



DOI:10.12404/j.issn.1671-1815.2403088

引用格式:邢昭,孟小军,袁晶晶,等.信息量支持下SVM-GBDT模型的滑坡危险性评价[J].科学技术与工程,2025,25(7):2712-2720.

Xing Zhao, Meng Xiaojun, Yuan Jingjing, et al. Landslides hazard assessment based on SVM-GBDT and gradient boosting decision tree model supported by information quantity[J]. Science Technology and Engineering, 2025, 25(7): 2712-2720.

信息量支持下 SVM-GBDT 模型的 滑坡危险性评价

邢昭¹, 孟小军², 袁晶晶³, 张迪³, 刘力³, 陈彦美^{1*}

(1. 长江大学资源与环境学院, 武汉 430000; 2. 中国地质大学(武汉)环境学院/武汉中地环科水工环科技咨询有限责任公司, 武汉 430000;
3. 湖北省地质局第七地质大队, 宜昌 443000)

摘要 采用机器学习方法在长阳土家族自治县研究区进行滑坡危险性评价,能够为地质灾害防治工作提供科学合理的依据。通过历史滑坡点选取研究区 12 个评价指标(平面曲率、地形起伏度、地表粗糙度、坡度、植被覆盖度、工程岩组、距断裂带距离、距水系距离、降雨量、土地利用类型、距房屋距离和距道路距离)相关性分析后均被选用。计算因子信息量,联合支持向量机(support vector machine, SVM)和梯度提升决策树(gradient boosting decision tree, GBDT)模型构建研究区的评价模型,将研究区危险性分为极高、高、中和低四个等级,生成危险性分区,并对评价模型进行评估。结果表明:极高危险区主要分布于研究区的西南部、中部和东部;I-SVM 和 I-GBDT 模型预测的极高危险区、高危险区、中危险区和低危险区的分区占比分别为 15.86%、21.29%、33.51%、28.68% 和 30.08%、7.41%、13.28%、49.22%, I-SVM 和 I-GBDT 模型 AUC(area under curve)值分别 0.859、0.829。结果表明 I-SVM 模型的预测危险性分区结果更合理可靠。

关键词 滑坡;信息量;危险性评价;支持向量机;梯度提升决策树

中图分类号 P642.2;

文献标志码 A

Landslides Hazard Assessment Based on SVM-GBDT and Gradient Boosting Decision Tree Model Supported by Information Quantity

XING Zhao¹, MENG Xiao-jun², YUAN Jing-jing³, ZHANG Di³, LIU Li³, CHEN Yan-mei^{1*}

(1. Resource and Environmental Engineering College, Yangtze University, Wuhan 430000, China;
2. School of Environmental Studies, China University of Geosciences(Wuhan)/Wuhan Zhongdi Huanke Water Engineering Technology Consulting Co., Ltd., Wuhan 430000, China;
3. The Seventh Geological Brigade of the Hubei Geological Bureau, Yichang 443000, China)

[Abstract] Machine learning methods have been employed in the study area of Changyang Tujia Autonomous County for landslide hazard assessment, it could provide a scientific basis for geological disaster prevention and control efforts. Through the correlation analysis of 12 evaluation indicators (planar curvature, terrain undulation, surface roughness, slope, vegetation coverage, engineering lithology, distance to fault zone, distance to water system, rainfall, land use type, distance to buildings, and distance to roads) in the study area selected by historical landslide points, they were selected. And the evaluation model of the study area was constructed by calculating the information content of factors and integrate support vector machine (SVM) and gradient boosting decision tree (GBDT) models. The hazard of the study area was classified into four levels: extreme high, high, medium, and low, to generate hazard zoning. Subsequently, an assessment of the evaluation model was conducted. The results indicated that the very high hazard zone was mainly distributed in the southwest, central, and eastern parts of the research area. The distribution percentages of very high, high, medium, and low hazard zones predicted by the I-SVM and I-GBDT models were 15.86%, 21.29%, 33.51%, 28.68%, and 30.08%, 7.41%, 13.28%, 49.22%, respectively. The prediction of hazard zones by the I-SVM model aligned more closely with reality. The AUC values for the I-SVM and I-GBDT models were 0.859 and 0.829, respectively. The prediction of risk zones by the I-SVM model is deemed more reasonable and reliable.

[Keywords] landslides; information quantity; hazard assessment; support vector machine; gradient boosting decision tree

收稿日期: 2024-04-26 修订日期: 2024-07-30

基金项目: 湖北省地质局第七地质大队(DZX2022-1)

第一作者: 邢昭(1998—),女,汉族,宁夏银川人,硕士研究生。研究方向:地质灾害与气象预警。E-mail:; elpidaxz@163.com。

* 通信作者: 陈彦美(1984—),女,汉族,河北行唐人,博士,副教授。研究方向:干扰条件下地下水系统演化、地下水环境评价及污染防控。
E-mail: 331769495@qq.com。

投稿网址: www.stae.com.cn

中国地形多变,地壳断裂活动普遍发生,因此诱发的地质灾害数量庞大,其中滑坡是最为常见且分布广泛危害性大的地质灾害之一^[1]。滑坡的发生使人的生命和社会经济财产受到巨大威胁。现阶段根据滑坡的数据,从定量的角度出发对特定区域进行危险性评价,以点推面更加精准地对不同危险性等级的地区制定相应防治对策,对危险性等级高的地区在滑坡发生前的采取有效预防措施,有效减少滑坡造成的损失。

