https://doi.org/10.3799/dqkx.2023.029



# 基于集成学习与贝叶斯优化的岩石抗压强度预测

吴禄源1,李建会1,马 丹2,王自法3,4,张建伟1\*,袁 超5,冯 义6,李 辉6

1. 河南大学土木建筑学院,河南开封 475004

2. 中国矿业大学煤炭资源与安全开采国家重点实验室,江苏徐州 221116

3. 中国地震局工程力学研究所,黑龙江哈尔滨 150080

4. 中震科建(广东)防灾减灾研究院,广东韶关 512026

5. 西安科技大学建筑与土木学院,陕西西安 710054

6. 中铁第一勘察设计院集团有限公司,陕西西安 710043

摘 要:岩石抗压强度是评估岩体工程稳定性的重要力学参数,传统统计回归方法对于岩石抗压强度预测存在一定的局限性.为此,提出了一种利用简单岩石力学参数实现岩石抗压强度智能预测的方法,首先收集了 620 组含不同类型岩石的三轴试验数据,然后分别采用随机森林(Random Forest,RF)、极限梯度提升树(XGBoost,XGB)和轻量梯度提升机(LightGBM,LGB)3种主流的集成学习算法建立了岩石抗压强度预测模型,使用贝叶斯优化算法在模型训练过程中进行超参数优化,最后利用决定系数(R<sup>2</sup>)、平均绝对百分比误差(MAPE)和均方根误差(RMSE)对优化后模型的泛化能力进行了综合评估和对比分析.此外,利用LGB模型对输入特征进行重要性分析,以评估不同输入特征对模型泛化性能的影响重要程度.研究结果表明:所建立的3种模型对岩石抗压强度均取得了较好的预测结果,其中LGB模型泛化性能的影响重要程度.研究结果表明:所建立的3种模型对岩石抗压强度均取得了较好的预测结果,其中LGB模型泛化性能的影响重要程度.研究结果表明:
RMSE=5.58,MAPE=9.70%),且运行耗时相对最少.弹性模量(E)、围压(σ<sub>3</sub>)和密度(ρ)对模型的泛化性能影响较大,泊松比(v)影响较小.提出的预测模型对于岩石抗压强度预测有良好的适用性,为机器学习与岩土工程的结合提供了新的思路.
关键词:岩石强度;集成学习;贝叶斯优化;随机森林;极限梯度提升树;轻量梯度提升机;工程地质.
中图分类号:P64

## Prediction for Rock Compressive Strength Based on Ensemble Learning and Bayesian Optimization

Wu Luyuan<sup>1</sup>, Li Jianhui<sup>1</sup>, Ma Dan<sup>2</sup>, Wang Zifa<sup>3,4</sup>, Zhang Jianwei<sup>1\*</sup>, Yuan Chao<sup>5</sup>, Feng Yi<sup>6</sup>, Li Hui<sup>6</sup>

1. School of Civil Engineering and Architecture, Henan University, Kaifeng 475004, China

 $2. State Key Laboratory of Coal \, Resources \, and \, Safe \, Mining \, , China \, University \, of \, Mining \, and \, Technology \, , Xuzhou \, \, 221116 \, , \, China \, University \, of \, Mining \, and \, Technology \, , Xuzhou \, \, 221116 \, , \, China \, University \, of \, Mining \, and \, Technology \, , Xuzhou \, \, 221116 \, , \, China \, University \, of \, Mining \, and \, Technology \, , Xuzhou \, \, 221116 \, , \, China \, University \, of \, Mining \, and \, Technology \, , Xuzhou \, \, 221116 \, , \, China \, University \, of \, Mining \, and \, Technology \, , Xuzhou \, \, 221116 \, , \, China \, University \, of \, Mining \, and \, Technology \, , Xuzhou \, \, 221116 \, , \, China \, University \, of \, Mining \, and \, Technology \, , Xuzhou \, \, 221116 \, , \, China \, University \, of \, Mining \, and \, Technology \, , Xuzhou \, \, 221116 \, , \, China \, University \, of \, Mining \, and \, Technology \, , Xuzhou \, \, 221116 \, , \, China \, University \, of \, Mining \, and \, Technology \, , Xuzhou \, \, 221116 \, , \, China \, University \, of \, Mining \, and \, Technology \, , Xuzhou \, \, 221116 \, , \, China \, University \, of \, Mining \, and \, Technology \, , Xuzhou \, \, 221116 \, , \, China \, University \, of \, Mining \, and \, Technology \, , Xuzhou \, \, 221116 \, , \, China \, University \, of \, Mining \, and \, Technology \, , Xuzhou \, \, 221116 \, , \, China \, University \, and \, Mining \, and \, Technology \, , Xuzhou \, \, 221116 \, , \, China \, University \, and \, Mining \, and \, Technology \, , Xuzhou \, \, 221116 \, , \, China \, University \, and \, Mining \, and \, Technology \, , Xuzhou \, \, 221116 \, , \, China \, University \, and \, Mining \, and \, Technology \, , Xuzhou \, \, China \, Mining \, , Xuzhou \, \, China \, Mining \, , Xuzhou \, Mining \, , Xuzhou \, \, China \, Mining \, , Xuzh$ 

4. Institute of Disaster Prevention and Mitigation of China Earthquake Engineering (Guangdong), Shaoguan 512026, China

6. China Railway First Survey and Design Institute Group Co. Ltd., Xi'an 710043, China

作者简介:吴禄源(1989-),男,博士,讲师,研究方向为岩石力学及岩土工程.ORCID:0000-0001-8403-9268. E-mail:wulymp@henu.edu.cn \* 通讯作者:张建伟, E-mail:zjw101\_0@163.com

<sup>3.</sup> Institute of Engineering Mechanics, China Earthquake Administration, Harbin 150080, China

<sup>5.</sup> College of Architecture and Civil Engineering, Xi' an University of Science and Technology, Xi' an 710054, China

基金项目:国家自然科学基金项目(Nos.41977238,51978634);河南省自然科学基金青年基金项目(No.232300421331);河南省高等学校重点科 研项目(No.23A440005).

