

DOI:10.13870/j.cnki.stbxb.2024.01.030

张兴存, 蒋玉琳, 王玉杰, 等. 基于多目标优化方法的滑坡易发性评价[J]. 水土保持学报, 2024, 38(1): 104-112, 121.

ZHANG Xingcun, JIANG Yulin, WANG Yujie, et al. Evaluation of landslide susceptibility based on multi-objective optimization method[J]. Journal of Soil and Water Conservation, 2024, 38(1): 104-112, 121.

基于多目标优化方法的滑坡易发性评价

张兴存, 蒋玉琳, 王玉杰, 祁子寒, 王云琦

(北京林业大学水土保持学院重庆缙云山三峡库区森林生态系统国家定位观测研究站,
北京林业大学水土保持学院重庆三峡库区森林生态系统教育部野外科学观测研究站, 北京 100083)

摘要: [目的] 在滑坡易发性评价中, 滑坡预测模型的选取和优化对运算过程的高效性和预测结果的准确性至关重要。针对现有单目标遗传优化算法(genetic algorithm, GA)易陷入早熟、局部搜索能力差、全局优化速度慢等问题, 拟提出一种新的优化算法框架, 将多目标遗传算法中的经典算法—带精英选择策略的非支配排序算法(the nondominated sorting genetic algorithm with an elite strategy, NSGA-II)与常用机器学习模型[随机森林(random forest, RF)、支持向量机(support vector machine, SVM)]相结合, 进行滑坡易发性预测。与单目标优化不同的是, NSGA-II 算法可同时进行特征选择和超参数优化, 并使预测模型同时实现最优准确度、召回率、精密度和 AUC(area under curve, AUC)。[方法] 以三峡库区重庆段为研究区, 从模型精度评价、滑坡灾害易发性分区图、分区统计 3 个方面对 4 种优化模型(RF-GA、SVM-GA、RF-NSGA-II、SVM-NSGA-II)进行对比分析。[结果] NSGA-II 较 GA 优化效果更明显, 在模型评价和滑坡易发性分区方面, RF-NSGA-II 模型具有更高的预测性能, 4 项评价价值分别为 80.91%, 81.89%, 80.07%, 88.60%, 证明 NSGA-II 优化算法的有效性; 极低至极高危险区面积占比依次为 23.06%, 22.46%, 22.96%, 19.99%, 11.53%, 验证了 RF-NSGA-II 模型的可靠性。由 RF-NSGA-II 模型预测得到的易发性图表明, 高和极高易发性区集中在研究区北部, 且由东向西呈带状分布。[结论] 研究采取的基于多目标选择的 RF-NSGA-II 模型, 为滑坡易发性评价中机器学习模型调优提供新思路。

关键词: 滑坡易发性; 多目标优化; 随机森林; 支持向量机

中图分类号: P642.22

文献标识码: A

文章编号: 1009-2242-(2024)01-0104-10

Evaluation of Landslide Susceptibility Based on Multi-objective Optimization Method

ZHANG Xingcun, JIANG Yulin, WANG Yujie, QI Zihan, WANG Yunqi

(National Positioning Observatory of Forest Ecosystem in Three Gorges Reservoir Area, Jinyun Mountain, Chongqing, School of Soil and Water Conservation, Beijing Forestry University, Ministry of Education Field Scientific Observatory of Forest Ecosystem in Three Gorges Reservoir Area, Chongqing, School of Soil and Water Conservation, Beijing Forestry University, Beijing 100083, China)

Abstract: [Objective] In the landslide susceptibility assessment, the selection and optimization of the landslide prediction model are very important to the efficiency of the calculation process and the accuracy of the prediction results. Aiming at the problems that the existing single-objective genetic optimization algorithm (Genetic Algorithm, GA) is prone to premature maturity, poor local search ability, and slow global optimization speed, this paper develops a new optimization algorithm framework, which integrates the classic algorithm in the multi-objective genetic algorithm-Non-dominated sorting method with elite selection strategy (NSGA-II) combined with common machine learning algorithms, random forest (RF), and support vector machine (SVM) to predict landslide susceptibility. Different from single objective optimization, NSGA-II algorithm can perform feature selection and hyper-parameter optimization simultaneously, and make the prediction model achieve optimal accuracy, recall, precision and AUC (area under curve, AUC) at the same time. [Methods] Taking the Chongqing section of the Three Gorges reservoir area as the study area,

收稿日期: 2023-07-18

修回日期: 2023-10-27

录用日期: 2023-11-10

网络首发日期(www.cnki.net): 2024-01-02

资助项目: 国家自然科学基金项目(31971726)

第一作者: 张兴存(1997—), 男, 硕士研究生, 主要从事山地灾害防治研究。E-mail: zhangxingcun@bjfu.edu.cn

通信作者: 王云琦(1979—), 女, 教授, 主要从事水土保持工程研究。E-mail: wangyunqi@bjfu.edu.cn

http://stbxb.alljournal.com.cn

the four optimized models (RF-GA, SVM-GA, RF-NSGA-II, and SVM-NSGA-II) were compared and analyzed in three aspects; model accuracy evaluation, landslide hazard susceptibility zoning map, and zoning statistics. [Results] NSGA-II was more effective than GA optimization, and in terms of model evaluation and landslide susceptibility zoning, the RF-NSGA-II model had higher predictive performance, with four evaluation values of 80.91%, 81.89%, 80.07% and 88.60% respectively, proving the effectiveness of the NSGA-II optimization algorithm; the area share of very low to very high hazard zones were in the order of 23.06%, 22.46%, 22.96%, 19.99%, and 11.53%, which verified the reliability of the RF-NSGA-II model. The susceptibility map predicted by the RF-NSGA-II model showed that the high and extremely high susceptibility areas were concentrated in the north and distributed in bands from east to west. [Conclusion] RF-NSGA-II algorithm based on multi-objective selection provides a new idea for the optimization of machine learning model for landslide risk assessment.

