



基于机器学习的边坡安全稳定性评价及防护措施

武梦婷 陈秋松 齐冲冲

Slope safety, stability evaluation, and protective measures based on machine learning

WU Meng-ting, CHEN Qiu-song, QI Chong-chong

引用本文:

武梦婷, 陈秋松, 齐冲冲. 基于机器学习的边坡安全稳定性评价及防护措施[J]. *工程科学学报*, 2022, 44(2): 180–188. doi: 10.13374/j.issn2095-9389.2021.06.02.008

WU Meng-ting, CHEN Qiu-song, QI Chong-chong. Slope safety, stability evaluation, and protective measures based on machine learning[J]. *Chinese Journal of Engineering*, 2022, 44(2): 180–188. doi: 10.13374/j.issn2095-9389.2021.06.02.008

在线阅读 View online: <https://doi.org/10.13374/j.issn2095-9389.2021.06.02.008>

您可能感兴趣的其他文章

Articles you may be interested in

基于机器学习的北京市PM2.5浓度预测模型及模拟分析

Machine-learning-based model and simulation analysis of PM2.5 concentration prediction in Beijing
工程科学学报. 2019, 41(3): 401 <https://doi.org/10.13374/j.issn2095-9389.2019.03.014>

基于支持向量回归与极限学习机的高炉铁水温度预测

Prediction of blast furnace hot metal temperature based on support vector regression and extreme learning machine
工程科学学报. 2021, 43(4): 569 <https://doi.org/10.13374/j.issn2095-9389.2020.05.28.001>

基于监督学习的前列腺MR/TRUS图像分割和配准方法

Prostate MR/TRUS image segmentation and registration methods based on supervised learning
工程科学学报. 2020, 42(10): 1362 <https://doi.org/10.13374/j.issn2095-9389.2019.10.10.006>

三维软硬互层边坡的破坏模式与稳定性研究

Numerical analysis of the failure modes and stability of 3D slopes with interbreeding of soft and hard rocks
工程科学学报. 2017, 39(2): 182 <https://doi.org/10.13374/j.issn2095-9389.2017.02.003>

油气资源开发的大数据智能平台及应用分析

Big data intelligent platform and application analysis for oil and gas resource development
工程科学学报. 2021, 43(2): 179 <https://doi.org/10.13374/j.issn2095-9389.2020.07.21.001>

基于机器学习的边坡安全稳定性评价及防护措施

武梦婷, 陈秋松, 齐冲冲[✉]

中南大学资源与安全工程学院, 长沙 410083
✉通信作者, E-mail: chongchong.qi@csu.edu.cn

摘要 为了更加快捷、高效地判定边坡稳定与否, 基于机器学习, 融合主成分分析法(PCA)、参数调整、影响因素权重分析等, 建立了一种边坡安全稳定性评价体系。研究发现, 运用 PCA 可以在保留 80% 数据原信息的前提下将输入变量维度从六维降至三维, 但此时模型效果有所下降; 随机森林及梯度提升(XGBoost)两种学习算法均可搭建有效的边坡安全稳定性评估模型, 通过对其预测效果的对比分析, 确定 XGBoost 为最佳评价模型。与此同时, 采取卡方检验、*F* 检验以及互信息法 3 种相关性检验手段, 并通过计算评价因子的重要程度且加以可视化展示, 明确了容重、坡高、内摩擦角以及内聚力 4 个内在因素的重要性, 最终将评估结果与实际结合提出了边坡安全防护措施。

关键词 边坡稳定性评价; 机器学习; 随机森林; XGBoost; 防护措施

分类号 X936

Slope safety, stability evaluation, and protective measures based on machine learning

WU Meng-ting, CHEN Qiu-song, QI Chong-chong[✉]

School of Resources and Safety Engineering, Central South University, Changsha 410083, China
✉ Corresponding author, E-mail: chongchong.qi@csu.edu.cn

ABSTRACT In recent years, the slope instability has brought immeasurable costs to production and life of human. As a result, it is essential to correctly understand, analyze, and design the slope reasonably, and implement appropriate protective measures to minimize the loss and harm caused by its instability. By far, slope stability can be investigated using theoretical analysis, numerical modeling and machine learning prediction, among them machine learning prediction has been the most encouraging one. Many studies have been performed using machine learning algorithms to predict the slope stability. However, these methods suffers from poor accuracy and poor generalisation capability, so its real-life application has been limited. In the current study, a machine learning-based slope safety and stability evaluation system is established by integrating principal component analysis, parameter adjustment, and influence factor weight analysis. It is shown that PCA can reduce the dimensions of the input variables from six to three while retaining 80% of the information; however, at the cost of the model's effectiveness. The random forest and XGBoost (eXtreme Gradient Boosting) learning algorithms can both be employed to develop effective evaluation models for slope safety and stability. The comparative analysis of algorithms' prediction effects established XGBoost as the best evaluation model, which can achieve the average accuracy of 92%, precision of 91%, recall of 96%, and the area under the receiver operating characteristic curve (AUC) of 0.95. In addition, this study employs three types of test methods: the chi-square test, *F* test correlation, and mutual information method, meanwhile by calculating and visualizing the importance of influencing factors, the influence of unit weight, slope height, internal friction angle and cohesion on slope stability is demonstrated. It has been shown that the unit weight is the most influencing factor for the slope stability. Finally, the slope safety protection measures are proposed by combining the evaluation results with the actual project.

