

引用本文:张欣怡,戴成元,李微雨,等. 基于 TPE-XGBoost 算法的再生粗骨料混凝土抗压强度预测模型[J]. 建筑科学与工程学报, 2024, 41(6):100-110.

ZHANG Xinyi, DAI Chengyuan, LI Weiyu, et al. Prediction model of compressive strength of recycled coarse aggregate concrete based on TPE-XGBoost algorithm[J]. Journal of Architecture and Civil Engineering, 2024, 41(6):100-110.

DOI:10.19815/j.jace.2022.10013

基于 TPE-XGBoost 算法的再生粗骨料混凝土抗压强度预测模型

张欣怡¹, 戴成元^{1,2}, 李微雨¹, 陈 阳¹, 刘 兵¹

(1. 桂林理工大学 土木工程学院, 广西 桂林 541004;

2. 桂林理工大学 广西建筑新能源与节能重点实验室, 广西 桂林 541004)

摘要:为了更好地预测再生粗骨料混凝土的抗压强度,提出了基于极限提升树(XGBoost)算法的再生粗骨料混凝土抗压强度预测模型;利用再生粗骨料混凝土数据库,对数据进行预处理,利用树结构概率密度估计贝叶斯优化(TPE-BO)方法优化模型参数;通过实例对再生粗骨料混凝土抗压强度预测模型进行对比验证。结果表明:数据预处理和 TPE-BO 超参数优化方法均能在一定程度提升模型性能;与随机森林、K 邻近回归、支持向量机回归、梯度提升决策树模型相比,提出的模型有更高的预测精度和泛化能力;高性能抗压强度预测模型可为再生粗骨料混凝土的研究和实践提供依据,同时也为再生混凝土性能预测提供新途径。

关键词:XGBoost 算法;再生粗骨料混凝土;抗压强度;贝叶斯优化

中图分类号:TU502

文献标志码:A

文章编号:1673-2049(2024)06-0100-11

Prediction model of compressive strength of recycled coarse aggregate concrete based on TPE-XGBoost algorithm

ZHANG Xinyi¹, DAI Chengyuan^{1,2}, LI Weiyu¹, CHEN Yang¹, LIU Bing¹

(1. School of Civil Engineering, Guilin University of Technology, Guilin 541004, Guangxi, China;

2. Guangxi Key Laboratory of New Energy and Building Energy Saving, Guilin University of Technology, Guilin 541004, Guangxi, China)

Abstract: In order to better predict the compressive strength of recycled coarse aggregate concrete, a compressive strength prediction model for recycled coarse aggregate concrete based on extreme gradient boosting (XGBoost) algorithm was proposed. Taking the recycled coarse aggregate concrete database as the research data set, the data set was preprocessed, and the Bayesian optimization (BO) method was used to estimate the tree-structured parzen estimator (TPE) to optimize the model parameters. The comparative verification of compressive strength prediction models for recycled coarse aggregate concrete was carried out through examples. The results show that data preprocessing and TPE-BO hyperparameter optimization methods can both

收稿日期:2023-11-07

基金项目:国家自然科学基金项目(52108201);广西自然科学基金项目(2021GXNSFBA220049);

广西建筑新能源与节能重点实验室基金项目(桂科能 22-J-21-28)

通信作者:戴成元(1974-),男,副教授,E-mail:dcy366@126.com.

improve model performance to a certain extent. Compared with random forest, K-nearest neighbor regression, support vector machine regression, and gradient boosting decision tree models, the proposed model has higher prediction accuracy and generalization ability. The high performance compressive strength prediction model provides a basis for the research and practice of recycled coarse aggregate concrete, and also provides a new approach for predicting the performance of recycled concrete.

Key words: XGBoost algorithm; recycled coarse aggregate concrete; compressive strength; Bayesian optimization

Author resume: DAI Chengyuan(1974-), male, associate professor, E-mail: dcy366@126.com.

0 引言

随着建筑业可持续发展意识的觉醒,绿色低碳发展成为行业共识。近几十年来,中国经济和城市化高速发展,新建筑的建设和老旧建筑的拆除显著增加。在城市化进程中,自然资源逐渐短缺和建筑垃圾日益增多成为亟需解决的问题。中国每年产生的城市建筑垃圾超过 20×10^8 t。主要处理方式为外运、填埋和露天堆放,不仅占用土地资源,而且容易造成各种环境问题。因此,探索建筑垃圾处理的新途径是建筑业绿色转型的关键。

混凝土作为使用最广泛的建筑材料之一,其天然骨料的过度开采会导致自然资源的枯竭,废弃混凝土更是会造成土地资源浪费,严重污染环境。因此,构建混凝土原料→产品→废料→再生的闭环,将废弃混凝土加工成再生原料替代砂石等天然粗骨料是解决上述问题的有效途径^[1]。国内外诸多学者研究表明再生粗骨料和天然粗骨料性能存在显著差异^[2],其中抗压强度是影响再生粗骨料混凝土承载力和耐久性的关键指标。由于再生粗骨料来源比较复杂,对于再生粗骨料混凝土抗压强度得出的结论存在一定差异。在任意再生粗骨料替代率下,再生粗骨料混凝土抗压强度均低于普通混凝土^[3-4]。当再生粗骨料取代率较低时,再生混凝土抗压强度略高于普通混凝土,随着粗骨料替代率的增加,混凝土的抗压强度降低^[5-6]。随着水灰比的增大,再生粗骨料混凝土抗压强度降低^[7-8]。一般而言,抗压强度是通过对试样进行规定时间的养护并进行抗压试验来获取的,试验周期长且耗材多、成本高。李秋义等^[9]考虑了再生粗骨料和再生细骨料的特性,在再生混凝土鲍罗米公式的基础上进行了修正研究,用以指导再生骨料混凝土配合比设计。王伟等^[10]采用半值强度指数建立了再生粗料混凝土抗压强度-龄期数学模型。李马力等^[11]基于普通混凝土配合比设

计方法,建立再生粗骨料混凝土强度经验模型,为再生粗骨料混凝土配合比设计提供参考。

再生粗骨料混凝土的抗压强度受诸多不确定因素的影响,如粗骨料替代率、水灰比、粗骨料类型、粗骨料压碎率等。再生粗骨料混凝土抗压强度与影响因素之间存在复杂的非线性关系,传统的经验公式不能准确反映和阐明各因素之间的相关性。

