

基于不同样本比例与超参数优化的滑坡易发性评价* ——以重庆市武隆区为例

张军以^{1,2}, 丁悦凯^{1,3}, 孙德亮^{1,3}

(1. 重庆师范大学 地理与旅游学院; 2. 重庆师范大学 三峡库区地表过程与环境遥感重庆市重点实验室;
3. 重庆师范大学 GIS 应用研究重庆市高校重点实验室, 重庆 401331)

摘要:【目的】以重庆市武隆区滑坡与非滑坡样本作为研究对象,探究不同滑坡与非滑坡样本比例与超参数优化对滑坡易发性模型准确性的影响。【方法】选取距河流距离、地形湿度指数、多年平均降水量、坡向、曲率、地形起伏度、距道路距离、坡度、POI核密度、归一化植被指数、高程等11个因子作为影响因子,在1:1~1:10作为正负样本比例范围内,采用网格搜索法与贝叶斯优化,基于XGBoost模型对研究区样本的滑坡易发性进行评价。【结果】随着样本数量中非滑坡数量的逐渐增加,XGBoost模型准确率逐步提升,曲线下面积(AUC)未发生明显改变,且较高、高易发区面积逐步减少,低、较低易发区面积逐步增加;基于超参数优化后XGBoost模型AUC值均得到提升,且贝叶斯优化后的XGBoost模型的滑坡易发性评价精度与运行速率更优。【结论】以1:10作为正负样本比例,通过贝叶斯优化所构建的基于XGBoost模型的武隆区滑坡易发性模型具有更好的预测能力与预测稳定性。

关键词:武隆区;样本比例;超参数优化;滑坡易发性

中图分类号:X43

文献标志码:A

文章编号:1672-6693(2022)05-0047-11

滑坡是中国最为常见的自然灾害之一,具有分布广泛、发生频繁、破坏性大等特点^[1]。仅在2019年,中国境内就发生滑坡4 220起,累计造成211人死亡及数亿元的经济损失^[2]。如何对滑坡进行科学合理的预测与防治愈加为人们所关注。随着人工智能技术的不断发展,基于机器学习的滑坡易发性区划逐渐受到重视^[3-5]:该方法基于高程、地形地貌、坡度、岩性、降水、耕地利用等多种自然和人为因子,采用机器学习等模型对不同研究地区的滑坡易发性进行预测,为预防滑坡的发生提供了重要的科学支持^[6-7]。

机器学习方法是滑坡易发性评价中使用最为广泛的方法之一^[8-9],滑坡正负样本的选取是滑坡易发性评价中最重要的一步。在滑坡易发性研究中,通常将历史滑坡点作为正样本,非滑坡区域作为负样本^[10-11]。目前,不同学者在滑坡易发性评价研究中应用不同诊断模型时,往往采用不同比例的正负样本进行研究。刘睿等人^[12]以1:1作为正负样本比例,采用随机森林模型对重庆市奉节县进行滑坡易发性区划;刘艳辉等人^[13]采用1:2作为正负样本比例,对四川省青川县进行滑坡易发性区划;任敬^[14]则采用1:10作为正负样本比例,对贵州省进行滑坡灾害的风险评价。不同比例的正负样本是否对滑坡易发性区划的模型精度存在影响是一个需要明确的问题。同时,为了提升滑坡易发性模型的评价精度与运行效率,基于不同方法的模型超参数优化法逐渐得到广泛应用^[15],其中网格搜索法与贝叶斯优化是最常见的两种优化算法。周昭飞^[16]采用网格搜索法对支持向量机模型进行优化,对陕西省延安市宝塔区进行了滑坡易发性区划,结果表明超参数优化后的模型预测精度达87.26%;杨灿等人^[17]采用贝叶斯优化算法对4种常用滑坡易发性模型进行优化,并对湖南省中部地区进行了滑坡易发性评价,结果表明优化后模型精度提高了3.5%~7.1%。然而有关两种优化算法对滑坡易发性模型的提升效果与运行效率对比的研究目前较为缺乏。

* 收稿日期:2022-05-26 修回日期:2022-07-10 网络出版时间:2022-09-19 10:53

资助项目:国家自然科学基金(No. 41901214);重庆英才计划“包干制”项目(No. CSTC2021ycjhbzxm0109);重庆市教育委员会人文社会科学研究重点项目(No. 22SKGH090)

第一作者简介:张军以,男,副教授,博士,研究方向为乡村转型发展与区域生态环境,E-mail:hellojunyi@cqnu.edu.cn;通信作者:孙德亮,男,副教授,博士,E-mail:67130612@qq.com

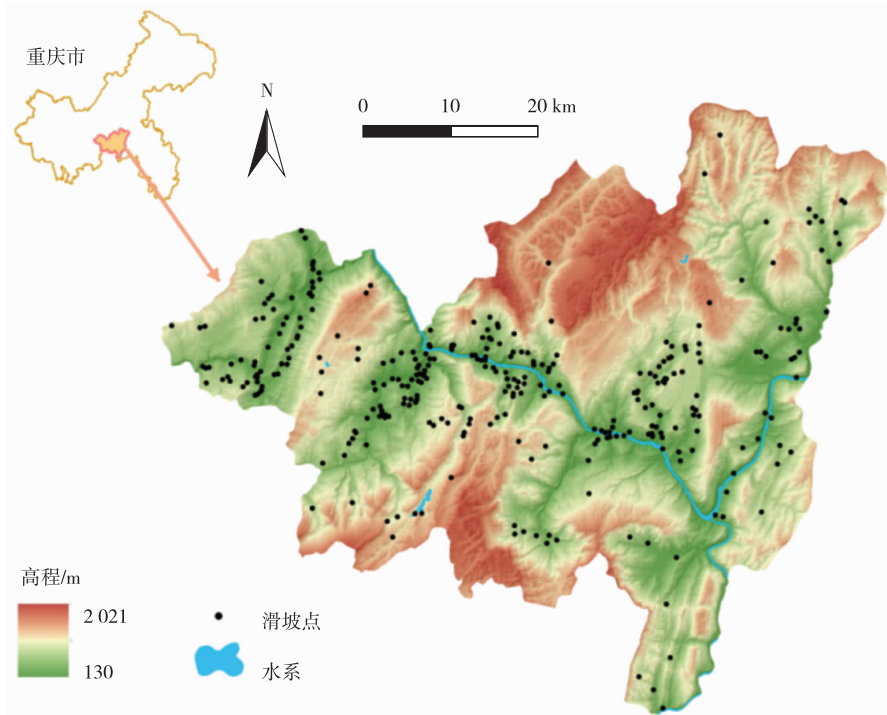
网络出版地址:https://kns.cnki.net/kcms/detail/50.1165.N.20220916.1822.025.html