**引用格式**:吴禄源,李建会,马丹,王自法,张建伟,袁超,冯义,李辉,2023.基于集成学习与贝叶斯优化的岩石抗压强度预测.地球科学,48(5): 1686-1695.

**Citation:** Wu Luyuan, Li Jianhui, Ma Dan, Wang Zifa, Zhang Jianwei, Yuan Chao, Feng Yi, Li Hui, 2023. Prediction for Rock Compressive Strength Based on Ensemble Learning and Bayesian Optimization. *Earth Science*, 48(5):1686-1695.

**Abstract:** Rock compressive strength is an important mechanical parameter to evaluate the stability of rock mass engineering. The traditional statistical regression method has some limitations on the prediction of rock compressive strength. To this end, in this paper it proposes a method for intelligent prediction of rock compressive strength using simple rock mechanics parameters. Firstly, 620 sets of triaxial test data containing different types of rocks were collected and preprocessed. Then, three main stream ensemble learning algorithms, Random forest, XGBoost and LightGBM, were used to establish a rock compressive strength prediction model, and Bayesian optimization algorithm was used to optimize the hyperparameters during model training. Finally, the coefficient of determination ( $R^2$ ), mean absolute percentage error (MAPE) and root mean square error (RMSE) were used to evaluate and compare the generalization ability of the optimized model. In addition, the importance of input features was analyzed by LGB model, to evaluate the importance of input features on the generalization ability of the LGB model is slightly better than that of the other two models ( $R^2 = 0.978$ , RMSE=5.58, MAPE=9.70%), and the running time is relatively minimum. Elastic modulus (E), confining pressure ( $\sigma_3$ ) and density ( $\rho$ ) have great influence on generalization ability of model, while Poisson's ratio(v) has little influence. The prediction model has good applicability to rock strength prediction, and provides a new idea for the combination of machine learning and geotechnical engineering.

Key words: rock strength; ensemble learning; Bayesian optimization; random forest; XGBoost; LightGBM; engineering geology.

## 0 引言

在实际岩体工程中,合理确定岩石的极限承载 力是从事设计、施工、支护的前提,岩石抗压强度是 岩石最基本和最重要的力学指标之一,可以很好地 评估岩石的承载能力(Bieniawski, 1974; Grima and Babuška, 1999; Gokceoglu, 2002). 因此, 准确、 经济和快速获取岩石抗压强度对于合理研究岩体 工程的稳定性和安全性具有重要意义(Culshaw, 2015; Jahed Armaghani, 2016). 对岩石进行不同围 压下的力学试验是获取岩石抗压强度最直接的方 式(Edelbro, 2003),然而力学试验往往对岩石试件 的制作标准要求较高,特别是对于裂隙岩石取样是 非常困难的,因此对岩石进行直接力学试验往往费 用较高且耗时耗力(Sarkar, 2010).许多研究人员 尝试用更简单、更经济的间接试验去获取岩石抗压 强度,如施密特锤试验(Goudie, 2006; Wang and Wan, 2019)、声波测试法(Çobanoğlu and Çelik, 2008; Wang, 2020) 和点荷载试验法等(Singh, 2012; 张春玲等, 2015), 这些方法所需仪器便携、 易操作可以对岩石强度进行快速的测定,已被广泛 用于各类岩石强度及其他力学参数的测定.此外, 国内外许多学者致力于从试验数据中挖掘出岩石 抗压强度与其力学参数之间的相关关系,并利用统 计回归的方法建立经验公式或回归模型,进而利用 简单的力学参数间接估计岩石的抗压强度(Cargill and Shakoor, 1990; Arjmandpour and Hosseinitoudeshki, 2013; 杨科等, 2013; Aladejare, 2020). 然而由于影响岩石抗压强度的因素具有极强的 非线性和不确定性,导致很难用一个特定的经验 公式或回归方程准确描述岩石抗压强度与其影 响因素间的关系.此外,函数形式和特征变量的 选取没有统一标准,多依赖研究人员的主观性, 其适用范围限制了相关研究的进一步发展.

近些年来随着人工智能理论的快速发展,许多 学者将善于处理复杂非线性问题和具有良好泛化 能力的机器学习方法用来研究工程领域问题(郭子 正等,2019;黄发明等,2021;Lietal.,2021;李文 彬等,2021; 仉文岗等,2021a,2023; 黄鑫怀等, 2022).机器学习可以从试验数据中挖掘变量之间 的隐藏关系,不需要预先假定方程形式,从而可以 突破传统经验回归模型局限性,为预测岩石抗压强 度提供一种新兴路径(Miah, 2020; Mahmoodzadeh, 2022). 李文和谭卓英(2016)采用径向基函数 (RBF)作为LS-SVM的核函数,在MATLAB环境 下建立了以岩石密度、点荷载强度、纵波波速作 为输入特征建立了页岩单轴抗压强度预测模型. Mohamad(2018)收集整理了38组包括岩石密度、 含水率、点荷载强度、纵波波速和岩石耐崩解性 指数的数据集,提出了基于粒子群优化算法和 BP神经网络的砂岩单轴抗压强度预测模型;He (2019)以钻井数据、粘聚力和内摩擦角作为输入 特征,采用深度卷积网络算法对岩石单轴抗压 强度进行了预测.以上研究均展现出了机器学 习构建岩石抗压强度预测模型的优势,总结前 人所建立的机器学习模型主要存在以下限制:

(1)所建立模型多针对于一种类型岩石的 强度预测,而不同类型岩石差异性较大,其适用 范围受到了限制.