随着计算机技术的高速发展,机器学习开始运用于预测混凝土的各项性能。韩越等^[12]以水泥强度和用量、水、砂、坍落度等7个参数为输入,采用L-M算法改进BP神经网络,预测再生混凝土强度。Duan等^[13]利用人工神经网络预测天然骨料混凝土和再生骨料混凝土的抗压强度和弹性模量,并将两种混凝土进行比较,以此来评估再生骨料的质量。黄炜等^[14]利用PSO-BP和GA-BP神经网络模型预测再生砖骨料混凝土强度,寻找最佳配合比,提高试验效率。上述预测模型大多基于神经网络,而神经网络对数据量要求高,模型的中间过程难以分析和控制,在预测准确率上仍然存在一定的局限。鉴于此,为开发性能更为优越的再生粗骨料混凝土抗压强度预测模型,本文基于XGBoost算法对再生粗骨料混凝土抗压强度进行预测,以净胶水比、天然骨料压碎率、再生粗骨料压碎率、再生粗骨料替代率为输入,利用树结构概率密度估计贝叶斯方法优化超参数,获得预测精度高、泛化能力较强的预测模型,满足预测再生混凝土抗压强度的需求。

1 基本原理

1.1 极限提升树基本原理

极限提升树(XGBoost)^[15]是基于决策树的集成算法,是梯度提升树算法的升级。通过构建基学习器,不断迭代减少上次迭代的残差,并把多个基学习器的结果累加作为最终的预测输出。XGBoost通过损失函数和正则项,实现了树模型精确性与复

杂性之间的平衡。在保证模型性能的同时,提升了模型的运行效率。

1.1.1 加法模型

集成学习的核心思想是反复训练基学习器,并将基学习器通过一定的方式组合,最终形成强大的集成学习模型。XGBoost 是基于加法模型框架的机器学习算法,采用前向分布算法,回归树为基学习器。对每一棵回归树寻求一个最优解,逐步逼近优化损失函数,即

$$\hat{y}_i^{(k)} = \sum_{j=1}^k f_j(x_i) = \sum_{j=1}^{k-1} f_j(x_i) + f_j^{(k)}(x_i) \quad (1)$$

式中: $\hat{y}_i^{(k)}$ 为第 i 个样本的预测值; $f_j(x_i)$ 是第 j 棵回归树; k 为基学习器的个数; x_i 为第 i 个样本的特征值; $f_j^{(k)}(x_i)$ 为当前需要优化的第 k 个回归树。

1.1.2 目标函数

XGBoost 算法中,目标函数[式(2)]是针对每一棵树的,对任意树 f_j 来说,目标函数有两个组成部分:一部分是任意可微的损失函数,它控制模型的经验风险;另一部分是控制模型复杂度的 $\Omega(f_j)$,它控制当前树的结构风险。

$$O_{bj}^{(t)} = \sum_{i=1}^N l(y_i, \hat{y}_i^{(t)}) + \sum_{j=1}^t \Omega(f_j) = \sum_{i=1}^N l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_j) \quad (2)$$

式中: $O_{bj}^{(t)}$ 为第 t 轮迭代的目标函数; y_i 为真实值; N 为样本数; $l(y_i, \hat{y}_i^{(t)})$ 为损失函数; $\sum_{j=1}^t \Omega(f_j)$ 为正则项, $\Omega(f_j) = \gamma + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2$, λ 为超参数控制惩罚力度, T 为当前回归树的叶子节点个数, w_j 为第 j 个叶子节点的值。

将目标函数以样本作为目标拆分为以叶子节点作为目标,将二阶泰勒展开公式代入目标函数得

$$O_{bj}^{(t)} = \sum_{j=1}^T \left[\left(\sum_{i \in I_j} g_i \right) w_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) w_j^2 \right] + \gamma T \approx \sum_{j=1}^T \left[w_j G_j + \frac{1}{2} w_j^2 (\lambda + H_j) \right] + \gamma T \quad (3)$$

式中: $g_i = \partial_{\hat{y}}^{(t-1)} l(y_i, \hat{y}^{(t-1)})$; $h_i = \partial_{\hat{y}}^2^{(t-1)} l(y_i, \hat{y}^{(t-1)})$; $G_j = \sum_{i \in I_j} g_i$; $H_j = \sum_{i \in I_j} h_i$ 。

通过最小化目标函数求解得到叶子节点预测值 w_j^* 和最小化目标函数 $O_{bj}^{(t)*}$ 。

$$w_j^* = -\frac{G_j}{H_j + \lambda} \quad (4)$$

$$O_{bj}^{(t)*} = \min(O_{bj}^{(t)}) = \gamma T - \frac{1}{2} \sum_{j=1}^T \frac{G_j^2}{H_j + \lambda} \quad (5)$$

1.1.3 增益值的计算

节点如何分裂是决策树生长过程的关键问题,XGBoost 算法采取贪心策略。针对每一个叶子节点的划分,计算结构分数增益,选取增益最大点进行分枝。增益值计算公式如式(6)所示。

$$g = \frac{1}{2} \left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \gamma \quad (6)$$

式中: g 为增益值; G_L 为左子节点的梯度和; G_R 为右子节点的梯度和; H_L 为左子节点的二阶梯度和; H_R 为右子节点的二阶梯度和; $\frac{G_L^2}{H_L + \lambda}$ 为左节点结构分数; $\frac{G_R^2}{H_R + \lambda}$ 为右节点结构分数; $\frac{(G_L + G_R)^2}{H_L + H_R + \lambda}$ 为父节点的结构分数。

1.2 基于树结构概率密度的贝叶斯优化基本原理

1.2.1 贝叶斯优化原理

贝叶斯优化(Bayesian Optimization, BO)是机器学习自动化中一种先进的超参数优化(Hyper-parameter Optimization, HPO)方法。针对超参数调优问题,贝叶斯优化可求得最优参数组合,其中 x^* 为最优超参数组合, χ 为参数空间, $f(x)$ 为目标函数,可得

$$x^* = \arg \max_{x \in \chi} [f(x)] \quad (7)$$

贝叶斯优化的关键步骤是概率代理模型和采集函数。概率代理模型根据已有的调参历史数据,建立概率分布模型;采集函数是用来挑选下一个参数组合的依据,不断将新的参数组合加入到概率分布模型,直到达到最大迭代次数。

1.2.2 树结构概率密度估计原理

不同贝叶斯优化方法的主要区别在于概率代理模型和采集函数不同。本文选择树结构概率密度估计(Tree-structured Parzen Estimator, TPE)作为概率代理模型[式(8)]。

$$p(x|y) = \begin{cases} l(x) & y^* > y \\ g(x) & y \geq y^* \end{cases} \quad (8)$$

式中: $p(x|y)$ 为条件概率密度函数; x 为观测值; y 为观测阈中的最优值, $y^* = \min\{(x_1, f(x_1)), \dots, (x_1, f(x_1))\}$; $l(x)$ 为观测值 x 的损失函数小于 y^* 的密度估计; $g(x)$ 为观测值 $x_{(i)}$ 的损失函数大于等于 y^* 的密度组成。