KEY WORDS slope stability evaluation; machine learning; random forest; XGBoost; protective measures

收稿日期: 2021-06-02

基金项目: 中央高校基本科研业务费专项资金资助项目(202045009); 国家自然科学基金资助项目(52074351)

关于边坡灾害事故的防治, 首先要做的就是对边坡的安全稳定状态进行合理分析与评价, 这也是边坡防治工程的核心与重点。目前, 评估边坡是否安全稳定最常采取的还是定性分析以及定量求解等传统手段。但边坡作为一个动态开环系统, 其难以确定的安全影响因子很多, 且大多具有参数模糊以及随机可变等特点, 因此, 传统手段应用起来都有相对的局限性。而随着研究的深入, 确定性与不确定分析相联合、通过物理模拟构建离心模型等也成为边坡研究的新途径。但上述方法尚存在一些未解决的问题, 方法结果的实际可信度有所降低。

随着机器学习技术的快速兴起, 为边坡安全稳定性研究提出了一种新思路。国内外学者开始将决策树^[1]、随机森林、支持向量机^[2-4]以及朴素贝叶斯^[5]等算法广泛应用于边坡研究中。Neuland通过收集分析250个边坡数据, 明确了31处变量, 并借助主成分分析方法(PCA)解释了其间的函数关系, 从而构建了一个评估边坡是否稳定的算法模型, 且在测试数据中泛化能力较好^[6]。Pradhan与Li借助逆向传播的神经网络, 对巴生谷地区的边坡进行了研究, 计算了坡率、坡高等11个相关影响因素的权重值, 计算结果显示人工神经网络(ANN)优于先前采用手段^[7]。Alavi与Gandomi利用遗传算法对岩土边坡的具体问题进行了研究^[8]。Martins与Miranda进行了多次逻辑回归, 并采用随机森林、K近邻算法以及决策树等多种方法途径来探索边坡的稳定状态^[9]。赵洪波基于极限平衡创建学习样本, 并将一阶二次矩应用于支持向量机中对边坡可靠性进行分析研究^[10]。陈善攀在对边坡进行可靠性分析时, 提出了能够精确计算边坡可靠评估标准并同步搜寻最危滑体的算法, 并借助面向对象进行了边坡稳定系统编码^[11]。

尽管上述学者应用机器学习在边坡领域已取得了一些成果, 但其研究尚存在以下不足: ①由于评价因子数量较少, 采用压缩、转换数据构建模型的手段并不常见, 且对于PCA法在边坡问题中的适用性与可行性也从未进行详细探究; ②文献中使用诸如梯度提升算法(XGBoost)等快速更新的先进评估器对边坡稳定性进行预测的尝试不多, 且构建多种模型进行对比研究的方式也相对较少; ③影响因素的权重分析不够全面, 缺少基于相关性检验的权重分析与基于智能模型的权重分析的综合对比; ④未对评价结果进行深入考虑及合理部署, 缺少智能驱动的边坡安全防护措施。

针对以上问题, 本文在使用归一化处理对数据质量进行改善后, 采用了随机森林以及XGBoost两种学习算法搭建边坡安全稳定性评估模型, 在对其预测效果进行对比后确定最终评价模型。此外, 还增加了建模前的特征工程以及建模后的特征重要性分析步骤, 得出评价因子的影响权重后, 将评估结果与实际结合提出了边坡安全防护措施。

1 方法原理

1.1 PCA

PCA是一种使用线性代数来转换压缩数据的技术手段。其通过将线性相关的多个指标重组为一组数量较少且相互独立的综合指标, 来实现在减少特征数量的同时, 保留大量有用信息^[12-14]。降维过程中所使用的衡量指标是样本方差, 方差越大, 说明特征含有的有效信息越多, 对于模型的创建能够有所贡献。方差的计算公式见式(1):

$$\text{Var} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{x})^2 \quad (1)$$

其中, Var代表方差; n 代表样本总数; x_i 代表每一个特征的第 i 个样本值; \hat{x} 代表对应某一个特征的所有样本的平均值。

PCA降维的核心原理是在确定方差数值排列前二的方向后, 使用类比的方法将全部主成分的方向予以确定, 然后通过分解协方差矩阵得出其特征向量及特征值, 从而实现将初始数据压缩映射至新的低维空间, 并保证总信息量损失较少。具体实施步骤如下:

(1) 输入 b 条 a 维数据构成样本集 $\mathbf{X}=(x^{(1)}, x^{(2)}, \dots, x^{(b)})$, 其中 a 代表特征个数;

(2) 将 $x^{(1)}, x^{(2)}, \dots, x^{(b)}$ 进行归一化处理, 得到矩阵 \mathbf{X}^1 ;

(3) 求解出协方差矩阵 $\mathbf{C} = \frac{1}{b} \mathbf{X} \mathbf{X}^T$, 以及其特征值和特征向量;

(4) 将特征值由大到小排序后, 取出与之对应的前 q 个特征向量组成新特征空间 V ;

(5) $\mathbf{Y} = \mathbf{V}^T \mathbf{X}^1$ 即为特征降到 q 维的新的样本集。

1.2 XGBoost

XGBoost是一种利用正则项等手段来简化模型, 使其具有高精度、高效率以及较高预测准确率的提升算法。该算法以梯度提升算法为基础, 通过不断添加决策树并进行特征分解来学习新函数, 从而拟合残留的预测误差; 当 k 棵树训练完毕后可根据样本特征计算叶子节点分数, 通过累加得

到样本的预测值. XGBoost 模型构建主要分为以下四个步骤:

(1) 建立目标函数并求取最优解. 如式(2)所示, 目标函数=传统损失函数+模型复杂度.

$$\text{Obj} = \sum_{i=1}^m l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (2)$$

其中, i 表数据集中的第 i 个样本; m 表示导入第 k 棵树的数据总量; K 代表建立的所有树; $l(y_i, \hat{y}_i)$ 是模型的损失函数, 用于评估分类器的概率输出, y_i 表示真实标签; \hat{y}_i 表示预测值; $\sum_{k=1}^K \Omega(f_k)$ 是全部 K 棵树的复杂度求和, f_k 表示第 k 棵决策树, Ω 表示树模型的某种变换.

(2) 利用泰勒公式将目标函数展开后对其进行整合、重组, 使其转化为与预测残差相关的多项式, 如式(3)所示:

$$\text{Obj} \approx \sum_{i=1}^m [l(y_i', \hat{y}_i^{(t-1)} + f_i(x_i))] \approx \sum_{i=1}^m [l(y_i', \hat{y}_i^{(t-1)}) + f_i(x_i)g_i + \frac{1}{2}(f_i^2(x_i)h_i)] + \Omega(f_i) \quad (3)$$

其中, $l(y_i', \hat{y}_i^{(t-1)} + f_i(x_i))$ 是 $f(x)$ 对应的损失函数, 其中 y_i' 表示第 t 棵树的真实标签, $\hat{y}_i^{(t-1)}$ 是第 $t-1$ 步模型给出的预测值, $f_i(x_i)$ 是第 t 棵树的模型; g_i 和 h_i 分别是损失函数的一阶和二阶导数.