在滑坡危险性评价的过程中,选择相适宜的预测模型尤为重要^[2-3],而现阶段机器学习被广泛运用在地灾易发性、危险性和风险性评价中,是许多国内外学者重点探讨的研究方法^[4]。常被应用的机器学习包含了决策树^[5]、随机森林^[6]、逻辑回归^[7]、支持向量机^[8]和人工神经网络^[9]等。其中支持向量机(support vector machine, SVM)可以通过核方法进行非线性分类,它在解决小样本和高维模式识别问题中,具有良好的泛化能力的优势^[10]。韩玲等^[11]、李家艳等^[12]、王念秦等^[13]均采用支持向量机的方法对滑坡进行易发性、危险性分区,结果显示支持向量机模型适合开展滑坡预测研究,同时经过前人的研究验证采用径向基核函数(radial basis function, RBF)的支持向量机模型在对滑坡评价预测具有更高精度^[14]。

决策树则能够直接呈现数据的特点,可对大型数据样本进行解释,但泛化能力较差,而梯度提升决策树(gradient boosting decision tree, GBDT)较决策树提高了数据泛化能力^[15-16],可以处理连续值和离散值。Rong 等^[17]研究结果表明优化的 GBDT 方法在处理不平衡滑坡数据时具有良好的性能;Dong 等^[18]、郭衍昊等^[19]、董张玉等^[20]、李勇等^[21]用梯度提升决策树对滑坡易发性进行评价,也表现出了较好的可靠性。

以上表明机器学习的方法在滑坡危险性评价中的应用是可靠的,因此在采用信息量法对评价因子做数据处理的基础上,结合支持向量机和梯度提升决策树的算法对研究区滑坡进行危险性评价,得到更准确的危险性分区结果,为滑坡的防治工作提供科学理论依据。

1 研究区概况

长阳土家族自治县(以下简称“长阳县”)研究区位于湖北省宜昌市,地理坐标为 110°20′46″E ~ 111°20′17″E, 30°11′53″N ~ 30°46′34″N,海拔为 52 ~ 2 225 m,总面积 3 391.608 km²。属亚热带大陆性潮湿气候,由于地形复杂,海拔高差悬殊,形成独特的“立体型”气候。在河谷中,气温最高,雨量最小,为 1 100 ~ 1 200 mm,向两岸山顶随着地面高程的增高,则

气温递减,雨量递增,可增加至 1 700 mm。清江自西向东横穿长阳腹地,流程长达 148 km,切割深达 100 ~ 1 000 m。

研究区地势自西向东逐渐降低,为构造溶蚀侵蚀中山区,岩性以滨海-浅海相的碳酸盐岩为主,碎屑岩类及内陆湖相的碎屑岩类和第四系松散堆积物为次。研究区地属扬子准地台八面山褶皱带,东西向主要受南北向压力作用产生的一系列近东西向的褶皱和断裂,北西向构造主要以断裂构造为主,规模较大的有仙女山断裂、天阳坪断裂。由于地质条件复杂,区内地质灾害多发,以滑坡为主,1950 年距今共发生 263 处。

2 数据来源及研究方法

2.1 数据来源

图 1 所示为研究区位置及滑坡分布。本次研究所采用的基本数据包括:①宜昌市地质灾害隐患点数据库,用于提取滑坡信息;②地理空间数据云平台(<http://www.gscloud.cn/home>),用于提取数字高程模型(digital elevation model, DEM)、高差和坡度等;③中国地质调查局 1:50 000 中国地质图(<https://www.cgs.gov.cn/>),用于提取工程地质岩组和地质构造信息;④全国地理信息资源服务系统(<https://www.webmap.cn/main.do?method=index>)全国 1:250 000 基础地理信息数据库,用于提取河流水系和道路信息。

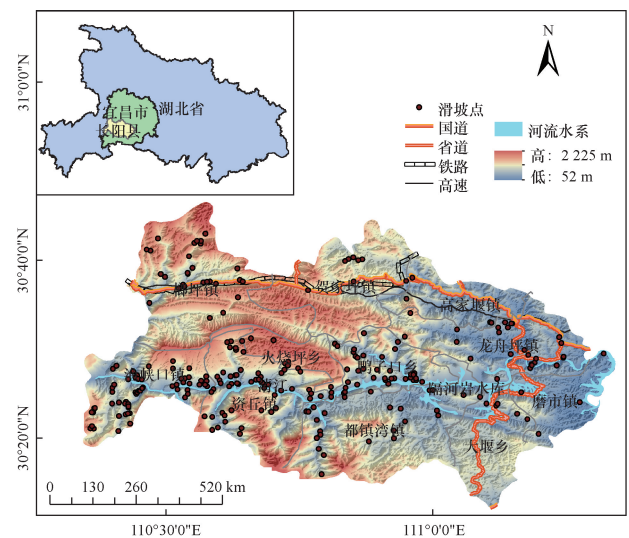


图 1 研究区位置及滑坡分布图

Fig. 1 Location of the study area and distribution of landslide map

2.2 研究方法

2.2.1 信息量模型

信息量模型是地质灾害危险性评价最常用的

模型,信息量模型^[22]的理论认为地质灾害的产生与预测过程中所获取的评价因子信息的数量和质量有关,信息量越大,表明产生地质灾害的可能性越大^[23]。采用面积比来计算信息量值,表达式为

$$I = (y, x_1 x_2 \cdots x_n) = \log_2 \frac{S_0/S}{A_0/A} \quad (1)$$

式(1)中: I 为预测区某单元信息量预测值; y 为滑坡灾害现象; x_i 为区域内对滑坡发生有影响的评价因子($i=1,2,\dots,n$); A 为区域内单元总面积; A_0 为已经发生滑坡灾害的单元面积之和; S 为具有相同因素组合的单元总面积; S_0 为具有相同因素组合单元中发生滑坡灾害的单元面积之和。