TPE 模型选择的采集函数为 E_1 , 定义为

$$E_{1y^*} = \int_{-\infty}^{y^*} (y^* - y) p(y|x) dy = \int_{-\infty}^{y^*} (y^* - y) \frac{p(y|x)p(y)}{p(x)} dy \quad (9)$$

令

$$\gamma = p(y < y^*) \int_R p(y | x) p(y) dy = \gamma l(x) + (1 - \gamma) g(x),$$

则

$$E_{1,y^*} = \frac{\gamma y^* - l(x) \int_{-\infty}^{y^*} p(y) dy}{\gamma l(x) + (1 - \gamma) g(x)} \propto [(\gamma + \frac{g(x)}{l(x)} (1 - \gamma))]^{-1} \tag{10}$$

确定下一个采样点的过程通过最大化 E_1 函数实现。当超参数 x 具有最大概率 $l(x)$ 和最小概率 $g(x)$ 时,取得最大化 E_1 值^[16]。在 TPE 模型中,最大化 E_1 过程即最大化 $g(x)/l(x)$ 的过程, $g(x)/l(x)$ 越大,则下一个采样点 x 在目标函数取

值大于 y^* 可能性越大,超参数性能越好。

2 预测模型的建立

基于 XGBoost 的再生粗骨料抗压强度预测模型框架如图 1 所示。再生粗骨料抗压强度预测模型的实质是利用 Python 调用构建的再生粗骨料抗压强度数据库,对 XGBoost 模型进行训练,通过交叉验证和贝叶斯优化算法在指定超参数空间内对超参数进行优化,使再生粗骨料混凝土抗压强度预测模型的精度最高,模型拟合效果最好。

步骤 1:通过文献[11]提供的再生粗骨料混凝土强度与净胶水比、天然骨料压碎率、再生骨料压碎率和再生骨料替代率数据库,对数据进行预处理。

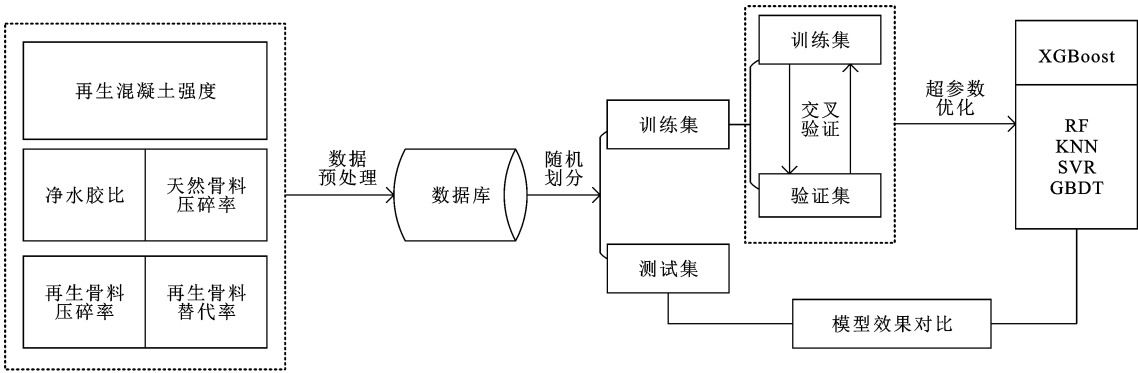


图 1 再生粗骨料强度预测模型框架

Fig. 1 Framework of strength prediction model of recycled coarse aggregate

步骤 2:利用 Python 调用机器学习库中 XGBoost 模块,在不调整超参数的情况下,将步骤 1 中的数据信息输入到 XGBoost 模型中进行训练,然后使用模型拟合测试集数据,查看预测效果。

步骤 3:对模型进行超参数调优。在训练集中使用 3 折交叉验证结合贝叶斯优化对 XGBoost 模型和随机森林(RF)模型、K 邻近回归(KNN)模型、支持向量机回归(SVR)模型、梯度决策提升树(GBDT)模型进行超参数优化。选取最优参数值,使回归模型的决定系数 R^2 最大。

步骤 4:将优化后的超参数输入到各个模型中进行训练。将测试集数据输入到训练好的模型中,比较不同模型的性能,验证该模型的准确性和泛化能力。

3 实例研究

3.1 数据分析

本研究利用数据库中 89 个试验数据样本对机器学习模型进行训练和测试。以影响再生粗骨料混

凝土强度的 4 个参数(净胶水比、天然骨料压碎率、再生骨料压碎率和再生骨料替代率)为模型输入,最后输出再生粗骨料混凝土强度模型预测值。表 1 列出了本研究使用的输入变量均值、中位数、最大值、最小值。本文使用数据的分布如图 2 所示。净胶水比主要集中在 2~3 之间;天然骨料压碎率取值主要为 4%左右、6%左右、7%左右、9%左右、10%左右;再生骨料压碎率普遍大于天然骨料压碎率,集中在 10%~18%;再生粗骨料替代率主要取值在 0%左右、30%左右、50%左右、70%左右、100%左右。混凝土抗压强度在 24.21~73.26 MPa 之间。

表 1 研究使用的输入变量

Table 1 Input variables used in study

输入变量	数量	均值	中位数	最小值	最大值
净胶水比	89	2.53	2.46	1.65	4.13
天然骨料压碎率/%	89	8.22	8.82	3.82	10.20
再生骨料压碎率/%	89	11.60	13.49	0.00	17.70
再生骨料替代率/%	89	60.00	70.00	0.00	100.00
抗压强度/MPa	89	46.88	47.61	24.21	73.26