(3) 基于目标函数, 采取某种手段将树的结构优化, 最终转变为由结构分数如式(4)所示:

$$\text{Obj} = -\frac{1}{2} \sum_{j=1}^T \frac{G_j^2}{H_j + \lambda} + \gamma T \quad (4)$$

其中, γ 和 λ 为提前设定的超参数; T 表示叶子数量; G_j 和 H_j 由损失函数和特定结构下树的预测结果共同决定; 结构分数 Obj 越小表示树越优化.

(4) 树的结构确定后, 借助贪心算法来实现树的分裂并求取算法模型的局部最优解^[15].

1.3 随机森林

为验证 XGBoost 算法的高精度与高效率, 基于先前研究成果^[16], 本文选取了预测准确率较高的随机森林算法进行对比.

随机森林是集成学习并行模式中的一种十分典型的机器学习算法. 该算法以单独的决策树作为弱学习器, 且弱学习器间互不依赖, 其原理是将 Bagging 袋装集成作为理论基础, 在对决策树进行独立训练时利用随机种子来进行抉择^[17]. 针对袋装法, 由于集成结果依赖于少数服从多数及平均值准则, 所以独立决策树的预测正确与否直接决

定了随机森林算法预测准确性的高低.

随机森林模型搭建的基本思路如下: ①从数据集中有放回的随机抽取 b 个样本^[18]; ②随机选取 a 个特征并利用这些特征对所选取的样本进行决策树的构建; ③将步骤①和②重复 z 次生成 z 棵决策树, 组合形成随机森林; ④利用随机森林中的 z 棵树对新数据进行判断, 最后投票确认所属类别.

2 边坡稳定性预测模型

2.1 样本数据集数据分析

收集整理数据集是建立机器学习预测模型的第一步. 在实际工程领域中, 对于一个具体的问题, 使用网络公开的一些合理且具有代表性的数据集, 其预测结果更具有说服力. 因此, 本文收集了 168 个坡度剖面的数据集, 该数据集由文献 [19-23] 中的不同论文汇编而成, 并已广泛用于边坡稳定性预测, 其统计特征如图 1 所示. 其中, N 代表边坡样本的总数量, Mean 和 Std 分别代表着每一个特征的均值和标准差, Min 和 Max 为对应特征的最小值和最大值, Med 为特征的中位数.

2.2 预测模型建立流程

(1) 数据预处理. 数据的质量影响着机器学习模型的性能, 决定了算法所能达到的模型上限^[24]. 分析图 1 可知, 原始数据集的孔隙水压力因子存在异常值, 因此, 应将该样本进行删除清理, 最终保留 167 条数据. 又因原始数据为连续型变量, 存在数据尺度、量纲不统一等问题. 对此本文使用归一化处理的手段对数据进行转换, 并将处理完毕的数据应用于算法, 来提高模型算法的准确度.

(2) 建模因子的初步确定及筛选. 本文收集的数据集中包含了对边坡稳定性影响最大的一些参数: 几何形状、重力、地质力学参数和土壤中的水量. 对这些参数进行分析, 初步确定了边坡高度、边坡角、孔隙压力比、容重、内聚力和内摩擦角 6 个建模因子后, 借助方差过滤手段来删除方差值为 0 且对样本区分没有价值的特征. 然后再使用卡方检验、 F 检验以及互信息法 3 种相关性检验的方法对建模因子进行筛选.

(3) PCA 降维. 为了对比不同特征工程手段对于预测模型准确率的影响, 本文还采用 PCA 降维算法对特征进行处理, 创造具有不可读性的全新特征. 此外, 考虑到保留特征过多达不到降维效果, 以及留下特征较少, 新特征向量无法容纳初始数据的大部分信息等弊端, 本文借助累积可解释方差贡献率曲线及最大似然估计自动选取降维参数 $n_components$ 的最佳取值, 并基于此查看模型