实际研究分析过程中由于作用于滑坡灾害的因素很多,样本统计数量会受到限制,采用简化的单因素信息量模型的综合叠加分析,相应的信息量模型为

$$I = \sum_{i=1}^n I_i = \sum_{i=1}^n \log_2 \frac{S_0^i/S^i}{A_0^i/A^i} \quad (2)$$

式(2)中: S^i 为因素 x_i 为所占单元总面积; S_0^i 为因素 x_i 单元中发生滑坡灾害的单元面积之和。 I_i 值的大小表明滑坡发生的可能性, I_i 越大时表示因素 x_i 对预测滑坡越有利。

2.2.2 梯度提升决策树

GBDT是一种将弱学习器与 boosting 方法结合的基于决策树的集成算法^[24-25]。通过不断地迭代,在上一次建立预测结果的弱学习器基础上,构建新的学习器,拟合上一次学习结果产生的残差以提高模型精度,直至函数残差达到预定值或达到最大迭代次数。其完整算法如下:

(1)对第一个学习器 $f(x)$ 初始化,公式为

$$f_0(x) = \lg \frac{P(Y=1|x)}{1-P(Y=1|x)} \quad (3)$$

式(3)中: $P(Y=1|x)$ 为训练样本中 $Y=1$ 的比例, Y 为样本标签, $Y=1$ 表示滑坡。

(2)对 $m=1,2,\dots,M$ 的 M 棵决策树进行以下步骤:

①对 $i=1,2,\dots,N$,计算第 m 棵树的伪残差,公式为

$$r_{m,i} = - \left\{ \frac{\partial L[y_i, f(x_i)]}{\partial f(x)} \right\}_{f(x)=f_{m-1}(x)} = y_i - \frac{1}{1 + e^{-f(x_i)}} \quad (4)$$

②对 $i=1,2,\dots,N$,拟合残差 $r_{m,i}$ 的一个回归树,得到第 m 颗回归树,其对应的叶子节点区域为 $R_{m,j}$,其中 $j=1,2,\dots,j_m$ 。

③对 $j=1,2,\dots,j_m$,计算最佳拟合值,公式为

$$c_{m,j} = \frac{\sum_{x_i \in R_{m,j}} r_{m,i}}{\sum_{x_i \in R_{m,j}} (y_i - r_{m,i})(1 - y_i + r_{m,i})} \quad (5)$$

(3)更新学习器 $f_m(x)$:

$$f_m(x) = f_{m-1}(x) + \sum_{j=1}^{j_m} c_{m,j} I, \quad x \in R_{m,j} \quad (6)$$

(4)得到最终分类模型 $f_M(x)$:

$$f_M(x) = f_0(x) + \sum_{m=1}^M \sum_{j=1}^{j_m} c_{m,j} I, \quad x \in R_{m,j} \quad (7)$$

2.2.3 支持向量机(SVM)

假设滑坡样本数据集 X 有一组样本 $X_i(i=1,2,\dots,n)$, X_i 为12个滑坡危险性评价因子的输入向量, $y_i \in \{-1,1\}$ 为输出值(-1为非滑坡,1为滑坡), n 为滑坡危险性评价因子的数量。SVM是寻找一个最优超平面解决滑坡与非滑坡分类问题的方法,表达式为

$$\omega^T X + b = 0 \quad (8)$$

式(8)中: ω 为超平面的法向量, b 为截距。当 ω 和 b 达到最优值时,即确定了滑坡、非滑坡样本间距在二分类中达到最大的最优分类超平面。

为了将所有样本点考虑到表达式中,避免因寻找所有正负类样本之间的最大几何间隔,所产生噪声的情况造成问题的无解,在约束条件中增加一个松弛项参数 ε ,解决了松弛变量允许一些点到分类平面的距离不满足原先的要求的问题,其表达式为

$$y_i(\omega^T X_i + b) \geq 1 - \varepsilon_i, \quad i = 1,2,\dots,n \quad (9)$$

当 ε 变大到一定程度时,所有训练点便会满足条件。为了使约束条件达到最好效果,在目标函数中引入惩罚参数 c ,得到

$$\min \frac{1}{2} \|\omega\|^2 + c \sum_{i=1}^n \varepsilon_i \quad (10)$$

$$y_i(\omega^T X_i + b) \geq 1 - \varepsilon_i, \quad i = 1,2,\dots,n \quad (11)$$

$$\varepsilon_i \geq 0, \quad i = 1,2,\dots,n \quad (12)$$

惩罚参数 c 代表着对误差的容忍度, c 越大,则对误差的容忍度越低,也容易出现过拟合的情况,而当 c 过小时,则会出现欠拟合的情况。

本次研究问题为非线性可分问题,表达式为

$$\omega^T \varphi(x) + b = 0 \quad (13)$$

φ 为映射函数具有复杂的形式,难以计算其内积,因此使用核函数(kernel function) $k(X_1, X_2) = \varphi^T(X_1)\varphi(X_2)$ 回避内积的显式计算。

3 评价因子的选取及分析

3.1 滑坡评价因子的选取

滑坡是在多种因素共同作用下产生,且在研究

区呈现出具有一定规律的分布特征。在对滑坡进行危险性评价时,选取合适的环境因子才能得出相对准确的结果。本次研究从地形地貌、地质、气象水文和人类工程活动四大类考虑,选取了12个环境因子(平面曲率、地形起伏度、地表粗糙度、坡度、植被覆盖度、工程岩组、距断裂带距离、距水系距离、降雨量、土地利用类型、距房屋距离和距道路距离)作为此次评价要素。

