

基于SOM-I-SVM耦合模型的滑坡易发性评价

贾雨霏, 魏文豪, 陈 稳, 杨清卓, 盛逸凡, 徐光黎

Landslide susceptibility assessment based on the SOM-I-SVM model

JIA Yufei, WEI Wenhao, CHEN Wen, YANG Qingzhuo, SHENG Yifan, and XU Guangli

在线阅读 View online: <https://doi.org/10.16030/j.cnki.issn.1000-3665.202206041>

您可能感兴趣的其他文章

Articles you may be interested in

基于滑坡分类的西宁市滑坡易发性评价

Landslide susceptibility assessment in Xining based on landslide classification

孙长明, 马润勇, 尚合欣, 谢文波, 李焱, 刘义, 王彪, 王思源 水文地质工程地质. 2020, 47(3): 173-181

基于逻辑回归信息量的川藏交通廊道滑坡易发性评价

Landslide susceptibility mapping in the Sichuan-Tibet traffic corridor using logistic regression-information value method

杜国梁, 杨志华, 袁颖, 任三绍, 任涛 水文地质工程地质. 2021, 48(5): 102-111

基于自组织神经网络的污染场地多监测指标相关性分析

Correlation analysis of multiple monitoring indicators of contaminated site based on self-organizing map

马春龙, 施小清, 许伟伟, 任静华, 王佩, 吴吉春 水文地质工程地质. 2021, 48(3): 191-202

基于遗传算法-支持向量机的滑坡渗透系数反演

Estimation of hydraulic conductivity of landslides based on support vector machine method optimized with genetic algorithm

胡鹏, 文章, 胡新丽, 张玉明 水文地质工程地质. 2021, 48(4): 160-168

考虑基质吸力作用的Newmark改进模型在地震滑坡风险评价中的应用

Application of Newmark improved model considering matrix suction in earthquake landslide risk assessment

冯卫, 唐亚明, 赵法锁, 陈新建 水文地质工程地质. 2019, 46(5): 154-160

基于逻辑回归的四川青川县区域滑坡灾害预警模型

A early warning model of regional landslide in Qingchuan County, Sichuan Province based on logistic regression

方然可, 刘艳辉, 苏永超, 黄志全 水文地质工程地质. 2021, 48(1): 181-187



关注微信公众号, 获得更多资讯信息

DOI: 10.16030/j.cnki.issn.1000-3665.202206041

贾雨霏, 魏文豪, 陈稳, 等. 基于 SOM-I-SVM 耦合模型的滑坡易发性评价 [J]. 水文地质工程地质, 2023, 50(3): 125-137.

JIA Yufei, WEI Wenhao, CHEN Wen, *et al.* Landslide susceptibility assessment based on the SOM-I-SVM model[J]. Hydrogeology & Engineering Geology, 2023, 50(3): 125-137.

基于 SOM-I-SVM 耦合模型的滑坡易发性评价

贾雨霏, 魏文豪, 陈 稳, 杨清卓, 盛逸凡, 徐光黎
(中国地质大学(武汉)工程学院, 湖北 武汉 430074)

摘要: 在使用机器学习模型对滑坡进行易发性评价时, 通常会在滑坡影响范围之外随机选取非滑坡样本点, 具有一定的误差。为了提高滑坡易发性评价的精度, 将自组织映射(self-organizing map, SOM)神经网络、信息量模型(information, I)以及支持向量机模型(support vector machine, SVM)进行耦合, 提出一种基于 SOM-I-SVM 模型的滑坡易发性评价方法, 并将 SOM 神经网络与 K 均值聚类算法进行对比, 验证模型的可靠性。以十堰市茅箭区为例, 首先通过对环境因子的相关性及其重要性分析, 筛选出距水系距离、坡度、降雨量、距构造距离、相对高差、距道路距离、地层岩性等 7 个因子, 建立滑坡易发性评价指标体系, 在此基础上计算出各因子的分级信息量值, 并作为模型的输入变量进行滑坡易发性评价。分别采用 SOM 神经网络和 K 均值聚类算法选取非滑坡样本, 然后将样本数据集代入 I-SVM 模型预测滑坡易发性。将 SVM、I-SVM、KMeans-I-SVM、SOM-I-SVM 等 4 种模型预测精度进行对比, 其 ROC 曲线下面积(AUC)分别为 0.82, 0.88, 0.90, 0.91, 说明 SOM-I-SVM 模型能有效提高滑坡易发性预测准确率。

关键词: 滑坡; 易发性评价; 信息量模型; 支持向量机; 自组织映射神经网络

中图分类号: P642.22

文献标志码: A

文章编号: 1000-3665(2023)03-0125-13

Landslide susceptibility assessment based on the SOM-I-SVM model

JIA Yufei, WEI Wenhao, CHEN Wen, YANG Qingzhuo, SHENG Yifan, XU Guangli
(Faculty of Engineering, China University of Geosciences (Wuhan), Wuhan, Hubei 430074, China)

Abstract: When using machine learning models for landslide susceptibility evaluation, the non-landslide sample points are usually selected randomly outside the landslide influence area, leading to a certain error. To improve the accuracy of landslide susceptibility evaluation, this paper couples the self-organizing map (SOM) neural network, information (I) model, and support vector machine (SVM) model, and proposes a SOM-I-SVM model-based method of landslide susceptibility evaluation, comparing with K-means clustering to verify the reliability of this model. The Maojian District of the city of Shiyan is taken as an example, and seven factors of the distance from water system, slope, rainfall, distance from structure, relative height difference, distance from road, stratigraphic lithology are selected by correlation and importance analyses of environmental factors to establish a landslide susceptibility evaluation system. Based on these, the graded information values of each factor are calculated and used as input variables for landslide susceptibility evaluation. The SOM neural network and K-means clustering

收稿日期: 2022-06-19; 修订日期: 2022-10-03

投稿网址: www.swdzgcdz.com

基金项目: 湖北省科技厅研发项目(2021BCA219)

第一作者: 贾雨霏(1998-), 女, 硕士研究生, 主要从事地质灾害分析与防治的研究。E-mail: jiayufei@cug.edu.cn

通讯作者: 徐光黎(1963-), 男, 教授, 主要从事地质工程与地质灾害方面的教学研究工作。E-mail: xu1963@cug.edu.cn

are used to select non-landslide samples, and the sample data set is substituted into the I-SVM model to predict landslide susceptibility. The prediction accuracies of the four models, SVM, I-SVM, KMeans-I-SVM and SOM-I-SVM, are compared, and the area under the ROC curve (AUC values) are 0.82, 0.88, 0.90 and 0.91, indicating that the SOM-I-SVM model can effectively improve the accuracy of landslide susceptibility prediction.

Keywords: landslide; susceptibility assessment; information model; SVM; SOM

地质灾害易发性评价是以地质环境条件为基础,结合地质灾害孕灾因子和诱发因子预测其在一定区域内发生的可能性大小^[1]。Cui 等^[2]将地质灾害易发性分析方法分为统计学方法、概率方法以及数据挖掘方法。常用的统计学方法包括逻辑回归^[3]、多元线性回归^[4]等;概率方法包括信息量法^[5]、熵指数法^[6]、确定性系数法^[7]等;数据挖掘方法包括人工神经网络^[8]、支持向量机^[9]、随机森林^[10]等。目前,国内外许多学者采用不同方法建立滑坡灾害易发性评价模型并取得了较好的结果:如 Yesilnacar 等^[11]采用逻辑回归模型和神经网络模型对土耳其 Hendek 地区进行滑坡灾害易发性评价;万洋等^[12]建立 MaxEnt 模型对中尼交通廊道进行滑坡易发性分析;黄武彪等^[13]基于层数自适应的加权卷积神经网络进行川藏铁路沿线滑坡易发性评价,解决了卷积神经网络权重初值、网络层数等超参数难以优化的问题。