Keywords: landslide susceptibility; multi-objective optimization; random forest; support vector machine

Received: 2023-07-18

Revised: 2023-10-27

Accepted: 2023-11-10

Online(www.cnki.net): 2024-01-02

重庆作为中国最典型的山地城市,其脆弱的地质条件、高强度的人类活动、周期性水位涨落等,导致滑坡频繁发生,呈现“点多,面广”的特点。滑坡灾害往往导致住宅受损、铁路中断等,对居民生命财产安全构成巨大威胁^[1]。未来,极端天气和城镇建设等因素的增加也会增加滑坡的发生概率。因此,建立精准的潜在滑坡预测模型,进行滑坡易发性分析,对防灾减灾和城镇区划至关重要^[2]。

绘制滑坡易发性图(landslide susceptibility map, LSM)是识别潜在滑坡空间分布的关键方法。目前,国内外绘制 LSM 的方法已经十分成熟,根据工作原理可以分为基于物理的方法、基于统计的方法和基于数据驱动的机器学习方法 3 种^[3]。物理方法从数学原理的角度分析滑坡诱发条件对边坡稳定性影响的机理,即使用极限平衡边坡稳定性模型等来计算安全系数,确定边坡稳定性,预测山体滑坡易发区。常见的模型包括无限边坡稳定性模型、分布式边坡稳定性模型等^[4-5]。然而,受地形、地质、气候、人为活动等控制和触发因素的耦合诱导,滑坡发生机制是非线性、动态和异质的,大量的数据输入需求使得基于物理的模型难以在区域尺度上生成高精度的易发性图^[6]。随着 GIS 技术的发展与应用,基于统计的方法,如频率比、证据权重、熵指数、信息值、逻辑回归、证据信念函数等^[7-8],可以通过滑坡清单和相关影响因素的地理空间数据,确定数据之间的内在规律,建立关系模型,从而实现研究区 LSM 绘制。然而,由于统计模型结构简单,复杂的非线性数据和滑坡异质性都会导致模型精度降低。相比于统计方法,机器学习并不依赖统计函数,而是基于输入数据,通过训练数据构建模型规则,进而揭示滑坡发生与影响因素之间复杂的高维非线性关系。机器学习模型可以在大

尺度研究区内快速、准确地绘制 LSM,实现对滑坡易发区的有效识别,成为目前绘制 LSM 的主要工具。相关文献中使用最多的机器学习模型为人工神经网络、支持向量机和随机森林^[9-10]。

机器学习模型的预测精度严重依赖于数据质量和超参数设置。学者一方面通过优化特征选择提升滑坡预测精度,常用方法有用于搜索初始特征的模糊算法(fuzzy algorithm)和遗传算法(GA)等,选择具有可解释的最佳特征组合,以增强模型的泛化能力;另一方面,通过优化算法确定模型超参数,如贝叶斯优化、粒子群优化和鲸鱼优化算法等进化、群智能算法可用于搜索最优超参数解,以提高模型收敛性和准确率^[11]。

以上研究存在 2 个可以改进的方向:第 1,在滑坡易发性评价的背景下,高性能模型应具有较强的特异性、敏感性和鲁棒性。上述优化方法具有单目标优化的特点,即在达到一个优化指标后,其他指标的性能可能会下降,无法有效达到全局最优。第 2,特征选择和参数优化是 2 个独立过程,耗费时间较多,目前还没有研究将其作为整体进行优化。为解决以上问题,本研究引入多目标优化算法 NSGA-II,选取常见的随机森林(RF)、支持向量机(SVM)作为基础模型,将最小特征数量、最大准确度、最大召回率、最大精密度和最大 AUC 作为目标,对机器学习模型进行优化,并与 GA 优化效果进行对比,以全面验证 NSGA-II 的优化性能^[12]。

基于对过去研究工作的总结和补充,本研究旨在:(1)建立基于 RF、SVM 基础机器学习模型和 NSGA-II、GA 优化算法的融合预测模型,并评估其性能,以填补滑坡敏感性测绘中多目标优化技术的空白;(2)绘制滑坡易发性图,划分滑坡易发区,为城市规划提供科学指导。

1 研究区域与数据来源

1.1 研究区域

本研究选取三峡库区重庆段——丰都至巫山 8 个县级行政区作为研究区。研究区位于 107°30′40″—110°12′12″E,30°02′20″—31°44′00″N,覆盖面积 30 698 km²(图 1)。研究区属亚热带季风湿润气候,年平均气温 16~18 ℃;年平均降水量 1 000~14 00 mm,降雨呈现东部高于西部,山区大于河谷的特点,且集中在夏季,多为灾害性天气和暴雨。研究区河流众多,长江干流自西向东贯穿,支流交错,具有流量大和水位变幅大的特点。地貌类型以丘陵和山地为主,总体地势西高东低,岭谷相间,海拔起伏较大,高差最大可达 2 600 m。研究区地质构造复杂,主要有大巴山弧形褶皱、川东褶皱等;以沉积岩为主,少数岩浆岩和变质岩。沉积岩层主要由陆生沉积碳酸盐岩和碎屑岩组成;奉节以西,主要由侏罗纪红碎屑岩组成。奉节以东,主要由古生代和中生代碳酸盐岩组成,并伴有碎屑岩和煤层,是长江流域地质灾害高发地区。

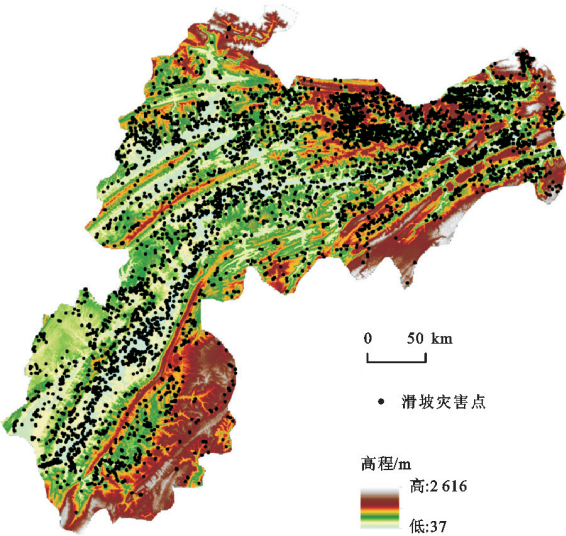


图 1 研究区滑坡灾害点分布

Fig. 1 Distribution of landslide disaster points in the study area

1.2 数据准备

1.2.1 滑坡清单图 本研究根据重庆市地质调查局统计资料编制滑坡清单图,使用 2000—2020 年共 20 年的滑坡信息,绘制了包含 4 961 次滑坡的山体滑坡分布图。在机器学习建模中,滑坡易感性评价可以视为二元分类问题,基于此,利用 ArcGIS 10.6 软件在河流外和滑坡点 100 m 缓冲区外随机选取相同数量的非滑坡样本,将所有样本随机按照 7 : 3 构建训练集和测试集^[13]。将滑坡样本指定为 1,非滑坡样本指定为 0。

1.2.2 滑坡易发性评价指标选取 滑坡发生机制具有区域异质性,因此没有固定的评价指标标准。综合已有研究,初步选择 12 个评价指标建立滑坡易发性指标评价体系,包括地形因素:海拔、坡度、坡向、平面曲率、剖面曲率;水文和地质因素:岩性、距离断层距离、年均降雨量、距离河流的距离,以及人类活动因素^[14]。数据来源和特征见表 1。评价指标的空间分布见图 2。