(2)包含小样本数据集的机器学习模型虽然预测结果相对较好,但泛化能力有待进一步提高.

(3)自然界中的岩石在复杂的地质构造运动 作用下将产生不同形式的裂隙,所建立的机器学 习模型多针对于完整岩石的强度预测,对含裂隙 岩石强度预测模型的研究较少.

相较于传统的神经网络算法,集成学习算法具 有建模简便以及学习效率高等优点(Sagi and Rokach, 2018).因此,针对不同类型岩石,本文首 先收集了大量的完整岩石及裂隙岩石试验数据,并 基于3种主流集成算法建立了岩石抗压强度预测 模型,在模型训练过程中使用贝叶斯优化算法完成 超参数优化,然后利用多项评价指标对模型的泛化 能力进行了评估,最后对输入特征进行了重要性 分析以评估各特征对模型泛化性能的影响重要程 度,对岩石力学性能研究提供一定的参考价值.

## 1 数据准备

## 1.1 数据来源

具有良好泛化能力的机器学习模型依赖于大 量且高质量的数据,本文通过室内压缩试验与文献 调研相结合的方式对岩石试验数据进行了收集整 理.试验数据来源见附图1所示,其中室内压缩试验 所用试样取自于陕西省宝鸡市园子沟煤矿现场,通 过地质勘探孔获取不同粒径大小的砂岩岩心,根据 组成沙粒粒径大小不同可分为粗砂岩、中砂岩和细 砂岩3种.然后按照《工程岩体分级标准》(GB/ T50218-2014)规程要求加工成标准圆柱体试件,对 试件进行密度测试后采用MTS 815电液伺服试验 系统进行不同围压下的加载试验,该试验系统可自 动完成数据的采集与处理.以此,收集整理得到137 组包含岩石密度 $(\rho)$ 、弹性模量(E)、泊松比(v)、围  $\mathbb{E}(\sigma_3)$ 、轴向峰值应变( $\varepsilon_{10}$ )和抗压强度( $\sigma_{10}$ )特征变 量的完整岩石三轴压缩试验数据.为确保机器学习 模型具有良好的泛化能力,通过公开发表的文献获 取相关试验数据,整理得到483组包括花岗岩、大理 岩和泥岩等多种岩石类型的完整岩石和裂隙岩石 试验数据,在完整岩石特征变量基础上以裂隙长度  $(F_{L})$ 、裂隙数量 $(F_{N})$ 和裂隙倾角 $(F_{a})$ 作为裂隙岩石 的特征变量.基于以上两种方式,共收集整理得到 620组含多种岩石类型的试验数据,其中完整岩石 328组,裂隙岩石 292组,每组数据含9个特征 变量,以ρ,E,v,σ<sub>3</sub>,ε<sub>1</sub>,F<sub>L</sub>,F<sub>N</sub>,F<sub>a</sub>作为模型的输入特征,σ<sub>1</sub>,作为模型的输出特征.

### 1.2 数据统计分析

数据来源多样,不同类型及不同地点的岩石 特征参数具有一定差异性,因此需要了解试验数 据的分布、特征参数之间相关性和离散情况.数 据统计分析结果见附图2所示,右上为矩阵散点 图,其中直方图代表数据分布特征情况,从附图2 中可以看出 $\rho$ 、v、E、 $\varepsilon_{lp}$ 和 $\sigma_{lp}$ 符合正态分布,利用 皮尔逊系数( $P_{ces}$ )衡量特征之间的相关性,可以 看出 $\rho$ 和E, $\sigma_{lp}$ 具有较高的线性相关性,E和 $\sigma_{lp}$ 具有更强的线性相关性,其中 $F_L$ 、 $F_N$ 和 $F_a$ 与岩 石抗压强度存在不同程度的负线性相关性,其 他特征之间存在较弱的相关性.饼状图表示了 数据集中不同类型岩石所占比例,表格中显示 了数据集的特征统计情况,限于篇幅,本文所收 集的数据集将以补充材料的方式进行展示.

## 2 集成学习原理

集成学习通过构建多个基学习器进行学习 并采用某种集成策略将基学习器训练结果进行 结合来完成回归或分类任务,相较于单一的学 习器,集成学习往往能获得更为显著优越的泛 化性能(Zhou, 2021).按照基学习器间是否存在 强依赖关系可将集成学习大致分为两类,即存 在强依赖关系的串行序列化方法和不存在强依 赖关系的并行法,两者的代表分别是提升法 (Boosting)和装袋法(Bagging)(周志华, 2016). 本文分别采用基于Bagging集成思想的RF算法 以及基于Boosting 集成思想XGBoost和Light GBM算法建模然后进行验证比较,旨在获得一 种具有良好泛化性能的岩石抗压强度预测模型. 2.1 随机森林

