

DOI: 10.16031/j.cnki.issn.1003-8035.2021.04-01

基于机器学习的区域滑坡危险性评价方法综述

方然可^{1,2}, 刘艳辉¹, 黄志全³

(1. 中国地质环境监测院(自然资源部地质灾害技术指导中心), 北京 100081; 2. 华北水利水电大学地球科学与工程学院, 河南 郑州 450046; 3. 洛阳理工学院, 河南 洛阳 471023)

摘要:我国滑坡灾害分布范围广, 危害严重。区域滑坡危险性评价一直都是滑坡灾害防灾减灾的重要内容之一。近年来, 随着大数据和人工智能技术的飞速发展, 机器学习技术逐渐在滑坡灾害危险性评价方面得到广泛应用, 并取得了较好效果。在大量研读文献的基础上, 系统阐述了基于机器学习技术的滑坡危险性评价方法研究现状。综述从评价因子选择与量化归一化、数据清洗与样本集构建、模型选取与训练评价等三个关键环节对现有研究成果进行分析评述, 最后对机器学习滑坡危险性评价方法的发展趋势提出讨论意见。

关键词:滑坡; 危险性评价; 机器学习; 方法综述; 模型研究

中图分类号: P642.22

文献标志码: A

文章编号: 1003-8035(2021)04-0001-08

A review of the methods of regional landslide hazard assessment based on machine learning

FANG Ranke^{1,2}, LIU Yanhui¹, HUANG Zhiquan³

(1. China Institute of Geo-Environment Monitoring (Guide Center of Prevention Technology for Geo-hazards, MNR), Beijing 100081, China; 2. College of Geosciences and Engineering, North China University of Water Resources and Electric Power, Zhengzhou, Henan 450046, China; 3. Luoyang Institute of Science and Technology, Luoyang, Henan 471023, China)

Abstract: The landslide disaster in China is widespread and serious. Regional landslide risk assessment has always been one of the most important contents of landslide disaster prevention and mitigation. In recent years, with the rapid development of big data and artificial intelligence technology, machine learning technology has gradually been widely used in landslide hazard assessment and achieved good results. Based on a large number of literatures, this paper systematically expounds the research status of landslide risk assessment methods based on machine learning technology. This paper reviews and analyzes the existing research results from three key links: evaluation factor selection and quantization normalization, data cleaning and sample set construction, model selection and training evaluation, and finally puts forward some suggestions on the development trend of machine learning landslide risk evaluation methods.

Keywords: landslide; risk assessment; machine learning; method survey; model research

0 引言

中国是世界上滑坡灾害影响范围最广、危害最严

重的国家之一。滑坡灾害遍及全国山地丘陵地区, 已知数量近百万处之多, 平均每年造成数百人死亡, 数十亿元直接经济损失。区域滑坡灾害危险性评价, 也称为敏

收稿日期: 2020-09-04; 修订日期: 2020-09-14

基金项目: 国家重点研发计划(2018YFC1505503); 国家科技支撑计划子课题(2015BAK10B021); 国家自然科学基金项目(41202217); 中原科技创新领军人才计划资助项目(214200510030)

第一作者: 方然可(1996-), 男, 河南郑州人, 硕士研究生, 主要从事滑坡灾害预警相关研究工作。E-mail: 1361853780@qq.com

通讯作者: 刘艳辉(1978-), 女, 博士, 教授级高级工程师, 主要从事滑坡灾害预警与防治、工程地质等方面的研究工作。

E-mail: 392990563@qq.com

感性评价、易发性区划,是在区域尺度上(可能是一个地区甚至一个国家),对影响滑坡灾害发生发展的地质环境因素开展系统分析,从区域上根据滑坡可能发生的危险性等级进行区划,进而划分轻重缓急区别对待,已成为主动有效预防和减轻滑坡灾害损失的重要手段之一。自 20 世纪 90 年代以来,国内外研究者针对滑坡危险性评价开展了卓有成效的工作。DIJKE 等^[1]基于 GIS 对山区地质灾害进行分析,建立了一套完整数据库,并开发了地质灾害分析评价模型;ALEOTTI 等^[2]总结回顾了滑坡地质灾害危险性评价研究进展,并对主要的危险性评价方法进行了分类等。同期国内学者在该领域也开展了深入全面的研究工作,比如运用多种数学模型开展的长江三峡库区库岸稳定性分区制图^[3];刘传正^[4-5]等提出基于发育度、潜势度、危险度、风险度和危害度的地质灾害空间评价体系,给出了系统评价模型,并在三峡库区、汶川地震区等进行了有效实践;同期很多学者也开展了相关工作^[6-8]。综上评价方法中,大多以数学统计分析方法为主,通过构建滑坡灾害影响因子评价体系,开展各影响因子与滑坡之间的统计规律,最后基于 GIS 开展滑坡危险性评价区划。

