

基于XGBoost和云模型的地质灾害易发性评价

张 威, 胡舫瑞, 蔡 巍, 彭 琳, 王咏林, 陈 枫

Susceptibility assessment of geological hazard based on XGBoost and cloud model

ZHANG Wei, HU Fangrui, QI Wei, PENG Lin, WANG Yonglin, and CHEN Feng

在线阅读 View online: <https://doi.org/10.16031/j.cnki.issn.1003-8035.202210041>

您可能感兴趣的其他文章

Articles you may be interested in

北京山区突发性地质灾害易发性评价

Assessment on the susceptibility of sudden geological hazards in mountainous areas of Beijing

罗守敬, 王珊珊, 付德荃 中国地质灾害与防治学报. 2021, 32(4): 126-133

基于RBF神经网络信息量耦合模型的滑坡易发性评价

Landslide susceptibility assessment by the coupling method of RBF neural network and information value: A case study in Min Xian, Gansu Province

黄立鑫, 郝君明, 李旺平, 周兆叶, 贾佩钱 中国地质灾害与防治学报. 2021, 32(6): 116-126

基于遥感影像多尺度分割与地质因子评价的滑坡易发性区划

Landslide susceptibility assessment based on multi-scale segmentation of remote sensing and geological factor evaluation

李文娟, 邵海 中国地质灾害与防治学报. 2021, 32(2): 94-99

基于GIS和信息量模型的安徽池州地质灾害易发性评价

王雷, 吴君平, 赵冰雪, 姚志强, 张乐勤 中国地质灾害与防治学报. 2020, 31(3): 96-102

香丽高速公路边坡地质灾害发育特征与易发性区划

Development characteristics and susceptibility zoning of slope geological hazards in Xiangli expressway

廖小平, 徐风光, 蔡旭东, 周文皎, 魏家旭 中国地质灾害与防治学报. 2021, 32(5): 121-129

基于自然灾害风险评价框架的省级地质灾害风险区划方法探讨——以吉林省为例

Provincial geological disaster risk zoning method based on natural disaster risk assessment framework: a case study in Jilin Province

张以晨, 郎秋玲, 陈亚南, 张继权, 田书文 中国地质灾害与防治学报. 2020, 31(6): 104-110



关注微信公众号, 获得更多资讯信息

DOI: 10.16031/j.cnki.issn.1003-8035.202210041

张威, 胡舫瑞, 蔡巍, 等. 基于 XGBoost 和云模型的地质灾害易发性评价[J]. 中国地质灾害与防治学报, 2023, 34(6): 136-145.

ZHANG Wei, HU Fangrui, QI Wei, et al. Susceptibility assessment of geological hazard based on XGBoost and cloud model[J]. The Chinese Journal of Geological Hazard and Control, 2023, 34(6): 136-145.

基于 XGBoost 和云模型的地质灾害易发性评价

张 威, 胡舫瑞, 蔡 巍, 彭 琳, 王咏林, 陈 枫
(辽宁省自然资源事务服务中心, 辽宁 沈阳 110044)

摘要:传统地质灾害易发性评价中,存在着易发性因子权重选取主观性强、因子分级具有随机性和模糊性等问题。采用单一评价模型只能对地质灾害的易发性进行定性评估,无法量化评价。针对这一问题,文章基于改进集成算法(XGBoost)和云模型,在辽宁省朝阳市 189 个灾害隐患点中选择坡度、多年平均降水量、归一化植被指数、高程等 12 个易发性因子,通过 XGBoost 分类算法确定了易发性因子权重,拟合准确率为 96.5%,达到了较高的精度。在此基础上利用云模型将因子分级的模糊性问题转化为定量问题,建立了朝阳市地质灾害易发性评价指标体系。以朝阳市大东山为评价单元对该评价体系进行验证。结果表明该评价单元的易发程度为高易发,与实际情况吻合,应用文章提出的方法进行地质灾害易发性评价的精度较高。

关键词: XGBoost; 易发性因子权重; 云模型; 易发性评价; 地质灾害

中图分类号: P642.22

文献标志码: A

文章编号: 1003-8035(2023)06-0136-10

Susceptibility assessment of geological hazard based on XGBoost and cloud model

ZHANG Wei, HU Fangrui, QI Wei, PENG Lin, WANG Yonglin, CHEN Feng
(Liaoning Province Natural Resources Affairs Service Center, Shenyang, Liaoning 110044, China)

Abstract: In the conventional process of geological hazard assessment, issues such as subjectivity in selecting susceptibility factor weights, randomness, and fuzziness in factor grading are prevalent. The application of a single assessment model can only provide qualitative evaluation of geological hazard susceptibility, lacking quantitative analysis. To overcome these challenges, this study employs an enhanced integrated algorithm (XGBoost) and cloud model. Among 189 disaster potential points in Chaoyang City, twelve susceptibility factors including slope, meteorological conditions, vegetation coverage and elevation were selected. The XGBoost classification algorithm was used to determine susceptibility factor weights. The results showed that the algorithm classification achieved high performance with fitting accuracy of 96.5%. On this basis, the cloud model was employed to transform the fuzzy factor grading into a quantitative problem, establishing a susceptibility evaluation index system for geological hazards in Chaoyang City, thereby assessing their susceptibility. To validate the evaluation index system, the Dadongshan landslide in Chaoyang City was selected as the assessment unit. Results indicate a high susceptibility level for this evaluation unit, consistent with actual conditions. The methodology proposed in this study is promising and can offers reference for evaluating geological hazard susceptibility.

Keywords: XGBoost; susceptibility factor weights; cloud model; susceptibility assessment; geological disaster

收稿日期: 2022-09-25; 修订日期: 2023-07-17

投稿网址: <https://www.zgdzhyfzxb.com/>

第一作者: 张 威(1982-), 男, 辽宁丹东人, 本科, 高级工程师, 主要从事自然灾害风险方面的研究。E-mail: 5869365@qq.com

通讯作者: 胡舫瑞(1988-), 男, 辽宁营口人, 硕士, 工程师, 主要从事地质灾害风险评价与管理方面的研究。E-mail: hufrcug@163.com

0 引言

我国是世界上地质灾害最严重、受威胁人口最多的国家之一。以 2021 年为例^[1],全国发生地质灾害 4 772 起,造成 91 人死亡失踪,成功预报地质灾害 905 起,涉及可能伤亡人员 2.5 万余人。地质灾害多发频发,已严重威胁人民群众生命财产安全。因此,地质灾害易发性评价工作尤为重要。

地质灾害易发性评价是对地质灾害活动程度和危害能力的综合判别,受多种因素影响。地质灾害易发性因素可分为直接因素(如地貌因素、地质因素等)和间接因素(如气象因素、人类活动因素等)。间接因素促进了地质灾害的发生,而直接因素是区域是否具备地质灾害发生的基本条件。