1.3 数据相关分析

为了避免冗余数据对建模的影响,采用多重共线性分析方法检验评价指标之间的相关性^[15]。本文通过方差膨胀系数和公差进行多重共线性分析^[16]。若 VIF 大于 10 且公差小于 0.1 则表示相关因子具有共线性,应在建模中剔除^[17]。计算方法为:

VIF = 1 / (1 - R^2) (1)

TOL = 1 / VIF (2)

式中:VIF 为方差膨胀系数,表示回归系数估计量的方差与假设自变量间非线性相关时方差相比的比值;TOL 为容忍度,是方差膨胀系数的倒数,表示共线性的强弱程度; R^2 为影响因子拟合的决定系数,反映影响因之间的变动关系。

表 1 数据来源
Table 1 Data source

数据类型	数据名称	数据来源	分辨率
地形因素	坡度	ASTER(高级星载热发射和反射辐射计)	30 m
	坡向		
	海拔	全球 DEM(数字高程模型)	
	剖面曲率	(http://gdem.ersdac.jspacesystems.or.jp)	
	平面曲率		
水文和地质因素	年平均降雨量(2000—2020 年)	国家地球系统科学数据中心、国家科技基础设施资源与环境科学与数据中心	1 : 1 000
	距离河流的距离	国家地质数据中心(https://www.cgs.gov.cn/)	1 : 1 000
	地质		1 : 50 000
	距离断层的距离		
人类活动因素	NDVI(2000—2020 年)	资源与环境科学与数据中心 (https://www.resdc.cn/Default.aspx)	1 : 1 000
	土地利用类型(2020 年)		
	距离道路的距离		

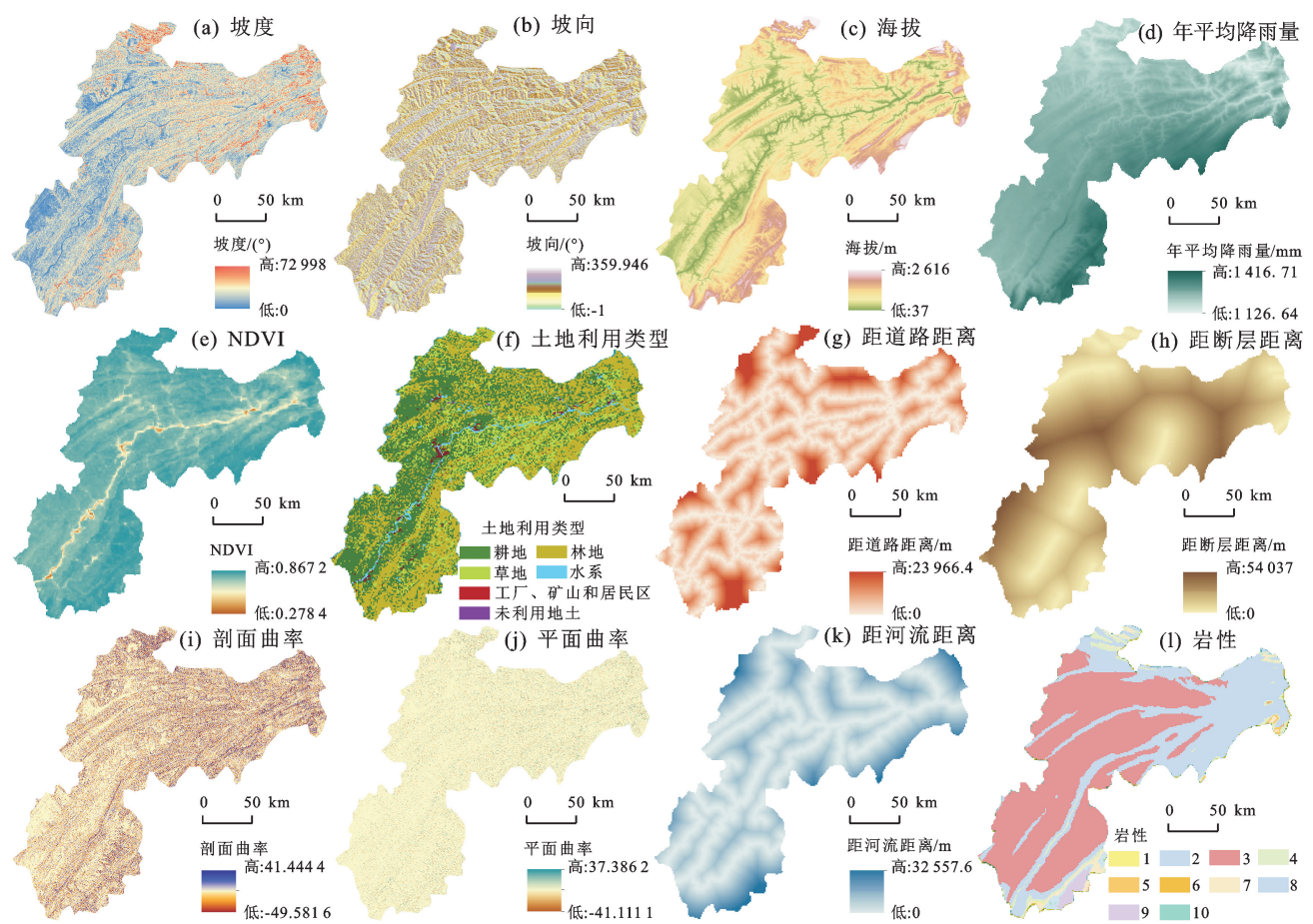


图 2 评价指标
Fig. 2 Feature factors

2 研究方法

2.1 机器学习模型

2.1.1 支持向量机(SVM) SVM 基于统计学原理的分类预测模型^[18]。该方法核心是通过最优化问题解,确定高维特征数据的最优分类超平面,以解决复杂数据分类问题。SVM 通过引入核函数实现低维空间向量集向高维空间向量的映射,同时降低分类误差,最大程度提升分类器的泛化能力^[11,19-20]。SVM 在解决小样本、高维度、非线性问题方面具有优势。

2.1.2 随机森林(RF) RF 是一种基于套袋集成和随机特征选择的算法。在解决分类问题时,该算法构造了一个基于自采样策略的决策树训练集。训练子集随机选择特征,计算最优解作为决策树上节点分割的基础,生成多个决策树,并将其构建为一个随机森林。随后,未抽样的样本纠正由于随机性产生的偏差。与传统的决策树算法相比,RF 算法通过在训练过程中随机抽样和随机改变各种因素的组合来增加分类树的多样性,减少过拟合。