近年来,随着人工智能和大数据的蓬勃发展^[9],机器学习算法在区域滑坡危险性评价中逐步得到广泛应用^[10]。例如神经模糊^[11]和人工神经网络^[12]、随机森林^[13]、决策树^[14]、增强回归树^[15]、支持向量机^[16]、朴素贝叶斯^[17]、核逻辑回归^[18]等多种模型算法被应用到区域滑坡危险性评价中,并取得了较好效果。

本文通过研读文献,系统总结基于机器学习的区域滑坡危险性评价方法,从评价因子选取与量化归一化、数据清洗与样本集构建、模型选取与训练评价等三个关键环节对现有研究成果进行分析总结,最后对其发展趋势提出讨论意见。

1 基于机器学习的滑坡危险性评价研究现状

1.1 评价因子选取与量化

影响滑坡灾害的地质环境因素纷繁复杂,评价因子选取对危险性评价模型的构建至关重要,同时各影响因素普遍存在量纲不同和描述方法的差异,在分析之前必须根据研究区的具体情况,针对性选取评价因子,并对各影响因素进行有效的量化处理和归一化处理。

1.1.1 评价单元确定

评价单元是在区域上开展滑坡危险性评价与区划的最小制图单位。一般为规则单元或不规则的单元^[19]。和传统易发性评价指标相比,单元指标涵盖的范围更

广、结果更精确、数据量更大^[20]。常用评价单元一般归纳为 5 类,即网格单元^[21]、斜坡单元^[22]、地域单元^[23](地形地貌单元)、地形单元^[24]和均一条件单元^[25]。但由于地域、地形单元和均一条件单元的人为主观性较强,所以在确定评价单元时经常使用网格单元和斜坡单元^[26]。

张晓东^[26]考虑到研究区的评价因子多基于 DEM 数据基 Landsat 数据,所以使用网格单元对盐池县进行地质灾害危险性评价。网格大小与敏感性评价单元相同,都是规则正方形 30 m×30 m 网格,便于快速划分和后续计算。霍艾迪等^[27]基于 DEM 数据和 GIS 水文分析模型对黄土残塬区采用了斜坡单元划分法。先根据地形把整个研究区分为多个倾向和角度大致相同的适宜评价单元,提供最小单元计算参数,最后通过大量随机计算得到区域内易发程度分布,划分为不同的水文区域。

1.1.2 评价因子选取

评价因子选取是滑坡危险性评价的第一环节,对评价结果至关重要。评价因子选取需要在研究区范围内滑坡灾害发生发展规律系统掌握基础上进行。常用方法有:

(1) 相关性分析确定

朱吉龙^[28]以滑坡分布和成灾规律为基础,从发育机理的角度选取灾害评价因子后,再依据共线性检测和 Pearson 相关性分析进行筛选,剔除高共线性因子和低影响性因子。于宪煜等^[29]把 SVM 模型结合粗糙集理论来选择滑坡评价因子,简约模型得到核因子集,通过主成分变换将评价因子筛选且合并,其余分为控制与影响因素两类,避免了因子相关性影响。王念秦等^[30]以陕西省长安区为研究对象,以地质条件和发育特征为基础,根据 Spearman 对选取的 12 个因子评价指标进行相关性分析,计算因子间的相关度指标,即相关性系数 R,计算后剔除了相关性系数较高的高程因子,其余因子都用于 SVM 的模型训练。

(2) 随机森林特征重要性排序确定

宋庆武等^[31]通过每次随机打乱任意一个滑坡因子取值,分析前后随机森林模型的 ACC 变化,变化程度和特征重要性成正比,再通过专家赋值法赋值相应判别指标重要性,构造因子判断矩阵 U。吴孝情等^[32]使用随机森林模型来计算变量分割节点减少的最大不纯度均值,特征重要性与平均 Gini 指数成反比,量化、分析特征重要性可提供新方法计算指标权重。张福浩等^[33]以湘西自治州为研究区,使用了特征分裂数、特征平均覆盖率和特征平均增益值等方法来计算评价因子特征度量指

标,反应了各个因子对RF模型的解释程度,根据模型输出的混淆矩阵分析准确度和因子特征重要性,最后得出坡度的特征重要性最高,即特征贡献最大。

1.1.3 因子量化与归一化

在选择因子时难免会有一些高维特征因子,使得数据多余,增加了计算成本,从而影响机器学习的评价精度,而且各因子对滑坡灾害的影响程度也不相同^[34]。所以需要评估并筛选原始灾害因子,从而选择合适的灾害因子^[35]。

可以先通过离散化连续型的灾害因子并统一格式,再根据信息值量法把致灾因子和因子区间影响程度量化,从而达到筛选因子的目的^[36]。胡旭东^[37]以泸水县为研究区,同样使用信息价值量法,先离散化各个因子,将因子和因子区间发生灾害的影响程度量化,即算出每个区间灾害点的信息价值量,从而量化灾害评价因子的重要程度,该方法也被称为统计法量化重要程度,是一种非监督方法。结果表明初始的16个因子只有9个具有较高贡献度。

