

DOI:10.11784/tdxbz202408006

# 土石方开挖钻孔效率预测可解释超级学习器集成学习模型

王晓玲<sup>1</sup>, 胡亦宁<sup>1</sup>, 张君<sup>1</sup>, 衣传宝<sup>2</sup>, 张捷<sup>3</sup>, 李希稷<sup>2</sup>

(1. 天津大学水利工程智能建设与运维全国重点实验室, 天津 300350;

2. 山东潍坊抽水蓄能有限公司, 潍坊 261000;

3. 中国电建集团北京勘测设计研究院有限公司, 北京 100024)

**摘要:** 动态预测钻孔效率并探究不同因素对钻孔效率的影响程度, 对土石方开挖进度分析和风险管理具有重要意义。然而, 现有土石方钻孔效率分析大都依赖人工经验, 少数机器学习模型无法解释不同因素对钻孔效率的影响程度。针对上述问题, 本研究提出土石方开挖钻孔效率预测可解释超级学习器(SL)集成学习模型。通过强化学习中的 Q 学习改进猎人猎物优化算法局部搜索过程与全局信息进行交互的能力, 提出 Q 学习改进的猎人猎物优化(QIHPO)算法对 SL 的  $n\_estimators$ 、 $learning\_rate$ 、 $max\_depth$  等超参数进行优化, 进而利用 SL 能够通过具有互补特征的异构基学习器捕捉样本特征差异性的优势, 建立基于 QIHPO 优化的超级学习器土石方开挖钻孔效率预测 QIHPO-SL 模型, 以揭示地质、作业、环境和机械特性等众多因素与钻孔效率的复杂非线性映射关系。进一步将 QIHPO-SL 集成学习算法与可解释机器学习框架下的沙普利加性解释(SHAP)理论相结合, 挖掘影响钻孔效率的关键特征, 并解释不同因素对钻孔效率的影响程度。案例分析表明: QIHPO-SL 具有较高的预测精度, 相较于 QIHPO-XGB、QIHPO-RF 和 SL 等基准模型, 本文所提方法的预测精度分别提高了 12.94%、12.02%和 1.58%, 且 SHAP 理论提高了模型的可解释性和预测结果的可信度, 为钻孔效率预测及成因分析提供了新思路和新途径。

**关键词:** 钻孔效率; 沙普利加性解释; 超级学习器; 强化学习

中图分类号: TV53

文献标志码: A

文章编号: 0493-2137(2025)04-0331-12

## Interpretable Super Learner Ensemble Learning Model for Earth-Rock Excavation Drilling Efficiency Prediction

Wang Xiaoling<sup>1</sup>, Hu Yining<sup>1</sup>, Zhang Jun<sup>1</sup>, Yi Chuanbao<sup>2</sup>, Zhang Jie<sup>3</sup>, Li Xiji<sup>2</sup>

(1. State Key Laboratory of Hydraulic Engineering Intelligent Construction and Operation, Tianjin University,

Tianjin 300350, China; 2. Shandong Weifang Pumped Storage Co., Ltd., Weifang 261000, China;

3. Beijing Engineering Co., Ltd., Power Construction Corporation of China, Beijing 100024, China)

**Abstract:** The dynamic prediction of the drilling efficiency and the exploration of the degree of influence of different factors on the drilling efficiency are of significance for the earth-rock excavation schedule analysis and risk management. However, most of the existing earth-rock drilling efficiency analysis methods rely on manual experience, and a few machine learning models cannot explain the degree of influence of different factors on the drilling efficiency. To address these problems, an interpretable super learner (SL) ensemble learning model for earth-rock excavation drilling efficiency prediction is proposed in this paper. The capability of the local search process of a hunter-prey optimization algorithm interacting with the global information is improved by Q-learning in reinforcement learning, and a Q-learning improved hunter-prey optimization (QIHPO) algorithm is put forward to optimize the hyper-parameters of

收稿日期: 2024-08-09; 修回日期: 2024-11-03.