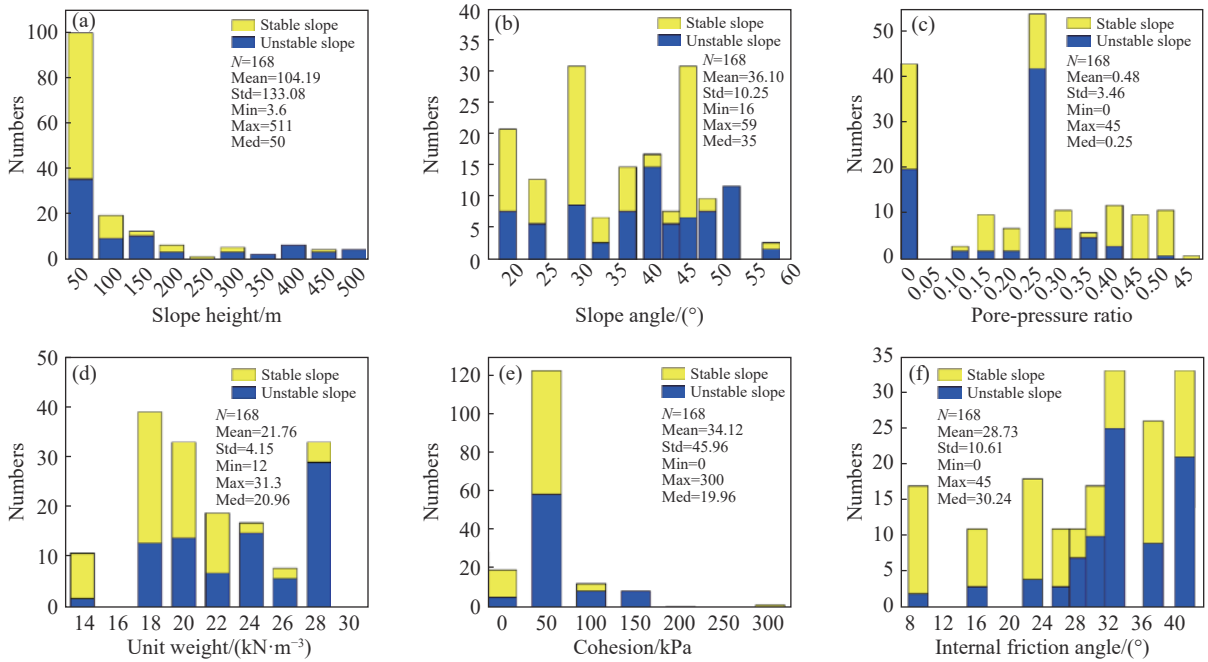


图1 影响因素的特征统计。(a)坡高分布；(b)坡角分布；(c)孔隙压力比分布；(d)容重分布；(e)内聚力分布；(f)内摩擦角分布

Fig.1 Characteristic statistics of influencing factors: (a) distribution of slope height; (b) slope angle distribution; (c) pore pressure ratio distribution; (d) unit weight distribution; (e) distribution of cohesion; (f) internal friction angle distribution

预测准确率。下文将详细介绍 PCA 技术的研究成果。

(4)模型的初步建立及调参。本文首先将数据集中的 167 个样本按照 3 : 7 的比例随机划分产生测试集与训练集。然后在完整数据集上,利用筛选后的建模因子作为输入指标,同时采用随机森林以及 XGBoost 两种学习算法搭建边坡安全稳定性评估模型。并使用 10 折交叉验证作为上述两种分类器的评估指标,在不断增加弱学习器数量的基础上通过改变超参数来提升模型的分类效果。

2.3 模型评估

最佳超参数组合确定后,本文将调参完毕的两种分类算法重新使用原始训练集进行训练并用测试集进行评估,选取了基于混淆矩阵的 4 个评估指标:准确率(Accuracy)、查准率(Precision)、召回率(Recall)^[25]以及接受者操作特性曲线(ROC)和该曲线下围成的面积(AUC, Area under curve),对模型性能进行衡量。

其中,混淆矩阵可将真实类别与预测类别的误分及比重情况进行可视化,并将样例分为真、假、正、反例 4 种情况,分别用 TP、FP、TN 及 FN 表示。而准确率是指在全部样本中被预测正确的样本所占的比例,其取值越接近 1 代表预测效果越好。如公式(5)所示:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}} \quad (5)$$

精确度,又被称为查准率,其分母为所有被判断为正例的样本,而分子是真正的正例样本。召回率,又叫作查全率、敏感度,其含义是在所有真实为正例 1 的样本中,被预测为 1 的样本的占比。计算定义如式(6)和(7)所示:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (6)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (7)$$

ROC 是一条以假正率(FPR)为横坐标, Recall 为纵坐标的受试者工作特征曲线。曲线下方围成的面积由 AUC 表示,其有效评价范围一般为 [0.5,1]。面积越大代表 ROC 曲线越陡,则模型性能越优良。当 AUC 的值超过 0.9 时,代表该模型的预测非常准确;而当其低于 0.5 时,便不再具有判断价值。

3 结果分析

3.1 PCA 结果分析

以降维后保留的特征个数作为横坐标,新特征矩阵捕捉到的累积可解释方差贡献率作为纵坐标,绘制曲线如图 2 所示。分析可知,降维后的最佳维度范围处于 [3,4] 之间,其所占的信息量占原始数据总信息量的 84% 左右。特征减少一半的同时保留的信息量仍超过 80%,在某种程度上证明了 PCA 算法的可取性。继续缩小范围绘制细化的学习曲线,如图 3 所示,结合最大似然估计法可确

定降维最佳维度为 3, 此时随机森林分类模型的预测准确率达到 87.46%, XGBoost 分类模型的预测准确率达到 88.68%。但使用原始数据集构建两种模型的预测准确率分别为 92.90% 和 91.14%, 由此可见使用 PCA 降维后的模型预测效果有所下降。此外, 基于 6 个特征对运算速度影响较小的考虑, PCA 算法不适用于最终模型的构建。

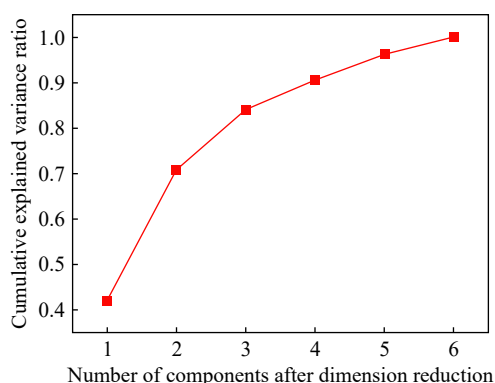


图 2 累积可解释方差贡献率曲线

Fig.2 Cumulative explanatory variance contribution curve

3.2 模型参数调整及评估结果分析

如图 4 所示, 训练集上的表现展示了模型的学习能力, 测试集上的表现展示了模型的泛化性能。

而初步建立的两个预测模型在训练集上都表现较为优秀, 其预测准确率随着样本的增加逐渐接近 100%, 但在验证集上却表现相对不佳。因此, 本文通过参数调整来提高模型的泛化能力。

针对两个分类模型, 基于学习曲线能预测弱学习器数量 $n_estimators$ 的变化趋势, 并且具有判断该参数是否能一直推动模型预测准确率上升的优势。本文借助学习曲线初步划定范围, 并在确定好的范围内, 将其进一步细化, 以此来对超参数 $n_estimators$ 进行优化。并在弱学习器数量确定后将泛化误差作为出发点, 借助网格搜索以及偏差-方差图对随机森林模型的最大深度 max_depth 、分枝时所需考虑的特征数 $max_features$ 以及生成森林的随机种子树 $random_state$ 参数进行调整, 从而提升模型的复杂度。最终确定随机森林分类算法的最佳参数组合为 [$'random_state':4, 'n_estimators':35, 'max_depth':8, 'max_features':3$]。