采用单一方法建立的模型在进行滑坡灾害易发性评价时能取得不错的结果,但也存在一定不足^[14]。如信息量(information, I)模型操作简单,实用性强,在工作生产中已被广泛应用,但在评价过程中没有考虑不同因子对地质灾害产生影响的差异性^[15]。为了减小单一信息量模型造成的误差,陈飞等^[16]将信息量与神经网络模型进行耦合,对江西省上犹县进行滑坡易发性评价,结果表明耦合模型预测精度更高。支持向量机(support vector machine, SVM)模型以结构风险最小为原则,能够较好地处理小样本、高维数据和非线性问题,具有很好的泛化能力^[17-18],但在实际应用中受因子量纲不统一等问题影响较大,且超参数的选择直接影响着模型的准确度和泛化能力。徐胜华等^[19]将熵指数融入支持向量机进行滑坡灾害易发性评价,并采用 PSO 算法求解 SVM 最优解,提高了预测精度。由此可以看出,通过多种方法的耦合以及算法的优化,可以有效提高模型预测的可靠性。

使用机器学习算法虽然能够提高滑坡易发性评价的准确性,但模型的预测效果受训练样本的影响较大,因此许多专家学者通过聚类方法对滑坡易发性进行初评,从低易发区中选择非滑坡负样本进行机器学

习训练,如:刘福臻等^[20]使用 K 均值聚类算法优化了神经网络模型,将训练精度由 73% 提升到 97%;鲍帅等^[21]联合 DBSCAN 聚类及 SVM 方法对四川省绵阳市进行滑坡易发性评价,比传统 SVM 得到了更高的精度。目前,常用的聚类算法包括 K 均值聚类、层次聚类以及 SOM(self-organizing map)神经网络等。

为了减小信息量模型和支持向量机模型在滑坡易发性评价时带来的误差,提高模型的预测精度,以十堰市茅箭区为例,综合考虑滑坡灾害发生的孕灾条件和诱发因素,通过影响因子的相关性分析,选择合适的评价因子,并对各评价因子进行分级。分别通过 K 均值聚类和 SOM 神经网络进行滑坡负样本选取,结合信息量模型和支持向量机模型,建立 SVM、I-SVM、KMeans-I-SVM、SOM-I-SVM 共 4 种模型进行滑坡灾害易发性评价,并采用受试者工作特征曲线(ROC 曲线)评价 4 种模型的准确性。通过对比分析得到茅箭区的滑坡灾害易发性区划,从而为地质灾害风险评估和预防治理提供依据。

1 研究区概况

湖北省十堰市茅箭区位于武当山的西北麓,属秦岭、大巴山的东延余脉,区域总面积 490.71 km²,研究区地理位置如图 1 所示。该区多年平均气温 15.4 °C,多年平均降水量为 769.6 mm。区内地形地貌受地质构造及岩性的控制,由于长期的剥蚀和河流下切,形成了构造剥蚀低山丘陵地形和构造剥蚀低中山地形和侵蚀堆积地形 3 种地貌类型,总体地势南高北低中间凹,海拔约 200 ~ 1 720 m。地质灾害发育较广泛,其中滑坡灾害较为突出,区内滑坡共 78 处,占地质灾害总数的 59.85%。按滑坡规模分类,大型滑坡共 1 处、中型滑坡共 7 处,小型滑坡共 70 处,区内滑坡规模以小型为主,占滑坡总数的 89.74%。

研究区位于新华夏系第三隆起带,属秦岭褶皱系南秦岭印支冒地槽褶皱带武当山复背斜区,为武当山复背斜中的两郧背斜和武当山背斜的一部分,根据构造形成的展布方向和力学性质,可分为北西向构造体系 and 新华夏构造体系。区域性变质作用总体以浅至

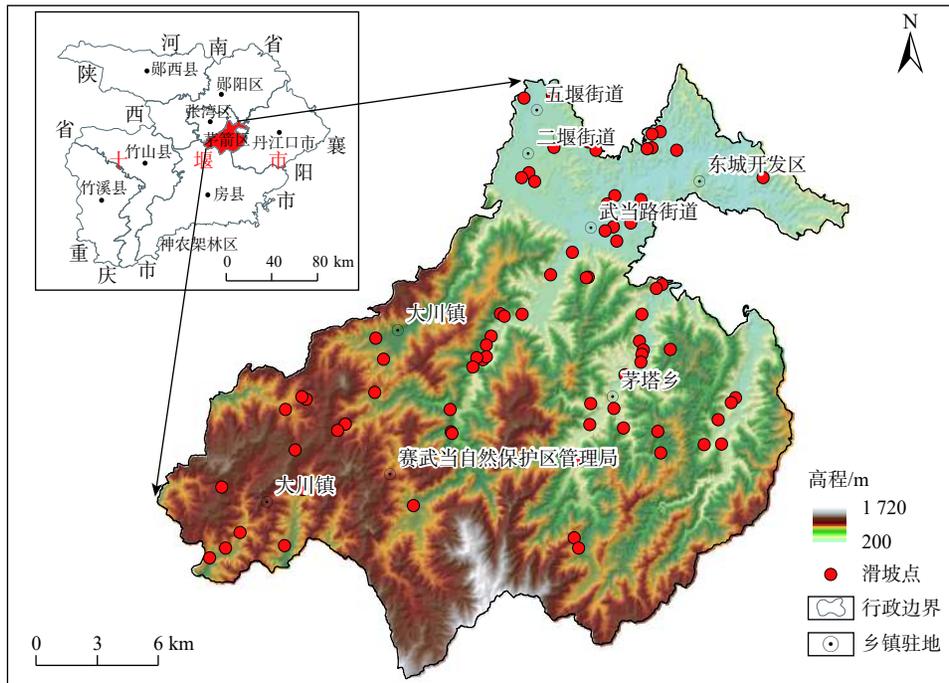


图 1 茅箭区地理位置示意图

Fig. 1 Location of the Maojian District

中等变质为主, 表现为片理发育。区内出露地层由新至老主要为第四系、元古界武当山群, 伴有岩浆岩侵入。侵入岩为早古生代岩浆岩, 分散分布于调查区, 多为脉状及顺层贯入的小型侵入体, 大部分与围岩同时变质, 其岩性主要为变辉绿岩。

2 研究方法及数据源

2.1 信息量模型

信息量模型最早由 Shannon 提出, 是基于信息论发展而来的预测评价方法, 早期应用于矿产领域^[22], 近年来在地质灾害领域得到广泛应用。信息量模型利用地质灾害发生频率或密度将影响因子(x_i)对地质灾害现象(Y)的影响转化为信息熵, 信息熵越大表明该因子的影响作用越大, 最终将各因子的信息熵进行空间叠加, 实现区域滑坡易发性评价。其中, 第 i 个影响因子(x_i)对地质灾害事件发生(Y)提供的信息量 $I(x_i, Y)$ 可表示为:

$$I(x_i, Y) = \ln \frac{N_i/N}{S_i/S} \quad (1)$$

式中: x_i ——评价单元内所取的因子等级;

$I(x_i, Y)$ ——因子 x_i 对地质灾害所贡献的信息量;

S ——研究区面积;

S_i ——研究区内含有因子 x_i 的面积;

N ——研究区内地质灾害总数;

N_i ——发生地质灾害区域中含有因子 x_i 的数量。

根据单一指标因子的信息量值可以计算出整个研究区内评价单元的信息量, 表达式为:

$$I = \sum_{i=1}^n I(x_i, Y) = \sum_{i=1}^n \ln \frac{N_i/N}{S_i/S} \quad (2)$$

式中: I ——评价单元的所有因子提供的总信息量值;

n ——影响因子的数量。

2.2 K 均值聚类

K 均值聚类又称 KMeans 聚类算法, 最早由 Lloyd 提出^[23], 该方法是一种无监督分类方法。将 N 个样本划分为 K 类, 每一类样本称为一个簇, KMeans 算法的核心是找到 K 个无交集的簇中最优的质心 u_{ij} 。假设数据集 N 有 j 个样本, 将样本分为 k 簇, 那么一个簇中所有样本点到质心距离的平方和就称为簇内平方和, 公式为:

$$CSS = \sum_{j=1}^m \sum_{i=1}^n (x_i - u_{ij})^2 \quad (3)$$

式中: m ——一个簇内样本的个数;

j ——每个样本的编号;

n ——每个样本的因子数目;