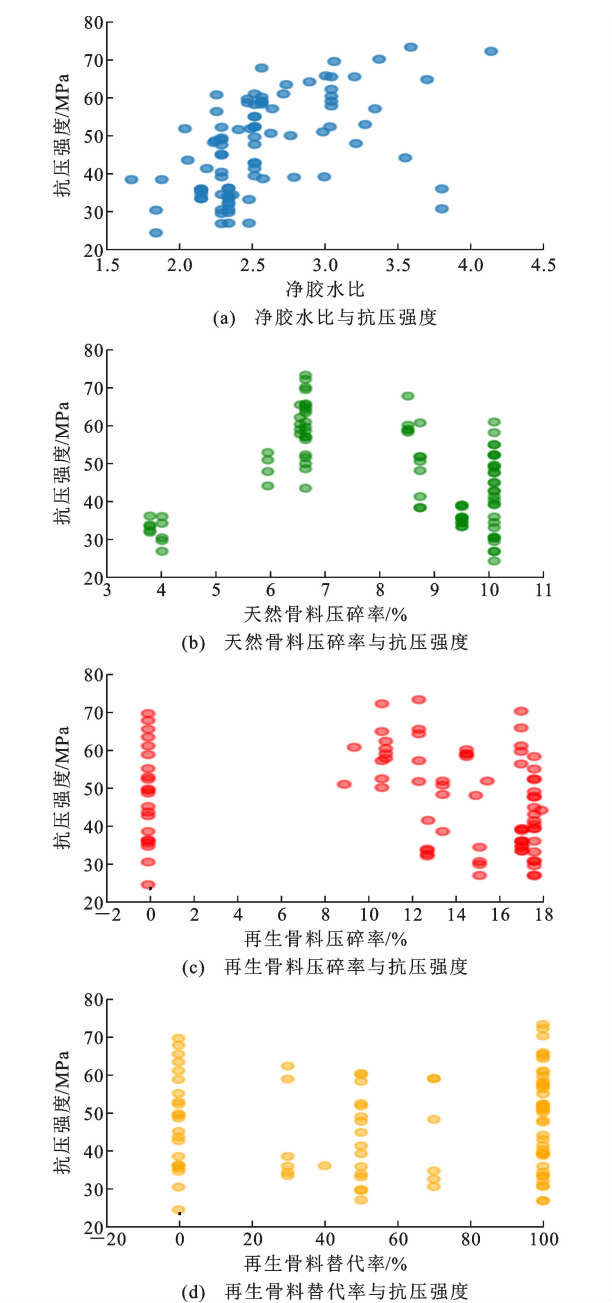


图 2 输入参数数据分布

Fig. 2 Input parameter data distribution

图 3 为本研究使用的输入参数和输出参数的多重相关矩阵,颜色的深浅表示相关性的 大小。在输入变量中再生骨料压碎率和再生骨料替代率之间的相关性最高($R=0.76$),其次是再生粗骨料替代率与净胶水比的相关性($R=0.37$),再生粗骨料替代率越大,骨料吸水率越大,高骨料吸水率会降低水胶比^[17]。在输入变量与输出变量之间,净胶水比与再生粗骨料混凝土抗压强度之间的相关性最大,总体而言输入变量与输出变量之间相关性相对较低,因此选取所有输入变量以提高预测模型的准确性。

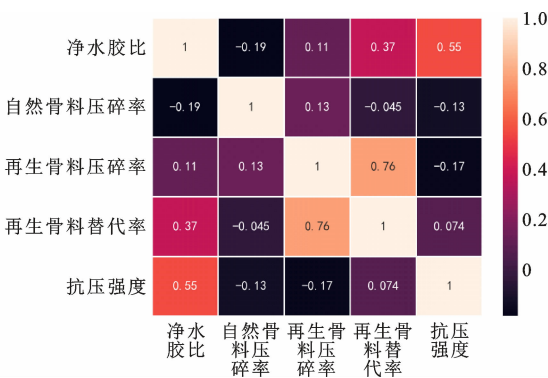


图 3 输入参数和输出参数多重相关矩阵

Fig. 3 Multiple correlation matrix between input parameters and output parameters

3.2 数据预处理

数据库中再生骨料压碎率、抗压强度存在缺失值。针对再生骨料压碎率缺失值,不进行特殊处理。其原因是:XGBoost 模型将缺失值当作稀疏矩阵来对待,子节点分裂时不考虑缺失值,直接在包含空值的数据上进行构建模型;针对抗压强度缺失值,将其所在样本删除。抗压强度为该模型标签特征,XGBoost 为监督学习算法,需要标签特征进行机器学习。

训练样本数据均通过标准化处理,消除数据单位的限制,将其转化为量纲一的数值,提高迭代收敛的速度,使特征值有均等的权重。

原始输入参数:根据已发表文献记录的再生粗骨料混凝土强度与净胶水比、天然骨料压碎率、再生骨料压碎率和再生骨料替代率数据,作为参数输入。

预处理参数:为了寻求更加高精度、高泛化能力的预测模型,消除不同数据组之间基准普通混凝土配合比的影响,以量纲一的参数再生粗骨料压碎率与普通混凝土骨料压碎率比值 R_s 、再生混凝土净胶水比与普通混凝土净胶水比的比值 R_{cw} 、再生粗骨料替代率 ρ 、再生混凝土抗压强度与普通混凝土强度比值 f_r/f_c 作为输入参数。

3.3 数据集划分

将 89 条数据按照 9 : 1 数量比例分为训练集和测试集,训练集用于训练预测模型,测试集用来评估最终的模型。训练集采用 3 折交叉验证,即将除测试集外的数据划分成 3 份,组合成多种不同的训练集和验证集。

3.4 超参数优化

超参数是指模型开始训练前人为进行调整的参数,超参数的选择直接影响到模型的预测效果。合

理选择超参数有利于平衡模型的结构风险和经验风险,提高模型的泛化能力。在超参数优化中,基础的方法是经验调参、网格搜索、随机网格搜索和贝叶斯优化算法^[18]。超参数调节开始之前,将需要调节的超参数的备选值列出,多个不同超参数的不同取值之间排列组合,最终将组成一个参数空间,如表 2 所示。

表 2 XGBoost 超参数含义及其范围

Table 2 Meaning and scope of XGBoost hyperparameter		
XGBoost 超参数	含义	取值范围
n_estimators	决策树数量	(30,100)
learning_rate	学习率	(0.01,1)
gamma	T 的正则化参数	(0,20)
reg_lambda	L2 正则项的权重	(0,2)
max_depth	数的最大深度	(2,10)
subsample	子样本占训练集比例	(0.8,1)

网格搜索算法将参数空间中的所有参数组合代入模型进行训练,网格搜索算法遍历每一个参数组合,需要消耗大量的计算资源;随机网格搜索以不放回抽样的方式选出部分参数组合作为“子空间”,并只对这些参数组合进行搜索。贝叶斯优化方法是先进的自动化调参方法,其步骤如下:

(1)构建基于 TPE-XGBoost 的再生混凝土预测模型,根据确定的参数空间,在超参数取值范围内随机生成超参数组合 $X=[x_1,x_2,x_3,x_4,x_5,\cdots,x_n]$,代入数据集,训练预测模型得到参数组合相应的目标函数,即交叉验证评估指标 R^2 ,可得 $Y=[f(x_1),f(x_2),f(x_3),\cdots,f(x_n)]$,形成初始的超参数与预测精度数据集。

(2)基于预测精度数据集,将样本点划分为优和差,建立代理模型。

(3)基于 E_1 采集函数从代理模型中寻找下一组超参数采样点 x_{n+1} ,并将该超参数组合代入数据集进行训练,得到再生混凝土抗压强度预测值与交叉验证评估指标 $f(x_{n+1})$ 。