与此同时, 本文使用学习率参数 $learning_rate$ 来干涉 XGBoost 树模型的迭代步长并进行剪枝调参, 通过调整损失函数下降的最小 $gamma$ 值及最大深度 max_depth 参数来增强模型的泛化能力, 最

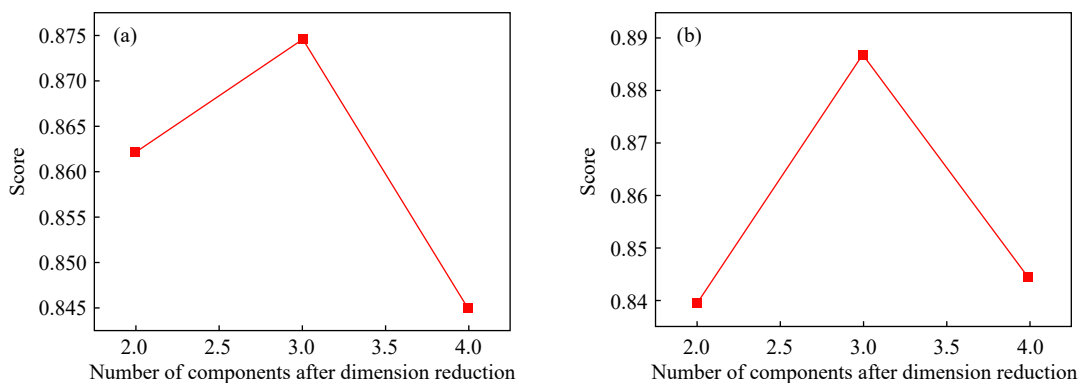


图 3 降维后的模型表现。(a) 随机森林模型; (b) XGBoost 模型

Fig.3 Model performance after dimensionality reduction: (a) random forest model; (b) XGBoost model

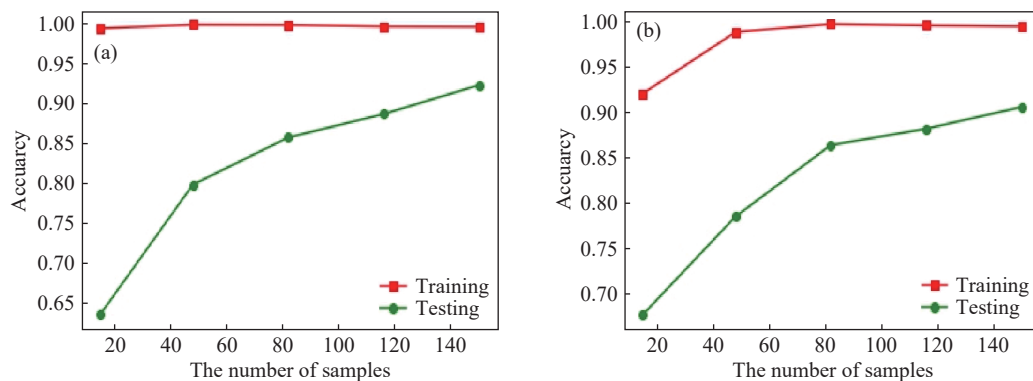


图 4 训练集和验证集的结果对比。(a) 随机森林模型; (b) XGBoost 模型

Fig.4 Comparison of the results between the training set and the verification set: (a) random forest model; (b) XGBoost model

终确定了 XGBoost 分类算法的最佳参数组合为 ['n_estimators':77,'learning_rate':0.95,'gamma':0.3,'max_depth':5].

将确定最佳参数的两个算法模型分别用 70% 的数据训练完毕后应用于测试集进行测试, 10 次预测结果表现如图 5 所示. 分析可知, XGBoost 模型表现相对较好, 其平均准确率达到 92%, 而随机森林的预测平均值为 86%.

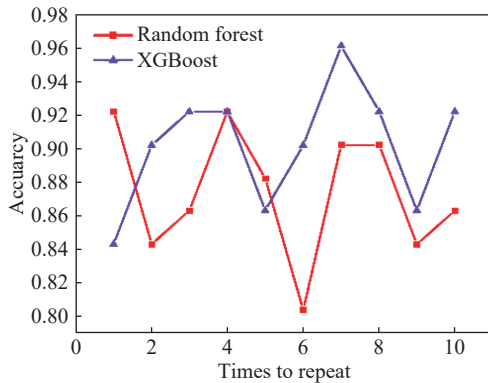


图 5 测试集的 10 次预测准确率

Fig.5 Ten times prediction accuracy of the test set using random forest and XGBoost models

此外, 鉴于数据集划分时的随机性, 为保证评估结果的合理准确, 本文利用准确率、查准率、召回率以及 AUC 4 个评估指标进行了 10 次均值求解并加以对比分析, 如表 1 所示. 通过不断调整 ROC 阈值, 确定了当该参数分别为 0.5867 以及 0.3186 时, 两种分类模型的综合评估结果最佳, 最终的 ROC 曲线如图 6 所示.

分析表 1 可知, XGBoost 算法所得各项值都略高于其他项, 说明 XGBoost 分类算法在数据集上表现效果较好, 其平均准确率达到 92%、精确度达到 91.0%、查准率达到 96% 以及 AUC 可达到 0.95. 因此 XGBoost 确定为本边坡数据集的最佳预测模型.