3.1.1 地形地貌

本次选取平面曲率、地形起伏度、地表粗糙度、坡度4个评价因子作为地形地貌的评价指标。这4

个因子均从DEM中提取,其中平面曲率是在坡度的基础上计算得出,二者呈正相关关系,平面曲率越大,地势越陡峭则坡度越大,而坡度和高差决定由重力作用产生下滑力的大小,进而对滑坡的规模和运动速度产生影响。该研究区滑坡多发生在坡度20°以上平面曲率0~35.37范围内,表明当斜坡坡度一定时并不是陡峭程度越高越容易发生滑坡,坡度为主要控制因素。地形起伏度由反映了地形的相对高差,如图2所示滑坡主要分布在起伏度为16.95~61.03区域内。地表径流切割作用和流水侵蚀作用与地表粗糙度息息相关,研究区属中高山

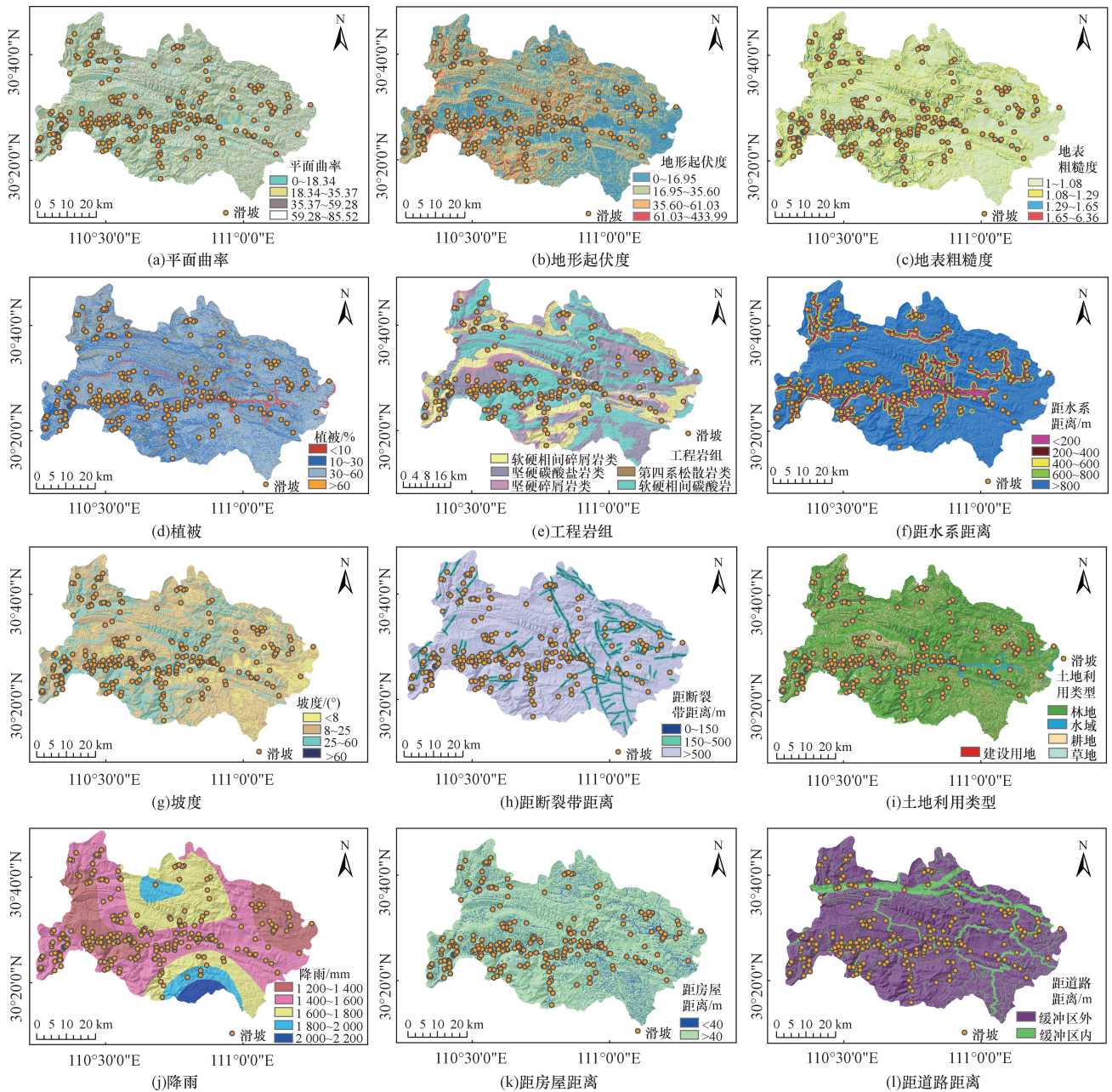


图2 评价因子图

Fig. 2 Evaluation factor graph

地区,河流的下切和侵蚀作用强烈,从而为滑坡提供孕灾条件,研究区的滑坡主要分布在粗糙度为1.08~1.29的区域内。

3.1.2 地质

选取工程岩组和距断裂带距离作为地质类的评价指标。研究区的岩性主要分为第四系松散岩类、坚硬碎屑岩类、坚硬碳酸盐岩类、软硬相间碎屑岩类和软硬相间碳酸岩五类,松散岩类结构较松散力学强度低,遇水易软化,碎屑岩类在地质营力作用下岩体结构构造易被破坏,碳酸盐岩类抗风化能力差,易出现较强烈的错动带,滑坡主要分布在软硬相间碎屑岩类。研究区发育两条较大的断裂带,断裂带内应力复杂,两侧地层岩性因强烈挤压,裂隙发育,岩体易被破坏,为滑坡发育提供了有利孕灾条件,研究区内的滑坡主要分布在距离断裂带500 m 范围外。