(4)若交叉验证评估指标满足要求,则以 x_{n+1} 作为最佳超参数并终止程序。

3.5 模型评价指标

各预测模型性能评价指标分别为:决定系数 R^2 、均方误差 M_{SE} 、平均绝对误差 M_{AE} 。其中, R^2 判断模型的可行性和线性关系, M_{SE} 反映预测值与实际值之差平方的期望, M_{AE} 反映预测值与实际值之间误差绝对值的平均值,计算公式为

$$R^2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})^2} \tag{11}$$

$$\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i \tag{12}$$

$$M_{SE} = \frac{1}{N} \sum_{i=1}^N (y_i - f_i)^2 \tag{13}$$

$$M_{AE} = \frac{1}{N} \sum_{i=1}^N |y_i - f_i| \tag{14}$$

式中: y_i 为样本实际值; f_i 为第 i 个样本的模型预测值。

4 结果分析

4.1 数据预处理对模型性能的影响分析

为验证数据预处理对预测效果的影响,利用模型在原始数据集(data1)和预处理后数据集(data2)进行建模和性能对比。图 4、5 为在不同数据集下的模预测模型学习曲线,反映模型在训练集和交叉验证集上的得分情况。训练样本数为横坐标, R^2 为纵坐标,XGBoost 预测模型在原始数据集(data1)训练集 R^2 得分在 0.85 上下浮动,模型在预处理数据集(data2)上训练集 R^2 得分均接近 1;XGBoost 预测模型在原始数据集(data1)交叉验证最高 R^2 得分为 0.7 左右,在预处理数据集(data2)上交叉验证最高得分为 0.8 左右;XGBoost 模型在预处理数据集(data2)性能更优。

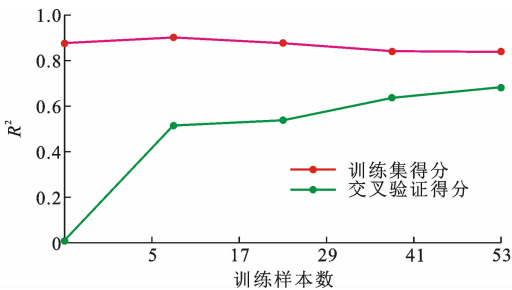


图 4 XGBoost 模型在原始收集数据集上的学习曲线
Fig. 4 Learning curve of XGBoost model on original collected data set

预测模型均经过多次超参数调优,以 3 折交叉验证 R^2 得分的均值为超参数选取指标,最终建立模型。其他预测模型在预处理数据集上的训练集、测试集、验证集上的模型性能(表 3)表现大多优于在原始收集数据集上的表现。

4.2 超参数调优对模型性能的影响分析

网格搜索超参数调优方法在相同的参数空间下耗费更多的计算资源。因此,表格中网格搜索超参

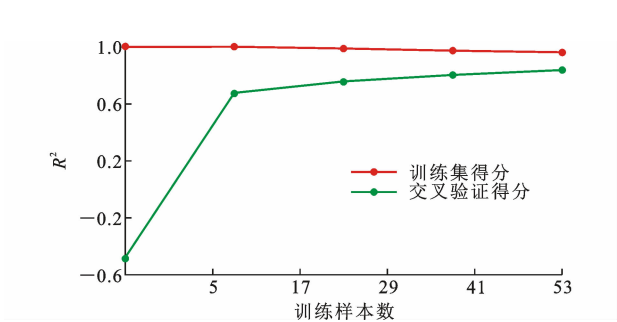


图5 XGBoost模型在预处理数据集上的学习曲线

Fig. 5 Learning curve of XGBoost model on preprocessed data set

表3 模型在两种不同数据集上的性能对比

Table 3 Performance comparison of models in two different data sets

数据集	数据	XGBoost	RF	KNN	SVR	GBDT
训练集	Data1	0.803 2	0.879 3	0.910 7	0.801 0	0.904 1
	Data2	0.947 8	0.933 8	0.976 1	0.764 2	0.933 3
测试集	Data1	0.591 2	0.919 3	0.844 7	0.674 9	0.753 4
	Data2	0.957 6	0.908 9	0.924 6	0.855 6	0.945 3
交叉验证集	Data1	0.736 1	0.619 4	0.515 3	0.563 4	0.728 7
	Data2	0.845 7	0.833 9	0.711 6	0.751 8	0.844 3

数空间相较于贝叶斯优化超参数空间更小。为了更好地对比不同超参数调优方法对模型性能的影响,网格搜索超参数空间与随机搜索超参数空间一致,基于高斯过程回归(Gaussian Process Regression, GPR)的贝叶斯优化与基于TPE的贝叶斯优化迭代次数相同,为300次。

网格搜索方法与随机搜索方法本质都为枚举法,搜索的超参数组合相同,预测模型验证集分数相同(表4)。由于随机搜索的搜索空间仅为全域空间的子空间,超参数搜索时间相较于网格搜索大幅度减少,本次试验中随机搜索速度比网格搜索快3倍~7倍。基于GPR的贝叶斯超参数优化和基于TPE的贝叶斯优化确定的预测模型性能均优于网

表4 不同超参数优化方法下预测模型性能对比

Table 4 Performance comparison of prediction models under different hyperparameter optimization methods

算法	HPO方法	网格搜索	随机搜索	TPE-BO	GPR-BO
XGBoost模型	交叉验证分数	0.834 7	0.834 7	0.847 5	0.839 3
	测试集分数	0.937 3	0.937 3	0.957 6	0.926 1
	搜索时间/s	126.395	18.006	33.000	1 156.000
RF模型	交叉验证分数	0.822 0	0.828 4	0.833 9	0.833 8
	测试集分数	0.908 9	0.908 9	0.908 9	0.908 9
	搜索时间/s	258.572	89.504	62.000	1 116.000
GBDT模型	交叉验证分数	0.844 3	0.844 3	0.844 5	0.840 8
	测试集分数	0.945 3	0.945 3	0.934 7	0.933 8
	搜索时间/s	319.599	87.701	23.000	1 053.000

格搜索方法和随机网格搜索方法,XGBoost、RF、GBDT模型拟合程度依次上升0.012 8、0.005 5、0.000 2。基于TPE的贝叶斯超参数优化方法确定的预测模型性能均优于基于GPR的贝叶斯超参数优化方法确定的预测模型,XGBoost、RF、GBDT模型拟合程度依次上升0.008 2、0.000 1、0.003 7,且速度更快,TPE方法的平均搜索时间约为39 s,GPR方法的平均搜索时间为1 108 s。表5为各预测模型在不同超参数优化方法下确定的最优超参数取值。