基于上述研究背景,本研究以重庆市武隆区作为研究区域,以 XGBoost 模型作为滑坡易发性模型,采用高程、距道路距离、坡度、POI 核密度、归一化植被指数(NDVI)、地形起伏度等 11 个因子,以 10 种不同滑坡正负样本比例(1:1~1:10)构建模型,并通过网格搜索法与贝叶斯优化分别对表现最优的模型进行超参数优化,探究不同的滑坡正负样本比例与超参数优化对于滑坡易发性模型精度的影响作用,以便为滑坡灾害的治理与预测提供更为有效的科学支持。

1 研究区与数据来源

1.1 研究区概况

武隆区(东经 $107^{\circ}14' \sim 108^{\circ}05'$ 、北纬 $29^{\circ}02' \sim 29^{\circ}40'$)位于重庆市东南部、乌江下游地区,属于武陵山区与大娄山区的峡谷地带区,总面积为 $2\,901\text{ km}^2$ 。该区域地形以山地丘陵为主,平均海拔 987 m ;地势呈东北高、西南低;主要地貌类型为喀斯特地貌。武隆区气候温湿,四季分明,降水丰富,年平均降水量为 $1\,136\text{ mm}$,气候为亚热带湿润季风气候。整体上看,武隆区地形地貌高差变化明显,是重庆市滑坡、泥石流等地质灾害重要发生区,全区累计拥有 319 个历史滑坡(图 1)。



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

图 1 研究区位置及滑坡分布

Fig. 1 Location and landslide distribution in the study area

1.2 数据来源

本研究的主要数据来源如下:1) 重庆市地质环境监测站所提供的研究区滑坡历史数据;2) ASTER GDEM 数字高程模型,用于提取坡度、坡向等滑坡相关影响因子;3) 重庆市水利局所提供的河网、水系等数据,用于提取滑坡距河流距离;4) 重庆市气象局所提供的降水等数据,用于分析降水对滑坡的影响作用;5) 重庆市交通委员会所提供的路网数据,用于提取滑坡距道路距离;6) Landsat 8 遥感影像,用于提取不同波段数据,计算研究区的 NDVI;7) 基于爬虫技术获取 2016 年研究区的 POI 数据,用于分析人类活动对滑坡的影响作用。

2 地理空间数据库

滑坡的发生受区域地质地貌、气象水文、岩性等多方面条件的综合影响,诱发机理较为复杂,选取科学合理的评价因子体系对于准确进行滑坡易发性评价具有重要的作用^[2]。已有研究表明,滑坡易发性评价因子的数量

并非越多越好,且合理的评价因子体系因研究地环境的不同而有所差异^[18]。因此,本研究综合考虑武隆区历史滑坡的空间分布规律以及内在特征,并参考已有相关研究成果^[19-21],选取 11 个滑坡影响因子构建滑坡易发性影响因子(图 2)评价指标体系,其中:高程、地形起伏度、曲率、坡向、坡度等为地形地貌因子;地形湿度指数、距河流距离、多年平均降水量、NDVI 等为环境条件因子;距道路距离和 POI 核密度为人类活动因子。