某些情况下,每个样本的具体特征值并不重要,但是每个样本特征值的占比很重要,这时就需要进行归一化处理,从而统一因子量纲,减小评价结果误差。梁万杰等^[38]为了解决评价因子的量纲差异问题,先对选取的7个危险性评价因子进行归一化操作,但由于使用的模型对连续变量处理不完善,同样也将归一化之后的数据采取了离散化处理,即归一化后结果乘10,将评价指标限定在0~10。结果显示空间分布更合理,适合大尺度上的灾害危险性评价。戴福初等^[39]在进行滑坡危险性评价时将空间数据进行了归一化处理,即把全部数据转化成可比的数值空间和坐标,重新分类各个因子值,使得空间数据都在同一个坐标空间内,从而消除量纲差异,实现归一化。

1.2 数据清洗与样本集构建

对于机器学习模型来说,原始数据集的好坏直接决定了评价精度的准确与否,所以在选定灾害因子后,还需要进行样本数据集的构造和清洗,才能有效地提升评价效果和预测精度。

1.2.1 原始样本选择

YAO等^[40]以香港天然地形为研究区,将1990年前发生的2031次滑坡作为模型训练集;之后发生的229次滑坡为验证集。此外在滑坡点40m的缓冲区之外随机产生负样本(未滑坡点),设置随机点之间的最小距离是20m。将可能影响滑坡的因素作为环境参数并

训练两类SVM模型。KUMAR等^[41]利用GIS收集曼达基尼河流域的滑坡数据,随机划分数据为两部分:第一部分包含50%的正样本(滑坡点),从无滑坡区域随机抽取等量的负样本(未滑坡点)作为训练集;第二部分包含用于验证SVM模型(测试集)剩余50%的数据,解决了像素点和滑坡数量不对应的问题。CHEN等^[42]以三峡秭归县为研究区,根据野外调查和航空影像及雷达地形影像的目视解译得到的滑坡库存量,将所有像元划分为PLS和PNLS,使用包含20%PLS和PNLS的训练集对树模型进行训练,并使用训练后的模型与测试集(即其余80%像素)。最后选择50%的PLS和PNLS作为训练集对研究地点进行灾害评价。刘坚等^[43]以三峡库区为研究区,转点前先把滑坡面降采样格网。为了减少把潜在滑坡视为负样本的错误,使用了ArcGIS生成随机点,再以正负样本1:2来构造随机森林原始数据集来进行危险性评价。CHEN等^[44]以陕西省太白县为研究区,在无滑坡区内随机生成和滑坡点相同的点,再按照7:3分类成两个子集,最后把正样本(滑坡点)赋值1,负样本(未滑坡)赋值0。

1.2.2 数据清洗与样本集构建

一般情况下,数据清洗是指对数据进行缺失值插补和异常值(噪音值)识别^[45]。人工错误、数据传输误差、设备故障、地质信息模糊等都会影响原始数据集,所以必须进行数据清洗。刘乾坤^[46]以四川省为研究区,由于滑坡数据存在结构不统一,人为误差等原因,通过数据筛选,字段提取,格式转换,误差修正等方法将原始数据进行清洗。朱力^[47]通过将原始数据进行标准化来处理数据属性不同,对于空值问题,如果空值较少可直接删除,若所占比例较大,则可用均值代替来补充。其中离群值并不一定是错值,而是由于这类数据很少出现,在一开始数据收集的时候比较困难,一般通过降低权重或是增加此类样本方式来进行清洗^[48-49]。结果表明使用清洗后的数据集模型评价精度更高。

1.3 模型选取与训练评价

常用机器学习模型主要解决回归、聚类和分类三大问题。对于滑坡危险性灾害评价问题,属于分类问题,机器学习的发展使得滑坡灾害评价模型识别精度逐步提高,单一模型逐渐升级为集成模型与耦合模型的混合模型^[50]。主要的滑坡危险性评价可选用的机器学习模型见表1。

1.3.1 模型选取与训练调参

李娟^[50]使用神经网络等模型对威州镇进行危险性

表 1 滑坡危险性评价可选用的机器学习模型
Table 1 An optional machine learning model for landslide hazard assessment