4.3 不同模型性能对比分析

不同模型预测值对比如图6所示,SVR预测模型预测值明显偏离真实值;KNN预测模型在样本2、4、9上的预测精度较低,预测效果不稳定;GBDT模型在测试样本点3、7、9上预测效果较差,XGBoost预测模型相较于其他预测模型预测效果表现最佳,在各测试样本点上的预测值与真实值偏差最小。XGBoost预测模型在测试集上预测精度最高(表6),对比RF、KNN、SVR、GBDT模型在测试集上的表现,其 R^2 分别高出0.048 7、0.033、0.102、0.012 3,其 M_{SE} 、 M_{AE} 均低于其他预测模型。

XGBoost预测模型在交叉验证上 R^2 指标高于其他预测模型,对比RF、KNN、SVR、GBDT模型在验证集表现,其 R^2 分别高出0.013 6、0.135 9、0.095 7、0.003, M_{AE} 低于其他预测模型;相较于其他模型,XGBoost具有更强的泛化能力,其次是GBDT模型和RF模型。

预测模型性能对比如图7所示,XGBoost交叉验证的平均 R^2 值最大,模型具有更好的性能,KNN模型的交叉验证 R^2 值最差;各预测模型的交叉验证集的平均 R^2 均小于训练集和验证集,平均 M_{SE} 和平均 M_{AE} 均大于训练集和验证集,其可能的原因是:再生粗骨料混凝土样本数据规模较小;给定关于再

表 5 不同超参数优化方法下确定的预测模型最优超参数组合

Table 5 Optimal combination of hyperparameters of prediction models determined under different hyperparameter optimization methods					
算法	超参数	网格搜索	随机搜索	TPE-BO	GPR-BO
XGBoost 模型	gamma	0	0	0	0
	learning_rate	0.200 0	0.200 0	0.200 4	0.238 0
	max_depth	2	2	2	2
	n_estimators	32	32	30	49
	reg_lambda	0.490 0	0.490 0	0.001 5	0.038 6
	subsample	0.800 0	0.800 0	0.817 1	0.888 6
RF 模型	max_depth	5	5	5	5
	max_leaf_nodes	None	None	None	None
	max_samples	None	None	None	0.998 7
	min_samples_split	2	2	2	2
	n_estimators	31	31	62	31
GBDT 模型	criterion	mse	mse	mse	friedman_mse
	learning_rate	0.200 0	0.200 0	0.198 0	0.215 5
	max_depth	2	2	2	2
	min_impurity_decrease	0	0	0	0
	n_estimators	32	32	28	28
	subsample	0.900 0	0.900 0	0.900 0	0.898 8
KNN 模型	algorithm	brute	brute	brute	brute
	leaf_size	32	32	32	32
	n_neighbors	4	4	4	4
	p	2	2	2	2
	weights'	distance			
SVR 模型	kernel	rbf	rbf	rbf	rbf
	gamma	0.01	0.01	0.01	0.01
	C	100	100	100	100

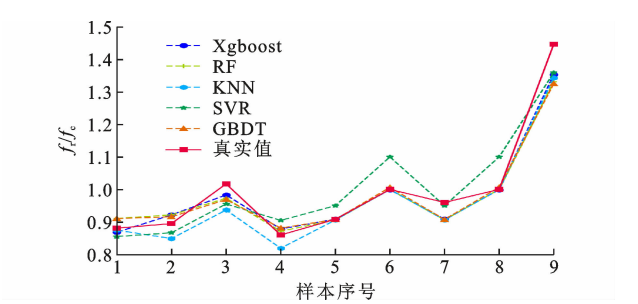


图 6 模型预测值对比

Fig. 6 Comparison of predicted values of models

生粗骨料混凝土的相关特征不够;收集数据中存在异常值。XGBoost 模型、RF 模型、GBDT 模型训练集 R^2 和测试集 R^2 与其他模型相比相差较小,其泛化误差与经验误差平衡度比其他模型高。单一模型学习相较于集成学习(XGBoost、RF、GBDT)有更高的经验风险,易过拟合;KNN 模型训练集 R^2 大于测试集,SVR 模型训练集 R^2 小于测试集 R^2 ,SVR 模型可能存在结构风险。

表 6 不同再生骨料混凝土预测模型性能对比

Table 6 Performance comparison of prediction models of different recycled aggregate concrete

指标	数据集	XGBoost	RF	KNN	SVR	GBDT
R^2	训练集	0.940 3	0.933 3	0.976 1	0.764 2	0.921 6
	测试集	0.957 6	0.908 9	0.924 6	0.855 6	0.945 3
	交叉验证集	0.847 5	0.833 9	0.711 6	0.751 8	0.844 5
M_{SE}	训练集	0.001 9	0.002 1	0.000 8	0.007 5	0.007 5
	测试集	0.001 5	0.002 2	0.002 1	0.004 1	0.002 0
	交叉验证集	0.036 5	0.041 1	0.036 1	0.038 2	0.040 6
M_{AE}	训练集	0.030 1	0.032 5	0.007 8	0.071 0	0.071 0
	测试集	0.026 6	0.031 5	0.033 5	0.055 5	0.034 6
	交叉验证	0.093 8	0.119 8	0.113 0	0.135 4	0.100 4

4.4 小样本数据对预测模型性能影响分析

预测模型的性能在很大程度上取决于足够多的准确、干净和相关的数 据。基于少量观测数据的模型容易产生过度拟合且预测结果存在偏差。小样本数据中存在异常值也会对模型的预测结果产生较大