国内外学者对地质灾害易发性进行了大量的研究。MOHAMMADY 等^[2]基于信息量法采用坡度、坡向、高程等 13 个因素对伊朗 Golestan 地区地质灾害易发性进行评价。易靖松等^[4]基于信息量法和层次分析法,采用地形地貌、工程地质岩组、结构类型等 7 个因素分别对四川省阿坝县段的地质灾害易发性进行评价。黄艳婷等^[5]基于层次分析法采用公路距离、流域距离、断层距离等 11 个因素对藏东南地区泥石流地质灾害的易发性进行评价。许冲等^[6]基于逻辑回归模型采用坡度、坡向、高程、与水系距离等 8 个影响因子对汶川地震区地震滑坡易发性进行评价。陈水满等^[7]基于人工神经网络采用坡度、坡高、岩性、岸坡结构类型等 11 个影响因子构成基本评价体系对福建南平市滑坡易发性评价。HONG 等^[8]基于随机森林模型采用高程、坡向、地形湿度指数等 15 个因子对江西省广昌县滑坡易发性评价。

总体上看,基于信息量法和层次分析法^[3-5]进行定量与定性分析评价地质灾害易发性的方法虽然操作简便易行,但是主观性较强,结果也比较抽象,难以反映各因子之间的权重关系。与传统的研究方法相比,基于机器学习的评价模型^[6-8],能很好地解决传统模型中对非线性关系表达的问题,提高了地质灾害识别的精度,但是不能很好地反映各因子权重,可解释性不强。且地质灾害易发性判别有其自身的随机性和模糊性,以上单一评价模型无法获得很好的效果。针对这些问题,本文提出基于 XGBoost 和云模型进行地质灾害易发性评价。

XGBoost 是一种新兴的基于决策树的改进集成算法,被地质、气象、能源和医学等领域广泛使用^[9-14]。

该算法对数据质量要求不高,将多个学习器进行组

合,通过优势互补达到强学习器的效果,可直观输出特征因子对结果的贡献值分数,具有准确率高、解释性强、不易过拟合等优势,能很好的反映各因子权重,避免人为干扰。云模型是李德毅等^[15]首先提出的,可以将定性描述的随机性和模糊性问题转化为量化的评价模型,用以解决因子分级中带来的随机性和模糊性问题^[16-19]。本文根据辽宁朝阳市地质灾害资料,通过 XGBoost 算法确定地灾因子权重,在此基础上利用云模型将因子分级的模糊性问题转化为定量问题,建立了朝阳市地质灾害易发性评价指标体系,进而评价地质灾害的易发性。

1 模型方法

1.1 XGBoost 模型

XGBoost 模型是 CHEN 等^[20]在集成算法的基础上提出,解决了集成算法中各个学习器之间只能串行,时间开销大的问题。

XGBoost 是一个寻找损失函数最小化的过程,集成了多棵分类决策树,对于一棵有 t 个分枝决策树,可表示为:

$$\hat{y}_i^{(t)} = \sum_{k=1}^t f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i) \quad (1)$$

式中: $\hat{y}_i^{(t)}$ ——迭代 t 轮后预测值;

$\hat{y}_i^{(t-1)}$ ——迭代 $t-1$ 轮后预测值;

$f_t(x_i)$ ——决策树模型。

XGBoost 模型构建目标函数如下:

$$\begin{aligned} Obj^{(t)} &= \sum_{i=1}^n L(y_i, \hat{y}_i^{(t)}) + \sum_{j=1}^t \Omega(f_j) \\ &= \sum_{i=1}^n L(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \sum_{j=1}^t \Omega(f_j) \end{aligned} \quad (2)$$

式中: i ——数据集中样本号;

n ——为所建导入第 j 颗树的数据总量;

t ——建立全部树数量;

$\Omega(f)$ ——正则惩罚项,其目的是防止模型过拟合。

$\Omega(f)$ 可表示为:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|\omega\|^2 \quad (3)$$

式中: γ 、 λ ——正则系数;

T ——末端叶子节点的个数;

ω ——末端叶子节点的权重。

损失函数 L 作二阶泰勒展开可得:

$$f(x + \Delta x) \approx f(x) + f(x)' \Delta x + f(x)'' \Delta x^2 \quad (4)$$

令损失函数中 $\hat{y}_i^{(t-1)}$ 项为泰勒公式中的变量 x , 目标函数可简化为:

$$Obj^{(t)} \approx \sum_{i=1}^n \left[L(y_i, \hat{y}_i^{(t-1)}) + g_i f_i(x_i) + \frac{1}{2} h_i f_i(x_i)^2 \right] + \Omega(f_i) \quad (5)$$

式中: g_i 、 h_i ——损失函数 L 关于变量 x 的一阶与二阶导数。

$$\begin{cases} g_i = \frac{\partial L(y_i, \hat{y}_i^{(t-1)})}{\partial \hat{y}_i^{(t-1)}} \\ h_i = \frac{\partial^2 L(y_i, \hat{y}_i^{(t-1)})}{\partial \hat{y}_i^{(t-1)^2}} \end{cases} \quad (6)$$

通过迭代找到使目标函数 f_i 取值最小, 即完成模型的训练过程。

1.2 云模型

1.2.1 云模型及云发生器

云模型使用期望 (Ex)、熵 (En) 和超熵 (He) 3 个数字特征来定量描述一个不确定性概念。云发生器是云的具体实现方法。其中由云的数字特征 $C(Ex, En, He)$ 生成云滴算法称为正向发生器^[21], 用来实现定量数值和其定性语言值之间的不确定性转换。反之将一系列的精确数据有效转换为由数字特征表示的定性概念的算法称为逆向云发生器^[22]。本文通过 python 语言实现了正、逆向云发生器的编译。云模型 ($Ex=0, En=1, He=0.1$) 的示意图如图 1 所示。

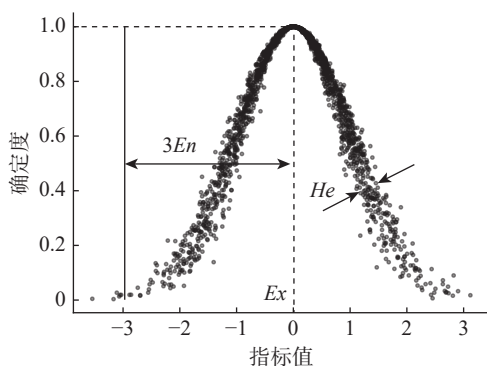


图 1 云模型示意图

Fig. 1 Schematic diagram of cloud model

1.2.2 综合云模型

假设集合中有 n 朵云 $\{C_1(Ex_1, En_1, He_1), C_2(Ex_2, En_2, He_2), \dots, C_n(Ex_n, En_n, He_n)\}$, 权重向量为 $\omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ 。若指标间相互独立, 相关性较小, 则运用虚拟云理论中综合云算法^[23], 综合云 $C_n(Ex_n, En_n, He_n)$ 计算公式如下:

$$\begin{cases} Ex = \frac{Ex_1 \omega_1 + Ex_2 \omega_2 + \dots + Ex_n \omega_n}{\omega_1 + \omega_2 + \dots + \omega_n} \\ En = \frac{\omega_1^2 En_1 + \omega_2^2 En_2 + \dots + \omega_n^2 En_n}{\omega_1^2 + \omega_2^2 + \dots + \omega_n^2} \\ He = \frac{\omega_1^2 He_1 + \omega_2^2 He_2 + \dots + \omega_n^2 He_n}{\omega_1^2 + \omega_2^2 + \dots + \omega_n^2} \end{cases} \quad (7)$$