2.2 优化算法

2.2.1 GA 遗传算法 GA 是一种搜索启发式算法。算法的结构由一个群体组成,其中每个个体都被认为

是问题的解决方案。个体被称为染色体,由不同的问题变量组成,这些变量在算法中起着基因的作用。该算法运用选择、交叉、变异 3 种运算方法产生下一代群体。运算产生的新的后代染色体被作为下一代亲代,用于产生新的子代染色体。运算将持续迭代,直到满足终止条件。

2.2.2 NSGA-Ⅱ 多目标优化算法 NSGA-Ⅱ 基于快速且非支配性的排序技术来区分基于帕累托支配性和最优性概念的解决方案^[21]。NSGA-Ⅱ 算法的核心是快速非支配排序,即依据个体的非劣解水平对种群分层,指引搜索向 Pareto 最优解集方向进行。同时,NSGA-Ⅱ 采用了精英策略,将父代个体与子代个体合并后进行非支配排序,使得搜索空间变大,生成下一代父代种群时按顺序将优先级较高的个体选入,并在同级个体中采用拥挤度进行选择,保证了优秀个体能够更大概率被保留。

2.3 精度评价

本研究采用准确度、召回率、精密度和 AUC 4 个常见指标来评估模型的性能。准确度[公式(3)]能够正确分类滑坡和非滑坡栅格的比例,表征算法的分类精度;精密度[公式(4)]是被预测为滑坡的栅格在真正滑坡栅格中的比例;召回率[公式(5)]是滑坡栅格

中正确预测为滑坡栅格的比例,并结合精度,表征了算法的分类识别能力。ROC 曲线表征敏感性和特异性的相互关系,其中模型的特异性(假阳性率)为 x 轴,敏感度(真阳性率)为 y 轴。AUC 值的范围为 0.5~1.0, AUC 值越高,模型预测性能越好。本研究使用的评估公式为:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (3)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (5)$$

式中: FP 、 FN 为错误分类网络的像素数; TP 、 TN 为正确分类的像素数。

2.4 基于 GA 和 NSGA-II 优化机器学习模型建模过程

基于优化算法机器学习模型寻优的本质是根据研究目标确定模型最佳超参数组合。基于 GA 和 NSGA-II 算法的优化流程(图 3、图 4),其中 GA 算法以常用的最大 AUC 为目标函数,NSGA-II 算法以最小特征数、最大精度、最大召回率、最大准确度和最大 AUC 为目标函数,参数设置见表 2, RF 模型和 SVM 模型需要优化超参数见表 3。

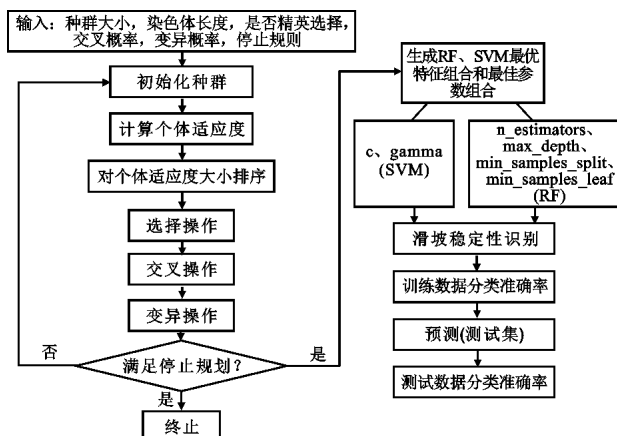


图 3 GA 算法优化机器学习模型的流程

Fig. 3 The process of GA algorithm to optimize machine learning models

基于 GA 算法的优化过程见图 5。当迭代次数为 38 时,模型适应度恒定达到最大,表明超参数达到最优,即模型最优,此时 AUC 值为 85.89;当迭代次数为 42 时, SVM 模型达到最优, AUC 值为 76.89。基于 NSGA-II 算法优化模型时,为扩大特征的搜索范围并防止陷入局部最优,运行 10 次模型优化程序,产生 200 组 Pareto 最优解,最优解集结果见图 6。RF-NSGA-II 模型所获得的最优指标组合为:准确率 80.91%, 召回率 81.89%, 精准度 80.07%, AUC 88.60%, 特征数为 6, 特征组合为年平均降雨量、坡向、海拔、与河流的距离、与道路的距离,以及与断层

的距离。SVM-NSGA-II 模型所获得的最优指标组合为:准确率 70.13%, 召回率 72.69%, 精准度 68.59%, AUC 76.92%。特征数为 10, 特征组合为海拔、坡向、平面曲率、剖面曲率、与断层距离、年平均降雨量、与河流的距离、平均 NDVI、土地利用类型,以及与道路的距离。基于以上优化过程,各模型最优参数见表 4。

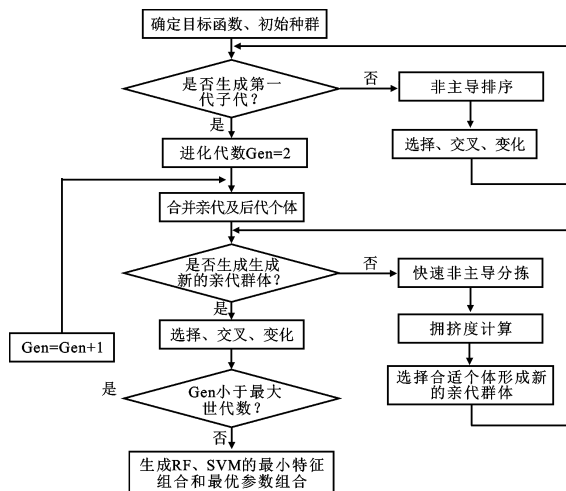


图 4 NSGA-II 算法优化机器学习模型的流程

Fig. 4 The process of NSGA-II algorithm to optimize machine learning models

表 2 GA 和 NSGA-II 优化算法参数

Table 2 Parameters of GA and NSGA-II optimization algorithms

参数	值
种群大小	20
迭代数	50
变异概率	0.8
交叉概率	0.001

表 3 RF 和 SVM 模型参数及取值范围

Table 3 Parameters of RF and SVM models and the range of values

模型	参数	范围	
		最小值	最大值
RF	n_estimators	2	500
	max_depth	2	40
	min_samples_leaf	2	30
	min_samples_split	1	30
SVM	c	0	30
	gamma	0	1

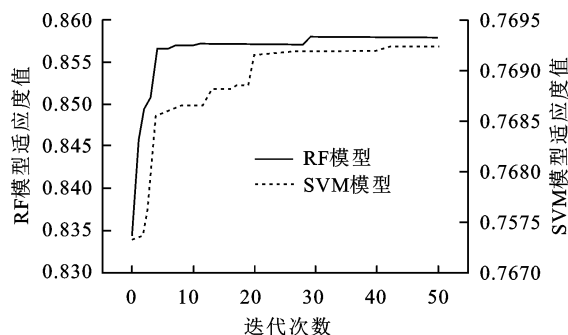


图 5 基于 GA 算法优化机器学习模型的过程

Fig. 5 The process of optimizing machine learning models based on the GA algorithm