表 1 两种算法评估结果差异表

Indicators	Random forest	XGBoost
Accuracy	0.86	0.92
Precision	0.90	0.91
Recall	0.85	0.96
AUC	0.92	0.95

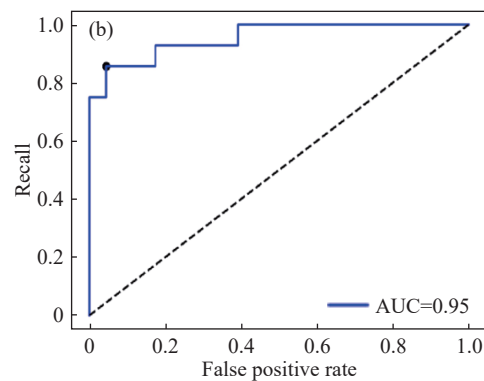
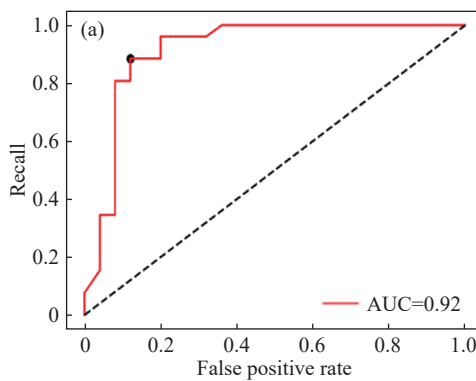


图 6 调整后的 ROC 与 AUC 展示. (a)随机森林模型; (b)XGBoost 模型

Fig.6 ROC and AUC display after adjustment: (a) random forest model; (b) XGBoost model

3.3 影响因素权重分析

3.3.1 相关性检验

本文基于地形地貌、地质条件、水的作用以及人为干扰 4 个要素对于边坡安全稳定的影响分析, 采用卡方检验、 F 检验以及互信息法 3 种特征工程手段来评判特征与标签之间的相关性, 并将结果通过卡方值 χ^2 、 P 值、 F 值以及互信息量估计值 4 个统计变量进行呈现. 其中, χ^2 表示观察值与理论值之间的吻合程度, P 值为结果可信水平的一个递减指标, 通常被认为 $P=0.05$ 是样本变量关联可接收过错的边界; χ^2 越大 P 值越小, 两个变量的

关联性越强; F 值是用来表示拟合方程显著性的两个均方的比值(效应项/误差项), F 越大, 拟合程度越好.

对表 2 进行分析可知, 在卡方检验及 F 检验的结果中, 坡高、容重还有内摩擦角这三个评价因子的卡方值 χ^2 与 F 值较高. 其中, 通过卡方计算, 得出容重与内摩擦角两个因子的 P 值结果均大于临界值 0.05, 说明该评价因子与标签无明显的线性关系, 但基于上述影响因素分析可知, 其对边坡的安全稳定仍具有某种干扰. 同样地, F 检验中, 这 3 个评价因子的 P 值都小于临界值, 且容重的 P 值

结果为 0, 说明此时容重的 P 位于拒绝域, 从而证明了该特征与标签有很强的相关性, 对边坡的安全稳定干扰性最强. 此外, 互信息法的评估值都大

于 0, 表示全部特征都和标签之间满足线性相关, 从而说明了内聚力对于边坡的稳定状态也有一定的影响.

表 2 三种特征选择方法下的评价因子重要性

Table 2 Importance of evaluation factors under the three feature selection methods

Feature selection method	Important evaluation factors and corresponding index scores		
	Slope height	Unit weight	Internal friction angle
Chi-square (χ^2) test	$\chi^2=6.9396$ $P=0.0084$	$\chi^2=3.5136$ $P=0.0609$	$\chi^2=2.5307$ $P=0.1117$
F test	$F=22.5560$ $P=0.00000441$	$F=49.7141$ $P=0$	$F=34.8498$ $P=0.00000002$
Mutual information method	0.2354	0.2173	0.2791 (Cohesion)

综上所述, 在特征工程中, 容重、坡高、内摩擦角以及内聚力 4 个评价因子对边坡稳定性影响最大.

3.3.2 评价因子重要性分析

随机森林和 XGBoost 分类算法都具有计算特征重要程度并将其进行输出的功能. 其中, 特征重要性 (feature_importance) 是一种在某种程度上

利于特征筛选并提高模型鲁棒性的特有属性. 此外, 针对边坡是否处于稳定状态的评估问题, 特征重要性也可以从侧面反映各因素对边坡稳定的影响程度, 从而为边坡安全防护方案的制定提供有效参考. 将重复 10 次计算所得的各评价因子对预测模型的重要性进行可视化展示, 如图 7 所示.

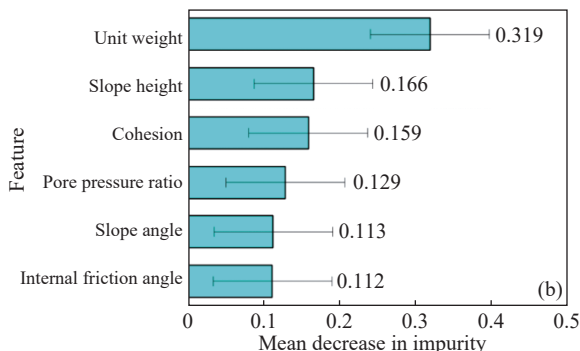
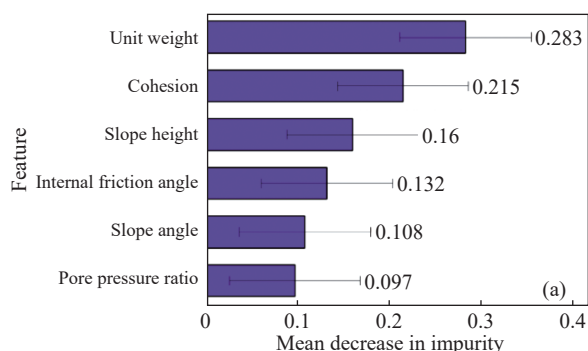


图 7 特征重要性排序. (a) 随机森林模型; (b) XGBoost 模型

Fig.7 Rank of feature importance: (a) Random forest model; (b) XGBoost model

其中, 横坐标为各评价因子的权重系数, 纵坐标为 6 个特征的英文表示. 分析图 7 可知, 容重、内聚力和坡高 3 个评价因子在随机森林与 XGBoost 算法模型中的重要程度最高.

结合 3.3.1 相关性检验可知, 容重、坡高、内摩擦角以及内聚力被证明是导致边坡失稳破坏的重要参数, 可尝试作为边坡防治的着手点进行分析研究.