式中: ω_i ——指标 i 的权重;

Ex_i 、 En_i 、 He_i ——相应指标云模型的数字特征参数。

1.2.3 云相似度

将评价对象云模型参数与评价指标云进行比较, 分别定义云向量 $\vec{V}_i = (Ex_i, En_i, He_i)$ 和 $\vec{V}_j = (Ex_j, En_j, He_j)$ 。它们之间的余弦夹角称为相似度 $\text{sim}(i, j)$ ^[24], 公式如下所示:

$$\text{sim}(i, j) = \cos(\vec{V}_i, \vec{V}_j) = \frac{\vec{V}_i \cdot \vec{V}_j}{\|\vec{V}_i\| \|\vec{V}_j\|} \quad (8)$$

利用式(8)来计算评价指标云与评价指标云的相似度, 并依照最大相似度原则, 相似度最高所对应的评价等级即为最终的评价等级。

1.3 易发性评价体系

基于 XGBoost 和云模型, 本文提出的地质灾害易发性评价体系的技术步骤如下:

(1)通过收集研究区的详细地质灾害调查资料, 选取易发性影响因素;

(2)利用 ArcGIS 平台提取各因子相关数据, 分析各因子之间的相关性和分布规律, 并对其进行归一化处理;

(3)使用网格搜索法, 选定 XGBoost 超参数取值, 根据归一化后的数据进行模型训练, 计算得到各因子的权重;

(4)根据规范、技术要求以及野外实践等分级标准构建评价指标体系, 合理划分指标体系结构;

(5)利用编译好的正向云发生器程序绘制评价指标云图;

(6)针对研究对象生成各因子的单朵云模型并确定每个因子指标权重 ω_{ij} , 结合指标权重进行综合云计算, 绘制评价对象综合云模型;

(7)通过评价对象综合云模型参数与评价指标云的相似度, 相似度最高所对应的评价等级即为最终的评价等级。

2 研究区概况及数据源

2.1 研究区概况

研究区朝阳市位于辽宁省西部。地处辽西山地丘陵地貌区, 北接内蒙古黄土高原, 东邻渤海沿岸和松辽

平原。

地势由北西向南东呈阶梯状降低,属于塔里木—华北板块。南北横跨 2 个二级构造单元,南部为华北陆块,北部为天山-赤峰造山系。区内 NE 向分布大量中生代盆地群,断裂均为倾向盆内的正断层(图 2)。其中侏罗纪盆地的盆缘断裂在盆地挤压褶皱时常反转为背向盆地的逆断层或逆冲断层。朝阳地区地层工程地质条件可以分为第四系松散土类、碳酸岩类、碎屑岩类、花岗岩类、片麻杂岩类、片岩、变粒岩、浅粒岩类等。

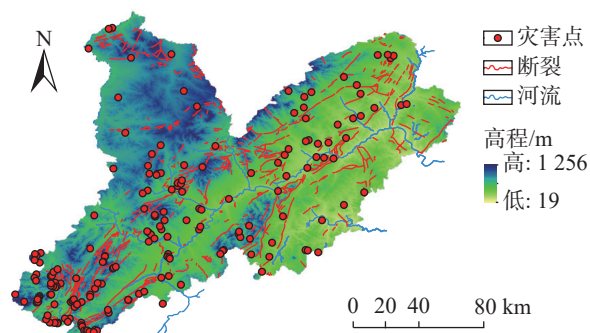


图 2 朝阳市地质情况示意图

Fig. 2 Location map of the study area

2.2 数据来源

研究区是辽宁省地质灾害较严重区域,地质灾害类型以崩塌、滑坡为主。依据《辽宁省朝阳市地质灾害风险调查评价成果报告》中最新的地质灾害风险普查资料,研究区内共有地质灾害隐患点 189 个。通过总结研究区地质灾害调查成果以及过往灾害发生历史情况,本文选取了 12 个易发性影响因素,分别为高程、坡度、坡向、归一化植被指数、多年平均降水量、工程地质岩

性、地质构造、人口密度、地下水涌水量、地质灾害点至公路、铁路和水系的距离。为便于统计与分析,根据研究区大小将整个研究区按照 500 m×500 m 的格网单元进行划分,共划分为 175 520 个栅格单元。各因子与研究区地质灾害点的空间位置如图 3 所示。

3 地质灾害易发性因子权重计算

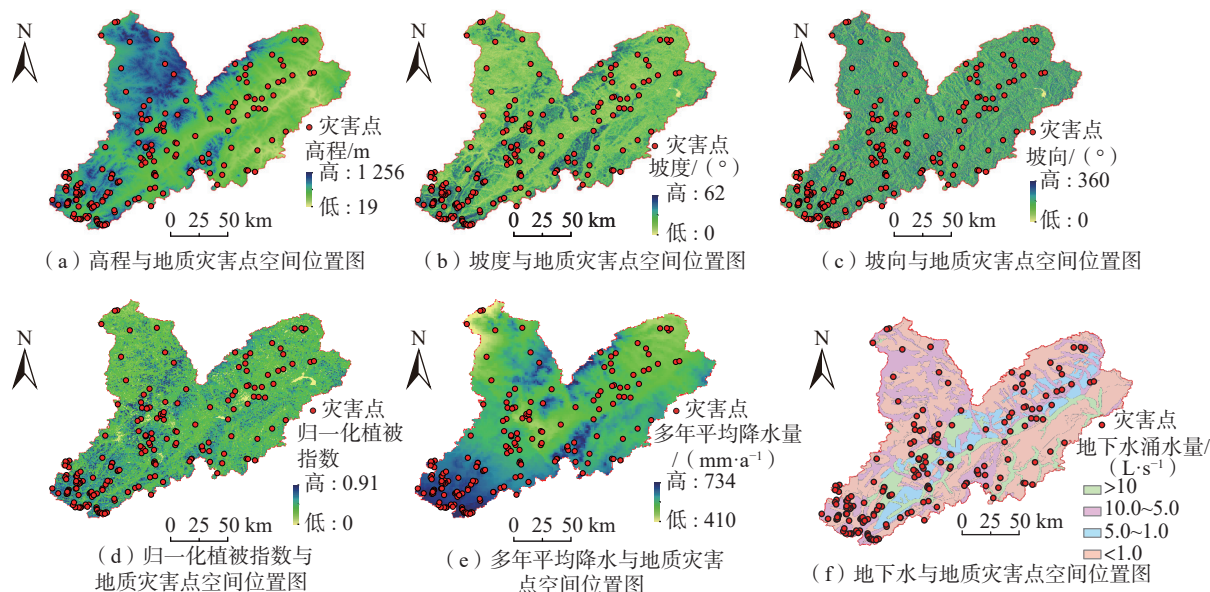
3.1 易发性因子选取与空间分析

对选取的 12 个因子进行相关性研究,结果如图 4 所示。12 个因子之间的相关性均小于 0.5,即各因子之间相对独立,无强关联性,满足 XGBoost 算法对因子独立性的要求。