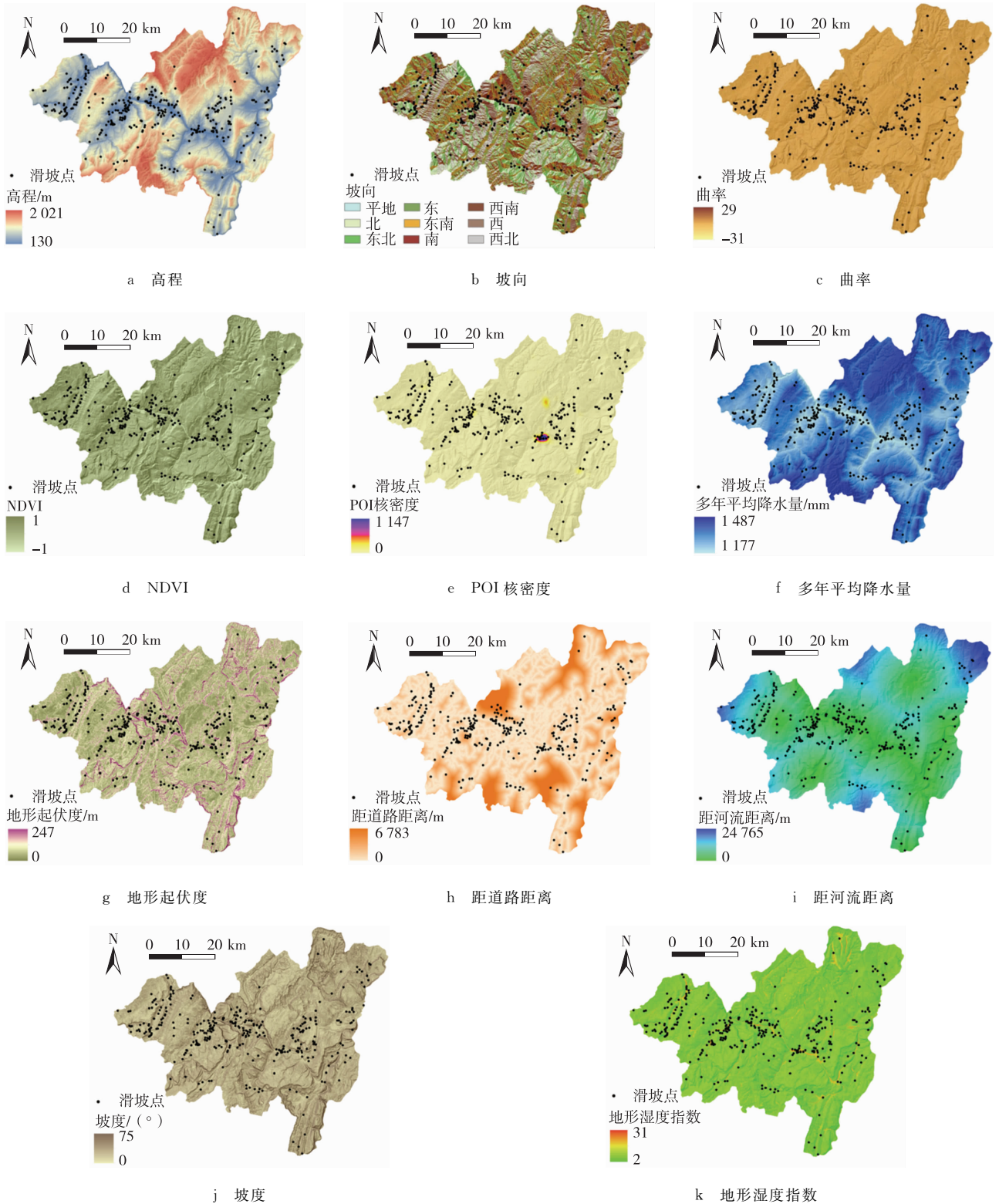


图 2 武隆区滑坡影响因子专题图层

Fig. 2 Thematic layer of landslide impact factors in Wulong district

3 研究方法

3.1 XGBoost 算法

XGBoost 算法是由 Chen 等人^[22]于 2016 年提出的一种新的机器学习算法,是一种高级的梯度增强算法。该算法通过对 Boosting 算法进行改进,在求解损失函数极值时采用牛顿法,并加入正则化项,使得算法的准确率与工作运行效率得到有效提升。相较于随机森林、逻辑回归等传统的机器学习分类算法,XGBoost 算法因正则化的加入使模型更加简洁高效,有效防止了过拟合的发生,且算法收敛速度更快,支持并行处理,计算效率更高。同时,XGBoost 算法可通过统计不同因子的贡献程度,从而获取不同因子的特征重要性。模型的目标函数($O^{(t)}$)如下:

$$O^{(t)} \approx \sum_{j=1}^K \left[G_j W_j + \frac{1}{2} (H_j + \lambda) w_j^2 \right] + \gamma T,$$

其中: T 为叶子结点个数; w_j 为叶子结点权重; K 为模型中树的总棵数; λ, γ 为相关参数,用于防止模型过拟合; G_j 与 H_j 分别是损失函数一阶与二阶偏导的累加值之和。

XGBoost 模型工作流程如下:首先初始化每个样本的预测值,并定义损失函数;其次,根据损失函数计算每个样本预测值的导数,从而根据导数信息建立决策树;最后,利用所建立的决策树预测样本并累加至原来的值上,进行多次循环运算,直至精度达到最高时停止。

3.2 超参数优化

3.2.1 网格搜索法 网格搜索法通过设定参数调整范围,对参数进行随机组合,并将每种组合以模型的形式进行实例化运行,尝试每种组合通过交叉验证,将平均得分最高的组合作为模型的最佳参数选择。

3.2.2 贝叶斯优化 贝叶斯优化是由 Snoek 等人^[23]于 2012 年提出超参数优化方法,该方法一经提出便被广泛运用于模型参数优化。贝叶斯优化以目标函数作为对象,通过添加样本点,不断更新目标函数;与网格搜索法类似,该方法将参数组合以模型的形式进行实例化运行,并采用交叉验证,将平均得分最高的组合作为模型最佳参数选择,有关公式如下:

$$x^{(*)} = \arg \min_{x \in S} f(x),$$

其中: $x^{(*)}$ 为最优参数组合, S 是 x 的候选集, $f(x)$ 是目标函数。

概率代理模型与采集函数是贝叶斯优化最重要的两部分。首先,随机读取 n 个不重复的点,并对它们进行初始化,获取它们的函数值,随后采用高斯过程(Gaussian process)^[24]构建代理模型,再根据历史经验构造采集函数,从而获取下一步的评估点,获取最佳的模型参数,使得先验知识在贝叶斯优化算法中被充分利用,进而提升了超参数优化效率。

3.3 评价指标

受试者工作特征(Receiver operating characteristic, ROC)曲线是机器学习领域运用最为广泛的模型评价指标^[25]。ROC 曲线下的面积(Area under curve, AUC)可用于评估模型的精度,AUC 值越接近 1,表明模型的诊断效果越好。本研究采用 AUC 值与模型准确率作为模型优劣的主要评价指标来评估模型稳定性与预测精度能力。

4 结果与分析

4.1 基于不同正负样本比例的滑坡易发性评价

以武隆区历史滑坡点作为正样本,并以滑坡点作为对象生成 500 m 缓冲区,将缓冲区与地表水体共同剔除作为非滑坡区域,运用 ArcGIS 10.4 的子集要素工具按比例随机提取非滑坡点作为负样本,同时按照 7 : 3 比例划分训练集与测试集,运用 XGBoost 模型预测研究区滑坡发生概率的空间分布情况。依据专家经验,按照 $[0, 0.15]$, $(0.15, 0.3]$, $(0.3, 0.45]$, $(0.45, 0.55]$ 和 $(0.55, 1.0]$ 作为分类区间,将研究区滑坡易发性图划分为低易发区、较低易发区、中易发区、高易发区、极高易发区等 5 个滑坡易发性等级,得到武隆区滑坡易发性区划图(图 3)。通过计算可以发现,随着非滑坡点负样本数量的增加,低易发区、较低易发区的面积逐渐增加,高易发区、较