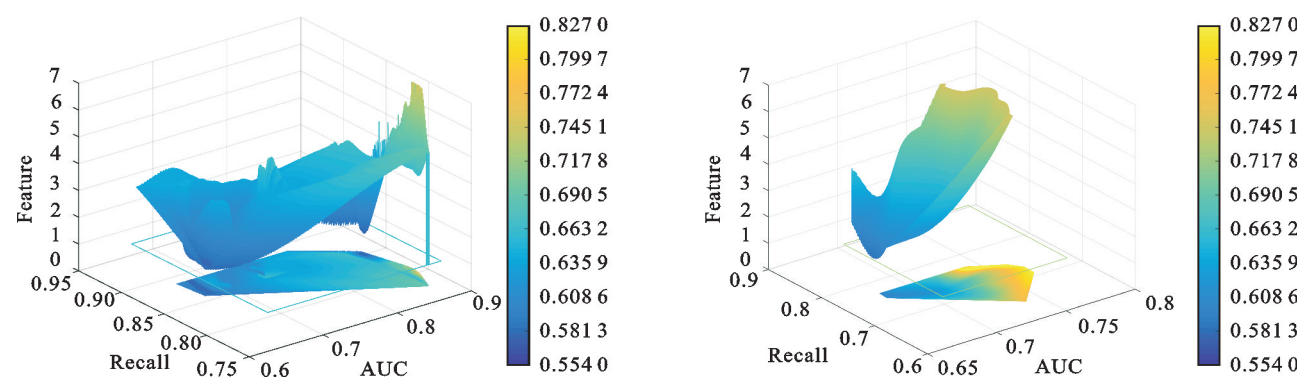


图 6 基于 NSGA-II 算法优化机器学习模型获取的帕累托最优解

Fig. 6 Pareto-optimal solutions obtained from optimized machine learning models based on NSGA-II algorithm

表 4 各模型参数

Table 4 Parameters of each model

模型类型	参数					
	n_estimators	max_depth	min_samples_leaf	min_samples_split	c	gamma
RF-GA	90	18	2	13	—	—
RF-NSGA-II	121	32	1	6	—	—
SVM-GA	—	—	—	—	10.6	0.03
SVM-NSGA-II	—	—	—	—	1.6	0.15

3 结果与分析

3.1 多重共线性分析

使用 SPSS 软件计算了 12 个评价指标的 VIF 和 TOL 值(表 5),12 个条件指标的 VIF 值均小于 10, TOL 值均大于 0.1,其中 VIF 最大值为 2.166(高程),最小值为 1.056(坡向);TOL 最大值为 0.947(坡向),最小值为 0.462(海拔)。综上所述,所有指标不具有共线性,可以作为模型输入进行训练。

表 5 多重线性分析结果

Table 5 Results of multiple linear analysis

变量	TOL	VIF
坡向	0.947	1.056
海拔	0.462	2.166
平面曲率	0.831	1.203
剖面曲率	0.828	1.208
与断层的距离	0.914	1.094
岩性	0.864	1.157
坡度	0.927	1.079
与河流的距离	0.656	1.523
平均 NDVI	0.559	1.788
年平均降雨	0.736	1.360
与离道路的距离	0.838	1.193
土地利用类型	0.888	1.126

3.2 滑坡空间建模

3.2.1 模型预测性能分析 4 种模型的预测性能结果见图 7。RF-GA 模型、RF-NSGA-II 模型、SVM-GA 模型和 SVM-NSGA-II 模型的准确度分别为 78.03%,80.91%,69.22%,70.13%,召回率分别为 73.21%,81.89%,73.51%,72.69%,精准度分别为 76.91%,80.07%,67.16%,68.59%。

由图 8 可知,4 种模型的 AUC 值分别为 85.59%, 88.6%,76.89%,76.92%。对比发现,整体上 RF 模型的预测效果优于 SVM 模型,SVM-NSGA-II 与 SVM-GA 的优化结果相近,RF-NSGA-II 模型的 4 项指标均高于其他模型,其次是 RF-GA 模型。综上所述,在 RF 模型中,NSGA-II 多目标优化算法的优化效果优于 GA 单目标优化算法,对比 4 个模型证明,RF-NSGA-II 模型在研究中为最佳预测模型。

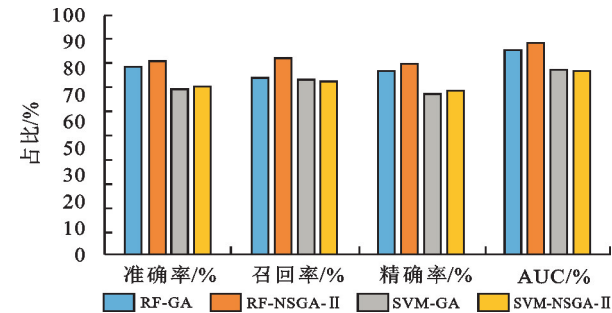


图 7 4 种算法的性能对比

Fig. 7 Performance comparison of four algorithms

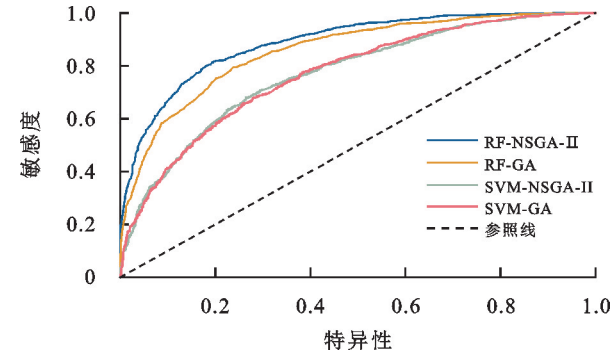


图 8 4 种模型的 ROC 曲线

Fig. 8 ROC curves of four models

3.2.2 滑坡易发性分析 将评价指标数据集输入优化后的机器学习模型,计算研究区所有栅格点的滑坡概率,以表示滑坡易发性。

在 ArcGIS 10.2 中利用自然断点法,通过寻找滑坡概率数据集内的裂点,将滑坡易发性分为极低、低、中、高和极高 5 个等级。由图 9 可知,4 种模型的历史滑坡

数随着易发性等级的提高呈现上升趋势,且历史滑坡的密度均随易发性等级提高而增大,在高易发区或者极高易发区达到最大。整体上各图的滑坡易发分区类似,高、极高危险区多呈东北—西南分布,且靠近水系、道路网和城镇,而低、极低危险区集中在研究区边缘地带。