3.2 数据处理及建模

研究表明在进行易发性区划中,样本灾害点与总样本点的比例在 1:5~1:10 时构建的模型效果最好^[25]。本文经过多次实验,最终选取样本灾害点的比例为 1:5,即选取 950 个非灾害点样本。对地质灾害点至地质构造、公路、铁路、水系等线性因子的距离依照 <500 m, 500~<1 000 m, 1 000~<1 500 m, ≥1 500 m 以上分为 4 个等级,距线性因子 500 m 内定义为第 4 等级,以此类推。为了消除坡度和人口密度的数据倾斜问题,对上述 2 个因子的数据进行了 Log 化处理,12 个易发性因子的数据分布如图 5 所示。

采用网格搜索法对 XGBoost 模型的超参数进行优化,并将数据按照 80% 和 20% 的比例将总样本划分为地质灾害训练集与测试集,并进行建模计算。样本与总体 ROC 曲线及 P-R 曲线如图 6 所示。结果表明得到的参数在地质灾害测试准确率 AUC=82.7%,地质灾害总



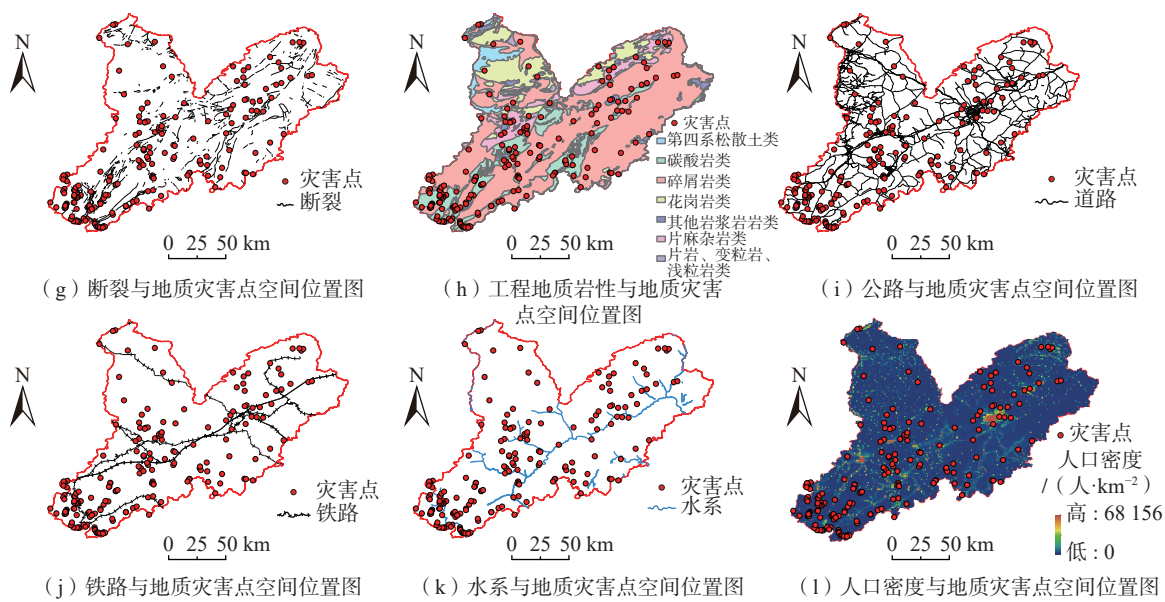


图 3 地质灾害易发性因子与地质灾害点关系图

Fig. 3 The relationship between geological hazard susceptibility factors and disaster points

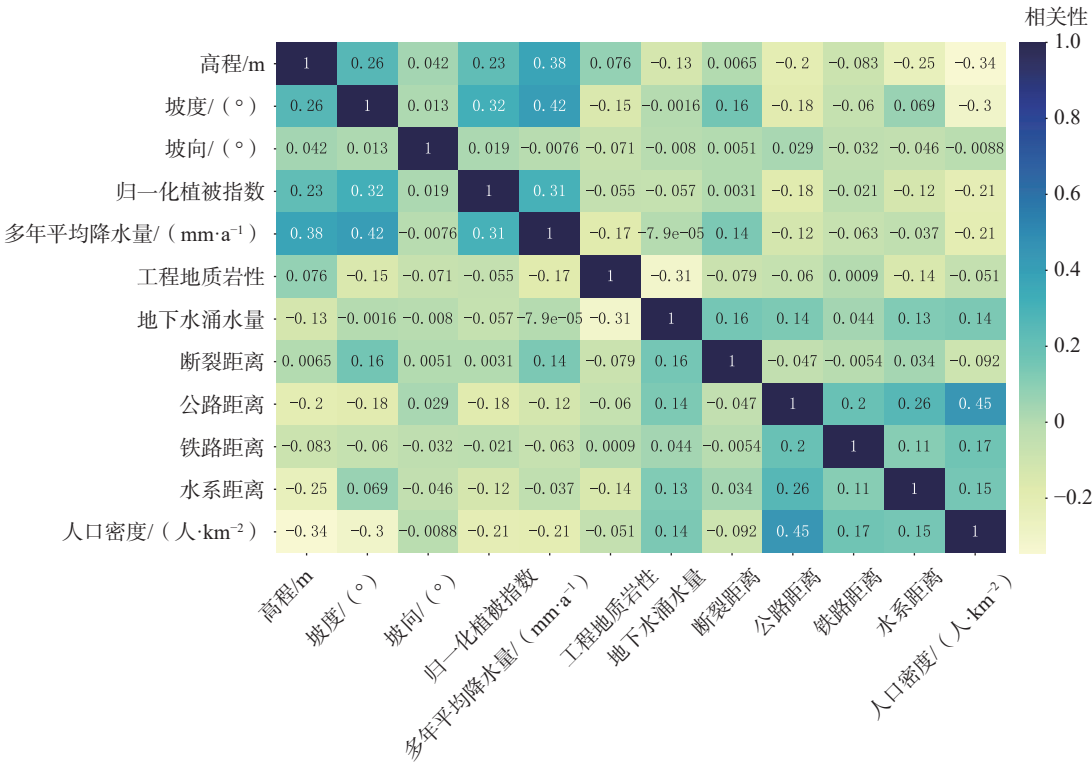


图 4 易发性因子相关性图

Fig. 4 Correlation diagram of susceptibility factors

体预测准确率 AUC 为 96.5%, 取得了良好的学习效果。

整理得到的易发性因子权重分数如图 7 所示。由图 7 可知影响研究区内地质灾害发生的易发性影响因子重要程度依次为: 坡度、多年平均降水量、归一化植被指数、高程、人口密度、坡向、地下水涌水量、公路距离、工程地质条件、断裂距离、水系距离、铁路距离。