随机森林(Random Forest, RF)是 Breiman 在 2001年提出的一种以决策树为基学习器的集成学 习算法,它改善了单一决策树的不足,使得整体模 型的结果具有较高的精确性和稳定性(Breiman, 2001).决策树是由根节点、决策节点和叶子节点构 成的树形数据结构,决策树遍历每个根部节点待决 策样本的属性值然后与决策节点值进行比较,以判 定该样本的分枝走向,直至持续达到某个叶子节点



Fig.1 Schematic diagram of random forest algorithm

已无样本可分时,代表决策过程的完成,所对应的 叶子节点即为该样本的最终决策结果.如图1所 示,RF在训练决策树的过程中,利用Bootstrap重 采样法从训练集S中随机抽取n个独立同分布的 训练子集,并根据不同的训练子集生成对应n棵 不同的决策树.在对决策树每个节点进行分裂时, RF随机选取样本的特征属性,然后利用Gini系数 或均方误差(MSE)来衡量特征属性的不纯度指 标以寻求最优属性用于分裂,并设置节点阈值和 满足停止分裂的标准,最后基于Bagging集成思 想将每棵决策树的训练结果相结合,通过投票或 者取均值的方法决定最后的分类或回归结果.

## 2.2 极端梯度提升树(XGBoost)

极端梯度提升树(XGBoost,XGB)是陈天奇在 2014年提出的一种基于梯度提升决策树(Gradient Boosting Decision Tree,GBDT)的改进优化算法 (Chen and Guestrin,2016).XGB与传统的GBDT 相同点在于两者都是基于Boosting集成思想的串 行序列化算法,基于上一棵决策树预测结果与实测 值之间的残差对决策树进行训练,通过多次迭代拟 合残差从而不断地逼近实测值,达到设定的阈值或 者迭代次数后训练停止,将各个决策树的预测结果 进行加权求和后就是这个样本在整个集成模型上 的最终预测结果.不同点在于求解损失函数时 XGB使用泰勒二阶展开用于简化计算,并在目标 函数中加入了正则化项用于控制模型的复杂度, 其计算式见式(1).相较于传统的GBDT算法, XGB算法具有更高的运算效率和防过拟合能力.

 $Obj = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k), \quad (1)$ 式中, *Obj*代表目标函数;等式右边第1项为损失函 数, *y<sub>i</sub>*和 *ŷ<sub>i</sub>*分别为样本*i*的实测值和预测值;等式右 边第2项为惩罚函数,K代表决策树的总数量,f<sub>k</sub>表示第k棵树的复杂度.

## 2.3 轻量梯度提升机

轻量梯度提升机(LightGBM,LGB)在2017年 由微软公司提出,与XGBoost原理类似,也是基于 GBDT算法的改进优化算法(Ke,2017).不同之处 在于,LGB利用直方图算法(Histogram)将连续的 特征值离散化到多个分箱(bin)中,并根据特征所 在的bin对其进行梯度累加及个数统计,最后遍历 所有的特征和数据寻找最优分割点,这样可以解决 分裂点数量过多的问题,提高决策树的训练效率以 及减少内存的占用.同时,LGB使用单边梯度采 样算法和互斥特征捆绑算法分别解决了样本数 量和特征数量过多的问题,进一步提高了模型的 运算效率.此外,在Histogram算法基础上,使用 了带有深度限制的叶子生长(leaf-wise)算法,每 次只在分裂增益最大的叶子节点进行分裂,在降 低模型复杂度的同时也提高了抗过拟合能力.

## 3 岩石强度回归预测模型

#### 3.1 模型训练流程

本研究采用基于 Python 语言的 Scikit-learn 机 器学习库作为模型建立的基础工具,其涵盖了多种 主流的机器学习算法,与 NumPy、Pandas 和 Matplotlib等 Python数据分析处理库相结合可以简单高 效地实现数据切分、数据预处理、模型建立及结果 可视化等环节.本研究所提出的岩石抗压强度预测 模型实现流程如下:Step1:指定岩石强度试验数据 文件所在路径,读取数据;Step2:对数据进行切片处 理,指定模型的输入特征及输出特征;Step3:设置模 型中训练集和数据集的所占比例,对数据集进行划 分;Step4:调用 Scikit-learn 库中的 RandomForestRegressor、XGBRFRegressor、lightgbm 函数,并 设置模型初始超参数,建立3种初始预测模型; Step5:对预测模型进行训练,利用贝叶斯优化 算法对模型进行超参数优化,寻找最优超参数 组合;Step6:调整超参数后获得最优预测模型, 对测试集进行预测并打印预测结果;Step7:通 过统计指标对模型预测结果进行评估,对3种 预测模型综合性能进行对比分析;Step8:利用 最优预测模型对输入特征进行重要性分析,评 估备特征对模型性能的重要性;Step9:调用 pickle 函数保存最优预测模型的权重函数.

#### 3.2 模型建立

模型建立前需要预先将数据集划分为训练集 和测试集,训练集用于模型的训练,测试集用于验 证与评估已训练好模型的可靠性及泛化性能(汤志 立和徐千军,2020).合理划分数据集对于模型最终 泛化性能有重要影响,当训练集比例过高而测试集 比例较少时,会导致模型出现过拟合,即在训练集 表现出较好的泛化性能而在测试集上表现较差,当 训练集比例较少而测试集比例过高时,又会导致模 型出现欠拟合,在训练集和测试集都表现出较差的 泛化性能.训练集和测试集的划分比例为7:3和8:2 时较为常见(仉文岗等, 2021b; 黄发明等, 2023), 经过验证对比后发现按照8:2比例对数据集进行划 分时泛化性能相对较好,因此本文将数据集按照8: 2的比例划分为训练集(496组)与测试集(124组)分 别用作模型的训练与验证.且为了保证训练集和测 试集之间相互独立而无任何交集,采用 Scikit-learn 中的train\_test\_split函数对数据集进行随机划分.