i ——该样本点的第 i 个因子。

KMeans 算法通过迭代计算出能够使簇内平方和最小的质心, 从而对样本进行分类。

2.3 SOM 神经网络

SOM 神经网络又称自组织映射网络,是一种无监督自主学习神经网络^[24],其层次结构一般包括输入层和输出层(图 2)。SOM 采用 Kohonen 算法,通过聚类学习寻找样本之间的规律从而实现分类,因此该算法不需要期望输出,也不需要输入层进行先验学习,而是分析输入变量的相似性进行分类。对于输入数据集 $D = \{x_1, x_2, \dots, x_n\}$, 变量的相似性可以通过输入向量的欧氏距离 d_i 计算得到:

$$d_i = \sqrt{\sum_{i=1}^n (x_i - w_i)^2} \quad (4)$$

式中: x_i ——输入向量;

w_i ——输入向量和输出向量的连接权值向量,在初始化网络时,随机为每个节点的权值向量赋值。

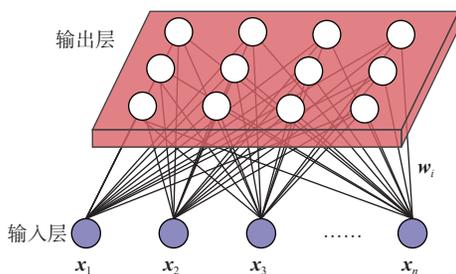


图 2 SOM 神经网络

Fig. 2 SOM Neural Network

若第 i 个节点的 d_i 最小,则称该节点为获胜神经元。以获胜神经元为中心,调整其邻近神经元的权值向量并计算学习率,当两次迭代的学习率小于设定的阈值时结束迭代,从而实现输入变量的分类。

2.4 支持向量机(SVM)模型

支持向量机算法的分类思想是基于训练集的样本空间分布,寻找一个最优的超平面作为决策边界,使模型的分类误差尽量减小。假设一组样本数据集 $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, $y_i \in \{-1, +1\}$, 决策函数可以表示为:

$$f(x) \sum_{i=1}^n w_i x_i + b = W^T X + b \quad (5)$$

式中: w ——超平面的法向量;

b ——标量。

“支持向量”是指超平面两侧距离最近的几个样本点,过两侧支持向量且平行于决策边界的两个平面被称为这条决策边界的边际(Margin),如图 3 所示,两个

边际之间的距离称为函数间隔,表示为:

$$\gamma = \frac{2}{\|w\|} \quad (6)$$

SVM 算法的最终目的就是找到最大边界的决策边界,从而实现样本数据的分类。根据式(4),将求解最大间隔问题转化为求解最小 w 的问题,并引入拉格朗日乘数 α 求极值,得到如下函数:

$$L = \frac{\|w\|^2}{2} - \sum_{i=1}^n \alpha_i [y_i (w_i \cdot x_i + b) - 1] \quad (7)$$

其约束条件为:

$$\alpha_i \geq 0, \quad \sum_{i=1}^n \alpha_i y_i = 0 \quad (8)$$

由于边际越大被分到错误类别的样本越多,因此引入非负的松弛变量 ξ 和惩罚力度系数 C , 将式(5)转化为:

$$L = \frac{\|w\|^2}{2} - C \sum_{i=1}^n \xi_i \quad (9)$$

对于非线性不可分问题,通过核函数将原始样本投影到高维空间中,使其在高维空间中线性可分,采用在多数模型中都能取得较好效果的高斯径向基函数作为核函数,如式(10)。

$$K(x, y) = e^{-\gamma \|x - y\|^2}, \gamma > 0 \quad (10)$$

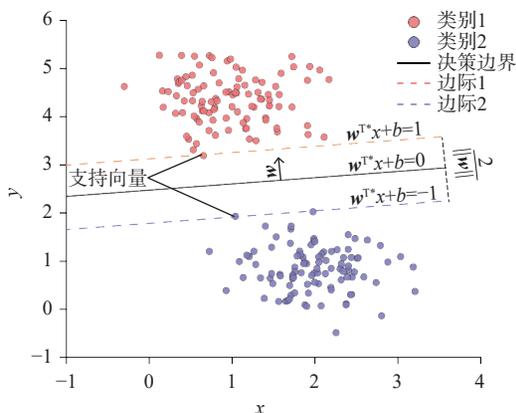


图 3 支持向量机模型

Fig. 3 Support vector machine model

2.5 SOM-I-SVM 耦合模型

SVM 模型可以反映出各指标因子之间的非线性关系,同时以信息量为载体可以解决因子量纲不统一的问题,充分结合两种模型的优势进行滑坡灾害易发性评价, SOM-I-SVM 耦合模型的步骤主要为: (1) 基于相关性分析选取的评价因子,用 SOM 神经网络进行初始易发性分区,在低易发区中随机生成非滑坡样本

点; (2)对评价因子进行分级并计算出每个因子的分级信息量值; (3)将各因子的信息量值作为 SVM 模型的

输入变量并对模型进行训练, 反复调整参数创建出 SOM-I-SVM 模型, 建模流程如图 4。

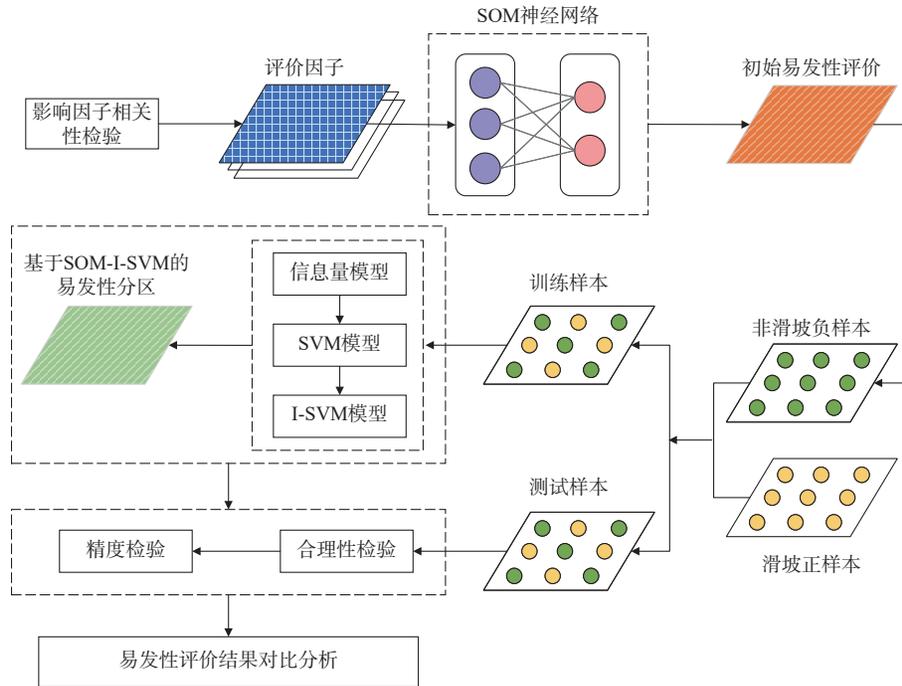


图 4 SOM-I-SVM 模型建模流程

Fig. 4 Process of the SOM-I-SVM model

2.6 数据来源

数据源于十堰市自然资源和规划局支撑项目“十堰市茅箭区第一次全国自然灾害综合风险普查”相关资料, 主要包括: (1)历史滑坡数据及地质灾害调查资料; (2)数字高程模型(DEM), 用于提取高程、坡度等评价因子; (3)1:50万《湖北省区域环境地质调查》, 用于提取地层岩性、地质构造等评价因子。选用分辨率为 20 m 的栅格单元作为评价单元, 利用 ArcGIS 软件将区域划分为 1 219 394 个栅格单元。

3 评价指标的选取与确定

3.1 评价因子的选择

滑坡灾害的发生受基础地质条件与外在环境因素共同影响^[25]。根据研究区概况, 结合影响滑坡发生的内在因子和诱发因素, 选取相对高差、坡度、距构造距离、地层岩性、距水系距离、降雨量、建筑指数、距道路距离共 8 个影响因子进行分析。