高易发区的面积逐渐减少。模型中非滑坡点数量的增加,使得模型正负样本比例更加趋向研究区真实的滑坡与非滑坡数量比例,可进一步提升预测结果的准确性和现实客观性。表 1 显示:随着非滑坡负样本数量的增加,模型的 AUC 值未发生明显改变,但均高于 50%,且模型的准确率逐步上升——正负样本比例为 1:10 的模型准确率比正负样本比例为 1:1 的模型准确提升了 18%。因此,对于研究区滑坡易发性的模型构建而言,1:10 的正负样本比例是构建滑坡易发性模型的最佳样本比例。

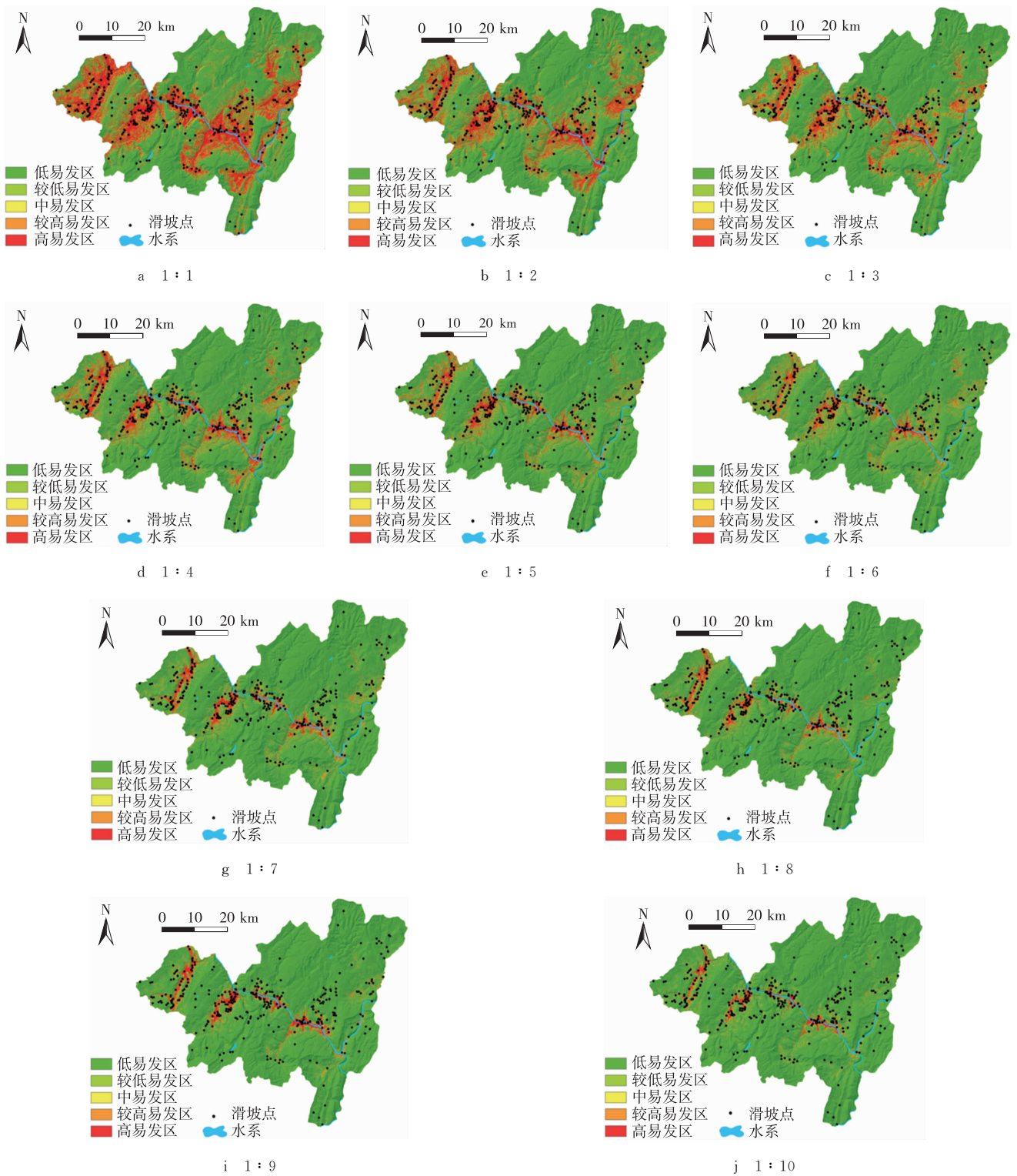


图 3 不同样本比例下的滑坡易发性区划

Fig. 3 Landslide susceptibility based on different sample proportions

表 1 不同样本比例下的模型精度与 AUC 值

Tab. 1 Model accuracy and AUC value under different sample proportions

参数	正负样本比例									
	1:1	1:2	1:3	1:4	1:5	1:6	1:7	1:8	1:9	1:10
非滑坡点数量/个	319	638	957	1 276	1 595	1 914	2 233	2 552	2 871	3 190
AUC 值	83%	79%	79%	84%	82%	82%	81%	81%	79%	80%
准确率	74%	74%	79%	82%	85%	88%	88%	88%	91%	92%

4.2 基于不同超参数优化的滑坡易发性评价

以 1:10 作为滑坡正负样本比例,分别采用网格搜索法和贝叶斯优化算法对 XGBoots 模型进行超参数优化,优化了模型的 8 个参数,可以发现网格搜索法和贝叶斯优化后的 XGBoots 模型参数值大部分均呈现较为明显的差异(表 2)。

