichuan-Tibet railway, China[J]. *Mountain Research*, 2019, 37(2): 240-251.
- [8] 郭子正, 殷坤龙, 付圣, 等. 基于 GIS 与 WOE-BP 模型的滑坡易发性评价[J]. *地球科学*, 2019, 44(12): 4299-4312.
GUO Z Z, YIN K L, FU S, et al. Evaluation of landslide susceptibility based on GIS and WOE-BP model[J]. *Earth Science*, 2019, 44(12): 4299-4312.
- [9] SUN D L, WEN H J, WANG D Z, et al. A random forest model of landslide susceptibility mapping based on hyperparameter optimization using Bayes algorithm[J]. *Geomorphology*, 2020, 362: 107201.
- [10] 孙德亮, 吴健平, 文海家, 等. 山地斜坡灾变韧性大数据分析模型与应用研究: 以城口县为例[J]. *重庆师范大学学报(自然科学版)*, 2019, 36(3): 64-71.
SUN D L, WU J P, WEN H J, et al. Damage resistance mapping of mountain slopes based on geospatial big data mining[J]. *Journal of Chongqing Normal University (Natural Science)*, 2019, 36(3): 64-71.

- [11] 刘新荣,景瑞,缪露莉,等. 巫山段消落带岸坡库岸再造模式及典型案例[J]. 岩石力学与工程学报,2020,39(7):1321-1332.
- LIU X R, JING R, MIAO L L, et al. The reconstruction model and typical case analysis of the fluctuation belt of reservoir bank slope in Wushan[J]. Chinese Journal of Rock Mechanics and Engineering, 2020, 39(7): 1321-1332.
- [12] 张俊,殷坤龙,王佳佳,等. 三峡库区万州区滑坡灾害易发性评价研究[J]. 岩石力学与工程学报,2016,35(2):284-296.
- ZHANG J, YIN K L, WANG J J, et al. Evaluation of landslide susceptibility for Wanzhou district of Three Gorges reservoir[J]. Chinese Journal of Rock Mechanics and Engineering, 2016, 35(2): 284-296.
- [13] 刘睿,施妮娟,孙德亮,等. 基于GIS与随机森林的巫山县滑坡易发性区划[J]. 重庆师范大学学报(自然科学版),2020,37(3):86-96.
- LIU R, SHI S X, SUN D L, et al. Based on GIS and random forest model for landslide susceptibility mapping in Wushan county[J]. Journal of Chongqing Normal University (Natural Science), 2020, 37(3): 86-96.
- [14] 赵艳南. 三峡库区蓄水过程中滑坡变形规律研究[D]. 武汉:中国地质大学,2015.
- ZHAO Y N. Study of landslide deformation regular based on the impoundment process in the Three Gorges reservoir[D]. Wuhan: China University of Geoscience, 2017.
- [15] 张佑祥. 巫山县地质灾害现状及防治[C]//全国突发性地质灾害应急处置与灾害防治技术高级研讨会论文集. [出版地不详]:中国灾害防御协会,2010:187-192.
- ZHANG Y X. The current situation and prevention of geological disasters in Wushan county[C]//Proceedings of the National Advanced Seminar on Emergency Geological Disaster Emergency Disposal and Disaster Prevention Technology. [S. l.]: China Disaster Prevention Association, 2010: 187-192.
- [16] AYALEW L, YAMAGISHI H J G. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, central Japan[J]. Geomorphology, 2005, 65: 15-31.
- [17] ALTHUWAYNEE O F, PRADHAN B, PARK HYUCK-JIN, et al. A novel ensemble bivariate statistical evidential belief function with knowledge-based analytical hierarchy process and multivariate statistical logistic regression for landslide susceptibility mapping[J]. Catena, 2014, 114: 21-36.
- [18] MOORE L D, WILSON J P. Length-slope factors for the revised universal soil loss equation: simplified method of estimation[J]. Journal of Soil and Water Conservation, 1992, 47: 423-428.
- [19] POURGHASEMI H R, MOHAMMADY M, BISWAJEET P. Landslide susceptibility mapping using index of entropy and conditional probability models in GIS: Safarood basin, Iran[J]. Catena, 2012, 97: 71-84.
- [20] SESTRAS P, BILASCO S, ROŞCA S, et al. Landslides susceptibility assessment based on GIS statistical bivariate analysis in the hills surrounding a metropolitan area[J]. Sustainability, 2019, 11(5): 1362.
- [21] 文海家,王桂林,黄雪浪,等. 基于地形地质图的边坡稳定性初步判别方法:中国, CN107346361A[P]. 2017-11-14.
- WEN H J, WANG G L, HUANG X L, et al. A preliminary evaluation method of slope stability based on topographic map and geological map: China, CN107346361A[P]. 2017-11-14.
- [22] 谢朋,文海家,胡东萍. 山区公路沿线地震滑坡易发性区划研究[J]. 中国公路学报,2018,31(2):106-114.
- XIE P, WEN H J, HU D P. Research on susceptibility mapping of earthquake-induced landslides along highway in mountainous region[J]. China Journal of Highway and Transport, 2018, 31(2): 106-114.
- [23] 文海家,李洋,薛靖元,等. 基于大数据挖掘的山区公路沿线滑坡易发性小区划[J]. 自然灾害学报,2018,17(4):159-165.
- WEN H J, LI Y, XUE J Y, et al. Landslides susceptibility microzoning along highway in mountainous region based on mining the big data[J]. Journal of Natural Disasters, 2018, 17(4): 159-165.
- [24] 薛蒙蒙,文海家,林渝,等. 山区公路沿线斜坡物理韧性随机森林评价模型:以四川省茂县为例[J]. 水土保持通报,2020,40(4):168-175.
- XU M M, WEN H J, LIN Y, et al. Random forest evaluation model for physical toughness of slopes along mountain roads: taking Maoxian county of Sichuan province as an example[J]. Bulletin of Soil and Water Conservation, 2020, 40(4): 168-175.
- [25] 赵鑫,宋英强,刘轶伦,等. 基于卫星遥感和POI数据的人口空间化研究:以广州市为例[J]. 热带地理,2020,40(1):101-109.
- ZHAO X, SONG Y Q, LIU Y L, et al. Population spatialization based on satellite remote sensing and POI data: taking Guangzhou as an example[J]. Tropical Geography, 2020, 40(1): 101-109.
- [26] XIE P, WEN H J, MA C C, et al. Application and comparison of logistic regression model and neural network model in earthquake-induced landslides susceptibility mapping at mountainous region, China[J]. Geomatics, Natural Hazards and Risk, 2018, 9(1): 501-523.