(1) 相对高差

相对高差是单元内海拔最高点与最低点之间的差值, 决定着坡体应力值的大小, 相对高差越大, 发生滑坡的可能性就越大。通过 ArcGIS 表面分析中的焦

点统计工具计算区域内的相对高差, 并将计算所得高差划分为 4 个等级, 如图 5(a)所示。

(2) 坡度

坡度与坡体内应力分布状态有着密切的关系, 随着坡度的增大, 坡脚的应力集中也随之增大。如图 5(b)所示, 在 DEM 数据基础上, 利用 ArcGIS 坡度工具得到坡度分布图, 并划分为 6 个等级。

(3) 距构造距离

一般来说, 距离构造近的区域, 地质体切割相对较为破碎, 在降雨等因素的作用下, 易于诱发滑坡的发生。根据 ArcGIS 缓冲区分析, 将构造缓冲距离划分为 4 个等级, 如图 5(c)所示。

(4) 地层岩性

不同类型的工程地质岩组对崩滑地质灾害形成的影响程度有明显差异, 根据岩石的坚硬程度、成分的纯度及软弱夹层等划分为四个岩组: ①松散土体; ②较坚硬-较软弱薄-厚层状变粒岩、石英片岩互层岩组; ③较坚硬中-厚层状变粒岩岩组; ④坚硬块状变辉绿岩岩组, 如图 5(d)所示。

(5) 距水系距离

水系主要表现为对河岸产冲刷破坏, 结合研究区

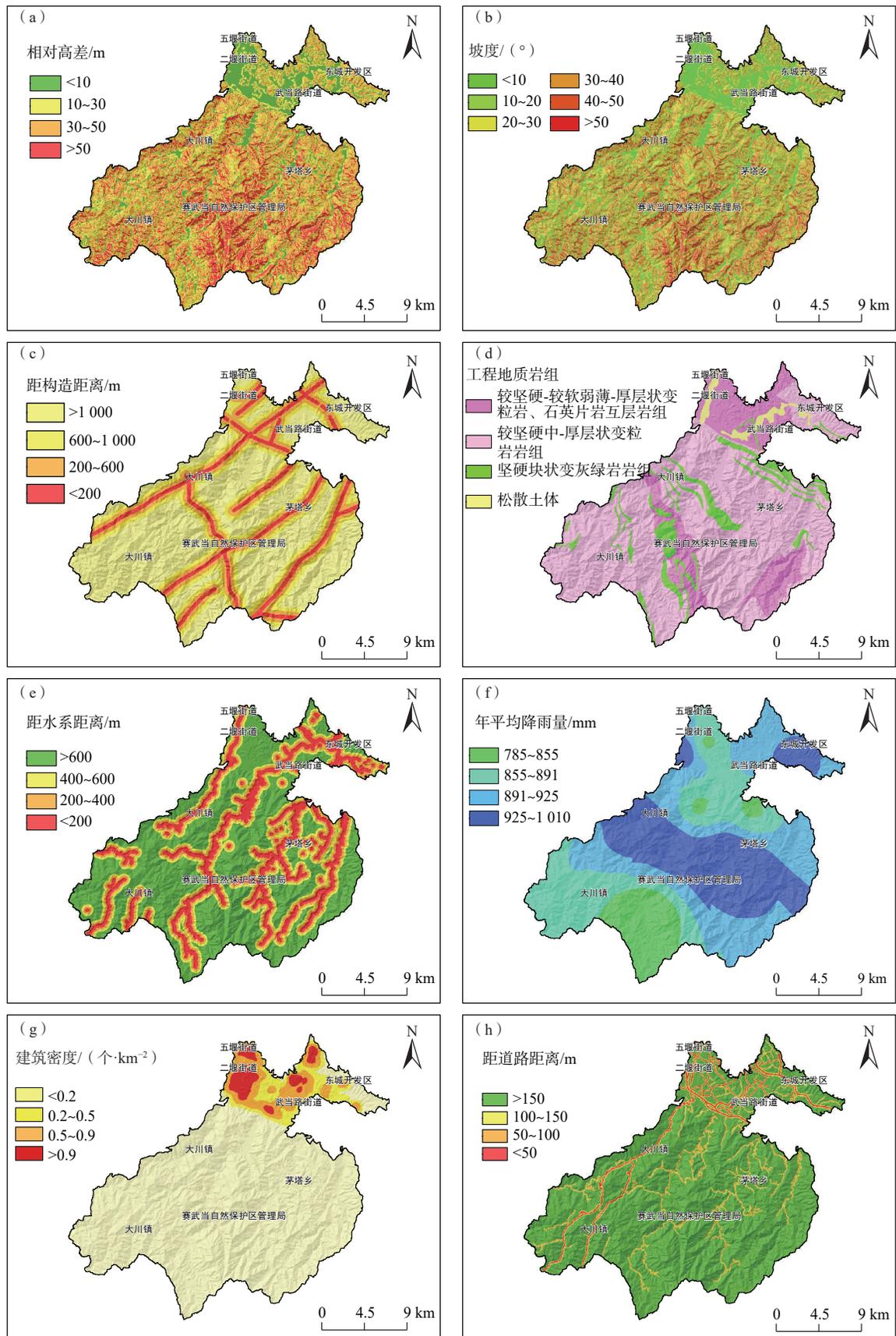


图 5 研究区滑坡灾害易发性评价指标因子

Fig. 5 Index factors of landslide hazard susceptibility evaluation in the study area

滑坡地质灾害的发育特征, 将水系影响范围划分为 4 个等级, 如图 5(e) 所示, 采用 ArcGIS 邻域分析的缓冲区工具对水系进行缓冲分析, 得出河流影响范围。

(6) 年降雨量

降雨是滑坡地质灾害发生的重要诱发因素, 根据研究区地质灾害野外调查统计分析, 绝大多数地质灾害点均受降雨的影响, 且地质灾害发生数量随降雨量的增加而增多。通过克里金插值法对收集到的年平均降雨量数据进行处理, 利用 Geostatistical Analyst 工具中的探索数据功能设置变换类型和趋势的移除阶数以及搜索领域等插值参数, 在交叉验证界面查看预测误差并不断调整参数, 得到降雨量分布图, 根据自然断点法将降雨结果划分为 4 个等级, 如图 5(f) 所示。

(7) 建筑指数

人类工程活动为研究区内地质灾害诱发的先决条件, 建筑指数是区内主要的人类工程活动之一, 地质灾害受切坡建房等类工程活动的影响较大。在 ArcGIS 中对区域内建筑进行核密度分析, 并将结果划分为 4 个等级, 如图 5(g) 所示。

(8) 距道路距离

研究区道路类型包括铁路、高速公路、国道、省道、城市道路和一般公路 6 种, 将距道路距离划分为 4 个等级, 在 ArcGIS 中使用缓冲区工具对道路进行缓冲分析, 如图 5(h) 所示。

3.2 评价因子相关性分析

对于初步选择的 8 个因子, 其两两之间可能存在不同程度的相关性, 如果不对评价因子进行相关性检验, 则模型在训练过程中会过多地选用相关性高的因子^[26], 进而降低模型预测精度。基于 SPSS 平台, 采用皮尔逊(Pearson)相关系数进行影响因子相关性分析。图 6 为影响因子的相关性热力图, 根据相关性检验结果, 建筑指数与其他多数因子表现为显著相关, 因此将该因子剔除, 最终选择距水系距离、坡度、降雨量、距构造距离、相对高差、距道路距离、地层岩性 7 个因子进行易发性评价, 表 1 为各因子分级信息量值, 由于分母不能为 0, 因此没有灾点发育的区段栅格数取 0.000 01。

3.3 评价因子重要性分析

为分析各评价因子对滑坡易发性产生的影响, 需要对其进行重要性排序。通过 Python 语言中的 Scikit-learn 模块建立随机森林模型, 将各因子信息量作为输入变量得到其重要性。图 7 为各因子重要性分布图, 从图中可以看出, 茅箭区滑坡灾害受地表水系影响最