3.1.3 气象水文

选取降雨量和距水系距离作为气象水文的评价指标。降水的发生使得土壤容重增加,岩土体软化,为滑坡发育提供有利条件,是滑坡产生的润滑剂,大多数滑坡都是由于强降水诱发。距离水系近的区域,地下水活动强烈,地下水位变化明显,加上水流冲蚀坡脚,岩土体长期处于不稳定的环境中使得斜坡失稳概率高升,从而引发滑坡。如图2所示滑坡主要分布在年均降雨1 400~1 600 mm,距离水系200~600 m 的区域内。

3.1.4 人类工程活动

选取土地利用类型、距房屋距离和距道路距离作为人类工程活动的评价指标。人类建设活动会对原有生态地貌产生较大的扰动,打破原有平衡状态,开挖、切割等使得岩土体破碎,也会使山体产生振动,斜坡的应力发生改变容易产生滑坡。如图2

所示在耕地,距离房屋40 m 范围内和距离道路缓冲区内(铁路1 000 m、高速800 m、国道500 m、省道400 m、县道300 m、乡道60 m)附近都多有滑坡发育。

除此之外,植被覆盖也是滑坡危险性评价的重要指标。良好的植被覆盖情况可以固土,减少降雨时产生的径流量,降低径流速率,为抑制滑坡发育创造环境条件。长阳县研究区的植被覆盖度集中在30%~60%,而滑坡多发生在10%~30%的覆盖度区域内。

3.2 因子相关性分析

在滑坡危险性评价中,由于滑坡成因复杂多样,选取的评价因子可能会相互产生影响,而具有较强相关性的因子会降低危险性评价模型的预测结果精度。因此对12个评价因子采用皮尔逊系数(Pearson correlation coefficient, PCC)检验各个因子间的相关性,通过相关性分析剔除相关性高的因子,降低模型出现过拟合现象的概率,达到提高模型精度和泛化能力的目的。当皮尔逊系数的绝对值越接近于0,表明因子间的相关性越弱,如表1所示,最大的皮尔逊系数绝对值为0.256,表明本文选取的评价因子间不存在明显相关性。

4 滑坡危险性评价结果及模型精度评价

4.1 滑坡危险性评价模型的构建

选取研究区滑坡点缓冲区外300 m 的非滑坡点263个以及263个滑坡点,计算其各个因子的信息量值(表2),并对其进行归一化处理,构成含有信息量特征值和分类的数据集,分为70%的训练集和30%的测试集,输入到支持向量机和梯度提升决策树训练模型中。为防止数据出现过拟合现象,保证

表1 评价因子相关系数

Table 1 Correlation coefficient of evaluation factors

因子	平面 曲率	地形起 伏度	地表粗 糙度	植被覆 盖度	工程 岩组	距水系 距离	坡度	距断裂 带距离	土地利 用类型	降雨	距房屋 距离	距道路 距离
平面曲率	1.000											
地形起伏度	0.008	1.000										
地表粗糙度	-0.117	-0.209	1.000									
植被覆盖度	0.190	0.036	0.018	1.000								
工程岩组	-0.127	0.023	0.008	-0.035	1.000							
距水系距离	0.218	0.029	0.050	0.120	-0.139	1.000						
坡度	0.202	0.096	0.207	0.034	-0.009	0.079	1.000					
距断裂带距离	0.029	0.083	-0.003	0.054	0.032	-0.030	0.038	1.000				
土地利用类型	-0.040	0.050	-0.007	-0.012	0.042	0.090	-0.101	-0.079	1.000			
降雨	0.090	-0.142	0.091	0.074	-0.022	0.100	0.074	-0.074	0.068	1.000		
距房屋距离	-0.101	0.093	-0.256	0.009	-0.006	-0.026	-0.216	0.149	0.106	0.016	1.000	
距道路距离	-0.052	0.001	0.059	0.061	-0.059	-0.025	0.015	-0.003	0.064	0.151	-0.109	1.000

表 2 评价因子信息量

Table 2 Evaluation factor information

因子	分级	信息量	因子	分级	信息量
平面曲率	0 ~ 18.34	-0.159	地形起伏度	0 ~ 16.95	1.430
	18.34 ~ 35.37	0.158		16.95 ~ 35.60	0.013
	35.37 ~ 59.28	-0.119		35.60 ~ 61.03	-0.111
	59.28 ~ 85.52	1.116		61.03 ~ 433.99	-0.060
地表粗糙度	1 ~ 1.08	-0.148	工程岩组	第四系松散岩类	5.535
	1.08 ~ 1.29	0.146		坚硬碎屑岩类	-0.368
	1.29 ~ 1.65	0.379		坚硬碳酸盐岩类	0.294
	1.65 ~ 6.36	0.355		软硬相间碎屑岩类	0.369
软硬相间碳酸岩				-0.337	
坡度	< 8°	1.092	降雨	1 200 ~ 1 400 mm	-0.352
	8° ~ 25°	-0.166		1 400 ~ 1 600 mm	0.109
	25° ~ 60°	0.314		1 600 ~ 1 800 mm	0.509
	> 60°	0.000		1 800 ~ 2 000 mm	0.454
植被覆盖度	< 10%	0.023		2 000 ~ 2 200 mm	0
	10% ~ 30%	0.509	距水系距离	< 200 m	0.467
	30% ~ 60%	-0.166		200 ~ 400 m	-0.121
	> 60%	0.624		400 ~ 600 m	0.049
600 ~ 800 m				-0.548	
距断裂带距离	> 500 m	-0.052		> 800 m	0.051
	150 ~ 500 m	0.492	土地利用类型	建设用地	-0.132
0 ~ 150 m	-0.087	林地		-0.096	
距房屋距离	< 40 m	0.142		水域	2.433
	-0.110	> 40 m		耕地	0.339
距道路距离	缓冲区外*	0.097		草地	0.900
	缓冲区内	-0.186			