表 2 主要优化的超参数

Tab. 2 Main optimized hyperparameters

名称	含义	网格搜索法	贝叶斯优化	名称	含义	网格搜索法	贝叶斯优化
gamma	损失减少阈值	0.71	0.011 1	n_estimators	学习器的数量	2	324
learning_rate	学习率	0.1	0.011	reg_alpha	L1 正则化	0.05	0.001
max_depth	树深度	3	3	subsample	样本子采样	0.8	0.3
min_child_weight	子集最小权重	3.3	5.04	colsample_bytree	列采样	0.999 9	0.999 9

将优化后的超参数代入 XGBoost 模型,比较两种方法对模型精度和效率的影响作用。由表 3 可知,基于贝叶斯优化后的 XGBoost 模型在 AUC 值和准确率的提升幅度以及运行时间方面均优于网格搜索法的 XGBoost 模型,特别是运行时间方面前者远远少于后者。

表 3 基于超参数优化的 XGBoost 模型效果对比

Tab. 3 Comparison of XGBoost model effects based on parameter optimization

评价指标	网格搜索法	贝叶斯优化	评价指标	网格搜索法	贝叶斯优化
AUC 值	81.96%	84.24%	AUC 值提升幅度	1.6%	3.80%
准确率	93.07%	93.16%	准确率提升幅度	0.76%	0.85%
运行时间/s	61 498	1 277			

4.3 滑坡易发性区划结果

选取贝叶斯优化后的 XGBoost 模型预测整个研究区,得到的武隆区的滑坡易发性区划图(图 4)。由该图可知:武隆区滑坡易发性区划以低易发区、较低易发区为主,且集中分布于武隆区的南部与北部;较高易发区、高易发区则集中分布于武隆区西部和乌江及它的支流两岸。同时,通过统计历史滑坡点与不同等级的栅格单元的滑坡点个数可知(表 4),武隆区滑坡易发性等级从低易发区至高易发区,滑坡点密度明显提高。可以看出,优化后 XGBoost 模型获得的滑坡易发性区划图较为合理。

4.4 滑坡影响因子重要性排序

不同的滑坡影响因子对滑坡的发生起着不同的影响作用,探究不同影响因子重要性对于掌握滑坡发生概率及滑坡点的空间分布规律有着重要意义。本文以每个滑坡影响因子的增益^[22]作为参考指标,用于分析不同滑坡影响因子对滑坡的影响作用,并对武隆地区滑坡各滑坡因子的重要性进行排序,如图 5 所示。

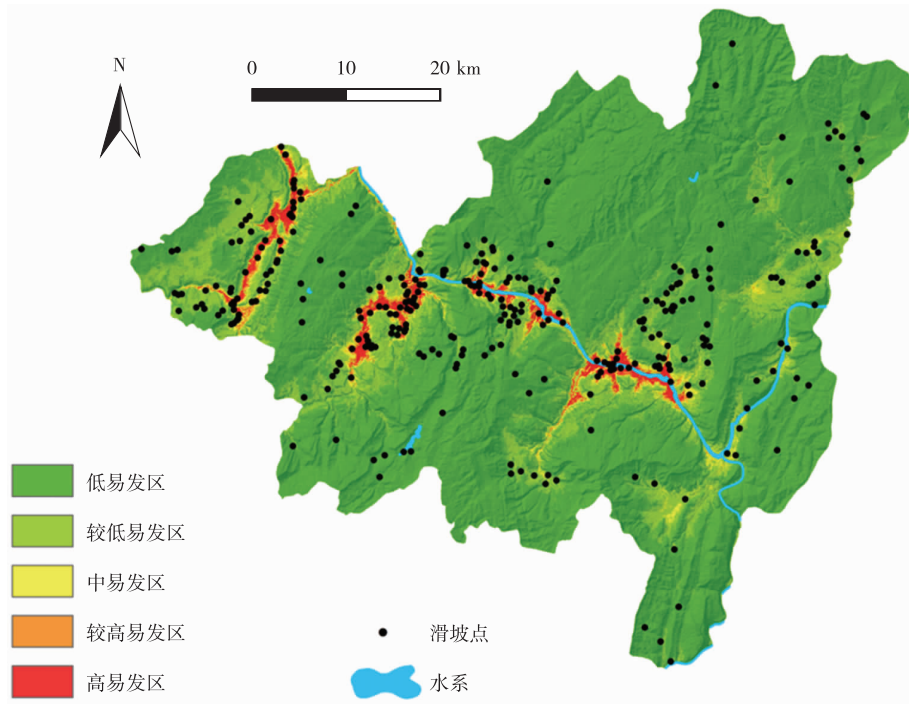


图 4 武隆区滑坡易发性区划图

Fig. 4 Landslide susceptibility of Wulong district

表 4 武隆区滑坡易发性分级统计

Tab. 4 Classification statistics of landslide susceptibility in Wulong district

滑坡概率	滑坡易发性等级	栅格数/个	面积/km ²	滑坡数/个	滑坡密度/(个·km ⁻²)
0~0.15	低易发区	2 535 618	2 282.056 0	100	0.043
>0.15~0.3	较低易发区	462 115	415.903 5	106	0.254
>0.3~0.45	中易发区	83 446	75.101 4	28	0.372
>0.45~0.55	高易发区	32 918	29.626 2	18	0.607
>0.55~1.0	极高易发区	68 442	61.597 8	67	1.087