以通过实验和文献调研方式得到的 620 组试 验数据为基础,建立岩石抗压强度预测数据集,然 后对数据集进行划分,以ρ,E,v,ε<sub>1</sub>,σ<sub>3</sub>,F<sub>L</sub>,F<sub>N</sub>,F<sub>a</sub> 作为模型的输入特征,σ<sub>1</sub>,作为模型的输出特征,通 过 Scikit-learn 调用机器学习包,建立基于 RF、 XGB和LGB三种集成算法的岩石强度回归预测 模型,其中数据集以.xls格式存储,采用 Pandas 库 中的 read\_excel()函数读取数据并写入到模型之 中.考虑到以决策树为基学习器的集成学习算法 对特征缩放的不敏感性,本文模型未对输入特征 进行归一化或标准化处理.限于篇幅,本文所建立 的预测模型源代码将以补充材料方式进行展示. 表1 基于贝叶斯优化后的3种模型最优超参数和运行耗时

 
 Table 1
 Optimal super parameters and running time of three models based on Bayesian optimization

模型	超参数名称	最优取值	运行耗时
			(s)
Random Forest	n_estimators	915	
	max_depth	13	
	max_features	7	13.625
	min_samples_split	2	
	min_samples_leaf	1	
LightGBM	n_estimators	615	
	max_depth	17	
	gamma	26.251	
	lambda	0.849	5.383
	alpha	24.513	
	num_leaves	82	
	learning_rate	0.201	
XGBoost	n_estimators	985	
	max_depth	3	
	gamma	25.802	24.558
	lambda	2.881	
	alpha	8.830	
	learning_rate	0.185	

#### 3.3 模型超参数优化

机器学习模型含有较多超参数,不同的超参数 组合对模型的运算效率和泛化性能均有较大影响, 因此为获得一个兼具良好运算效率和泛化性能的 模型,优化超参数是必要的.常用的超参数优化方 法包括网格搜索、随机搜索和贝叶斯优化算法等 (李亚茹等,2022).网格搜索是在给定的超参数范围 内,通过遍历循环查找超参数空间中网格上的所有 点来确定一组最优值作为模型的最终超参数组合, 但该方法在多参数和大数据集的情况下非常消耗 时间和计算资源.随机搜索的思想与网格搜索比 较相似,它通过在搜索范围中随机取样本点来寻 找全局最优点,节省了超参数优化时间,但无法确 保得到最佳的超参数组合.贝叶斯优化(Bayesian Optimization, BO)是一种十分有效的全局优化算 法(崔佳旭和杨博, 2018),与网格搜索和随机搜 索构建目标函数不同,BO利用有限的样本点构 造代理函数近似替代黑箱目标函数,降低了计 算目标函数的时间成本,并引入了采集函数,对 损失函数最小的评估点进行有效探索和利用, 能在较少的评估次数下取得复杂目标函数的最 优解,从而可以大大提高超参数的优化效率.

表 2 预测模型评价指标				
Table 2Evaluation index of prediction model				
评价指标	计算公式	评判标准		
决定系数	$R^{2} = 1 - \frac{\sum_{i=1}^{n} (y_{i} - \hat{y}_{i})^{2}}{\sum_{i=1}^{n} (y_{i} - \bar{y}_{i})^{2}}$	R <sup>2</sup> 数值越大,模型性能越好		
均方根误差	$\text{RMSE} = \sqrt{\frac{1}{n} \times \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$	RMSE数值越小,模型性能越好		
平均绝对百分比误差	$\text{MAPE} = \sum_{i=1}^{n} \left  \frac{y_i - \hat{y}_i}{y_i} \right  \times \frac{100\%}{n}$	MAPE 数值越小,模型性能越好		

注:式中,n为预测样本的数量, yi和 ŷi分别表示实测值 yi的平均值和预测值.

本文使用贝叶斯优化算法对已建立的模型进行超参数优化,设置参数空间对指定的超参数进行搜索,以寻求最优超参数组合.其最优参数及模型运行耗时见表1所示,LGB模型运行耗时最少,其次是RF模型,XGB模型耗时最长,但也仅需数十秒即可完成模型的运行,由此可见,利用贝叶斯优化算法的机器学习模型具有较高的训练效率.除本文所提到的超参数之外,模型的其他超参数取为默认值,将优化后的超参数代入模型之中即可得到训练完成的抗压强度预测模型.

## 3.4 模型性能评估

为评估所建3种预测模型的泛化性能,采用决 定系数(R<sup>2</sup>)、平均绝对百分比误差(MAPE)和均方 根误差(RMSE)作为衡量模型泛化能力的评价指 标.相应计算公式见表2所示,R<sup>2</sup>表示模型拟合数 据的准确程度,其值越接近1表明模型的拟合效果 越好.RMSE反映模型预测值与实测值的偏离程 度,MAPE表示预测值和实测值之间绝对误差的 平均值,两者数值越小表明模型的泛化性能越好.