注: * 缓冲区内指距离铁路 1 000 m、高速 800 m、国道 500 m、省道 400 m、县道 300 m、乡道 60 m 内。

模型精度,采用交叉验证,多次训练调整参数后分别求得最优结果训练模型。最后将研究区 377.65 万个栅格计算因子信息量后归一化的特征数据输入两个训练好的模型中,得到模型预测出每一栅格单元发生滑坡的概率,利用 ARCGIS 进行滑坡危险性可视化处理。

4.2 危险性结果评价分析

信息量支持下的 SVM 模型和 GBDT 模型预测研究区共 377.65 万个栅格滑坡发生的概率,输出预测概率范围在 0 ~ 1,并通过 ARCGIS 的自然间断点分类法将结果分为极高危险区、高危险区、中危险区和低危险区 4 个危险性分区等级,得到了两个模型的危险性分区预测结果图如图 3、图 4 所示。对比两个模型的预测结果,共同点在于极高和高危险区在研究区上分布的位置大致相同,主要在研究区的西部、中部和东部,且主要沿水系周围分布,I-SVM 模型和 I-GBDT 模型预测结果中分别有 64.89% 和 67.94% 的滑坡位于该区域。结合评价因子图分析(如图 2 所示),极高和高危险区内滑坡主要受地形地貌的控制,集中发育于坡度 25° ~ 60° 的地段,此坡度范围内岩石风化物易堆积,堆积层较厚,一般位于坡体中下部靠近坡脚,是降雨与坡面径流的汇集处,水流交换最强烈的地区,该区域

内的岩土体性质会被劣化,表部岩土体也会产生较高的动、静水压力,为滑坡的发育提供良好条件。且研究区内滑坡多为土质滑坡主要分布在碎屑岩类区,揭露的岩性以页岩、粉砂岩、泥岩为主,为软弱地层,软弱接触带处易出现较强烈的错动带和风化带。并且在极高和高危险区内,建房修路切坡等人类工程活动较频繁,在人类活动的影响下造成岩体出露,出露的碎屑岩受风化作用和降水作用的影响,使得强弱风化面被浸湿,加重土体重量,导致剪切强度减弱,造成破碎的风化岩石发生滑坡。

如表 3 所示 I-SVM 评价模型中,由极高到低 4 个危险性分区的面积占比分别为 15.86%、21.29%、33.51% 和 28.68%,其中中危险区的分区面积占比最大,极高危险区面积占比最小且灾害点最多。在 I-GBDT 评价模型中,由极高到低 4 个危险性分区的面积占比分别为 30.08%、7.41%、13.28% 和 49.22%,面积占比最大的为低危险区,最小的为高危险区。但二者所预测的危险性分区中灾害密度都呈现出由极高到低危险区逐渐降低的规律,随着等级变高灾害点增多灾害密度增大,这与实际相符,说明两个模型的预测评价结果在整体上是合理的。然而与 I-SVM 模型评价结果相比,I-GBDT 模型预测极高危险区的面积不是最小且相对

表3 模型危险性预测分区结果对比

Table 3 Comparison of model hazard prediction partition results

分区模型	I-SVM			I-GBDT		
	分区占比/%	灾害点/个	灾害密度/ [个·(100 km ²) ⁻¹]	分区占比/%	灾害点/个	灾害密度/ [个·(100 km ²) ⁻¹]
极高危险区	15.86	114	2.74	30.08	155	1.97
高危险区	21.95	56	0.97	7.41	23	1.18
中危险区	33.51	55	0.63	13.28	31	0.89
低危险区	28.68	37	0.49	49.22	53	0.41

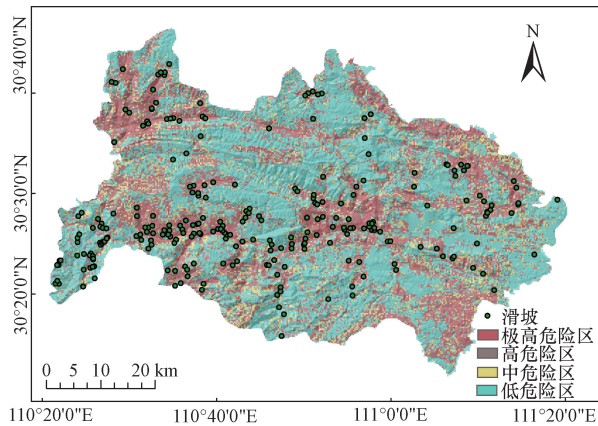


图3 I-GBDT模型评价结果

Fig. 3 Evaluation results of I-GBDT model

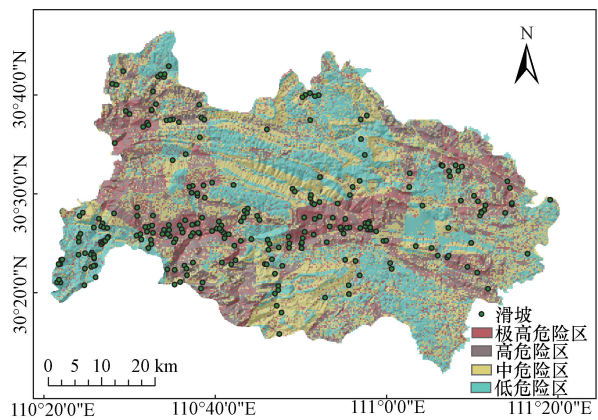


图4 I-SVM模型评价结果

Fig. 4 Evaluation results of I-SVM model

较大,致使灾点在不同危险性等级中的数量分布规律与实际相比可能存在误差,相对而言 I-SVM 模型的评价结果更符合事实规律。

4.3 模型评价结果检验

运用模型对滑坡危险性进行评价,很大程度上提高了评价结果的准确性、可靠性。直观的数据会使得模型更有信服力,本文中采用绘制模型 ROC 曲线(receiver operating characteristic curve)以检验模型的精确度,曲线能够分析所用模型特异性和灵敏性的关系,通过曲线下区域面积 AUC 值(area under curve)衡量模型的精确度,一般 AUC 取值 0.5 ~ 1,