- [27] STEGER S, BRENNING A, BELL R, et al. The influence of systematically incomplete shallow landslide inventories on statistical susceptibility models and suggestions for improvements[J]. *Landslides*, 2017, 14: 1767-1781.
- [28] HONG H Y, POURGHASEMI H R, POURTAGHI Z S. Landslide susceptibility assessment in Lianhua county (China): a comparison between a random forest data mining technique and bivariate and multivariate statistical models [J]. *Geomorphology*, 2016, 259: 105-118.
- [29] WANG Y, WU X, CHEN Z, et al. Optimizing the predictive ability of machine learning methods for landslide susceptibility mapping using SMOTE for Lishui city in Zhejiang province, China[J]. *International Journal of Environmental Research and Public Health*, 2019, 16(3): 368.

Resources, Environment and Ecology in Three Gorges Area

Landslide Susceptibility Mapping Based on Logistic Regression in Wushan County

XU Jiahui¹, ZHANG Hong¹, WEN Haijia², SUN Deliang¹

(1. The Key Laboratory of GIS Application Research, Chongqing Normal University, Chongqing 401331;

2. Key Laboratory of New Technology for Construction of Cities in Mountain Area, Ministry of Education, Chongqing University, Chongqing 400045, China)

Abstract: [Purposes] Based on previous studies, it constructed a landslide susceptibility mapping model on account of logistic regression, analyzed and compared its performance with the random forest model, and explored the most suitable landslide susceptibility mapping model in Wushan county. [Methods] 22 factors under the influence of geological conditions, geomorphology, environmental conditions, and human engineering activities were selected as the conditioning factors of landslide susceptibility, and combined with 963 historical landslide data in the study area, a geospatial database was constructed. Then, in the geospatial database, 10-fold cross-validation were used to train all the datasets (including landslide points and non-landslide points). The best optimized model was chose to simulate the landslide susceptibility in the study area. The results were divided into five classes, that is, very low susceptibility, low susceptibility, moderate susceptibility, high susceptibility and very high susceptibility region. [Findings] The area under the curve (AUC) value of receiver operating characteristic (ROC) curve showed that the prediction accuracy of training dataset, validation dataset were 0.803 and 0.787. 63.24% of the historic landslides occurred in the high-very areas. Elevation, annual average rainfall and POI kernel density did more for the occurring of landslide. [Conclusions] The landslide susceptibility mapping models were constructed on account of logistic regression and random forest both have great stability and predictive functions, but the performance of the random forest is better than that of the logistic regression, which is more suitable for the study of landslide susceptibility mapping in the Three Gorges reservoir area.

Keywords: landslide susceptibility mapping; logistic regression; Wushan county; Three Gorges reservoir; random forest

(责任编辑 许 甲)