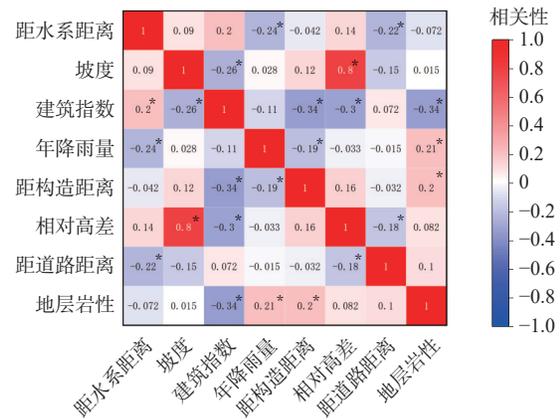


图 6 影响因子相关性热力图

Fig. 6 Correlation chart of impact factors

注: “*”表示两个因子之间相关性显著。

大, 其次为道路和地层岩性, 这与各因子信息量结果基本吻合, 该结果可以为滑坡灾害的防范工作提供参考依据。

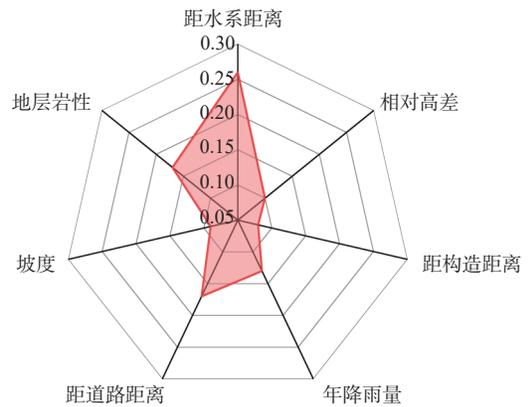


图 7 评价因子重要性分布图

Fig. 7 Importance chart of evaluation factors

4 滑坡易发性评价与精度分析

4.1 基于 SVM、I-SVM 模型的易发性评价

调查区内共有 78 处滑坡, 对于小型滑坡直接采用一个矢量点表示, 为避免对中大型滑坡描述不准确, 将灾害体实际面积大于 10 000 m² 的滑坡进行实体勾绘, 根据范围大小由滑坡不同位置的几个矢量点表示, 最终构成了 110 个滑坡正样本。利用 ArcGIS 随机点生成工具在非滑坡区域生成 110 个非滑坡点作为负样本, 将 110 个滑坡点与 110 个非滑坡点组成模型构建的样本空间, 选择其中的 70% 作为训练集, 其余 30% 作为测试集。

将 7 个评价因子的其信息量值提取至样本空间并导入模型进行训练, 其中 SVM 模型训练时输入评价

表 1 各评价因子分级信息量值
Table 1 Information values of each evaluation factor

因子	分段	灾点数	灾点栅格数	信息量	排序
相对高差/m	≤10	23	289	0.085 905 143	15
	11 ~ 30	25	593	0.234 929 441	8
	31 ~ 50	28	792	0.095 149 584	13
	>50	2	75	-1.368 920 571	28
坡度/(°)	≤10	27	398	0.090 011 487	14
	11 ~ 20	11	228	0.198 677 83	9
	21 ~ 30	24	733	0.440 412 967	7
	31 ~ 40	12	322	-0.467 576 242	23
	41 ~ 50	4	68	-0.995 249 727	27
	>50	0	0.000 01	-14.864 234 87	29
工程地质岩组	坚硬块状变辉绿岩岩组	4	98	-0.573 082 265	25
	较坚硬中-厚层状变粒岩岩组	50	1 113.75	-0.087 934 243	18
	较坚硬-较软弱薄-厚层状变粒岩、石英片岩互层岩组	24	537.25	0.518 090 502	6
	松散土体	0	0.000 01	-15.177 420 68	30
距构造距离/m	≤200	14	340	0.648 102 911	4
	201 ~ 600	9	273	-0.232 119 776	21
	601 ~ 1 000	15	255	-0.168 793 518	19
	>1 000	40	881	-0.048 684 404	17
距水系距离/m	≤200	42	738	0.783 625 576	3
	201 ~ 400	12	225	-0.214 301 889	20
	401 ~ 600	6	282	0.127 347 427	10
	>600	18	504	-0.562 859 754	24
年降雨量/mm	785 ~ 855	5	109	-0.633 53	26
	856 ~ 891	27	525	0.109 45	12
	892 ~ 925	24	526	-0.041 13	16
	926 ~ 1 010	22	589	0.115 10	11
距道路距离/m	≤50	3	54	0.555 417	5
	51 ~ 100	9	279	1.058 39	2
	101 ~ 150	20	351	1.366 02	1
	>150	45	1 065	-0.363 34	22

因子的分级赋值, I-SVM 模型训练时输入评价因子的分级信息量值。模型构建基于 Python 语言的 Scikit-learn 模块, 并采用 GridSearchCV 模块基于 10 折交叉验证方法对模型进行参数选择, 需要优化的参数有惩罚参数 C 、RBF 核函数参数 γ , 参数设定在 $[0.1, 10]$ 的区间范围, 以步长 0.1 进行搜索, 评分策略选择采用默认的 R^2 , 交叉验证的分数越高越好, 最终 SVM 模型及 I-SVM 模型的参数及评分如表 2 所示。

将全区的 1 219 394 个栅格单元转为点要素, 并提取评价因子及其信息量值带入训练好的模型计算滑坡易发性指数, 利用自然断点法将滑坡灾害易发程度划分为 4 个等级: 极高易发区、高易发区、中易发区和低易发区, 得到基于 SVM、I-SVM 模型的滑坡灾害易发性分区图(图 8、图 9)。对易发性分区结果及各等级滑坡点的分布数量进行统计(表 3), 其中 SVM 模型

表 2 SVM、I-SVM 模型参数表

Table 2 Parameters of the SVM and I-SVM models

模型类型	超参数	调试范围	步长	最佳参数	模型评分
SVM	C	0.1 ~ 10	0.1	1.1	0.773
	γ	0.1 ~ 10	0.1	0.123 28	
I-SVM	C	0.1 ~ 10	0.1	0.8	0.848
	γ	0.1 ~ 10	0.1	10.481 13	

的极高易发区占 14.6%、高易发区占 18.3%、中易发区占 26.4%、低易发区占 40.7%, I-SVM 模型的极高易发区占 13.0%、高易发区占 17.0%、中易发区占 24.8%、低易发区占 45.2%。

4.2 基于 KMeans-I-SVM、SOM-I-SVM 模型的易发性评价

4.2.1 基于 KMeans 聚类算法的易发性评价

直接将 7 个评价因子的信息量值提取至全区

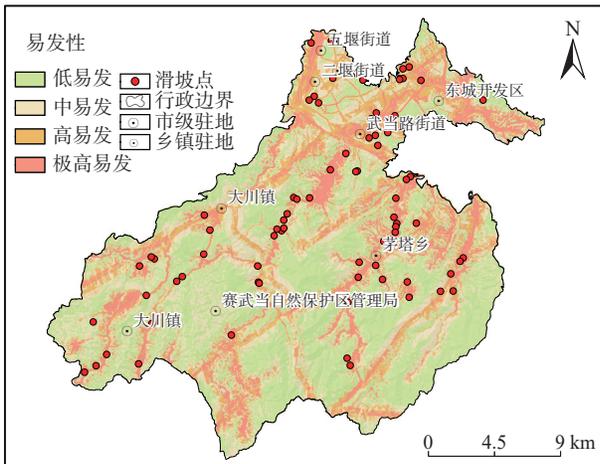


图 8 SVM 模型滑坡易发性分区

Fig. 8 Zoning map of landslide susceptibility based on the SVM model

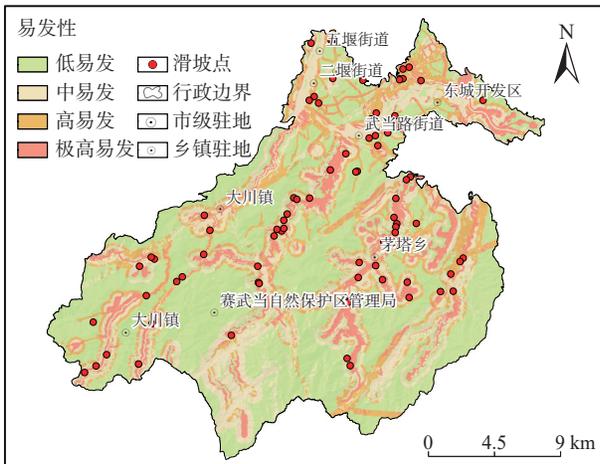


图 9 I-SVM 模型滑坡易发性分区

Fig. 9 Zoning map of landslide susceptibility based on the I-SVM model

表 3 SVM 及 I-SVM 模型分区结果

Table 3 Results of the SVM and I-SVM models

模型类型	易发性等级	分区面积 /km ²	所占比例 /%	滑坡数 /个	占滑坡总数 比例/%
SVM	低易发区	199.875	40.7	10	12.8
	中易发区	129.476	26.4	16	20.5
	高易发区	89.737	18.3	21	26.9
	极高易发区	71.622	14.6	31	39.7
I-SVM	低易发区	221.594	45.2	11	14.1
	中易发区	121.872	24.8	13	16.7
	高易发区	83.549	17.0	15	19.2
	极高易发区	63.695	13.0	39	50.0