类型	常用模型	优缺点	相关数学公式
分类 (判断类别已知的 离散型数据)	KNN最近邻算法	适用多分类评价; 准确度高, 对异常点不敏感。但计算量大, 过于依赖均衡训练数据。	欧式距离: $d(x, y) = \sqrt{\sum_{k=1}^n (x_k - y_k)^2}$ 曼哈顿距离: $d(x, y) = \sqrt{\sum_{k=1}^n x_k - y_k }$
	SVM支持向量机	核函数可映射至高维空间, 解决非线性分类评价。但对大规模和多分类训练样本难以进行评价。	高斯核函数: $K(X, Y) = \exp\left\{-\frac{\ X - Y\ ^2}{2\sigma^2}\right\}$
	神经网络 (线性、BP、卷积)	可高速寻找优化解。但需要大量参数, 学习时间过长, 评价结果不确定。	损失函数 $L = \frac{1}{2} \sum_{i=1}^{mK} (Y_i^{(K)} - T_i)^2 = \frac{1}{2} \sum_{i=1}^{mK} (\delta_i)^2$
	Logistic回归 (Sigmoid函数、梯度上升)	评价效率高。但不能观察学习过程。	逻辑函数: $y = \frac{1}{1 + e^{-x}}$
	决策树	适合评价离散小规模样本。但评价大量连续变量和多类别样本效果欠佳。	信息熵 $H(X) = -\sum_{x \in X} P(x) \log_2 P(x)$
集成算法 (bagging、随机森林RF、 boosting、stacking)	避免了强势样本对评价结果的影响。但在某些噪音值较大的样本来进行危险性评价时可能会发生过拟合现象。	Bagging $f(x) = 1/M \sum_{m=1}^M f_m(x)$	

评价, 先确定隐含层的神经元数量, 太多会导致训练时间过长, 太少则会导致模型性能较差, 初始网络权重会表现出多次训练结果的不同。该模型的优化可以通过结合其他优化算法来代替可能引起局部极值徘徊而导致训练失败的传统局部优化。李远远等^[51]通过支持向量机单一模型训练结合其他方法形成耦合模型来评价滑坡灾害。根据灾害隐患点, 使用 CF 方法算出支持向量机模型分类数据, 即影响因子的敏感性值, 按照 7:3 的比例划分正负样本单元格的训练集和测试集, 使用径向基函数为模型核函数, 进行训练, 最后再通过交叉验证得出最优参数组合。

模型参数的选取对于模型精度影响巨大, 模型调参在模型构建中至关重要。目前最常用的模型调参方法包括传统方法和超参数优化算法。传统参数优化方法是网格搜索法, 该方法优化精度和速度成反比, 为了更高效地进行调参, 逐步出现了贝叶斯优化算法^[52]。此方法采用了高斯过程, 通过增加样本数量来拟合目标函数分布, 目标函数通过交叉验证精度来进行优化, 每次迭代都会输出一次超参数, 在这个寻找最优值的过程中优化超参数。孙德亮^[53]采用贝叶斯优化算法对研究区逻辑回归、随机森林、BP 神经网络模型进行超参数优化。结果显示贝叶斯法比网格搜索法更高效, 优化精度

更高, 鲁棒性也更强。

1.3.2 模型比选与效果评价

机器学习模型的好坏, 取决于其评价精度和模型泛化能力。目前, 最常用的评价方法是模型精确度 (Accuracy)、ROC 曲线和 AUC 值。其中精确度等于分类正确的样本数和总样本数的比; ROC 曲线是根据概率阈值得到二分类结果, 再与实际结果比较, 曲线面积越大, 模型评价精确度就越高; AUC 值即是曲线包含的面积^[54]。邱维蓉等^[55]构造了 4 种机器学习模型, 训练完毕后对其进行评价, 分别绘制 ROC 曲线, 并对比各模型评价结果, 显示效果较好。管新邦^[56]以云南为研究区, 通过 ROC 曲线、Kappa 系数和 Sridevi Jadi 精度评估方法评价对比了 5 种模型, 最终综合这三种评价结果选取了最恰当的评价模型。

另外, 输出模型训练后的分类报告也是重要的评价手段。分类报告中不仅可以得到混淆矩阵, 还可以得到交叉验证查准率、召回率、f1 得分和精确度的结果, 可以分析出哪些样本属于异常样本。学习曲线可以判断模型的拟合问题, 也是一个重要的评价标准。郝国栋^[57]以商南县为研究区, 构造随机森林和逻辑回归模型, 计算模型的混淆矩阵, 通过矩阵输出结果调整参数, 从而提升了模型的精度。

2 发展趋势与问题讨论

机器学习在滑坡灾害危险性评价中得到广泛应用,但仍存在一些不足:(1)评价因子选取分析不足。由于研究区地质结构复杂,滑坡灾害影响因子间可能存在明显的相关性,或是筛选因子的控制因素与影响因素不足,有可能导致模型准确度降低。(2)评价因子归一化不足。影响因子和控制因子之间往往存在着不同类别的属性差异与量纲差异,这种不均衡性会导致重要的灾害评价因子缺失。(3)样本集构建不足,特别是负样本选取方法至关重要。数据集清洗是否到位、负样本选取规则、正负样本比例失衡,都会直接影响模型的评价精度。(4)模型训练和调参的人为差异性较大。对模型进行训练时,传统方法对于模型的训练时间和精度相矛盾,无法高效调参。