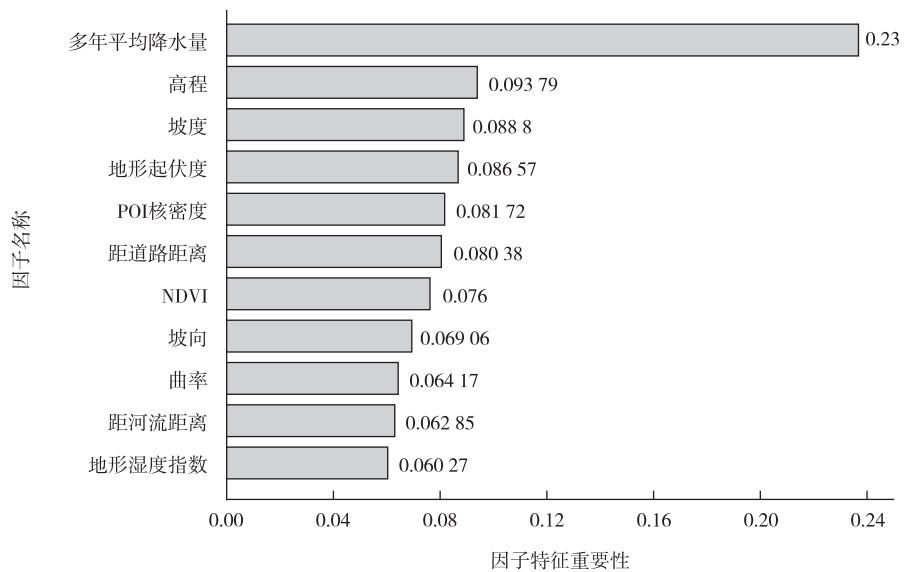


图 5 影响因子重要性排序

Fig. 5 Importance ranking of influencing factors

以滑坡影响因子中的多年平均降水量、高程、坡度为例,统计分析它们的量级变化对于滑坡的分布影响作用。由图 6 可知,滑坡密度的增加与多年平均降水量的升高呈现相反趋势,这一结果与重庆市已有研究结果^[12,26]以及国内其他地区相关研究成果^[27-28]基本一致,当多年平均降水量达到一定阈值后,该因子对应的滑坡密度急剧下降。多年平均降水量与滑坡发生关系主要是统计学上的相关性。降水作为滑坡发生的重要诱发因素,降水入渗导致斜坡内部组织的应力的变化,降低坡体的抗剪强度,超过阈值从而导致斜坡失稳。分析认为同一地区由于岩性具有同质性,在空间尺度较小范围内滑坡失稳时的抗剪强度阈值对应一定的年降水量。实际情况下,小时降水强度才是造成山洪、滑坡等地质灾害的主要原因。已有研究表明,重庆地区多年的小时强降水对年总降水量和降水频次的贡献率相对稳定,贡献率分别为 1%、15%^[29]。分析认为研究区滑坡发生的小时降水强度对应的多年平均降水量阈值在 1 200 mm 左右,故在这一阈值下研究区滑坡密度最高。滑坡发生后新的滑坡生成在地表堆积物积累、岩性结构变化等方面需要较长的生长过程;而研究区范围较小,岩性、土壤等空间差异性较小,新的滑坡不断发育,绝大部分在降水量达到 1 200 mm 左右时发生滑坡;因此当降水量再持续增加后滑坡的密度将持续降低。同时,降水量的增加促进了当地的植被发育,减少了水土流失,植被根系生产的固结合作用提高了边坡的抗剪程度和抗渗性,滑坡发生的可能性降低。已有研究表明重庆地区山前平坝和丘陵区易出现强降水,而在中山等区域反而不利于强降水的产生^[29]。因此,受小时强降水地形高低的分布差异的影响,高程与滑坡密度也呈现负相关关系。此外,自然环境中,高程较低的区域是人类活动的主要场所,该区域人口密度大,人类活动强烈,导致这部分区域地表环境受人类活动胁迫干扰严重,地表堆积物较为松散,提升了滑坡发生的概率。同时,坡脚分布着许多松散岩体,导致土质斜坡的形成,促进了滑坡的形成;而高程较高的区域受人类活动影响较少,并且富含大量植被,缺少形成滑坡的松散的岩石风化物等堆积,因而更不易发生滑坡^[30]。对于坡度因子而言,滑坡集中分布区与 10°~40°坡度区间相对应,该区域岩体受重力分力作用较大,岩体及表层堆积物易沿滑体面下滑。此外,坡度较小的区域地势平坦,不易发生滑坡,而坡度很大的区域则更易发生崩塌,而非出现滑坡^[31]。

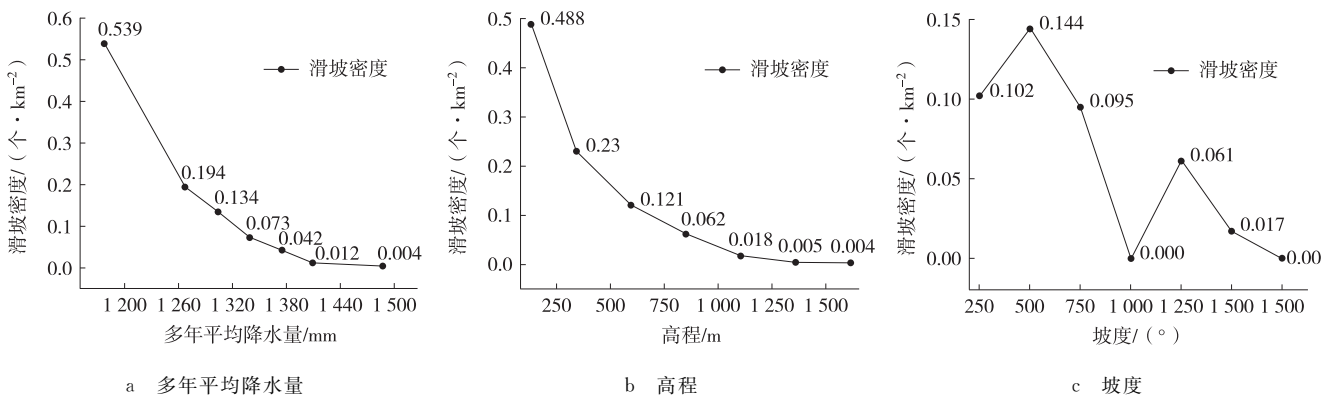


图 6 不同因子滑坡密度统计图

Fig. 6 Statistical diagram of different factor landslide density

5 结束语

本研究以重庆市武隆区为研究区,利用卫星遥感影像、水文气象、交通道路等多源数据,提取高程、多年平均降水量、坡向、距道路距离等 11 个滑坡影响因子,采用 XGBoost 模型构建了武隆区滑坡易发性模型,并比较不同的滑坡正负样本比例对 XGBoost 模型结果的影响以及网格搜索法与贝叶斯优化在该模型运用中的差别,主要研究结果如下:

1) 对于武隆区滑坡易发性区划而言,滑坡正负样本比例为 1 : 10 的 XGBoost 模型效果最佳,模型准确率为 92%。不同滑坡正负样本比例对 XGBoost 模型准确率影响较大,但对模型 AUC 值的影响较小。随着模型非滑坡点负样本数量增大,较高易发区、高易发区面积逐渐减少而低易发区、较低易发区面积逐渐增加。

2) 贝叶斯优化算法与网格搜索法均对 XGBoost 模型具有较好的优化效果,以武隆区为例,通过贝叶斯优化后的 XGBoost 模型 AUC 值提升了 3.8%,准确率提升 0.86%,通过网格搜索法优化后 XGBoost 模型 AUC 值提升了 1.6%,准确率提升了 0.76%。贝叶斯优化算法用于优化 XGBoost 模型更佳,更有助于模型精度与运行速率的提升。

滑坡发生的成因机理复杂,不同区域内致同一滑坡灾因子对滑坡的影响作用存在明显差异,而本研究采用 XGBoost 模型为黑箱模型,还无法对模型的评估结果进行基于物理机制的解释,因此后期有必要采用其他方法进一步提高滑坡易发性模型的可解释性。

参考文献:

- [1] 殷坤龙,朱良峰. 滑坡灾害空间区划及 GIS 应用研究[J]. 地学前缘,2001,8(2):279-284.
YIN K L,ZHU L F. Landslide hazard zonation and application of GIS[J]. Earth Science Frontier,2001,8(2):279-284.
- [2] 白志刚,刘启蒙,刘瑜. 基于熵指数与随机森林耦合的滑坡易发性评价[J/OL]. 人民长江. [2022-09-04]. <http://kns.cnki.net/kcms/detail/42.1202.TV.20220114.1634.002.html>.
- BAI Z G,LIU Q M,LIU Y. Landslide susceptibility evaluation based on coupling of entropy index and random forest[J/OL]. Yangtze River. [2022-09-04]. <http://kns.cnki.net/kcms/detail/42.1202.TV.20220114.1634.002.html>.
- [3] 罗路广,裴向军,崔圣华,等. 九寨沟地震滑坡易发性评价因子组合选取研究[J]. 岩石力学与工程学报,2021,40(11):2306-2319.
LUO L G,PEI X J,CUI S H, et al. Combined selection of susceptibility assessment factors for Jiuzhaigou earthquake-induced landslides[J]. Chinese Journal of Rock Mechanics and Engineering,2021,40(11):2306-2319.
- [4] 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 R Z,HU X D,MEI H B, 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.
- [5] 张虹,辜庆渝,孙诚彬,等. 基于可解释性机器学习的丘陵缓坡地区滑坡易发性区划研究[J]. 重庆师范大学学报(自然科学版),2022,39(3):78-92.
ZHANG H,GU Q Y,SUN C B, et al. Landslide susceptibility mapping in hilly and gentle slope region based on interpretable machine learning[J]. Journal of Chongqing Normal University (Natural Science),2022,39(3):78-92.
- [6] FLEUCHAUS P,BLUM P,WILDE M, et al. Retrospective evaluation of landslide susceptibility maps and review of validation practice[J]. Environmental Earth Sciences,2021,80(15):485.
- [7] OZTURK U,PITTORE M,BEHLING R, et al. How robust are landslide susceptibility estimates?[J]. Landslides,2021,18(2):681-695.
- [8] SUN D L,XU J H,WEN H J, et al. An optimized random forest model and its generalization ability in landslide susceptibility mapping:application in two areas of Three Gorges reservoir,China[J]. Journal of Earth Science,2020,31(6):1068-1086.
- [9] NGUYEN V V,PHAM B T,VU B T, et al. Hybrid machine learning approaches for landslide susceptibility modeling[J]. Forests,2019,10(2):157.
- [10] 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.
- [11] 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.
- [12] 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 Y 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.