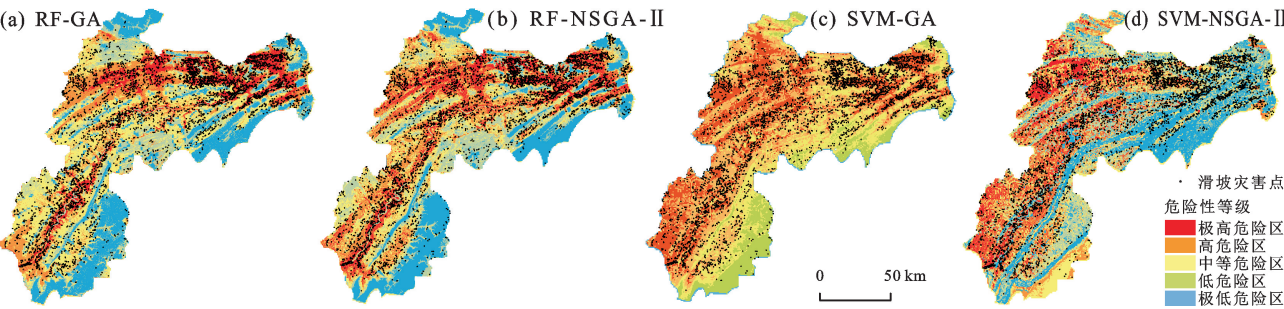


图 9 滑坡易发性分析评价结果

Fig. 9 Evaluation results of landslide susceptibility

为了进一步分析不同模型的性能,本文将滑坡易发性图与历史滑坡分布图进行叠加分析。由表 6 可知,4 个模型中,RF-NSGA-II 模型的极高风险区历史滑坡密度最大,为 0.757 0 个/km²;高风险区密度次之,为 0.283 0 个/km²。统计各模型滑坡易发性分区极高和高风险区的历史滑坡数密度,RF-NSGA-II 模型为 1.040 0 个/km²,RF-GA 模型为 0.880 0 个/km²,SVM-GA 为 0.544 8 个/km²,SVM-NSGA-II 模型为 0.473 7 个/km²,滑坡密度在 RF-NSGA-II 模型分区中最大。统计各模型滑坡易发性分区极低和低危险区的历史滑坡数密度,RF-NSGA-II 模型为 0.099 1 个/km²,RF-GA 模型为 0.091 0 个/km²,SVM-GA 为 0.085 5 个/km²,

SVM-NSGA-II 模型为 0.326 9 个/km²,滑坡密度在 SVM-GA 模型分区中最小。

对于 RF-GA 模型,中易发区面积占研究区总面积最大,达 29.56%。对于 SVM-GA 模型,高易发区和极高易发区占大部分面积,达 64.23%,低易发区和极低易发区仅占 13.03%。对于 SVM-NSGA-II 模型,极低、低、中、高易发区面积占比近似,不符合滑坡分布规律。相比之下,RF-NSGA-II 模型极低易发区至极高易发区面积与研究区总面积的比值逐渐降低,预测结果较为合理。

综合分区面积和分区密度考虑,RF-NSGA-II 模型对研究区滑坡易发性分区精度较高且比较合理。

表 6 滑坡易发性与历史滑坡分布叠加分析

Table 6 Superposition analysis of the landslide susceptibility and historical landslide distribution						
模型	易发性等级	面积/km ²	面积占比/%	历史滑坡数/个	历史滑坡数占比/%	历史滑坡数/面积比值
RF-GA	极低	4 862.62	19.26	116	2.34	0.023 9
	低	4 173.41	16.53	280	5.64	0.067 1
	中	7 461.35	29.56	1 075	21.67	0.144 1
	高	5 387.11	21.34	1 422	28.66	0.264 0
	极高	3 357.26	13.30	2 068	41.69	0.616 0
RF-NSGA-II	极低	5 818.79	23.06	133	2.68	0.022 9
	低	5 666.93	22.46	432	8.71	0.076 2
	中	5 794.76	22.96	764	15.40	0.131 8
	高	5 045.73	19.99	1 428	28.79	0.283 0
	极高	2 910.07	11.53	2 203	44.42	0.757 0
SVM-GA	极低	630.27	2.49	38	0.77	0.060 3
	低	2 663.60	10.54	67	1.35	0.025 2
	中	5 740.91	22.72	544	10.97	0.094 8
	高	9 183.64	36.35	2 036	41.04	0.221 7
	极高	7 044.53	27.88	2 276	45.88	0.323 1
SVM-NSGA-II	极低	5 368.21	21.25	769	15.50	0.143 3
	低	5 686.31	22.51	1 044	21.04	0.183 6
	中	5 496.28	21.76	1 082	21.81	0.196 9
	高	5 051.12	19.99	1 206	24.31	0.238 8
	极高	3 661.03	14.49	860	17.34	0.234 9

4 讨论

本研究的一大贡献是首次将多目标优化算法应用于滑坡易发性评价,并以准确度、召回率、精密度、AUC 为指标将其与单目标优化进行对比。结果显示,对于 RF 模型,RF-NSGA-II 的 4 项指标值较 RF-GA 分别提高 2.88%,8.68%,3.16%,3.01%,NSGA-II 的优化效果明显优于 GA。SVM-NSGA-II 与 SVM-GA 的优化结果相近,这是因为 SVM 模型的原理具有识别重要特征的属性,同时 2 种优化均达到模型的全局最优^[13],因此多目标优化算法的特征选择优势显现不明显。从分类统计角度来看,RF-NSGA-II 模型预测结果图中极低危险区至极高危险区面积的百分比依次分别为 23.06%,22.46%,22.96%,19.99%,11.53%;极高和高危险区的历史滑坡数占总数的 73.21%,密度为 0.757 0 个/km²,大于 RF-GA 模型(0.616 0 个/km²),SVM-GA 模型(0.323 1 个/km²)和 SVM-NSGA-II 模型(0.234 9 个/km²),具有最佳预测效果。目前,已有学者^[22]在研究区使用 k 聚类、RF 和 ANN 等机器学习模型预测山体滑坡易发性。但上述模型均未考虑输入特征和参数的优化。相比之下,考虑特征组合的 RF-NSGA-II 模型显示出最佳的泛化能力和预测精度。