1 219 394 个样本点中并进行归一化处理, 将处理好的样本数据作为 KMeans 聚类算法模型输入变量并得到滑坡易发性指数, 采用自然断点法将滑坡灾害易发程度划分为 4 个等级 (图 10), 各等级所占比例分别为:

极高易发区占 7.5%、高易发区占 22.7%、中易发区占 35.0%、低易发区占 34.8%, 其中高和极高易发区共有 77 个滑坡样本点, 占总数的 70%, 说明基于 K 均值聚类算法的初始易发性评价具有一定准确性。

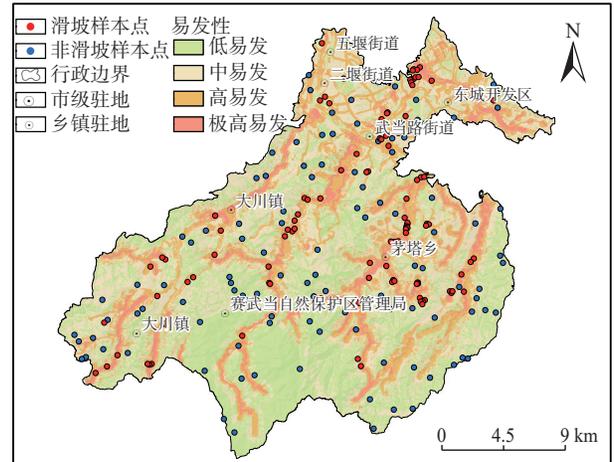


图 10 KMeans 聚类模型易发性分区及样本选择

Fig. 10 Landslide susceptibility zoning and sample selection of the KMeans model

4.2.2 基于 SOM 神经网络的易发性评价

同样对 7 个评价因子的信息量值进行归一化处理, 将处理好的样本数据作为 SOM 神经网络输入变量并得到滑坡易发性指数, 采用自然断点法将滑坡灾害易发程度划分为 4 个等级 (如图 11), 各等级所占比例分别为: 极高易发区占 13.5%、高易发区占 23.8%、中易发区占 26.3%、低易发区占 36.4%, 其中高和极高易发区共有 77 个滑坡样本点, 占总数的 70%, 说明 SOM 神经网络的初始易发性评价同样具有一定准确性。

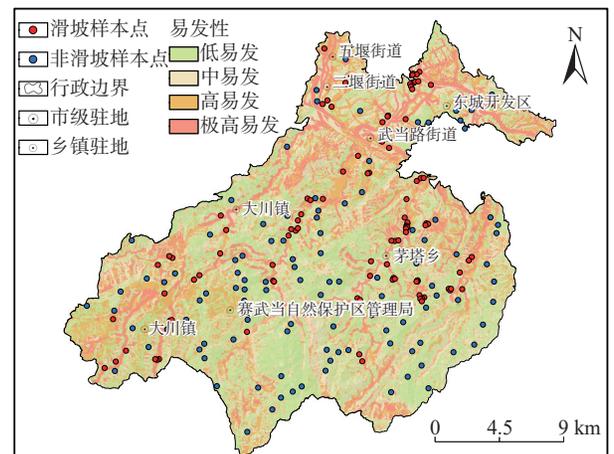


图 11 SOM 神经网络易发性分区及样本选择

Fig. 11 Landslide susceptibility zoning and sample selection of the SOM neural network

4.2.3 基于 SOM-I-SVM、KMeans-I-SVM 模型的易发性评价

基于 SOM 神经网络和 KMeans 算法的初始易发性评价,从两种方法的低易发区中随机生成 110 个非滑坡样本点,与 110 个滑坡正样本分别构成 SOM-I-SVM、KMeans-I-SVM 模型的样本空间,并以 7 : 3 的比例划分训练集和测试集。

将各因子的分级信息量值作为输入变量构建模型,同样使用 GridSearchCV 模块进行模型参数选择,两种模型的参数如表 4。

表 4 SOM-I-SVM、KMeans-I-SVM 模型参数表
Table 4 Parameters of the SOM-I-SVM and KMeans-I-SVM models

模型类型	超参数	调试范围	步长	最佳参数	模型评分
SOM-I-SVM	C	0.1 ~ 10	0.1	0.7	0.879
	γ	0.1 ~ 10	0.1	0.394 42	
KMeans-I-SVM	C	0.1 ~ 10	0.1	6.0	0.856
	γ	0.1 ~ 10	0.1	0.763 96	

将全区 1 219 394 个数据点带入模型计算出滑坡易发性指数,并采用自然断点法将其划分为 4 个等级,分别得到基于 KMeans-I-SVM 和 SOM-I-SVM 模型的滑坡灾害易发性分区图(图 12、图 13)。对易发性分区结果及各等级滑坡点的分布数量进行统计(表 5),其中 KMeans-I-SVM 模型极高易发区占 13.2%、高易发区占 18.0%、中易发区占 20.8%、低易发区占 48.0%;SOM-I-SVM 模型极高易发区占 12.7%、高易发区占 17.0%、中易发区占 24.7%、低易发区占 45.6%。从高和极高易发区中预测出的滑坡点数目来看,KMeans-I-SVM 模型预测出了 54 个滑坡点,占总数的 69.23%;

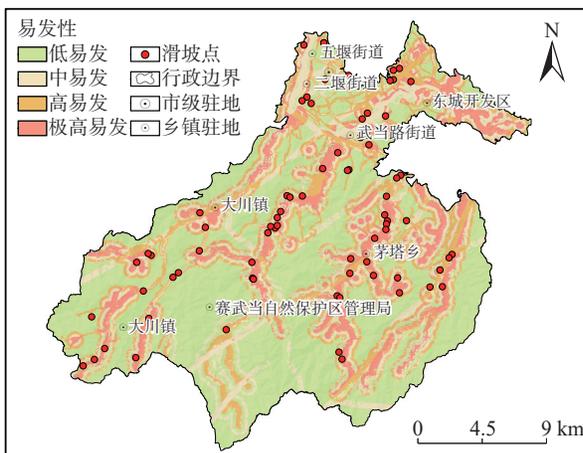


图 12 KMeans-I-SVM 模型滑坡易发性分区

Fig. 12 Zoning map of landslide susceptibility based on the KMeans-I-SVM model

SOM-I-SVM 模型预测出了 59 个滑坡点,占总数的 75.64%。

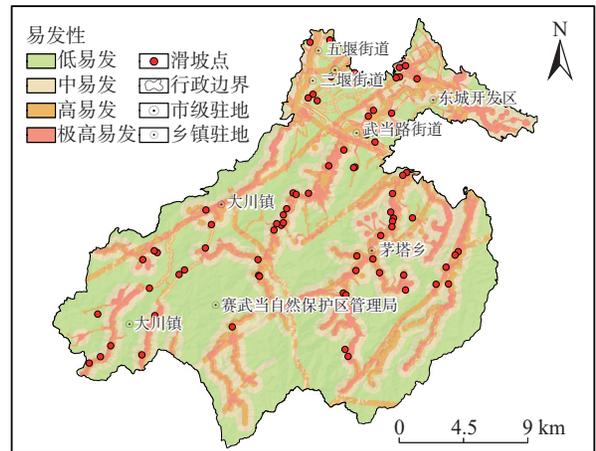


图 13 SOM-I-SVM 模型滑坡易发性分区

Fig. 13 Zoning map of landslide susceptibility based on the SOM-I-SVM model

表 5 KMeans-I-SVM、SOM-I-SVM 模型分区结果

Table 5 Results of the KMeans-I-SVM and SOM-I-SVM models

模型类型	易发性等级	分区面积 /km ²	所占比例 /%	滑坡数 /个	占滑坡总数 比例/%
KMeans-I-SVM	低易发区	235.54	48.0	14	18.0
	中易发区	101.93	20.8	10	12.8
	高易发区	88.31	18.0	17	21.8
	极高易发区	64.93	13.2	37	47.4
SOM-I-SVM	低易发区	223.59	45.6	9	11.5
	中易发区	121.10	24.7	10	12.8
	高易发区	83.54	17.0	14	18.0
	极高易发区	62.48	12.7	45	57.7