越接近 1,模型预测效果越好,即当曲线呈“凸”形态越靠近 y 轴时 AUC 值越大,表示模型精度越高。绘制两个评价模型的 ROC 曲线(图 5),得到 I-SVM 模型和 I-GBDT 模型的 AUC 值分别为 0.859 和 0.829,说明两个模型在本研究区的预测评价结果均有良好的精确性,但 I-SVM 模型的精度要略高于 I-GBDT 模型。

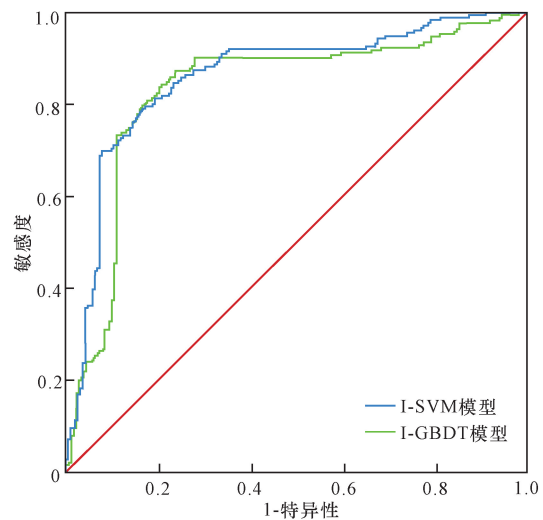


图5 评价结果 ROC 曲线

Fig. 5 Evaluation results ROC curve

5 结论

(1) 本文选取了平面曲率、地形起伏度、地表粗糙度、坡度、植被覆盖度、工程岩组、距断裂带距离、距水系距离、降雨量、土地利用类型、距房屋距离和距道路距离 12 个评价指标,经过相关性分析表明无明显相关关系均被采用,其中历史滑坡点分布规律主要受水系影响,工程岩组中的第四系松散岩类是模型滑坡预测贡献最大的指标。

(2) 通过 ROC 精度检验得到 I-SVM 模型和 I-GBDT 模型的 AUC 值分别为 0.859 和 0.829,均有较良好的精确性,但 I-GBDT 模型预测的滑坡发生的概率过于集中在过低和过高的范围,使得分区结果明显偏向极高危险区和低危险区,I-SVM 模型预测的不同危险等级分区占比更符合实际,对该研究