对于上面的不足,笔者认为在模型构建评价中要着重关注以下几个方面:(1)强调有监督的机器学习。可以选择机器学习与统计评价结合的混合模型,对因子进行多元共线性诊断后再选择因子;然后通过主成分变换,筛选初始灾害因子后分为控制因素与影响因素;最后使用粗糙集理论把因素的属性简约化,得到核因子集。(2)强调数据清洗的重要性。量纲差别需要进行量化、归一化处理,可以保证每个特征对评价结果影响程度相同,从而保证结果准确。(3)强调特征属性的重要性排序。在进行因子特征重要性选取的时候可以使用sklearn库里的feature_importances_来存储每个特征的重要性,根据每个特征划分前后的信息熵减少量就表示了该特征的重要程度,即为该特征重要性指标。(4)强调训练样本集的构建,注重样本的均衡性。对于原始数据集进行完善的删除空值、替换噪音值、插补异常值等清洗工作,规范选取负样本方式,以正样本为中心选择合适的缓冲区距离,使用ArcGIS等辅助软件随机生成一定比例的负样本,确保数据集输入的准确。(5)强调模型参数优化。可以使用高斯过程的贝叶斯优化算法代替网格搜索调参。

3 结论

(1)随着机器学习与大数据的蓬勃发展,基于机器学习的滑坡危险性评价得到广泛应用,在评价因子选取与量化、数据清洗与样本集构建、模型训练评价等方面都有一套较为成熟的方法,取得了较好效果。本文对其中的几个关键点进行了系统梳理和总结。

(2)评价因子的选取,应使用多元共线性诊断、粗

糙集理论、量化、归一化处理等把致灾因子分为控制因素与影响因素,避免因子相关性与量纲的差别。获得因子特征重要性可以使用sklearn库中相应的计算包,直观且方便。

(3)机器学习模型评价精度很大程度上取决于输入的数据集,所以数据清洗,正负样本的选择至关重要,根据具体问题进行具体分析,高质量的样本数据才能保证模型评价的精度。

(4)模型调参对模型评价结果影响重大,贝叶斯优化算法精确且高效,模型评价可以通过ROC曲线、AUC值、分类报告来衡量评价精度,学习曲线判断模型拟合问题。

机器学习应用于区域滑坡危险性评价尚有一些不足,笔者尝试提出了一些问题和下一步发展方向。

参考文献 (References) :

- [1] VAN DIJKE J J, VAN WESTEN C J. Rockfall hazard: a geomorphologic application of neighbourhood analysis with ILWIS [J]. ITC Journal, 1990: 40-44.
- [2] ALEOTTI P, CHOWDHURY R. Landslide hazard assessment: summary review and new perspectives [J]. *Bulletin of Engineering Geology and the Environment*, 1999, 58(1): 21-44.
- [3] 黄润秋,李曰国.三峡工程水库岸坡稳定性预测的逻辑信息模型[J].水文地质工程地质,1992,19(1):15-20. [HUANG Runqiu, LI Yueguo. LOGICAL message modei of stability predication ot bank slopes in Three Gorges [J]. *Hydrogeology & Engineering Geology*, 1992, 19(1): 15-20. (in Chinese with English abstract)]
- [4] 刘传正,刘艳辉,温铭生.长江三峡库区地质灾害成因与评价研究[M].北京:地质出版社,2007. [LIU Chuanzheng, LIU Yanhui, WEN Mingsheng. Research on the geo-hazards genesis and assessment in the Three Gorges reservoir of Changjiang River in China[M]. Beijing: Geological Publishing House, 2007. (in Chinese)]
- [5] 刘传正,温铭生,刘艳辉,等.汶川地震区重大地质灾害成因规律研究[M].北京:地质出版社,2017. [LIU Chuanzheng, LIU Yanhui, WEN Mingsheng, et al. Study on the law of formation of major geological disasters in Wenchuan Earthquake area[M]. Beijing: Geological Publishing House, 2017. (in Chinese)]
- [6] 唐川,朱静,张翔瑞. GIS支持下的地震诱发滑坡危险区预测研究 [J]. *地震研究*, 2001, 24(1): 73-81. [TANG Chuan, ZHU Jing, ZHANG Xiangrui. GIS based earthquake triggered landslide hazard prediction [J]. *Journal of Seismological Research*, 2001, 24(1): 73-81. (in Chinese)]