因子权重分数归一化之后可得其因子权重如表 1 所示。

4 构建评价指标体系

依照《滑坡崩塌泥石流灾害调查规范(1:50 000)》(DZ/T 0261—2014)、《地质灾害风险调查评价技术要

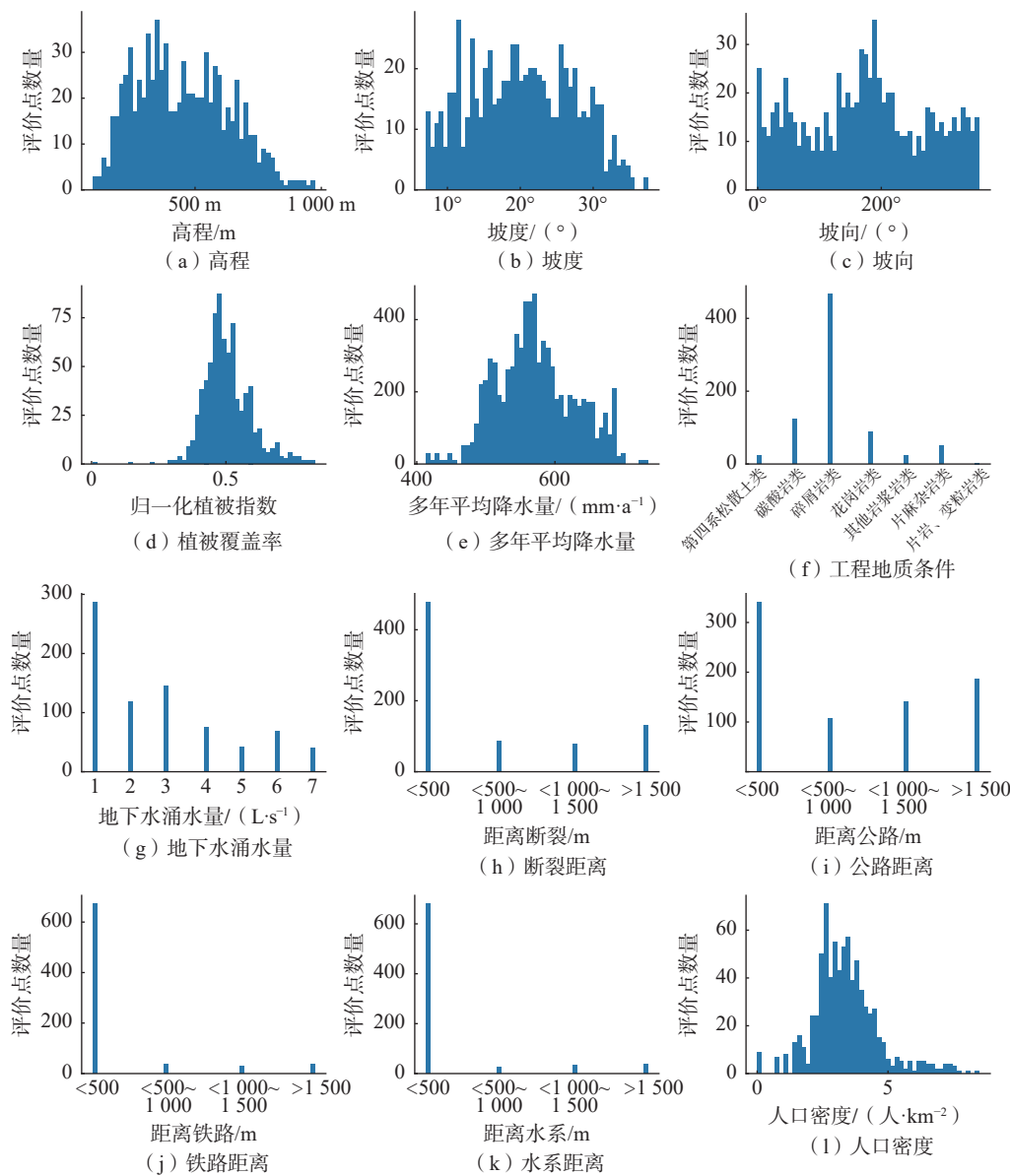


图 5 易发性因子数据分布图

Fig. 5 Data distribution of susceptibility factors

求(1 : 50 000)》以及地质灾害野外调查实践确定指标分级,形成评价指标体系。将指标分级划分为 4 级,如坡度分为平台、缓坡、陡坡、悬崖,多年平均降水量采取黄金分割法分为好、中等、较差、差。具体的因子分级见表 2。

对于“不易发”,“高易发”的判断较为精准,而对于“低易发”,“中易发”的概念认知较为模糊。将论域[0, 10]划分为 4 个评级区间,其中 0 和 10 分别为“不易发”与“高易发”的期望值。利用模糊逻辑概念结合黄金分割法^[26],可以计算出“中易发”和“低易发”的期望值为 6.91、3.09。本研究中取 $He=0.1$,利用正态云的性质可以得到各评价等级的 En 值,其计算公式为:

$$En_n=\frac{(Ex_{n+1}-Ex_n)}{3} \tag{9}$$

综上所述可计算出的指标分级模型表示为: 高易发 (10, 1.031, 0.1), 中易发 (6.91, 1.27, 0.1), 低易发 (3.09, 1.27, 0.1), 不易发 (0, 1.031, 0.1)。各评价等级对应的评价云图见图 8。

在地质灾害实地调查的基础上,每个评价单元的易发性因子进行评价,得到每个评价单元的云评价指标。根据表 1 中的易发性影响因子权重及式(7),计算出每个评价单元的综合云。在求得评价单元综合云的基础上,利用云相似度式(8)计算云模型隶属度关系,最终得到每个评价单元的易发性评价。