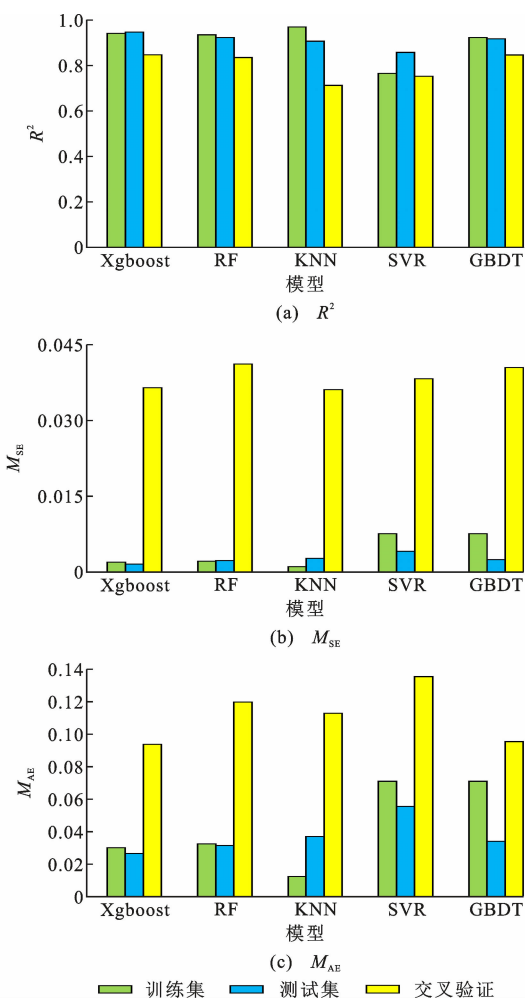


图7 模型预测性能对比

Fig. 7 Comparison of model prediction performance

影响。实际情况中,获取大量的优质数据成本较高,难度较大。XGBoost 模型在数据集较大时预测具有很好的准确性和稳定性,在小样本数据集下,预测精度会下降但仍然具有可接受的准确性^[19]。XGBoost 模型在少量再生粗骨料混凝土数据集上表现优于 RF、KNN、SVR、GBDT 模型,是由于 XGBoost 算法中的正则项技术降低了模型过拟合的风险。交叉验证结合贝叶斯优化的超参数调节方法尽可能提升模型的性能,增加模型的泛化能力。

4.5 不同特征对再生粗骨料混凝土抗压强度的影响分析

在建立单个决策树模型时,其核心思想是选择最合适的特征作为叶子节点的分割点。这个信息可以作为评估特征重要度的一个指标。在 XGBoost 和 RF 这类以决策树作为基学习器的集成学习算法都具有计算特征重要度并进行输出的功能^[20],其特征重要度为该特征在构建决策树中被采用次数的总和。一个特征被用来构建决策树的次数越多,其特

征重要度就越大。

再生粗骨料混凝土抗压强度受到多种因素的影响,且不同因素与抗压强度之间内在的联系复杂,定量计算不同特征对再生粗骨料抗压强度影响的重要性,有助于理解再生粗骨料混凝土抗压强度的内在机理。XGBoost 和 RF 两种模型对影响混凝土强度的不同特征重要度进行排序,其结果见图 8。

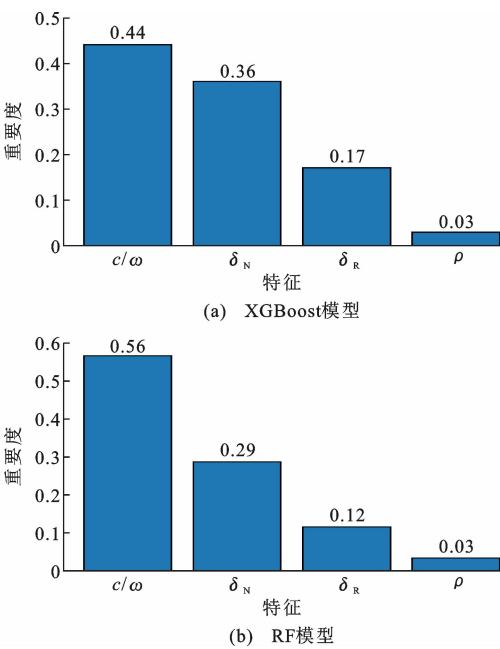


图8 模型特征重要度排序

Fig. 8 Ranking of feature importance of model

各特征在 XGBoost 和 RF 中重要度排序一致,由大到小依次为净胶浆比(c/w)、天然骨料压碎率(δ_N)、再生骨料压碎率(δ_R)、再生骨料替代率(ρ)。净胶浆比对再生粗骨料混凝土的抗压强度影响最大,若再生粗骨料混凝土内部的自由水越多,自由水蒸发后会形成小孔,使得再生粗骨料混凝土抗压强度变小^[21-22];压碎指标是骨料强度的重要参数,由于再生粗骨料表面粗糙、内部有裂纹等特性,使其压碎指标高于天然骨料,对再生混凝土的抗压强度贡献值低于天然骨料^[23];再生粗骨料替代率增加使得再生混凝土各项性能发生变化,不同取代率时抗压强度变化不同。

5 结 语

(1)采用数据预处理方法在一定程度上可以提高再生粗骨料抗压强度预测性能。XGBoost、RF、KNN、SVR 和 GBDT 模型在预处理后的数据集中均表现出了更佳的性能, R^2 在交叉验证中分别提高了 0.109 6、0.214 5、0.196 3、0.188 4、0.115 6。

(2)TPE-BO超参数优化方法有效提升再生粗骨料混凝土抗压强度预测模型性能,且效率高于其他超参数优化方法。

(3)在相同的训练样本条件下,本研究所提出的XGBoost预测模型相较于其他预测模型,预测再生粗骨料混凝土抗压强度的精度更高、离散性更小,表现出较好的预测性能和泛化能力,对比RF、KNN、SVR和GBDT模型在测试集表现,其 R^2 分别高出0.0487、0.033、0.102、0.0123。

(4)影响再生粗骨料混凝土抗压强度的特征重要度排序由大到小依次为净胶比、天然骨料压碎率、再生骨料压碎率、再生骨料替代率,该结论可为后续选取特征提供参考。

(5)未来研究需收集更多的样本数据,建立更加完善的非线性映射关系,以优化XGBoost模型,使预测模型精度更高,更具有可信力。

(6)实际影响再生粗骨料混凝土抗压强度的因素还包括混凝土材料性能、浇筑施工工艺等。模型特征的筛选有待扩展,预测模型的优化和泛化待进一步研究。

(7)XGBoost算法处理低维数据时有很好的精度和速度,但针对高维数据的处理效果一般且XGBoost算法超参数众多,调节参数需要熟练掌握算法原理。未来考虑更多影响再生骨料混凝土抗压强度的特征时,可借助深度学习等方法。

参考文献:

References:

- [1] 肖建庄,张航华,唐宇翔,等.废弃混凝土再生原理与再生混凝土基本问题[J].科学通报,2023,68(5):510-523.
- XIAO Jianzhuang, ZHANG Hanghua, TANG Yuxiang, et al. Principles for waste concrete recycling and basic problems of recycled concrete[J]. Chinese Science Bulletin, 2023, 68(5): 510-523.
- [2] XIAO J Z. Recycled aggregate concrete[M]//Recycled Aggregate Concrete Structures. Berlin: Springer, 2018: 65-98.
- [3] 李佳彬,肖建庄,孙振平.再生粗骨料特性及其对再生混凝土性能的影响[J].建筑材料学报,2004,7(4):390-395.
- LI Jiabin, XIAO Jianzhuang, SUN Zhenping. Properties of recycled coarse aggregate and its influence on recycled concrete[J]. Journal of Building Materials, 2004, 7(4): 390-395.
- [4] KISKU N, JOSHI H, ANSARI M, et al. A critical re-

view and assessment for usage of recycled aggregate as sustainable construction material[J]. Construction and Building Materials, 2017, 131: 721-740.