图 2 显示了 3 种模型预测值和实测值的比较 结果,图中白色和红色散点分别表示模型对训练 集和测试集中样本的预测结果,前者可以反映模 型对已知数据中复杂非线性关系的学习能力,后 者反映模型对未知数据预测的泛化能力.从图中 可以看出 3 种模型的散点皆均匀分布在斜率为1 的参考线附近,表明所建模型对岩石抗压强度都 取得了较好的预测结果.从评价指标数值来看, RF模型在训练集的表现结果优于LGB和XGB模 型,其 R<sup>2</sup>、RMSE和MAPE分别为:0.991、3.41和 4.89%,其次分别是LGB和XGB模型,表明RF在 训练集上的学习能力略优于另外两种模型.对于 测试集预测结果,LGB模型各项评价指标均高于 RF模型和 XGB 模型,其 R<sup>2</sup>和 RMSE 分别为 0.978 和 5.58,其整体平均误差在 10% 以内,其次分别 是 RF 模型和 XGB 模型, RF 模型略优于 XGB 模 型,两者 R<sup>2</sup>分别达到了 0.961 和 0.947, RMSE 分 别为 7.34 和 10.4,两者 MAPE 较为接近,整体平 均误差在 15% 左右,分别为 14.65% 和 15.7%.

综合运行耗时和泛化能力来看,相较于 RF 模型和 XGB 模型,基于 LGB 算法的岩石抗压强度预测模型准确度高、稳定性好、具有最高的运行效率和泛化能力,表明 LGB 算法在处理高维度数据问题时具有较大优势.RF 模型和 XGB 模型预测精度较为接近,但其运行耗时优于 XGB 模型.

#### 3.5 特征重要性分析

集成学习算法在决策树叶节点分裂时利用基 尼系数衡量特征的重要性来确定最优分裂特征,为 增加模型的可解释性,利用LGB模型对输入特征进 行重要性分析,以评估各特征对模型泛化性能的影 响重要程度.特征重要性分析结果见图3所示,对于 完整岩石而言,E是影响模型泛化性能最重要的因 素,ρ和σ<sub>3</sub>两者重要程度较为接近,是影响模型泛化 性能的关键因素,ε<sub>1</sub>次之,v对模型精度影响最小, 最终特征相对重要性排序为: $E > \rho > \sigma_3 > \varepsilon_{lp} > v,$ 研 究结果与前文数据统计分析结果较为吻合,进一步 验证了模型的可靠性.对于裂隙岩石而言,其E、p 和σ<sub>3</sub>的相对重要性和完整岩石较为一致,三者对裂 隙岩石强度的预测精度起着决定性作用,其次分别 是 $F_L$ 和 $F_a$ ,两者的重要性优于 $\varepsilon_{lp}$ 和v,从上文相关 性分析结果来看,两者与岩石抗压强度存在着较弱 的负线性相关性,但对预测结果具有较大影响,表 明FL和F。与岩石抗压强度之间存在着复杂的非线 性关系,F<sub>N</sub>影响重要程度与数据统计分析结果偏差 较大,分析其原因可能是本文所收集的292组裂隙



Fig.2 Comparison between predicted values and measured values under different models



Fig.3 Results of feature importance analysis

岩石数据中多为单裂隙和双裂隙试样,少部分为多裂隙试样,数据样本的不均衡弱化了该特征之间的

差异性,因此利用LGB模型进行重要性分析时F<sub>N</sub> 相较于其他特征相对重要性最低,最终特征相对重 要性排序为: $E > \sigma_3 > \rho > F_L > F_a > \varepsilon_{1p} > v > F_N$ .

## 4 结论

(1)通过室内试验和文献调研相结合的方式整 理 620组包含多种岩石类型的完整和裂隙岩石试验 数据,对影响因素之间进行相关性分析发现,弹性 模量、密度和围压对岩石抗压强度存在较强的线性 关系,裂隙长度、裂隙数量和裂隙倾角与岩石抗压 强度均呈现不同程度的负相关性,而各特征之间相 互作用、相互耦合存在着较为复杂的非线性关系.

(2)利用贝叶斯优化算法对训练的模型进行超参数优化,提高训练效率的同时也保证了模型具有较高的预测精度.利用多项评价指标对3种模型泛化性能评估,其中LGB模型在各项评价指标上均优于RF和XGB模型,且运行耗时最少,综合性能排序为:LGB>RF>XGB.

(3)利用 LGB 模型对输入特征进行重要性 分析以评估各个特征对模型泛化性能的影响 重要程度,其中完整岩石的特征重要性排序 为: $E > \rho > \sigma_3 > \varepsilon_{1p} > v$ ,裂隙岩石的特征重要性 排序为: $E > \sigma_3 > \rho > F_L > F_a > \varepsilon_{1p} > v > F_N$ .

综上所述,本文将机器学习与岩土工程相结 合,利用简单力学参数实现了岩石强度的智能预 测,相较于传统的试验方法,节约了大量的人力、财 力及时间,为岩石强度预测提供了一条可供参考的 新思路,此外,特征重要性分析结果也可为岩石力 学性能研究提供一定的参考价值.考虑到影响岩石 强度的因素较多,除文本所提及到的之外还有孔隙 率、含水率等特征,将来可考虑收集更多的数据及 特征用作模型训练,进一步提高模型的泛化性能.

附图见本刊官网(http://www.earth-science.net).

#### References

- Aladejare, A. E., 2020. Evaluation of Empirical Estimation of Uniaxial Compressive Strength of Rock Using Measurements from Index and Physical Tests. Journal of Rock Mechanics and Geotechnical Engineering, 12(2): 256-268. https://doi.org/10.1016/j.jrmge.2019.08.001
- Arjmandpour, J., Hosseinitoudeshki, V., 2013. Estimation of Tensile Strength of Limestone from Some of Its Physical Properties via Simple Regression. *Journal of Novel Applied Sciences*, 2: 1041-1044.