- with English abstract)]
- [7] 张春山, 张业成, 马寅生. 黄河上游地区崩塌、滑坡、泥石流地质灾害区域危险性评价 [J] . *地质力学学报*, 2003, 9(2): 143 – 153. [ZHANG Chunshan, ZHANG Yecheng, MA Yinsheng. Regional dangerous on the geological hazards of collapse, landslide and debris flows in the upper reaches of the Yellow River [J] . *Journal of Geomechanics*, 2003, 9(2): 143 – 153. (in Chinese with English abstract)]
- [8] JIANG H, EASTMAN J R. Application of fuzzy measures in multi-criteria evaluation in GIS [J] . *International Journal of Geographical Information Science*, 2000, 14(2): 173 – 184.
- [9] 周超, 方秀琴, 吴小君, 等. 基于三种机器学习算法的山洪灾害风险评价 [J] . *地球信息科学学报*, 2019, 21(11): 1679 – 1688. [ZHOU Chao, FANG Xiuqin, WU Xiaojun, et al. Risk assessment of mountain torrents based on three machine learning algorithms [J] . *Journal of Geo-Information Science*, 2019, 21(11): 1679 – 1688. (in Chinese with English abstract)]
- [10] TIEN BUI D, PRADHAN B, LOFMAN O, et al. Landslide susceptibility mapping at Hoa Binh Province (Vietnam) using an adaptive neuro-fuzzy inference system and GIS [J] . *Computers & Geosciences*, 2012, 45: 199 – 211.
- [11] CHEN W, POURGHASEMI H R, ZHAO Z. A GIS-based comparative study of Dempster-Shafer, logistic regression and artificial neural network models for landslide susceptibility mapping [J] . *Geocarto International*, 2017, 32(4): 367 – 385.
- [12] TRIGILA A, IADANZA C, ESPOSITO C, et al. Comparison of Logistic Regression and Random Forests techniques for shallow landslide susceptibility assessment in Giampieri (NE Sicily, Italy) [J] . *Geomorphology*, 2015, 249: 119 – 136.
- [13] HONG H Y, PRADHAN B, XU C, et al. Spatial prediction of landslide hazard at the Yihuang area (China) using two-class kernel logistic regression, alternating decision tree and support vector machines [J] . *CATENA*, 2015, 133: 266 – 281.
- [14] YOUSSEF A M, POURGHASEMI H R, POURTAGHI Z S, et al. Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia [J] . *Landslides*, 2016, 13(5): 839 – 856.
- [15] CHEN W, POURGHASEMI H R, KORNEJADY A, et al. Landslide spatial modeling: Introducing new ensembles of ANN, MaxEnt, and SVM machine learning techniques [J] . *Geoderma*, 2017, 305: 314 – 327.
- [16] TSANGARATOS P, ILIA I. Comparison of a logistic regression and Naïve Bayes classifier in landslide susceptibility assessments: The influence of models complexity and training dataset size [J] . *CATENA*, 2016, 145: 164 – 179.
- [17] TIEN BUI D, TUAN T A, KLEMPE H, et al. Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree [J] . *Landslides*, 2016, 13(2): 361 – 378.
- [18] 朱清华. 基于RF和SVM模型的灞桥区地质灾害易发性评价 [D] . 西安: 西安科技大学, 2019. [ZHU Qinghua. Geological hazard susceptibility assessment in Baqiao district based on RF and SVM model [D] . Xi'an: Xi'an University of Science and Technology, 2019. (in Chinese with English abstract)]
- [19] 刘永垚, 第宝锋, 詹宇, 等. 基于随机森林模型的泥石流易发性评价: 以汶川地震重灾区为例 [J] . *山地学报*, 2018, 36(5): 765 – 773. [LIU Yongyao, DI Baofeng, ZHAN Yu, et al. Debris flows susceptibility assessment in Wenchuan earthquake areas based on random forest algorithm model [J] . *Mountain Research*, 2018, 36(5): 765 – 773. (in Chinese with English abstract)]
- [20] WESTEN C J V. Application of geographic information systems to landslide hazard zonation [EB] . 1993.
- [21] CARRARA A, CARDINALI M, DETTI R, et al. GIS techniques and statistical models in evaluating landslide hazard [J] . *Earth Surface Processes and Landforms*, 1991, 16(5): 427 – 445.
- [22] MEIJERINK A M J. Data acquisition and data capture through terrain mapping units [EB] . 1988.
- [23] ROSSI M, GUZZETTI F, REICHENBACH P, et al. Optimal landslide susceptibility zonation based on multiple forecasts [J] . *Geomorphology*, 2010, 114(3): 129 – 142.
- [24] 邱海军. 区域滑坡崩塌地质灾害特征分析及其易发性和危险性评价研究: 以宁强县为例 [D] . 西安: 西北大学, 2012. [QIU Haijun. Study on the regional landslide characteristic analysis and hazard assessment: A case study of Ningqiang County [D] . Xi'an: Northwest University, 2012. (in Chinese with English abstract)]
- [25] 刘林通. 基于TRIGRS模型的降雨型浅表层滑坡易发性评价: 以秦州区教场坝沟为例 [D] . 兰州: 兰州大学, 2018. [LIU Lintong. Rainfall-induced shallow landslides susceptibility assesment based on TRIGRS model in jiaochangba valley [D] . Lanzhou: Lanzhou University, 2018. (in Chinese with English abstract)]
- [26] 张晓东. 基于遥感和GIS的宁夏盐池县地质灾害风险评价研究 [D] . 北京: 中国地质大学(北京), 2018. [ZHANG Xiaodong. Research on risk assessment of geological disasters in Yanchi county, Ningxia [D] . Beijing: China university of geosciences (Beijing), 2018. (in Chinese with English abstract)]
- [27] 霍艾迪, 张骏, 卢玉东, 等. 地质灾害易发性评价单元划分方法: 以陕西省黄陵县为例 [J] . *吉林大学学报(地*