- [5] 徐蔚.再生粗骨料取代率对混凝土基本性能的影响[J].混凝土,2006(9):45-47.
- XU Wei. Experimental study on influence of recycled coarse aggregates contents on properties of recycled aggregate concrete[J]. Concrete, 2006(9): 45-47.
- [6] 胡敏萍.不同取代率再生粗骨料混凝土的力学性能[J].混凝土,2007(2):52-54.
- HU Minping. Mechanical properties of concrete prepared with different recycled coarse aggregates replacement rate[J]. Concrete, 2007(2): 52-54.
- [7] 曾力,赵伟.再生粗骨料混凝土强度公式研究[J].混凝土,2009(5):80-82.
- ZENG Li, ZHAO Wei. Strength formula of recycled coarse aggregate concrete[J]. Concrete, 2009(5): 80-82.
- [8] 彭立港,赵羽习,曾维来,等.再生粗骨料混凝土界面参数研究[J].建筑材料学报,2022,25(7):737-743.
- PENG Ligang, ZHAO Yuxi, ZENG Weilai, et al. Interface parameter of recycled coarse aggregate concrete[J]. Journal of Building Materials, 2022, 25(7): 737-743.
- [9] 李秋义,岳公冰,郭远新.再生混凝土性能调控与配合比设计[M].北京:中国建筑工业出版社,2019.
- LI Qiuyi, YUE Gongbing, GUO Yuanxin. Property regulation and mix proportion design of recycled concrete[M]. Beijing: China Architecture & Building Press, 2019.
- [10] 王伟,周爱兆,冯丽,等.再生粗骨料混凝土抗压强度-龄期数学模型[J].建筑材料学报,2012,15(5):633-637.
- WANG Wei, ZHOU Aizhao, FENG Li, et al. Mathematical model for age-dependent compressive strength of recycled coarse aggregate concrete[J]. Journal of Building Materials, 2012, 15(5): 633-637.
- [11] 李马力,隋莉莉,周英武,等.再生粗骨料混凝土强度预测模型研究[J].防灾减灾工程学报,2016,36(1):132-137.
- LI Mali, SUI Lili, ZHOU Yingwu, et al. Study of strength model for recycled coarse aggregate concrete[J]. Journal of Disaster Prevention and Mitigation Engineering, 2016, 36(1): 132-137.
- [12] 韩越,张新东.再生混凝土强度预测的神经网络模型[J].混凝土,2008(4):22-23,26.
- HAN Yue, ZHANG Xindong. Neural network model of the recycled concrete strength[J]. Concrete,

- 2008(4):22-23,26.
- [13] DUAN Z H, POON C S, XIAO J Z. Using artificial neural networks to assess the applicability of recycled aggregate classification by different specifications[J]. Materials and Structures, 2016, 50(2):107.
- [14] 黄 伟, 周 煌, 葛 培, 等. 基于 PSO-BP 和 GA-BP 神经网络再生砖骨料混凝土强度模型的对比研究[J]. 材料导报, 2021, 35(15):15026-15030.
HUANG Wei, ZHOU Lang, GE Pei, et al. A comparative study on compressive strength model of recycled brick aggregate concrete based on PSO-BP and GA-BP neural networks [J]. Materials Reports, 2021, 35(15):15026-15030.
- [15] CHEN T, GUESTRIN C. XGBoost: a scalable tree boosting system[C]//KRISHNAPURAM B, SHAH M. Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. New York: ACM, 2016:785-794.
- [16] 林光伟, 王 珊, 张西亚, 等. 基于贝叶斯参数优化的无模型自适应硅单晶直径控制[J]. 人工晶体学报, 2022, 51(2):229-241, 247.
LIN Guangwei, WANG Shan, ZHANG Xiya, et al. Model-free adaptive diameter control of monocrystalline silicon based on Bayesian parameter optimization [J]. Journal of Synthetic Crystals, 2022, 51(2):229-241, 247.
- [17] 史才军, 曹芷杰, 谢昭彬. 再生混凝土力学性能的研究进展[J]. 材料导报, 2016, 30(23):96-103, 127.
SHI Caijun, CAO Zhijie, XIE Zhaobin. Research progress in the mechanical properties of recycled aggregate concrete[J]. Materials Reports, 2016, 30(23):96-103, 127.
- [18] 孙 晨, 文 龙, 李新宇, 等. 基于自动机器学习的不平衡故障诊断方法[J]. 计算机集成制造系统, 2021, 27(10):2837-2847.
SUN Chen, WEN Long, LI Xinyu, et al. New automated machine learning based imbalanced learning method for fault diagnosis[J]. Computer Integrated Manufacturing Systems, 2021, 27(10):2837-2847.
- [19] ZOU M, JIANG W G, QIN Q H, et al. Optimized XGBoost model with small dataset for predicting relative density of Ti-6Al-4V parts manufactured by selective laser melting[J]. Materials, 2022, 15(15):5298.
- [20] 武梦婷, 陈秋松, 齐冲冲. 基于机器学习的边坡安全性评价及防护措施[J]. 工程科学学报, 2022, 44(2):180-188.
WU Mengting, CHEN Qiusong, QI Chongchong. Slope safety, stability evaluation, and protective measures based on machine learning[J]. Chinese Journal of Engineering, 2022, 44(2):180-188.
- [21] 陈宗平, 占东辉, 徐金俊. 再生粗骨料含量对再生混凝土力学性能的影响分析[J]. 工业建筑, 2015, 45(1):130-135.
CHEN Zongping, ZHAN Donghui, XU Jinjun. Research on mechanical properties of recycled concrete using different recycled coarse aggregate replacement [J]. Industrial Construction, 2015, 45(1):130-135.
- [22] 张丽素, 乔京生, 张 弛, 等. 不同因素对再生混凝土抗压强度的影响[J]. 华北理工大学学报(自然科学版), 2018, 40(2):61-65.
ZHANG Lisu, QIAO Jingsheng, ZHANG Chi, et al. Effect of different factors on compressive strength of recycled concrete[J]. Journal of North China University of Science and Technology (Natural Science Edition), 2018, 40(2):61-65.
- [23] 郭远新, 李秋义, 汪卫琴, 等. 再生粗骨料品质提升技术研究[J]. 混凝土, 2015(6):134-138.
GUO Yuanxin, LI Qiuyi, WANG Weiqin, et al. Research on recycled coarse aggregate quality of enhancement technology [J]. Concrete, 2015(6):134-138.