Bieniawski, Z. T., 1974. Estimating the Strength of Rock

Materials. Journal of the South African Institute of Mining and Metallurgy, 74(8): 312-320.

- Breiman, L., 2001. Random Forests. *Machine Learning*, 45: 5-32.
- Cargill, J. S., Shakoor, A., 1990. Evaluation of Empirical Methods for Measuring the Uniaxial Compressive Strength of Rock. *International Journal of Rock Mechanics and Mining Sciences*, 27(6): 495-503. doi: https:// doi.org/10.1016/0148-9062(90)91001-N
- Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, 785-794. https://doi. org/10.1145/2939672.2939785
- Çobanoğlu, İ., Çelik, S. B., 2008. Estimation of Uniaxial Compressive Strength from Point Load Strength, Schmidt Hardness and P-Wave Velocity. Bulletin of Engineering Geology and the Environment, 67: 491-498.
- Cui, J.X., Yang, B., 2018. Survey on Bayesian Optimization Methodology and Applications. *Journal of Software*, 29 (10): 3068-3090 (in Chinese with English abstract).
- Culshaw, M.G., 2015. The ISRM Suggested Methods for Rock Characterization, Testing and Monitoring: 2007 – 2014. Bulletin of Engineering Geology and the Environment, 74: 1499–1500. https://doi.org/10.1007/978-3-319-007713-0
- Edelbro, C., 2003. Rock Mass Strength: A Review. Department of Civil Engineering Division of Rock Mechanics, Beijing.
- Gokceoglu, C., 2002. A Fuzzy Triangular Chart to Predict the Uniaxial Compressive Strength of the Ankara Agglomerates from Their Petrographic Composition. *Engineering Geology*, 66: 39-51. https://doi.org/10.1016/ S0013-7952(02)00023-6
- Goudie, A.S., 2006. The Schmidt Hammer in Geomorphological Research. Progress in Physical Geography, 30: 703-718. doi: https://doi.org/10.1177/ 0309133306071954
- Grima, M.A., Babuška, R., 1999. Fuzzy Model for the Prediction of Unconfined Compressive Strength of Rock Samples. International Journal of Rock Mechanics and Mining Sciences, 36: 339-349. doi: https://doi.org/ 10.1016/S0148-9062(99)00007-8
- Guo, Z. Z., Yin, K. L., Fu, S., et al., 2019. Evaluation of Landslide Susceptibility Based on GIS and WOE - BP Model. *Earth Science*, 44(12): 4299-4312 (in Chinese with English abstract).
- He, M., 2019. Deep Convolutional Neural Network for Fast

Determination of the Rock Strength Parameters Using Drilling Data. *International Journal of Rock Mechanics and Mining Sciences*, 123: 104084. https://doi.org/ 10.1016/j.ijrmms.2019.104084

- Huang, F. M., Cao, Y., Fan, X. M., et al., 2021. Effects of Different Landslide Boundaries and Their Spatial Shapes on the Uncertainty of Landslide Susceptibility Prediction. *Chinese Journal of Rock Mechanics and Engineering*, 40(S02): 3227-3240 (in Chinese with English abstract).
- Huang, F.M., Chen, B., Mao, D.X., et al., 2023. Landslide Susceptibility Prediction Modeling and Interpretability Based on Self-Screening Deep Learning Model. *Earth Science*, 48(5): 1696-1710 (in Chinese with English abstract).
- Huang, X.H., Li, Z.H., Deng, T., et al., 2022. Uranium Potential Evaluation of the Zhuguangshan Granitic Pluton in South China Based on Machine Learning. *Earth Science*, 1–23 (in Chinese with English abstract).
- Jahed Armaghani, D., 2016. Application of Several Non-Linear Prediction Tools for Estimating Uniaxial Compressive Strength of Granitic Rocks and Comparison of Their Performances. *Engineering with Computers*, 32: 189-206.
- Ke, G., 2017. Light GBM: A Highly Efficient Gradient Boosting Decision Tree. Advances in Neural Information Processing Systems, New Orleans, 30.
- Li, S., Chen, J., Liu, C., et al., 2021. Mineral Prospectivity Prediction via Convolutional Neural Networks Based on Geological Big Data. *Journal of Earth Science*, 32(2): 327-347. https://doi.org/10.1007/s12583-020-1365-z
- Li, W., Tan, Z.Y., 2016. Comparison on Rock Strength Prediction Models Based on MLR and LS-SVM. *Mining Research and Development*, 36(11): 36-40 (in Chinese with English abstract).
- Li, W. B., Fan, X. M., Huang, F. M., et al., 2021. Uncertainties of Landslide Susceptibility Modeling under Different Environmental Factor Connections and Prediction Models. *Earth Science*, 46(10): 3777-3795 (in Chinese with English abstract).
- Li, Y. R., Zhang, Y. L., Wang, J. C., 2022. Survey on Bayesian Optimization Methods for Hyper - Parameter Tuning. *Computer Science*, 49(S01): 86-92 (in Chinese with English abstract).
- Mahmoodzadeh, A., 2022. Machine Learning Techniques to Predict Rock Strength Parameters. *Rock Mechanics and Rock Engineering*, 55: 1721-1741.
- Miah, M. I., 2020. Machine Learning Approach to Model

Rock Strength: Prediction and Variable Selection with Aid of Log Data. *Rock Mechanics and Rock Engineering*, 53: 4691-4715.