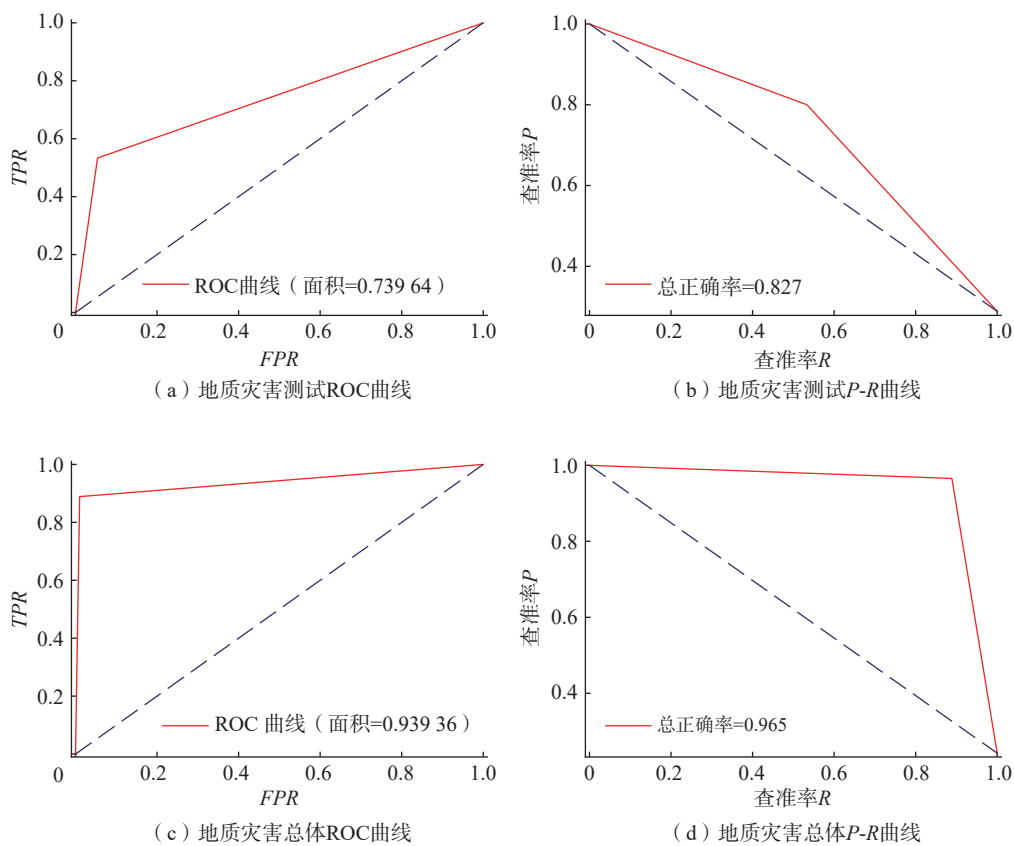


图 6 样本与总体 ROC 曲线及 P-R 曲线

Fig. 6 ROC curve and P-R curve for sample and overall population

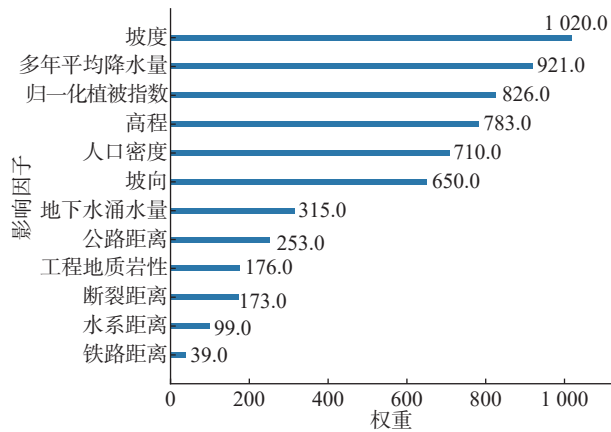


图 7 易发性因子权重分数图

Fig. 7 Weighted scores of susceptibility factors

5 工程应用

评价单元大东山位于朝阳市龙城区边杖子乡林杖子村东山组的斜坡地带(图 9),为龙城区西北城乡结合部。于 2017 年发生滑动,现场监测发现本点仍有变形发生,复滑可能性大。经过实地调查发现,滑坡南北长 175 m,东西宽 95 m。本点高程在 331 m,属于高丘陵

表 1 影响因子权重表		
Table 1 Weight table of impact factors		
序号	易发性因子	权重
1	坡度	0.169
2	多年平均降水量	0.151
3	归一化植被指数	0.136
4	高程	0.133
5	人口密度	0.122
6	坡向	0.115
7	地下水涌水量	0.053
8	公路距离	0.042
9	工程地质岩性	0.029
10	断裂距离	0.028
11	水系距离	0.017
12	铁路距离	0.006

区。多年平均降雨量为 549.3 mm,气象条件为较差。坡度为 26°,属于陡坡。威胁人口 20 人,滑坡主滑方向为 242°。归一化植被指数为 0.4,归一化植被指数差。地下水涌水量 10 L/s,富水性好。距公路距离为 1 000 m。工程岩性为泥质粉砂岩,灾害点距断裂较近,距水系、铁路距离较远。

表 2 影响因子分级表
Table 2 Grading table of impact factors

序号	易发性因子	分级
1	坡度	{平台,缓坡,陡坡,悬崖}
2	多年平均降水量	{好,较好,较差,差}
3	归一化植被指数	{好,中等,较差,差}
4	高程	{平原,低丘,高丘,低山}
5	人口密度	{好,较好,较差,差}
6	坡向/(°)	{45~135, 315~45, 135~225, 225~315}
7	地下水涌水量	{富水性差,富水性较差,富水性较好,富水性好}
8	公路距离	{远,较远,较近,近}
9	工程地质岩性	{碎屑岩类,花岗杂岩类、碳酸岩类,其他岩浆岩岩类,第四系松散土类、花岗岩类、片麻杂岩类}
10	断裂距离	{远,较远,较近,近}
11	水系距离	{远,较远,较近,近}
12	铁路距离	{远,较远,较近,近}

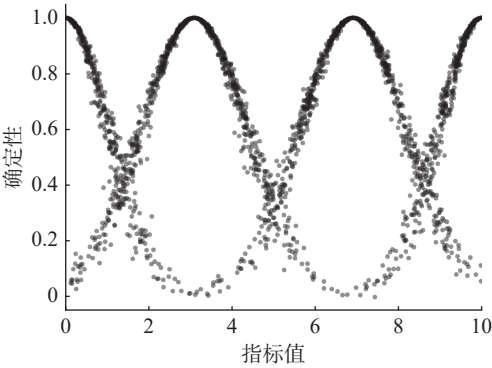


图 8 评价指标云图
Fig. 8 Cloud map of evaluation indicators

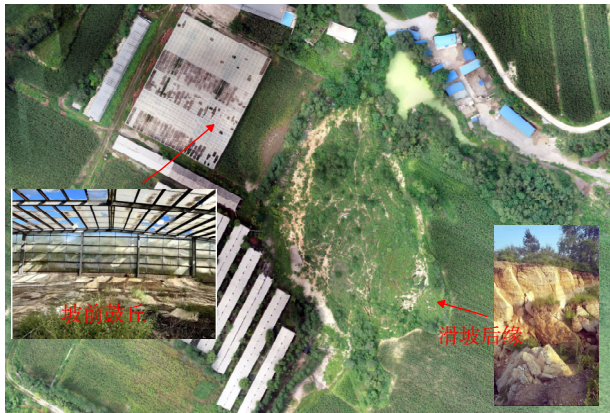


图 9 大东山滑坡正射影像
Fig. 9 Orthophoto image of Dadongshan landslide

根据调查,本评价单元的云模型评价价值如表 3 所示。可通过式(7)计算得到本评价单元综合评价的云模型数字特征分别为(8.57, 1.135, 0.1)。利用 Python 绘制综合评价云图,与评价指标云叠加,图 10 是将单元综合