- 球科学版), 2011, 41(2): 523–528. [HUO Aidi, ZHANG Jun, LU Yudong, et al. Method of classification for susceptibility evaluation unit for geological hazards: a case study of Huangling County, Shaanxi, China [J]. Journal of Jilin University (Earth Science Edition), 2011, 41(2): 523–528. (in Chinese with English abstract)]
- [28] 朱吉龙. 溪洛渡库区滑坡地质灾害风险评价研究[D]. 成都: 西南石油大学, 2019. [ZHU Jilong. Risk assessment of landslide geological hazard in Xiluodu Reservoir area[D]. Chengdu: Southwest Petroleum University, 2019. (in Chinese with English abstract)]
- [29] 于宪煜, 胡友健, 牛瑞卿. 基于RS-SVM模型的滑坡易发性评价因子选择方法研究 [J]. 地理与地理信息科学, 2016, 32(3): 23–28. [YU Xianyu, HU Youjian, NIU Ruiqing. Research on the method to select landslide susceptibility evaluation factors based on RS-SVM model [J]. Geography and Geo-Information Science, 2016, 32(3): 23–28. (in Chinese with English abstract)]
- [30] 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. (in Chinese with English abstract)]
- [31] 宋庆武. 葫芦岛市建昌县山区滑坡危险性评价 [J]. 黑龙江水利科技, 2019, 47(7): 217–221. [SONG Qingwu. Risk evaluation of landslide of mountainous area in Jiangchang County in Huludao City [J]. Heilongjiang Hydraulic Science and Technology, 2019, 47(7): 217–221. (in Chinese with English abstract)]
- [32] 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 Xiaoqing, 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. (in Chinese with English abstract)]
- [33] 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)]
- [34] COSTANZO D, ROTIGLIANO E, IRIGARAY C, et al. Factors selection in landslide susceptibility modelling on large scale following the gis matrix method: application to the river Beiro basin (Spain) [J]. Natural Hazards and Earth System Sciences, 2012, 12(2): 327–340.
- [35] CHEN W, PENG J B, HONG H Y, et al. Landslide susceptibility modelling using GIS-based machine learning techniques for Chongren County, Jiangxi Province, China [J]. Science of the Total Environment, 2018, 626: 1121–1135.
- [36] GHAILAN O, HODA M O, HEGAZY O. Improving credit scorecard modeling through applying text analysis [J]. International Journal of Advanced Computer Science and Applications, 2016, 7(4): 512–517.
- [37] 胡旭东. 基于集成学习的地质灾害易发性评价研究: 以云南省泸水县为例[D]. 武汉: 中国地质大学, 2019. [HU Xudong. Study on the geo-hazards susceptibility assessment based on A novel ensemble learning framework —application to the Lushui County, Yunnan Province[D]. Wuhan: China University of Geosciences(Wuhan), 2019. (in Chinese with English abstract)]
- [38] 梁万杰. 滑坡、泥石流地质灾害评价方法研究[D]. 南京: 南京农业大学, 2012. [LIANG Wanjie. Study on assessment methodology of landslide and debris flow geological hazards[D]. Nanjing: Nanjing Agricultural University, 2012. (in Chinese with English abstract)]
- [39] 戴福初, 李军. 地理信息系统在滑坡灾害研究中的应用 [J]. 地质科技情报, 2000, 19(1): 91–96. [DAI Fuchu, LI Jun. Applications of geographical information systems in landslide studies [J]. Geological Science and Technology Information, 2000, 19(1): 91–96. (in Chinese with English abstract)]
- [40] YAO X, THAM L G, DAI F C. Landslide susceptibility mapping based on Support Vector Machine: a case study on natural slopes of Hong Kong, China [J]. Geomorphology, 2008, 101(4): 572–582.
- [41] KUMAR D, THAKUR M, DUBEY C S, et al. Landslide susceptibility mapping & prediction using Support Vector Machine for Mandakini River Basin, Garhwal Himalaya, India [J]. Geomorphology, 2017, 295: 115–125.
- [42] CHEN W T, LI X J, WANG Y X, et al. Forested landslide detection using LiDAR data and the random forest algorithm: a case study of the Three Gorges, China [J]. Remote Sensing of Environment, 2014, 152: 291–301.
- [43] 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)]
- [44] CHEN W, XIE X S, PENG J B, et al. GIS-based landslide