- [13] 刘艳辉,方然可,苏永超,等.基于机器学习的区域滑坡灾害预警模型研究[J].工程地质学报,2021,29(1):116-124.
LIU Y H,FANG R K,SU Y C,et al. Machine learning based model for warning of regional landslide disasters[J]. Journal of Engineering Geology,2021,29(1):116-124.
- [14] 任敬.基于不同尺度的贵州省滑坡灾害发育分布规律与定量风险评价研究[D].成都:成都理工大学,2019.
REN J. Study about the development and distribution law and quantitative risk assessment of landslide disaster in Guizhou province based on different scales[D]. Chengdu:Chengdu University of Technology,2019.
- [15] WANG B H,GONG N Z Q. Stealing hyperparameters in machine learning[EB/OL]. (2019-09-07)[2022-06-22]. <https://arxiv.org/abs/1802.05351>.
- [16] 周昭飞.基于不确定多分类支持向量机在滑坡危险性预测的应用[D].赣州:江西理工大学,2016.
ZHOU Z F. Landslide hazard assessment based on uncertain multi-classification support vector machine method[D]. Ganzhou: Jiangxi University of Technology,2016.
- [17] 杨灿,刘磊磊,张遗立,等.基于贝叶斯优化机器学习超参数的滑坡易发性评价[J].地质科技通报,2022,41(2):228-238.
YANG C,LIU L L,ZHANG Y L,et al. Machine learning based on landslide susceptibility assessment with Bayesian optimized the hyperparameters[J]. Bulletin of Geological Science and Technology,2022,41(2):228-238.
- [18] 彭国骄.库岸消落带滑坡编目及危险性评价机器学习模型应用研究[D].重庆:重庆大学,2020.
PENG G J. Study on the landslide cataloguing and application of machine learning model for danger assessment in the fluctuating zone[D]. Chongqing:Chongqing University,2020.
- [19] 刘月,王宁涛,周超,等.基于 ROC 曲线与确定性系数法集成模型的三峡库区奉节县滑坡易发性评价[J].安全与环境工程,2020,27(4):61-70.
LIU Y,WANG N T,ZHOU C,et al. Evaluation of landslide susceptibility based on ROC and certainty factor method in Fengjie county,Three Gorges reservoir[J]. Safety and Environmental Engineering,2020,27(4):61-70.
- [20] 何太蓉,杨达源.重庆市武隆县地质灾害特征及防治对策研究[J].重庆师范大学学报(自然科学版),2005,22(1):62-65.
HE T R,YANG D Y. The characteristics and prevention countermeasures of geological hazards in Wulong county of Chongqing [J]. Journal of Chongqing Normal University (Natural Science),2005,22(1):62-65.
- [21] 姜树辉.重庆市武隆县滑坡地质灾害遥感调查与评价分析研究[D].重庆:重庆交通大学,2013.
JIANG S H. The interpretation and analysis about landslide disaster based on remote sensing in Wulong county,Chongqing[D]. Chongqing:Chongqing Jiaotong University,2013.
- [22] CHEN T Q,GUESTIN C. XGBoost: a scalable tree boosting system [C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco: Association for Computing Machinery, 2016:785-794.
- [23] SNOEK J,LAROCHELLE H,ADAMS R P,et al. Practical bayesian optimization of machine learning algorithms[EB/OL]. (2012-08-29)[2022-06-22]. <https://arxiv.org/abs/1206.2944v1>.
- [24] JAKKALA K. Deep Gaussian processes:a survey[EB/OL]. (2021-06-21)[2022-06-22]. <https://arxiv.org/abs/2106.12135>.
- [25] KOOTTE R S,LEVIN E,SALOJÄRV L J,et al. Improvement of insulin sensitivity after lean donor feces in metabolic syndrome is driven by baseline intestinal microbiota composition[J]. Cell Metabolism,2017,26(4):611-619.
- [26] 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.
- [27] 崔阳阳.基于不同评价单元的滑坡易发性评价方法研究:以陕西省洛南县为例[D].西安:西安科技大学,2021.
CUI Y Y. A comparative study on evaluation methods of landslide susceptibility based on different rvaluation units;a case study on Luonan county,Shaanxi province[D]. Xi'an:Xi'an University of Science and Technology,2021.
- [28] 丛佳伟.天水地区降雨型滑坡的降雨阈值研究[D].兰州:兰州大学,2020.
CONG J W. Rainfall threshold research of rainfall induced landslides in Tianshui area[D]. Lanzhou:Lanzhou University,2020.
- [29] 方德贤,董新宇,邓承之,等.2008—2016 年重庆地区降水时空分布特征[J].大气科学,2020,44(2):327-340.
Fang D X,Dong X N,Deng C Z,et al. Temporal and spatial distribution of precipitation in Chongqing during 2008—2016[J]. Chinese Journal of Atmospheric Sciences,2020,44(2):327-340.

- [30] SUN D L, SHI S X, WEN H J, et al. A hybrid optimization method of factor screening predicated on GeoDetector and random forest for landslide susceptibility mapping[J]. *Geomorphology*, 2021, 379(1): 107623.
- [31] 凌炳, 余敏. 滑坡灾害坡度坡向敏感性分析研究: 以云南大关县为例[J]. *城市地质*, 2015, 10(3): 66-68.
LING B, YU M. Sensitivity factors of slope gradient and slope direction of landslides in Daguan county, Yunnan[J]. *Urban Geology*, 2015, 10(3): 66-68.

Resources, Environment and Ecology in Three Gorges Area

Landslide Susceptibility Evaluation Based on Different Sample Proportion and Super Parameter Optimization: Take Wulong District of Chongqing Municipality as an Example

ZHANG Junyi^{1,2}, DING Yuekai^{1,3}, SUN Deliang^{1,3}

(1. School of Geography and Tourism, Chongqing Normal University;

2. Chongqing Key Laboratory of Surface Process and Environment Remote Sensing in the
Three Gorges Reservoir Area, Chongqing Normal University;

3. Chongqing Key Laboratory of GIS Application, Chongqing Normal University, Chongqing 401331, China)

Abstract: [Purposes] Taking the landslide and non landslide samples in Wulong district of Chongqing municipality as the research object, it explores the influence of different landslide and non landslide sample proportions and superparametric optimization on the accuracy of landslide susceptibility model. [Methods] 11 landslide factors including distance from river, TWI, average annual precipitation, aspect, curvature, relief amplitude, distance from road, slope, POI kernel density, NDVI, to elevation are selected as impact factors. Within the range from 1 : 1 to 1 : 10 as positive and negative sample proportions, grid search method and Bayesian optimization algorithm are used to evaluate the landslide susceptibility of samples in the study area with XGBoost algorithm as the main model. [Findings] With the gradual increase of the number of non landslides in the number of samples, the accuracy of XGBoost model gradually improved, the AUC value did not change significantly, and the area of high and high prone areas gradually decreased, and the area of low and low prone areas gradually increased; The AUC value of XGBoost model based on super parameter optimization is improved, and the accuracy and running speed of landslide susceptibility evaluation of XGBoost model based on Bayesian optimization are better. [Conclusions] Taking 1 : 10 as the positive and negative sample ratio, the landslide susceptibility model based on XGBoost model constructed by Bayesian optimization in Wulong district has better prediction ability and stability.

Keywords: Wulong district; sample proportion; superparametric optimization; landslide susceptibility

(责任编辑 方 兴)