此外,一些学者^[23-24]在超参数优化前通过地理探测器筛分、后向消除、GA 等特征选择方法选择重要特征,消耗大量时间。本研究首次尝试特征选择和超参数优化同时进行,与其他方法相比,优化效率较高。这是因为,在原理上 NSGA-II 算法可以通过权衡多个优化目标扩展到种子搜索空间,避免了单目标优化算法的局部最优,也提供了同时优化特征选择和参数的可能性。综上所述,本研究中 NSGA-II 算法对 RF 模型优化效果最好,但是其应用到其他机器模型的效果有待探索^[9],从评价指标看,SVM-GA 模型和 SVM-NSGA-II 模型表现相近,但绘制的滑坡易发性图差异较大,这由于不同超参数组合使模型架构不同,使得输出结构存在差异,也印证了超参数组合选择会影响基于机器学习的滑坡易发性评价效果,面对这种不确定性,下一步仍需使用 NSGA-II 算法对更多机器学习模型或深度学习模型进行优化,以全面了解多目标优化算法在滑坡易发性制图中的性能。

除了所提方法的适用性外,本研究中输入特征的选择也有证据支持。RF-NSGA-II 算法选取的 6 个重要特征:年平均降雨量、坡向、海拔、与河流的距离、与道路的距离及与断层的距离,与其他研究^[13,25]一致,具有可解释性。RF-NSGA-II 之所以没有选择坡

度、平面曲率、剖面曲率、NDVI、土地利用和岩性等因子,是因为它们具有空间自相关性以及非滑坡干扰特征,可能会影响模型预测精度。然而本研究还存在一些不足,例如,考虑到数据的可用性,本研究选择年平均降雨量作为气象因素。然而,详细的降雨因素,如最大月降雨量,将有助于提高所绘地图的准确性^[26],未来应考虑在滑坡易发性评价指标体系中加入详细影响因子以扩大特征选择空间,增强模型解释力和预测精度。

5 结论

(1)本研究提出的基于多目标选择的 RF-NSGA-II 模型,通过特征工程和参数优化,可以同时实现最优准确度、召回率、精准度和 AUC。在本研究区,这 4 项评价指标分别为 80.91%,81.89%,80.07%,88.60%。结果优于 RF-GA 模型,SVM-NSGA-II 模型和 SVM-GA 模型,表明 RF-NSGA-II 相对可以获得更好的拟合和预测性能。

(2)基于 RF-NSGA-II 模型预测研究区域滑坡易发性表明,高易发区与极高易发区的面积占比以及历史滑坡数占比依次为 19.99%,28.79%和 11.53%,44.42%。高易发区和极高易发区集中在北部,从东到西呈带状分布。

(3)虽然 NSGA-II 多目标优化算法较 GA 单目标遗传优化算法对 RF 模型的优化效果更明显,但在未来的研究工作中,应优化更多机器学习或深度学习模型,同时应增加滑坡易发性评价指标,充分挖掘重要影响特征,减少不确定性,增强预测精度。

参考文献:

- [1] 王迪,郝萌萌,陈帅,等.中国滑坡敏感性及危险因素评价[J].自然灾害,2021,108:3045-3059.
WANG D, HAO M M, CHEN S, et al. Sensitivity and risk factor evaluation of landslides in China[J]. Natural Calamities, 2021, 108: 3045-3059.
- [2] FERRER J, GUO Z Z, MEDINA V, et al. A framework to project future rainfall scenarios: An application to shallow landslide-triggering summer rainfall in Wanzhou County China[J]. Water, 2022, 14(6): e873.
- [3] REICHSTEIN M, CAMPS-VALLS G, STEVENS B, et al. Deep learning and process understanding for data-driven earth system science [J]. Nature, 2019, 566(7743): 195-204.
- [4] DOLOJAN N L J, MORIGUCHI S, HASHIMOTO M, et al. Mapping method of rainfall-induced landslide hazards by infiltration and slope stability analysis[J]. Landslides, 2021, 18(6): 2039-2057.

- [5] SHU H P, HÜRLIMANN M, MOLOWNY-HORAS R, et al. Relation between land cover and landslide susceptibility in Val d'Aran, Pyrenees (Spain): Historical aspects, present situation and forward prediction[J]. *The Science of the Total Environment*, 2019, 693: e133557.
- [6] ZHOU X Z, WEN H J, LI Z W, et al. An interpretable model for the susceptibility of rainfall-induced shallow landslides based on SHAP and XGBoost[J]. *Geocarto International*, 2022, 37(26): 13419-13450.
- [7] CHEN H Y, DENG T T, DU T, et al. An RF and LSSVM-NSGA-II method for the multi-objective optimization of high-performance concrete durability[J]. *Cement and Concrete Composites*, 2022, 129: e104446.
- [8] DENG D, LI Y X, MA J Q, et al. A comparative study for landslide susceptibility assessment using machine learning algorithms based on grid unit and slope unit[J]. *Frontiers in Environmental Science*, 2022, 10: e1009433.
- [9] HAKIM W L, REZAIE F, NUR A S, et al. Convolutional neural network (CNN) with metaheuristic optimization algorithms for landslide susceptibility mapping in Icheon, South Korea[J]. *Journal Environmental Management*, 2022, 305: e114367.
- [10] EKMEKCIOĞLU Ö, KOC K, ÖZGER M, et al. Exploring the additional value of class imbalance distributions on interpretable flash flood susceptibility prediction in the Black Warrior River Basin, Alabama, United States[J]. *Journal of Hydrology*, 2022, 610: e127877.
- [11] JAAFARI A, PANAHI M, MAFI-GHOLAMI D, et al. Swarm intelligence optimization of the group method of data handling using the cuckoo search and whale optimization algorithms to model and predict landslides[J]. *Applied Soft Computing*, 2022, 116: e108254.
- [12] SAĞ T, ÖRNEK H K. Classification rule mining based on Pareto-based Multiobjective Optimization[J]. *Applied Soft Computing*, 2022, 127: e109321.
- [13] CHENG J Y, DAI X A, WANG Z K, et al. Landslide susceptibility assessment model construction using typical machine learning for the Three Gorges Reservoir Area in China[J]. *Remote Sensing*, 2022, 14(9): e2257.
- [14] MERGHADI A, YUNUS A P, DOU J, et al. Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance[J]. *Earth Science Reviews*, 2020, 207: e103225.
- [15] 王毅, 方志策, 牛瑞卿, 等. 基于深度学习的滑坡灾害易发性分析[J]. *地球信息科学学报*, 2021, 23(12): 2244-2260.
WANG Y, FANG Z C, NIU R Q, et al. Landslide susceptibility analysis based on deep learning[J]. *Journal of Geo-Information Science*, 2021, 23(12): 2244-2260.
- [16] 刘纪平, 梁恩婕, 徐胜华, 等. 顾及样本优化选择的多核支持向量机滑坡灾害易发性分析评价[J]. *测绘学报*, 2022, 51(10): 2034-2045.
LIU J P, LIANG E J, XU S H, et al. Multi-kernel support vector machine considering sample optimization selection for analysis and evaluation of landslide disaster susceptibility[J]. *Acta Geodaetica et Cartographica Sinica*, 2022, 51(10): 2034-2045.
- [17] SAHA A, PAL S C, SANTOSH M, et al. Modelling multi-hazard threats to cultural heritage sites and environmental sustainability: The present and future scenarios[J]. *Journal of Cleaner Production*, 2021, 320: e128713.
- [18] 徐胜华, 刘纪平, 王想红, 等. 熵指数融入支持向量机的滑坡灾害易发性评价方法: 以陕西省为例[J]. *武汉大学学报(信息科学版)*, 2020, 45(8): 1214-1222.
XU S H, LIU J P, WANG X H, et al. Landslide susceptibility assessment method incorporating index of entropy based on support vector machine: A case study of Shaanxi Province[J]. *Geomatics and Information Science of Wuhan University*, 2020, 45(8): 1214-1222.
- [19] 王念秦, 郭有金, 刘铁铭, 等. 基于支持向量机模型的滑坡危险性评价[J]. *科学技术与工程*, 2019, 19(35): 70-78.
WANG N Q, GUO Y J, LIU T M, et al. Landslide susceptibility assessment based on support vector machine model[J]. *Science Technology and Engineering*, 2019, 19(35): 70-78.
- [20] 王倩, 薛云, 张维, 等. 基于支持向量机的滑坡易发性评价[J]. *湖南城市学院学报*, 2021, 30(1): 22-28.
WANG Q, XUE Y, ZHANG W, et al. Landslide susceptibility mapping based on support vector machine models[J]. *Journal of Hunan City University*, 2021, 30(1): 22-28.
- [21] TABATABAEI M, SALEHPOUR J A, AHMAD H S. Suspended sediment load prediction using non-dominated sorting genetic algorithm II[J]. *International Soil and Water Conservation Research*, 2019, 7(2): 119-129.
- [22] BAI Z G, LIU Q M, LIU Y. Landslide susceptibility mapping using GIS-based machine learning algorithms for the Northeast Chongqing Area, China[J]. *Arabian Journal of Geosciences*, 2021, 14(24): e2831.
- [23] LIU Y M, ZHANG W C, ZHANG Z J, et al. Risk factor detection and landslide susceptibility mapping using geo-detector and random forest models: The 2018 Hokkaido Eastern Iwate Earthquake[J]. *Remote Sensing*, 2021, 13(6): e1157.