- susceptibility evaluation using a novel hybrid integration approach of bivariate statistical based random forest method [J] . *CATENA*, 2018, 164: 135 - 149.
- [45] 韩帅, 孙乐平, 杨艺云, 等. 基于改进K-Means聚类 and 误差反馈的数据清洗方法 [J] . *电网与清洁能源*, 2020, 36(7): 9 - 15. [HAN Shuai, SUN Leping, YANG Yiyun, et al. A data cleaning method based on improved K-means clustering and error feedback [J] . *Power System and Clean Energy*, 2020, 36(7): 9 - 15. (in Chinese with English abstract)]
- [46] 刘乾坤. 基于SPARK的四川省滑坡灾害成因分析与临界阈值研究[D]. 成都: 电子科技大学, 2020. [LIU Qiankun. Analysis of causes and critical threshold of landslide hazards in Sichuan Province based on SPARK[D]. Chengdu: University of Electronic Science and Technology of China, 2020. (in Chinese with English abstract)]
- [47] 朱力. 决策树算法在山区公路地质灾害风险评估系统中的应用[D]. 重庆: 重庆师范大学, 2019. [ZHU Li. Application of decision tree algorithm in risk assessment system of mountain highway geological disaster[D]. Chongqing: Chongqing Normal University, 2019. (in Chinese with English abstract)]
- [48] ALBERTO BOSCHETTI, LUCA MASSARON. Python data science essentials (Second Edition)[M]. 2016. Packet Publishing.
- [49] 崔阳阳, 邓念东, 曹晓凡, 等. 基于集成学习的地质灾害危险性评价 [J] . *水力发电*, 2020, 46(10): 36 - 41. [CUI Yangyang, DENG Niandong, CAO Xiaofan, et al. Geological disaster risk assessment based on ensemble learning algorithm [J] . *Water Power*, 2020, 46(10): 36 - 41. (in Chinese with English abstract)]
- [50] 李娟. 汶川县威州镇高分影像滑坡信息提取及危险性评价研究[D]. 成都: 成都理工大学, 2019. [LI Juan. Landslide information extraction and risk assessment of high resolution imagery in Weizhou town, Wenchuan County[D]. Chengdu: Chengdu University of Technology, 2019. (in Chinese with English abstract)]
- [51] 李远远, 梅红波, 任晓杰, 等. 基于确定性系数和支持向量机的地质灾害易发性评价 [J] . *地球信息科学学报*, 2018, 20(12): 1699 - 1709. [LI Yuanyuan, MEI Hongbo, REN Xiaojie, et al. Geological disaster susceptibility evaluation based on certainty factor and support vector machine [J] . *Journal of Geo-Information Science*, 2018, 20(12): 1699 - 1709. (in Chinese with English abstract)]
- [52] SNOEK J, LAROCHELLE H, ADAMS R P. Practical Bayesian optimization of machine learning algorithms[C]//NIPS'12: Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 2.2012: 2951-2959.
- [53] 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. Mapping landslide susceptibility based on machine learning and forecast warning of landslide induced by rainfall[D]. Shanghai: East China Normal University, 2019. (in Chinese with English abstract)]
- [54] 姚雄, 余坤勇, 刘健, 等. 基于随机森林模型的降水诱发山体滑坡空间预测技术 [J] . *福建农林大学学报(自然科学版)*, 2016, 45(2): 219 - 227. [YAO Xiong, YU Kunyong, LIU Jian, et al. Application of random forest model on the landslide spatial prediction caused by precipitation [J] . *Journal of Fujian Agriculture and Forestry University (Natural Science Edition)*, 2016, 45(2): 219 - 227. (in Chinese with English abstract)]
- [55] 邱维蓉, 吴帮玉, 潘学树, 等. 几种聚类优化的机器学习方法在灵台县滑坡易发性评价中的应用 [J] . *西北地质*, 2020, 53(1): 222 - 233. [QIU Weirong, WU Bangyu, PAN Xueshu, et al. Application of several cluster-optimization-based machine learning methods in evaluation of landslide susceptibility in Lingtai County [J] . *Northwestern Geology*, 2020, 53(1): 222 - 233. (in Chinese with English abstract)]
- [56] 管新邦. 云南省滑坡地质灾害危险性评价研究[D]. 北京: 中国矿业大学(北京), 2018. [GUAN Xinbang. Study on risk assessment of landslide in Yunnan Province[D]. Beijing: China University of Mining & Technology (Beijing), 2018. (in Chinese with English abstract)]
- [57] 郝国栋. 基于随机森林模型的商南县滑坡易发性评价[D]. 西安: 西安科技大学, 2019. [HAO Guodong. Landslide susceptibility assessment based on random forest model in Shangnan County[D]. Xi'an: Xi'an University of Science and Technology, 2019. (in Chinese with English abstract)]