4.3 模型精度分析

根据滑坡易发性评价结果,将 4 种模型各易发等级的滑坡点分布数量进行对比分析,如图 14。其中 SVM、I-SVM、KMeans-I-SVM、SOM-I-SVM 模型高和极高易发区历史灾害点所占的比例分别为: 66.67%、69.23%、69.23% 和 75.64%。可以看出,当采用多种方法耦合时,模型的预测成功率有所提高,且滑坡个数由低易发区到极高易发区逐渐增多,说明模型具有一定的合理性。

在对评价结果进行分析的基础上,进一步采用受试工作者特征曲线(ROC)分析 4 种模型的预测精度,得到每种模型的 ROC 曲线并计算出曲线下面积(AUC)。当 AUC 值大于 0.7 时,说明模型预测效果较好。如图 15,SVM、I-SVM、KMeans-I-SVM、SOM-I-SVM 模型的 AUC 值分别为 0.82、0.88、0.90、0.91,说明四种模型都取得了较好的预测效果,并且 SOM-I-SVM

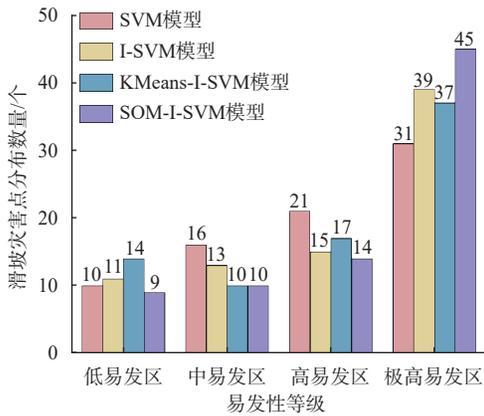


图 14 各易发等级历史滑坡点个数

Fig. 14 Number of historical landslide sites of each vulnerability level

耦合模型一定程度上提高了预测精度, 优于 SVM、I-SVM、KMeans-I-SVM 模型。

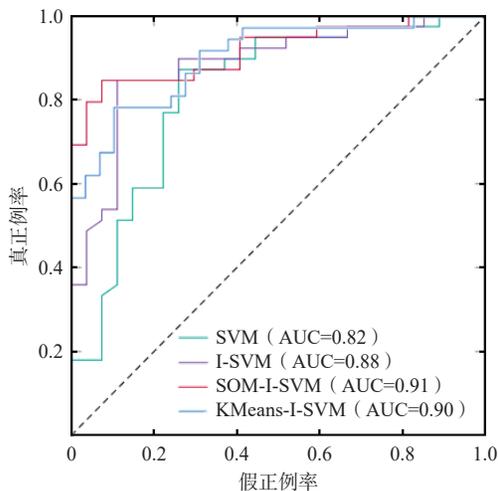


图 15 ROC 曲线对比图

Fig. 15 ROC curves of the four used models

4.4 易发性分区结果分析

综合上述 4 种模型的易发性评价结果可以分析出:

(1) 茅箭区滑坡灾害高和极高易发区受地表水系影响较大, 主要分布在马家河水库、茅塔河水库一带。由于地表水侵蚀河流两岸的边坡, 长时间作用会使边坡变陡形成临空面, 从而使其失去支撑。其次, 受区内茅塔河断裂和大坪断裂的影响, 岩体完整性较差, 次级小型断裂构造发育, 地层主要为较坚硬中-厚层状变粒岩类和较坚硬-较软弱薄-厚层状变粒岩类, 在风化、卸荷等内外营力长期作用下易形成滑坡灾害。在城镇区域, 由于人类工程活动较强烈, 高和极高易发区沿道路呈带状分布, 切坡修路等对地质环境破坏严重, 因此需要加强道路切坡的监测和预警。

(2) 滑坡灾害中和低易发区主要分布在相对高差小于 35m, 斜坡坡度小于 20°的区域内。该区域地层岩性主要为较坚硬中-厚层状变粒岩类, 具有较高的岩土体强度, 且远离地表水系, 人类工程活动强度较低, 这些都利于边坡的稳定, 不易发生滑坡灾害。

5 结论

(1) 以十堰市茅箭区为研究区域, 通过对滑坡灾害影响因子的相关性及其重要性分析, 选取距水系距离、坡度、降雨量、距构造距离、相对高差、距道路距离、地层岩性 7 个因子建立滑坡易发性评价指标体系, 采用 SVM、I-SVM、KMeans-I-SVM、SOM-I-SVM 等 4 种模型进行滑坡易发性评价, 其 AUC 值分别为 0.82, 0.88, 0.90, 0.91, 说明 SOM-I-SVM 模型能有效地提高滑坡易发性预测准确率。

(2) 通过 SOM 神经网络和 K 均值聚类对滑坡易发性进行预测, 两种模型的评价结果中, 历史滑坡点在高和极高易发区均达到 70%, 具有较高的准确率。且两种模型相较于 I-SVM 模型, 精度分别提高了 1.9% 和 3.5%。表明利用 SOM 模型和 KMeans 模型对区域进行初步易发性评价, 从低易发区中选取非滑坡样本点, 能够有效地提高模型预测精度。从 KMeans-I-SVM、SOM-I-SVM 两种模型的评价结果来看, SOM-I-SVM 模型取得了更好的效果, 相较于 KMeans-I-SVM 模型, 整体精度提高了 1.6%。

(3) 茅箭区滑坡灾害高和极高易发区多沿道路和地表水系呈带状分布, 其地层岩性主要为较坚硬中-厚层状变粒岩类和较坚硬-较软弱薄-厚层状变粒岩类, 受茅塔河断裂和大坪断裂等地质构造的影响易形成滑坡灾害。滑坡灾害低易发区主要分布在相对高差小于 35 m, 斜坡坡度小于 20°的区域内。该区域地层岩性主要为较坚硬中-厚层状变粒岩类, 具有较高的岩土体强度, 不易发生滑坡灾害。

参考文献 (References):

- [1] GUZZETTI F, REICHENBACH P, CARDINALI M, et al. Probabilistic landslide hazard assessment at the basin scale[J]. *Geomorphology*, 2005, 72(1/2/3/4): 272 - 299.
- [2] CUI Kai, LU Dong, LI Wei. Comparison of landslide susceptibility mapping based on statistical index, certainty factors, weights of evidence and evidential belief function models[J]. *Geocarto International*, 2017, 32(9): 935 - 955.