评价云与评价指标云叠加后所绘得,其中评价指标云图为红色,综合评价云图为黑色。

表 3 大东山滑坡影响因子云模型评价价值
Table 3 Cloud model evaluation values of impact factors for Dadongshan landslide

序号	易发性因子	影响因子云模型评价价值
1	坡度	(6.91, 1.270, 0.1)
2	多年平均降水量	(6.91, 1.270, 0.1)
3	归一化植被指数	(10.00, 1.031, 0.1)
4	高程	(10.00, 1.031, 0.1)
5	人口密度	(10.00, 1.031, 0.1)
6	坡向	(10.00, 1.031, 0.1)
7	地下水涌水量	(10.00, 1.031, 0.1)
8	公路距离	(3.09, 1.270, 0.1)
9	工程地质岩性	(10.00, 1.031, 0.1)
10	断裂距离	(10.00, 1.031, 0.1)
11	水系距离	(3.09, 1.270, 0.1)
12	铁路距离	(3.09, 1.270, 0.1)

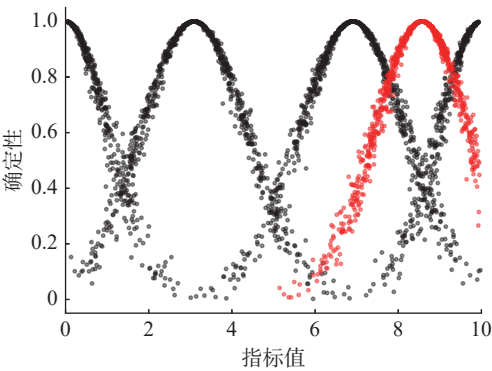


图 10 总体评估等级云相似度图
Fig. 10 Cloud similarity graph of overall evaluation grades

利用式(8)可以计算本评价单元综合云与评价指标云夹角值,结果如表 4 所示。

表 4 总体评估等级云相似度表
Table 4 Cloud similarity table of overall evaluation grades

云相似度	高易发	中易发	低易发	不易发
大东山滑坡	0.999 0	0.997 0	0.966 0	0.131 9

表 4 计算结果可以看出本评价单元易发性评价为高易发,与本灾害点的实际情况相符。

6 结论

(1)通过 XGBoost 算法获得了朝阳地区地质灾害易发性因子权重系数。结果表明地质灾害测试预测成功率 AUC 为 82.7%,总体地质灾害预测准确率 AUC 为 96.5%。XGBoost 模型取得了良好效果,得到的地质灾害易发性因子权重系数较为合理,重要程度依次为:坡

度、多年平均降水量、归一化植被指数、高程、人口密度、坡向、地下水涌水量、公路距离、工程地质岩性、断裂距离、水系距离、铁路距离。

(2)在获得地质灾害易发性因子权重系数的基础上,采用云模型创建了朝阳市地质灾害易发性评价指标体系。依据相关规范、调查技术要求以及野外调查实践将指标分级划分为 4 级,利用 Python 语言确定了评价指标云图,形成评价指标体系。

(3)选取朝阳市大东山滑坡作为评价单元对建立的朝阳市地质灾害易发性评价指标体系进行验证,发现该灾害评价单元的云模型数字特征为(8.57, 1.135, 0.1),与评价指标云图对比可判断该点的易发性为高易发,与实际情况吻合。验证结果表明本文提出评价指标体系可用于朝阳市的地灾易发性评价,该指标体系构建方法可对地质灾害易发性评价提供参考。

参考文献(References):

- [1] 中华人民共和国自然资源部.2021 年全国地质灾害灾情及 2022 年地质灾害趋势预测 [EB/OL]. (2022-01-13) [2022-09-13]. https://www.mnr.gov.cn/dt/ywbb/202201/t20220113_2717375.html. [Ministry of Natural Resources of the People's Republic of China. National geological disaster situation in 2021 and geological disaster trend forecast in 2022 [EB/OL]. (2022-01-13) [2022-09-13]. https://www.mnr.gov.cn/dt/ywbb/202201/t20220113_2717375.html. (in Chinese)]
- [2] 牛瑞卿,彭令,叶润青,等.基于粗糙集的支持向量机滑坡易发性评价[J].吉林大学学报(地球科学版),2012,42(2): 430 - 439. [NIU Ruiqing, PENG Ling, YE Runqing, et al. Landslide susceptibility assessment based on rough sets and support vector machine [J]. Journal of Jilin University (Earth Science Edition), 2012, 42(2): 430 - 439. (in Chinese with English abstract)]
- [3] MOHAMMADY M, POURGHASEMI H R, PRADHAN B. Landslide susceptibility mapping at Golestan Province, Iran: A comparison between frequency ratio, dempster-shafer, and weights-of-evidence models [J]. *Journal of Asian Earth Sciences*, 2012, 61: 221 - 236.
- [4] 易靖松,王峰,程英建,等.高山峡谷区地质灾害危险性评价——以四川省阿坝县为例[J].中国地质灾害与防治学报,2022,33(3): 134 - 142. [YI Jingsong, WANG Feng, CHENG Yingjian, et al. Study on the risk assessment of geological disasters in alpine valley area: A case study in Aba County, Sichuan Province [J]. The Chinese Journal of Geological Hazard and Control, 2022, 33(3): 134 - 142. (in Chinese with English abstract)]
- [5] 黄艳婷,郭永刚.考虑降雨敏感度的泥石流危险性评价——以藏东南地区为例[J].中国地质灾害与防治学报,2023,34(1): 129 - 138. [HUANG Yanting, GUO Yonggang. Debris flow risk assessment considering rainfall sensitivity: A case study of southeast Tibet [J]. The Chinese Journal of Geological Hazard and Control, 2023, 34(1): 129 - 138. (in Chinese with English abstract)]
- [6] 许冲,戴福初,徐素宁,等.基于逻辑回归模型的汶川地震滑坡危险性评价与检验[J].水文地质工程地质,2013,40(3): 98 - 104. [XU Chong, DAI Fuchu, XU Suning, et al. Application of logistic regression model on the Wenchuan earthquake triggered landslide hazard mapping and its validation [J]. Hydrogeology & Engineering Geology, 2013, 40(3): 98 - 104. (in Chinese with English abstract)]
- [7] 陈水满,赵辉龙,许震,等.基于人工神经网络模型的福建南平市滑坡危险性评价[J].中国地质灾害与防治学报,2022,33(2): 133 - 140. [CHEN Shuiman, ZHAO Huilong, XU Zhen, et al. Landslide risk assessment in Nanping City based on artificial neural networks model [J]. The Chinese Journal of Geological Hazard and Control, 2022, 33(2): 133 - 140. (in Chinese with English abstract)]
- [8] HONG Haoyuan, LIU Junzhi, BUI D T, et al. Landslide susceptibility mapping using J48 Decision Tree with AdaBoost, Bagging and Rotation Forest ensembles in the Guangchang area (China) [J]. *CATENA*, 2018, 163: 399 - 413.
- [9] 张福浩,朱月月,赵习枝,等.地理因子支持下的滑坡隐患点空间分布特征及识别研究[J].武汉大学学报(信息科学版),2020,45(8): 1233 - 1244. [ZHANG Fuhao, ZHU Yueyue, ZHAO Xizhi, et al. Spatial distribution and identification of hidden danger points of landslides based on geographical factors [J]. Geomatics and Information Science of Wuhan University, 2020, 45(8): 1233 - 1244. (in Chinese with English abstract)]
- [10] 赵楠,卢毅敏.基于XGBoost算法的近地面臭氧浓度遥感估算[J].环境科学学报,2022,42(5): 95 - 108. [ZHAO Nan, LU Yimin. Remote-sensing estimation of near-surface ozone concentration based on XGBoost [J]. Acta Scientiae Circumstantiae, 2022, 42(5): 95 - 108. (in Chinese with English abstract)]
- [11] 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 Deliang. Study on landslide susceptibility zoning and rainfall-induced landslide prediction and early warning based on machine learning [D]. Shanghai: East China Normal University, 2019. (in Chinese with English abstract)]
- [12] 赵晓东,徐振涛,刘福,等.基于极端梯度提升算法的滑坡易发性评价模型[J].科学技术与工程,2022,22(23): 10347 - 10354. [ZHAO Xiaodong, XU Zhentao, LIU Fu, et al. Evaluation model of landslide susceptibility based on extreme