3.4 边坡安全防护

边坡安全防护的本质就是预防边坡的变形、防止发生失稳破坏. 针对实际边坡工程, 可考虑从本文研究的数据集入手, 以容重、坡高、内摩擦角以及内聚力 4 个内在参数作为重点考虑对象, 结合先进技术手段以及工程社会效应来制定经济合理的防护方案, 不失为一种科学、理想的措施.

3.4.1 注浆支护与锚注加固

土质边坡内应力超过其剪切极限值时, 部分土体会出现相对滑动, 进而导致边坡变形破坏. 而内摩擦角是岩土体内部的抗剪强度指标之一, 其取决于摩擦强度和内聚力, 而内聚力又决定着土体的抗剪能力, 内聚力越大, 其抵抗变形的抗剪强度及刚度也会随之增强, 边坡的存在状态也会更加安全稳定^[26]. 注浆是一种有效提升内聚力的方式, 通过喷注浆液生成全新的高强度胶体, 并借助物化反应, 可提高界面的剪切强度和抗剪刚度^[27]. 此外, 在注浆完成后额外增加坡面防护, 不仅能对岩体软弱面进行直接调整, 还可避免单独使用锚杆所带来的问题, 使边坡由内至外的整体稳定性得以大幅提高^[28], 因此, 锚注加固方法应加以重视.

3.4.2 有效疏排水

地表水及地下水的渗入会造成岩土体含水量增加,进而使得孔隙水压力不断上升^[29],土的容重增加,内聚力及抗剪强度下降。采取有效的疏排水措施,可提高岩土体的内聚力与内摩擦角,降低孔隙水压力及容重值,进而提升边坡安全系数,使边坡的稳定状态得以改善。

边坡疏排水分为地表排水和地下排水两种。地表排水对各类边坡工程皆为适用,能够改善因地表水作用导致的边坡稳定性降低。一般在坍、滑体上方,按其汇水面积及降雨情况,结合地形设置一道或几道截水沟,使地表水全部汇入截水沟,引至路基边沟或涵洞排出。地下排水的整体方案应根据边坡所处位置、边坡与建筑物关系、工程地质和水文地质条件确定,可选用渗沟、盲沟、排水洞、排水孔及集水井等形式。坍、滑体内地下水丰富且层次较多时,可设支撑盲沟,用于排水和支撑。当坍、滑体上方有地下水时,在垂直于地下水流的方向设截水盲沟,将地下水引向两侧排出。

3.4.3 植被护坡

采用植被防护,就是利用植被对边坡的覆盖作用、植物根系对边坡的加固作用,保护路基边坡免受大气降水与地表径流的冲刷。植被防护的手段通常为种草、铺草皮和植树。

植被不仅能使坡面保持适宜湿度,防止岩土体含水量过高,其根系还具有高抗拉性、较大变形率及能将土壤和草根抗剪强度无形融合的固结能力,因此将其植入土中与工程材料结合使用,可有效控制土体变形并改善内聚力,使土体抗剪强度提升,增强边坡的安全稳定性^[27]。

3.4.4 减载

坡度较高的边坡力学条件特殊,破碎岩土体较多,容易在外界作用力的影响下发生崩塌、滑坡。而岩层间的大量裂隙又使得其渗透能力增强,雨季天气时水压的增大使得侵蚀严重,更易发生失稳事故。

因此,对于超出限定范围的高耸边坡,可采取削头减载与削坡减载两种手段。削头减载通过将边坡岩土体上部削减来实现降低边坡的总高度;而削坡减载在保持阻滑体不被破坏的情况下,尽量将处于不安全状态的岩土体剖面进行削减,从而提高边坡的稳定性。

4 结论

(1)借助 PCA 降维手段对本文数据集进行处

理,可以在保留 80% 信息的前提下将输入变量维度从六维降至三维,但模型预测效果有所下降,从而判定 PCA 算法不适用于最终模型的构建。

(2)采用随机森林和 XGBoost 两种学习算法构建边坡安全稳定性评价模型,将其评价结果对比分析可得 XGBoost 模型的评估效果较好,预测准确率的平均值为 92%,精确度为 91.0%、查准率达到 96% 以及 AUC 可达到 0.95。因此 XGBoost 确定为本边坡数据集的最佳预测模型。

(3)本文采取卡方检验、 F 检验以及互信息法 3 种相关性检验手段,对评价因子进行分析处理,并通过将两种分类算法下的评价因子的重要程度加以分析计算及可视化展示,明确了容重、坡高、内摩擦角以及内聚力 4 个内在因素对于模型构建的重要性,以及对边坡稳定存在的影响力。

(4)边坡安全防护工作可重点考虑容重、坡高、内摩擦角以及内聚力 4 种因素的影响,以此作为切入点并结合工程实际制定具体防护措施。

参 考 文 献

- [1] Thai Pham B, Tien Bui D, Prakash I. Landslide susceptibility modelling using different advanced decision trees methods. *Civ Eng Environ Syst*, 2018, 35(1-4): 139
- [2] Balogun A L, Rezaie F, Pham Q B, et al. Spatial prediction of landslide susceptibility in western Serbia using hybrid support vector regression (SVR) with GWO, BAT and COA algorithms. *Geosci Front*, 2021, 12(3): 101104
- [3] Peethambaran B, Anbalagan R, Kanungo D P, et al. A comparative evaluation of supervised machine learning algorithms for township level landslide susceptibility zonation in parts of Indian Himalayas. *CATENA*, 2020, 195: 104751
- [4] Samui P. Slope stability analysis: A support vector machine approach. *Environ Geol*, 2008, 56(2): 255
- [5] Jiang S H, Liu Y, Zhang H L, et al. Quantitatively evaluating the effects of prior probability distribution and likelihood function models on slope reliability assessment. *Rock Soil Mech*, 2020, 41(9): 3087
(蒋水华, 刘源, 章浩龙, 等. 先验概率分布及似然函数模型的选择对边坡可靠度评价影响的定量评估. *岩土力学*, 2020, 41(9): 3087)
- [6] Neuland H. A prediction model of landslides. *CATENA*, 1976, 3(2): 215
- [7] Pradhan B, Lee S. Landslide susceptibility assessment and factor effect analysis: Backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. *Environ Model Software*, 2010, 25(6): 747
- [8] Alavi A H, Gandomi A H. A robust data mining approach for formulation of geotechnical engineering systems. *Eng Comput*, 2011, 28(3): 242