- [3] 盛逸凡, 李远耀, 徐勇, 等. 基于有效降雨强度和逻辑回归的降雨型滑坡预测模型[J]. 水文地质工程地质, 2019, 46(1): 156 - 162. [SHENG Yifan, LI Yuanyao, XU Yong, et al. Prediction of rainfall-type landslides based on effective rainfall intensity and logistic regression[J]. Hydrogeology & Engineering Geology, 2019, 46(1): 156 - 162. (in Chinese with English abstract)]
- [4] SCHICKER R, MOON V. Comparison of bivariate and multivariate statistical approaches in landslide susceptibility mapping at a regional scale[J]. *Geomorphology*, 2012, 161/162: 40 - 57.
- [5] 金朝, 费雯丽, 丁卫, 等. 基于信息量模型和Logistic回归模型的地质灾害易发性评价——以十堰市郧阳区为例[J]. 资源环境与工程, 2021, 35(6): 845 - 850. [JIN Zhao, FEI Wenli, DING Wei, et al. Evaluation of geological disaster susceptibility based on information model and logistic regression model[J]. Resources Environment & Engineering, 2021, 35(6): 845 - 850. (in Chinese with English abstract)]
- [6] 李利峰, 张晓虎, 邓慧琳, 等. 基于熵指数与逻辑回归耦合模型的滑坡灾害易发性评价——以蓝田县为例[J]. 科学技术与工程, 2020, 20(14): 5536 - 5543. [LI Lifeng, ZHANG Xiaohu, DENG Huilin, et al. Assessment of landslide susceptibility based on coupling model of index of entropy and logistic regression: A case study of Lantian County[J]. Science Technology and Engineering, 2020, 20(14): 5536 - 5543. (in Chinese with English abstract)]
- [7] 覃乙根, 杨根兰, 江兴元, 等. 基于确定性系数模型与逻辑回归模型耦合的地质灾害易发性评价——以贵州省开阳县为例[J]. 科学技术与工程, 2020, 20(1): 96 - 103. [QIN Yigen, YANG Genlan, JIANG Xingyuan, et al. Geohazard susceptibility assessment based on integrated certainty factor model and logistic regression model for Kaiyang, China[J]. Science Technology and Engineering, 2020, 20(1): 96 - 103. (in Chinese with English abstract)]
- [8] 吴雨辰, 周晗旭, 车爱兰. 基于粗糙集-神经网络的IBURI地震滑坡易发性研究[J]. 岩石力学与工程学报, 2021, 40(6): 1226 - 1235. [WU Yuchen, ZHOU Hanxu, CHE Ailan. Susceptibility of landslides caused by IBURI earthquake based on rough set-neural network[J]. Chinese Journal of Rock Mechanics and Engineering, 2021, 40(6): 1226 - 1235. (in Chinese with English abstract)]
- [9] 丁茜, 赵晓东, 吴鑫俊, 等. 基于RBF核的多分类SVM滑坡易发性评价模型[J]. 中国安全科学学报, 2022, 32(3): 194 - 200. [DING Xi, ZHAO Xiaodong, WU Xinjun, et al. Landslide susceptibility assessment model based on multi-class SVM with RBF kernel[J]. China Safety Science Journal, 2022, 32(3): 194 - 200. (in Chinese with English abstract)]
- [10] 刘坚, 李树林, 陈涛. 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. Landslide susceptibility assesment based on optimized random forest model[J]. Geomatics and Information Science of Wuhan University, 2018, 43(7): 1085 - 1091. (in Chinese with English abstract)]
- [11] YESILNACAR E, TOPAL T. Landslide susceptibility mapping: A comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey)[J]. *Engineering Geology*, 2005, 79(3/4): 251 - 266.
- [12] 万洋, 郭捷, 马凤山, 等. 基于最大熵模型的中尼交通廊道滑坡易发性分析[J]. 中国地质灾害与防治学报, 2022, 33(2): 88 - 95. [WAN Yang, GUO Jie, MA Fengshan, et al. Landslide susceptibility assessment based on MaxEnt model of along Sino-Nepal traffic corridor[J]. The Chinese Journal of Geological Hazard and Control, 2022, 33(2): 88 - 95. (in Chinese with English abstract)]
- [13] 黄武彪, 丁明涛, 王栋, 等. 基于层数自适应加权卷积神经网络的川藏铁路沿线滑坡易发性评价[J]. 地球科学, 2022, 47(6): 2015-2030. [HUANG Wubiao, DING Mingtao, WANG Dong, et al. Evaluation of landslide susceptibility based on layer adaptive weighted convolutional neural network model along Sichuan-Tibet railway[J]. Earth Science, 2022, 47(6): 2015-2030. (in Chinese with English abstract)]
- [14] 安凯强, 牛瑞卿. 信息量支持下SVM模型滑坡灾害易发性评价[J]. 长江科学院院报, 2016, 33(8): 47 - 51. [AN Kaiqiang, NIU Ruiqing. Landslide susceptibility assessment using support vector machine based on weighted-information model[J]. Journal of Yangtze River Scientific Research Institute, 2016, 33(8): 47 - 51. (in Chinese with English abstract)]
- [15] 杜国梁, 张永双, 吕文明, 等. 基于加权信息量模型的藏东南地区滑坡易发性评价[J]. 灾害学, 2016, 31(2): 226 - 234. [DU Guoliang, ZHANG Yongshuang, LV Wenming, et al. Landslide susceptibility assessment

- based on weighted information value model in southeast Tibet[J]. *Journal of Catastrophology*, 2016, 31(2): 226 – 234. (in Chinese with English abstract)]
- [16] 陈飞, 蔡超, 李小双, 等. 基于信息量与神经网络模型的滑坡易发性评价[J]. *岩石力学与工程学报*, 2020, 39(增刊 1): 2859 – 2870. [CHEN Fei, CAI Chao, LI Xiaoshuang, et al. Evaluation of landslide susceptibility based on information volume and neural network model[J]. *Chinese Journal of Rock Mechanics and Engineering*, 2020, 39(Sup 1): 2859 – 2870. (in Chinese with English abstract)]
- [17] 武雪玲, 任福, 牛瑞卿, 等. 斜坡单元支持下的滑坡易发性评价支持向量机模型[J]. *武汉大学学报·信息科学版*, 2013, 38(12): 1499 – 1503. [WU Xueling, REN Fu, NIU Ruiqing, et al. Landslide spatial prediction based on slope units and support vector machines[J]. *Geomatics and Information Science of Wuhan University*, 2013, 38(12): 1499 – 1503. (in Chinese with English abstract)]
- [18] BALLABIO C, STERLACCHINI S. Support vector machines for landslide susceptibility mapping: The staffora river basin case study, Italy[J]. *Mathematical Geosciences*, 2012, 44(1): 47 – 70.
- [19] 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 Shenghua, LIU Jiping, WANG Xianghong, 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. (in Chinese with English abstract)]
- [20] 刘福臻, 王灵, 肖东升. 机器学习模型在滑坡易发性评价中的应用[J]. *中国地质灾害与防治学报*, 2021, 32(6): 98 – 106. [LIU Fuzhen, WANG Ling, XIAO Dongsheng. Application of machine learning model in landslide susceptibility evaluation[J]. *The Chinese Journal of Geological Hazard and Control*, 2021, 32(6): 98 – 106. (in Chinese with English abstract)]
- [21] 鲍帅, 刘纪平, 王亮. 联合DBSCAN聚类采样和SVM分类的滑坡易发性评价[J]. *震灾防御技术*, 2021, 16(4): 625 – 636. [BAO Shuai, LIU Jiping, WANG Liang. Landslide susceptibility evaluation based on combined DBSCAN cluster sampling and SVM classification[J]. *Technology for Earthquake Disaster Prevention*, 2021, 16(4): 625 – 636. (in Chinese with English abstract)]
- [22] 杜国梁, 杨志华, 袁颖, 等. 基于逻辑回归-信息量的川藏交通廊道滑坡易发性评价[J]. *水文地质工程地质*, 2021, 48(5): 102 – 111. [DU Guoliang, YANG Zhihua, YUAN Ying, et al. Landslide susceptibility mapping in the Sichuan-Tibet traffic corridor using logistic regression-information value method[J]. *Hydrogeology & Engineering Geology*, 2021, 48(5): 102 – 111. (in Chinese with English abstract)]
- [23] LLOYD S. Least Squares quantization in PCM. *IEEE Transactions on Information Theory*[J]. *The Chinese Journal of Geological Hazard and Control*, 1982, 28(2): 129 – 137.
- [24] 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. (in Chinese with English abstract)]
- [25] 周超, 殷坤龙, 曹颖, 等. 基于集成学习与径向基神经网络耦合模型的三峡库区滑坡易发性评价[J]. *地球科学*, 2020, 45(6): 1865 – 1876. [ZHOU Chao, YIN Kunlong, CAO Ying, et al. Landslide susceptibility assessment by applying the coupling method of radial basis neural network and adaboost: A case study from the Three Gorges Reservoir area[J]. *Earth Science*, 2020, 45(6): 1865 – 1876. (in Chinese with English abstract)]
- [26] WUBALEM A, METEN M. Landslide susceptibility mapping using information value and logistic regression models in Goncha Siso Eneses area, northwestern Ethiopia[J]. *SN Applied Sciences*, 2020, 2(5): 1 – 19.

编辑: 王支农