- gradient lifting algorithm [J]. *Science Technology and Engineering*, 2022, 22(23): 10347 – 10354. (in Chinese with English abstract)]
- [13] 李丽秋, 许成燕, 王晓丽, 等. 基于社区大数据的骨关节炎患病风险 XGboost 预测模型研究 [J]. *中华全科医学*, 2022, 20(12): 2080 – 2083. [LI Lili, XU Chengyan, WANG Xiaoli, et al. Study on XGboost prediction model of osteoarthritis risk based on community big data [J]. *Chinese Journal of General Practice*, 2022, 20(12): 2080 – 2083. (in Chinese)]
- [14] 钱宇, 何益丰, 谢斌鑫, 等. 基于特征工程的 XGboost 风速短期预测 [J]. *武汉大学学报(工学版)*, 2022, 55(10): 1057 – 1064. [QIAN Yu, HE Yifeng, XIE Binxin, et al. Short-term forecast of XGboost wind speed based on feature engineering [J]. *Engineering Journal of Wuhan University*, 2022, 55(10): 1057 – 1064. (in Chinese with English abstract)]
- [15] 李德毅, 刘常昱. 论正态云模型的普适性 [J]. *中国工程科学*, 2004, 6(8): 28 – 34. [LI Deyi, LIU Changyu. Study on the universality of the normal cloud model [J]. *Engineering Science*, 2004, 6(8): 28 – 34. (in Chinese with English abstract)]
- [16] 朱文举, 平建华, 侯俊山, 等. 安阳市地下水源热泵系统建设水资源管理区划研究 [J]. *水文地质工程地质*, 2022, 49(1): 200 – 208. [ZHU Wenju, PING Jianhua, HOU Junshan, et al. A study of the water resources management and division of the groundwater heat pump system construction in Anyang [J]. *Hydrogeology & Engineering Geology*, 2022, 49(1): 200 – 208. (in Chinese with English abstract)]
- [17] 阮永芬, 张虔, 乔文件, 等. 基于 C-V-T 模型的盾构穿越既有桥梁施工风险评估 [J]. *岩土力学*, 2023, 44(2): 552 – 562. [RUAN Yongfen, ZHANG Qian, QIAO Wenjian, et al. Construction risk assessment of shield crossing existing bridge based on C-V-T model [J]. *Rock and Soil Mechanics*, 2023, 44(2): 552 – 562. (in Chinese with English abstract)]
- [18] 何乐平, 罗舒月, 胡启军, 等. 基于理想点-可拓云模型的隧道围岩稳定性评价 [J]. *中国地质灾害与防治学报*, 2021, 32(2): 126 – 134. [HE Leping, LUO Shuyue, HU Qijun, et al. Stability evaluation of tunnel surrounding rock based on ideal point-extension cloud model [J]. *The Chinese Journal of Geological Hazard and Control*, 2021, 32(2): 126 – 134. (in Chinese with English abstract)]
- [19] 陈忠源, 戴自航. 基于指标变权重复合云模型的岩质边坡稳定性评价初探 [J]. *中国地质灾害与防治学报*, 2021, 32(6): 9 – 17. [CHEN Zhongyuan, DAI Zihang. A preliminary study on evaluation of rock slope stability based on index variable weight compound cloud model [J]. *The Chinese Journal of Geological Hazard and Control*, 2021, 32(6): 9 – 17. (in Chinese with English abstract)]
- [20] CHEN Tianqi, GUESTRIN C. XGBoost: A scalable tree boosting system [C] // *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. August 13 – 17, 2016, San Francisco, California, USA. New York: ACM, 2016: 785 – 794.
- [21] 刘桂花, 宋承祥, 刘弘. 云发生器的软件实现 [J]. *计算机应用研究*, 2007, 24(1): 46 – 48. [LIU Guihua, SONG Chengxiang, LIU Hong. Software implementation of cloud generators [J]. *Application Research of Computers*, 2007, 24(1): 46 – 48. (in Chinese with English abstract)]
- [22] 刘常昱, 冯芒, 戴晓军, 等. 基于云 X 信息的逆向云新算法 [J]. *系统仿真学报*, 2004, 16(11): 2417 – 2420. [LIU Changyu, FENG Mang, DAI Xiaojun, et al. A new algorithm of backward cloud [J]. *Acta Simulata Systematica Sinica*, 2004, 16(11): 2417 – 2420. (in Chinese with English abstract)]
- [23] BANDROWSKI J, HEHLMANN J, MERTA H, et al. Studies of sedimentation in settlers with packing [J]. *Chemical Engineering and Processing: Process Intensification*, 1997, 36(3): 219 – 229.
- [24] 黄琼桃, 刘瑞敏. 云模型的相似性度量综述 [J]. *数据通信*, 2019(6): 43 – 49. [HUANG Qiongtao, LIU Ruimin. A survey of similarity measurement of cloud models [J]. *Data Communications*, 2019(6): 43 – 49. (in Chinese with English abstract)]
- [25] DAS I, STEIN A, KERLE N, et al. Landslide susceptibility mapping along road corridors in the Indian Himalayas using Bayesian logistic regression models [J]. *Geomorphology*, 2012, 179: 116 – 125.
- [26] 李长春, 朱亚强, 吴京, 等. 基于组合赋权法—云模型的甘肃省水安全时空变化分析 [J]. *水电能源科学*, 2023, 41(2): 23 – 26. [LI Changchun, ZHU Yaqiang, WU Jing, et al. Spatial and temporal change analysis of water security in Gansu Province based on combined weighting method-cloud model [J]. *Water Resources and Power*, 2023, 41(2): 23 – 26. (in Chinese with English abstract)]