区进行滑坡危险性评价结果更合理。可以将 I-SVM 模型所预测的极高危险区和高危险区作为滑坡防治工作的参考依据,对重点地区优先加大防治力度。

参 考 文 献

- [1] 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 Xiaoping, Lai Chengguang, Chen Xiaohong, et al. A landslide hazard assessment based on random forest weight: a case study in the Dongjiang River Basin[J]. Journal of Natural Disasters, 2017, 26(5): 119-129.
- [2] 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.
- [3] 何万才, 赵俊三, 林伊琳, 等. 基于证据权和支持向量机模型的威信县滑坡易发性评价[J]. 科学技术与工程, 2023, 23(15): 6350-6360.
He Wancai, Zhao Junsan, Lin Yilin, et al. Landslide susceptibility assessment in Weixin county based on evidence weight and support vector machine model[J]. Science Technology and Engineering, 2023, 23(15): 6350-6360.
- [4] 盛明强, 刘梓轩, 张晓晴, 等. 基于频率比联接法和支持向量机的滑坡易发性预测[J]. 科学技术与工程, 2021, 21(25): 10620-10628.
Sheng Mingqiang, Liu Zixuan, Zhang Xiaoqing, et al. Landslide susceptibility prediction based on frequency ratio analysis and support vector machine[J]. Science Technology and Engineering, 2021, 21(25): 10620-10628.
- [5] Zhang T Y, Fu Q, Wang H. et al. Bagging-based machine learning algorithms for landslide susceptibility modeling[J]. Nat Hazards, 2022, 110: 823-846.
- [6] Leo B. Random forests[J]. Machine Learning, 2001, 45(1): 5-32.
- [7] 胡杨, 张紫昭, 林世河. 基于证据权与逻辑回归耦合的新疆伊犁河谷地区滑坡易发性评价[J]. 工程地质学报, 2023, 31(4): 1350-1363.
Hu Yang, Zhang Zizhao, Lin Shihe. Evaluation of landslides susceptibility in Ili Valley, Xing Jiang based on the coupling of WOE model and LOGISTIC regression[J]. Journal of Engineering Geology, 2023, 31(4): 1350-1363.
- [8] Vapnik V N. Statistical learning theory [M]. New York: Wiley, 1998.
- [9] Tian Y Y, Xu C, Hong H Y, et al. Mapping earthquake-triggered landslide susceptibility by use of artificial neural network (ANN) models: an example of the 2013 Minxian (China) Mw 5. 9 event[J]. Geomatics, Natural Hazards and Risk, 2019, 10(1): 1-25.
- [10] 傅文杰. GIS 支持下基于支持向量机的滑坡危险性评价[J]. 地理科学, 2008, 28(6): 838-841.
Fu Wenjie. Landslide hazard evaluation based on GIS and SVM [J]. Scientia Geographica Sinica, 2008, 28(6): 838-841.
- [11] 韩玲, 张庭瑜, 张恒. 基于 IOE 和 SVM 模型的府谷镇滑坡易发性分区[J]. 水土保持研究, 2019, 26(3): 367-372.
Han Ling, Zhang Tingyu, Zhang Heng. Landslide susceptibility mapping based on IOE and SVM model in Fugu town[J]. Research of Soil and Water Conservation, 2019, 26(3): 367-372.
- [12] 李家艳, 关艳丽, 徐永兵, 等. 基于 I-SVM 耦合模型的区域滑坡易发性评价[J]. 地质灾害与环境保护, 2024, 35(1): 19-25.
Li Jiayan, Guan Yanli, Xu Yongbing, et al. Regional landslides susceptibility evaluation based on I-SVM coupled model[J]. Journal of Geological Hazards and Environment Preservation, 2024, 35(1): 19-25.
- [13] 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 Nianqin, Guo Youjin, Liu Tieming, et al. Landslide susceptibility assessment based on support vector machine model[J]. Science Technology and Engineering, 2019, 19(35): 70-78.
- [14] 杨晨. 基于多种算法调优的支持向量机对滑坡易发性评价的精度影响研究[D]. 昆明: 昆明理工大学, 2024.
Yang Chen. A study on the impact of support vector machines on the accuracy of landslides susceptibility evaluation based on multiple algorithm tuning[D]. Kunming: Kunming University of Science And Technology, 2024.
- [15] Friedman J H. Stochastic gradient boosting [J]. Computational Statistics & Data Analysis, 2002, 38(4): 367-378.
- [16] 王恒, 姜亚楠, 张欣, 等. 基于梯度提升算法的岩性识别方法[J]. 吉林大学学报: 地球科学版, 2021, 51(3): 940-950.
Wang Heng, Jiang Yanan, Zhang Xin, et al. Lithology identification method based on gradient boosting algorithm[J]. Journal of Jilin University: Earth Science Edition, 2021, 51(3): 940-950.
- [17] Rong G Z, Alu S, Li K W, et al. Rainfall induced landslide susceptibility mapping based on bayesian optimized random forest and gradient boosting decision tree models: a case study of Shuicheng County, China[J]. Water, 2020, 12(11): 3066.
- [18] Dong J H, Niu R Q, Chen T, et al. Assessing landslide susceptibility using improved machine learning methods and considering spatial heterogeneity for the Three Gorges Reservoir Area, China [J]. Nat Hazards, 2024, 120: 1113-1140.
- [19] 郭衍昊, 窦杰, 向子林, 等. 基于优化负样本采样策略的梯度提升决策树与随机森林的汶川同震滑坡易发性评价[J]. 地质科技通报, 2024, 43(3): 251-265.
Guo Yanhao, Dou Jie, Xiang Zilin, et al. Optimized negative sampling strategies of gradient boosting decision tree and random forest for evaluating Wenchuan coseismic landslides susceptibility mapping[J]. Bulletin of Geological Science and Technology, 2024, 43(3): 251-265.
- [20] 董张玉, 张晋, 彭鹏, 等. 基于 GBDT-LR 和信息量模型耦合的滑坡易发性评价[J]. 水土保持通报, 2023, 43(1): 149-157, 166.
Dong Zhangyu, Zhang Jin, Peng Peng, et al. Landslide susceptibility evaluation based on coupling of GBDT-LR model and Information model landslides[J]. Bulletin of Soil and Water Conservation, 2023, 43(1): 149-157, 166.
- [21] 李勇, 宋英旭. 基于 GBDT 模型的广东阳春市地质灾害易发性评价研究[J]. 矿产勘查, 2023, 14(12): 2434-2446.

- Li Yong, Song Yingxu. Research on the susceptibility evaluation of geological disasters in Guangdong Yangchun City based on the GBDT model [J]. Mineral Exploration, 2023, 14 (12): 2434-2446.
- [22] 郭佳, 赵之星, 刘志奇, 等. 结合信息量与 AHP 模型的阳泉市矿区地质灾害风险评价 [J]. 测绘通报, 2022 (11): 101-105.
- Guo Jia, Zhao Zhixing, Liu Zhiqi, et al. Risk assessment of geological disaster based on information value and AHP model in mining area of Yangquan city [J]. Bulletin of Surveying and Mapping, 2022(11): 101-105.
- [23] 杨淞月, 翁茂芝, 田华, 等. 湖北省鹤峰县城区滑坡地质灾害风险评价 [J]. 安全与环境工程, 2021, 28(1): 144-155.
- Yang Songyue, Weng Maozhi, Tian Hua, et al. Landslide risk evaluation in the urban area in Hefeng county, Hubei Province [J]. Safety and Environmental Engineering, 2021, 28 (1): 144-155.
- [24] Mezaal M R, Pradhan B. An improved algorithm for identifying shallow and deep-seated landslides in dense tropical forest from airborne laser scanning data [J]. Catena, 2018, 167: 147-159.
- [25] 詹华炜. 基于机器学习的崩塌滑坡易发性评价方法研究——以广州市白云区为例 [D]. 广州: 广州大学, 2023.
- Zhan Huawei. Research on the assessment method of collapse landslide susceptibility based on machine learning: a case study of Baiyun district, Guangzhou city [D]. Guangzhou: Guangzhou University, 2023.