- [9] Martins F F, Miranda T F S. Application of data mining techniques to the safety evaluation of slopes // *Information Technology in Geo-Engineering: Proceedings of the 1st International Conference (ICITG)*. Shanghai, 2010: 84
- [10] Zhao H B. Reliability analysis of slope based on support vector machine. *Chin J Geotech Eng*, 2007, 29(6): 819
(赵洪波. 基于支持向量机的边坡可靠性分析. *岩土工程学报*, 2007, 29(6): 819)
- [11] Chen S P. *Genetic Algorithm and Program Design for Reliability Analysis of Soil Slope Stability* [Dissertation]. Changsha: Central South University, 2008
(陈善攀. 土质边坡稳定可靠度分析遗传算法方法及程序设计[学位论文]. 长沙: 中南大学, 2008)
- [12] Gou Q Q, Zhao M S, Chi E A, et al. Prediction and application of evaluation factors in blasting vibration based on PCA-BP neural network. *Min Res Dev*, 2018, 38(12): 97
(苟倩倩, 赵明生, 池恩安, 等. 基于PCA-BP神经网络在爆破振动评价要素中的预测及应用. *矿业研究与开发*, 2018, 38(12): 97)
- [13] Lin H M, Du Z F. Some problems in comprehensive evaluation in the principal component analysis. *Stat Res*, 2013, 30(8): 25
(林海明, 杜子芳. 主成分分析综合评价应该注意的问题. *统计研究*, 2013, 30(8): 25)
- [14] Chen G F, Lu Y F, Cheng S G. Principal component analysis of influence factors of slope stability. *Met Mine*, 2008(4): 123
(陈高峰, 卢应发, 程圣国. 边坡稳定性影响因素主成分分析. *金属矿山*, 2008(4): 123)
- [15] Mao Y K. *Stability Evaluation of Landslide Based on Machine Learning and System Development* [Dissertation]. Chengdu: University of Electronic Science and Technology of China, 2020
(毛宇昆. 基于机器学习的滑坡稳定性评价及系统研发[学位论文]. 成都: 电子科技大学, 2020)
- [16] Qi C C, Tang X L. Slope stability prediction using integrated metaheuristic and machine learning approaches: A comparative study. *Comput Ind Eng*, 2018, 118: 112
- [17] Zhang Y F, Lu Z Q. Remaining useful life prediction based on an integrated neural network. *Chin J Eng*, 2020, 42(10): 1372
(张永峰, 陆志强. 基于集成神经网络的剩余寿命预测. *工程科学学报*, 2020, 42(10): 1372)
- [18] Cong M, Wu T, Liu D, et al. Prostate MR/TRUS image segmentation and registration methods based on supervised learning. *Chin J Eng*, 2020, 42(10): 1362
(丛明, 吴童, 刘冬, 等. 基于监督学习的前列腺MR/TRUS图像分割和配准方法. *工程科学学报*, 2020, 42(10): 1362)
- [19] Li J H, Wang F. Study on the forecasting models of slope stability under data mining // *Proceedings of the Workshop on Biennial International Conference on Engineering*. Honolulu, 2010: 765
- [20] Lu P, Rosenbaum M S. Artificial neural networks and grey systems for the prediction of slope stability. *Nat Hazards*, 2003, 30(3): 383
- [21] Sah N K, Sheorey P R, Upadhyaya L N. Maximum likelihood estimation of slope stability. *Int J Rock Mech Min Sci Geomech Abstr*, 1994, 31(1): 47
- [22] Yan X M, Li X B. Bayes discriminant analysis method for predicting the stability of open pit slope // *2011 International Conference on Electric Technology and Civil Engineering (ICETCE)*. Lushan, 2011: 147
- [23] Zhou K P, Chen Z Q. Stability prediction of tailing dam slope based on neural network pattern recognition // *2009 Second International Conference on Environmental and Computer Science*. Dubai, 2009: 380
- [24] Fan Q. *Rapid Evaluation Method Research of Rock Slope Stability for Highway Based on Machine Learning Process* [Dissertation]. Chengdu: Chengdu University of Technology, 2019
(樊奇. 基于机器学习全流程的高速公路岩质边坡稳定性快速评价方法研究[学位论文]. 成都: 成都理工大学, 2019)
- [25] Wu S, Liu L, Lu D. Imbalanced data ensemble classification based on cluster-based under-sampling algorithm. *Chin J Eng*, 2017, 39(8): 1244
(武森, 刘露, 卢丹. 基于聚类欠采样的集成不平衡数据分类算法. *工程科学学报*, 2017, 39(8): 1244)
- [26] Luo S L, Wan W, Tang J Z, et al. Numerical study of factors influencing slope stability. *Miner Eng Res*, 2016, 31(4): 37
(罗世林, 万文, 唐劲舟, 等. 影响边坡稳定性因素数值研究. *矿业工程研究*, 2016, 31(4): 37)
- [27] Xiao D S. Quantitative theory of slope shear strength increased by vegetation. *Sichuan Archit*, 2004, 24(1): 63
(肖东升. 植被增加边坡抗剪强度的量化理论. *四川建筑*, 2004, 24(1): 63)
- [28] Xu W Z, Peng Z B, Yuan H P. Characteristic analysis of bolting grouting reinforcement in jointed and fractured slope. *Hydrogeol Eng Geol*, 2006, 33(5): 30
(许万忠, 彭振斌, 袁海平. 节理裂隙边坡锚注加固机制及特性研究. *水文地质工程地质*, 2006, 33(5): 30)
- [29] Zheng K H, Luo Z Q, Luo C Y, et al. Layered gravel soil slope stability of a waste dump considering long-term hard rain. *Chin J Eng*, 2016, 38(9): 1204
(郑开欢, 罗周全, 罗成彦, 等. 持续暴雨作用下排土场层状碎石土边坡稳定性. *工程科学学报*, 2016, 38(9): 1204)