(下转第 121 页)

- for large watersheds[J]. *Geoderma*, 2017, 308: 36-45.
- [20] 李天奇. 东北黑土区侵蚀沟成因与模型研究[D]. 北京: 中国科学院研究生院, 2012.
- LI T Q, A cause study and module simulation of gully erosion in black soil region[D]. Beijing: Chinese Academy of Sciences, 2012.
- [21] 蔡崇法, 丁树文, 史志华, 等. 应用 USLE 模型与地理信息系统 IDRISI 预测小流域土壤侵蚀量的研究[J]. 水土保持学报, 2000, 14(2): 19-24.
- CAI C F, DING S W, SHI Z H, et al. Study of applying USLE and geographical information system IDRISI to predict soil erosion in small watershed[J]. *Journal of Soil Water Conservation*, 2000, 14(2): 19-24.
- [22] 水利部. 土壤侵蚀分类分级标准: SL 190—2007[S]. 北京: 中国水利水电出版社, 2008.
- Ministry of Water Resources. Soil erosion classification standard: SL 190-2007 [S] Beijing: China Water Resources and Hydropower Press, 2008.
- [23] 黄振华, 余健, 房莉, 等. 基于 InVEST 模型的青弋江流域土壤侵蚀与影响因素研究[J]. 水土保持学报, 2023, 37(5): 189-197.
- HUANG Z H, YU J, FANG L, et al. Research on soil erosion and influencing factors in Qingyi River Basin based on InVEST model[J]. *Journal of Soil and Water Conservation*, 2023, 37(5): 189-197.
- [24] 何莎莎, 朱文博, 崔耀平, 等. 基于 InVEST 模型的太行山淇河流域土壤侵蚀特征研究[J]. 长江流域资源与环境, 2019, 28(2): 426-439.
- HE S S, ZHU W B, CUI Y P, et al. Study on soil erosion characteristics of Qihe Watershed in Taihang Mountains based on the InVEST model[J]. *Resources and Environment in the Yangtze Basin*, 2019, 28(2): 426-439.
- [25] 陕西省水利厅. 2020 年陕西省水土保持公报[Z]. 西安: 2020.
- Shaanxi Provincial Department of Water Resources. Shaanxi Provincial soil and water conservation bulletin in 2020 [Z]. Xi'an, 2020.
- [26] 汤国安, 杨勤科, 张勇, 等. 不同比例尺 DEM 提取地面坡度的精度研究: 以在黄土丘陵沟壑区的试验为例[J]. 水土保持通报, 2001, 21(1): 53-56.
- TANG G A, YANG Q K, ZHANG Y, et al. Research on accuracy of slope derived from DEMs of different map scales[J]. *Bulletin of Soil and Water Conservation*, 2001, 21(1): 53-56.
- [27] MA X F, ZHAO C Y, ZHU J T. Aggravated risk of soil erosion with global warming: A global meta-analysis[J]. *Catena*, 2021, 200(1): 105-129.
- [28] 朴世龙, 张新平, 陈安平, 等. 极端气候事件对陆地生态系统碳循环的影响[J]. 中国科学: 地球科学, 2019, 49(9): 1321-1334.
- PIAO S L, ZHANG X P, CHEN A P, et al. Effects of extreme climate events on carbon cycle of terrestrial ecosystem[J]. *Scientia Sinica (Terrae)*, 2019, 49(9): 1321-1334.
- (上接第 112 页)
- [24] PHAM B T, VAN D D, ACHARYA T D, et al. Performance assessment of artificial neural network using Chi-square and backward elimination feature selection methods for landslide susceptibility analysis[J]. *Environmental Earth Sciences*, 2021, 80(20): e686.
- [25] SUN D L, GU Q Y, WEN H J, et al. Assessment of landslide susceptibility along mountain highways based on different machine learning algorithms and mapping units by hybrid factors screening and sample optimization[J]. *Gondwana Research*, 2022, 123(2023): 89-106.
- [26] 邢昕芙, 吴承龙, 李金辉, 等. 采用修正逻辑回归方法对降雨滑坡的敏感性评价[J]. 自然灾害, 2021, 106(1): 97-117.
- XING X F, WU C L, LI J H, et al. Susceptibility assessment for rainfall-induced landslides using a revised logistic regression method[J]. *Natural Hazards*, 2021, 106(1): 97-117.