- Mohamad, E.T., 2018. Rock Strength Estimation: A PSO-Based BP Approach. Neural Computing and Applications, 30: 1635-1646.
- Sagi, O., Rokach, L., 2018. Ensemble Learning: A Survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8: e1249. https://doi.org/ 10.1002/widm.1249
- Sarkar, K., 2010. Estimation of Strength Parameters of Rock Using Artificial Neural Networks. Bulletin of Engineering Geology and the Environment, 69: 599-606.
- Singh, T., 2012. Correlation between Point Load Index and Uniaxial Compressive Strength for Different Rock Types. Rock Mechanics and Rock Engineering, 45: 259-264.
- Tang, Z. L., Xu, Q. J., 2020. Rockburst Prediction Based on Nine Machine Learning Algorithms. *Chinese Journal* of Rock Mechanics and Engineering, 39(4): 773-781 (in Chinese with English abstract).
- Wang, M., Wan, W., 2019. A New Empirical Formula for Evaluating Uniaxial Compressive Strength Using the Schmidt Hammer Test. International Journal of Rock Mechanics and Mining Sciences, 123: 104094. https:// doi.org/10.1016/j.ijrmms.2019.104094
- Wang, R., 2020. Application of Ultrasonic-Rebound Method in Fast Prediction of Rock Strength. *Geotechnical and Geological Engineering*, 38: 5915-5924.
- Yang, K., Yuan, L., Qi, L. G., et al., 2013. Establishing Predictive Model for Rock Uniaxial Compressive Strength of No. 11-2 Coal Seam Roof in Huainan Mining Area. *Chinese Journal of Rock Mechanics and Engineering*, 32(10): 1991-1998 (in Chinese with English abstract).
- Zhang, C. L., Zhang, C. P., Xu, J., 2015. Comparison Test of Rock Point Load Strength and Uniaxial Compressive Strength. *Chinese Journal of Underground Space and Engineering*, 11(S2): 447-451 (in Chinese with English abstract).
- Zhang, W.G., He, Y.W., Wang, L.Q., et al., 2023. Machine Learning Solution for Landslide Susceptibility Based on Hydrographic Division: Case Study of Fengjie County in Chongqing. *Earth Science*, 48(5):2024-2038 (in Chinese with English abstract).
- Zhang, W. G., Li, H. R., Wu, C. Z., et al., 2021a. Stability Assessment of Underground Entry-Type Excavations Using Data-Driven RF and KNN Methods. *Journal of*

Hunan University (Natural Sciences), 48(3): 164-172 (in Chinese with English abstract).

- Zhang, W. G., Tang, L. B., Chen, F. Y., et al., 2021b. Prediction for TBM Penetration Rate Using Four Hyperparameter Optimization Methods and Random Forest Model. *Journal of Basic Science and Engineering*, 29 (5): 1186-1200 (in Chinese with English abstract).
- Zhou, Z.H., 2016. Machine Learning. Tsinghua University Press, Beijing, 173(in Chinese).
- Zhou, Z.H., 2021. Ensemble Learning, Machine Learning. Springer, Berlin, 181-210.

## 附中文参考文献

- 崔佳旭,杨博,2018.贝叶斯优化方法和应用综述.软件学 报,29(10):3068-3090.
- 郭子正,殷坤龙,付圣,等,2019.基于GIS与WOE-BP模型的滑坡易发性评价.地球科学,44(12):4299-4312.
- 黄发明,曹昱,范宣梅,等,2021.不同滑坡边界及其空间形 状对滑坡易发性预测不确定性的影响规律.岩石力学 与工程学报,40(S02):3227-3240.
- 黄发明,陈彬,毛达雄,等,2023.基于自筛选深度学习的滑坡 易发性预测建模及其可解释性.地球科学,48(5): 1696-1710.
- 黄鑫怀,李增华,邓腾,等,2022. 基于机器学习的华南诸广山 花岗岩体铀矿潜力评价. 地球科学,1-23.

- 李文,谭卓英,2016. 基于 MLR 与 LS-SVM 的岩石强度预 测模型比较.矿业研究与开发,36(11):36-40.
- 李文彬,范宣梅,黄发明,等,2021.不同环境因子联接和预 测模型的滑坡易发性建模不确定性.地球科学,46 (10):3777-3795.
- 李亚茹,张宇来,王佳晨,2022.面向超参数估计的贝叶斯 优化方法综述.计算机科学,49(S01):86-92.
- 汤志立,徐千军,2020.基于9种机器学习算法的岩爆预测 研究.岩石力学与工程学报,39(4):773-781.
- 杨科,袁亮,祁连光,等,2013.淮南矿区11-2煤顶板岩石单 轴抗压强度预测模型构建.岩石力学与工程学报,32 (10):1991-1998.
- 张春玲,张传鹏,徐静,2015.岩石点荷载强度与单轴抗压强度的对比试验.地下空间与工程学报,11(S2): 447-451.
- 仉文岗,何昱苇,王鲁琦,等,2023.基于水系分区的滑坡易发 性机器学习分析方法——以重庆市奉节县为例.地球 科学:48(5):2024-2038.
- 仉文岗,李红蕊,巫崇智,等,2021a.基于RF和KNN的地 下采场开挖稳定性评估.湖南大学学报(自然科学版), 48(3):164-172.
- 仉文岗,唐理斌,陈福勇,等,2021b.基于4种超参数优化 算法及随机森林模型预测TBM 掘进速度.应用基础与 工程科学学报,29(5):1186-1200.
- 周志华,2016.机器学习.北京:清华大学出版社,173.