作者简介: 王晓玲(1968—), 女, 博士, 教授, wangxl@tju.edu.cn.

通信作者: 张君, zhangdajun@tju.edu.cn.

基金项目: 国家自然科学基金青年科学基金资助项目(52309165); 国网新源集团有限公司及国网新源山东潍坊抽水蓄能有限公司科技资助项目(SGXYKJ-2020-076).

**Supported by** the Youth Program of the National Natural Science Foundation of China (No. 52309165), the Scientific and Technological Program of the State Grid Xinyuan Company Limited and the State Grid Xinyuan Shandong Weifang Pumped Storage Co., Ltd. (No. SGXYKJ-2020-076).

the SL, such as  $n\_estimators$ ,  $learning\_rate$  and  $max\_depth$ . Accordingly, the QIHPO algorithm can take advantage of the SL which is capable of capturing the variability of sample features through the heterogeneous base-learner with complementary characteristics, and a QIHPO-SL model based on QIHPO optimization and SL for predicting the earth-rock excavation drilling efficiency is established, thereby revealing the complex nonlinear mapping relationship between the drilling efficiency and various factors including the geological, operation, environmental and mechanical characteristics. Furthermore, the QIHPO-SL ensemble learning algorithm is combined with the theory of Shapley additive explanation (SHAP) under the interpretable machine learning framework to excavate the key features affecting the drilling efficiency and explain the degree of influence of different factors on the drilling efficiency. The analysis of a case study shows that QIHPO-SL has a high prediction accuracy. Compared with those of the benchmark models such as QIHPO-XGB, QIHPO-RF and SL, the prediction accuracy of the proposed method is improved by 12.94%, 12.02% and 1.58%, respectively, and the SHAP theory improves the model interpretability and the credibility of prediction results, providing a new idea and a new way for the drilling efficiency prediction and the corresponding causal analysis.

**Keywords:** drilling efficiency; Shapley additive explanation (SHAP); super learner (SL); reinforcement learning

料场开挖是土石坝施工的关键组成部分,开挖进度直接制约工程总进度,是施工进度管理的重点<sup>[1-2]</sup>.料场开挖中多种工序在开挖块上进行流水作业,具体涉及场地清理、钻孔、装药连线、爆破、出渣运输等循环操作,其中以钻孔工序最为重要,钻孔耗时时长约占开挖总工期的 70%,钻孔效率直接影响开挖进度<sup>[3]</sup>.钻孔效率的精确预测可为完善料场开挖设计方案、节省开挖成本以及后续开挖施工控制和工程量优化提供依据.此外,钻孔效率受到地质、作业、环境和机械特性等多种复杂因素影响.因此,开展多种因素影响下钻孔效率预测模型研究,实现钻孔效率的准确预测,并量化分析不同因素对钻孔效率的影响,对于钻孔施工质量和成本控制具有重要意义.

现有研究在钻孔效率预测及分析方面做出了广泛的探索.传统钻孔效率分析主要依赖人工经验确定影响钻孔效率的原因<sup>[4-5]</sup>,一些学者通过建立数学模型来预测钻孔效率,并依赖人工经验确定对预测钻孔效率具有重要意义的参数<sup>[6-8]</sup>.这些模型具有附加参数,包括钻孔中涉及的机械、岩性和地质方面,如钻头特性、泥浆特性、岩石孔隙度和地层抗压强度等,通过多元回归分析研究多个钻孔参数与钻孔效率之间的关系.然而,由于钻孔参数与钻孔效率之间存在高度非线性和复杂的关系,这些数学模型无法在施工中动态变化,且无法全面准确地估计钻孔效率<sup>[9]</sup>.随着人工智能的发展,一些学者探索了机器学习算法在钻孔效率预测方面的应用.机器学习方法相比数学模型,在预测精度方面得到较大提升.然而,由于机器学习可解释性较差,预测模型无法解释不同因素对钻孔效率的影响程度.Zhou 等<sup>[10]</sup>采用支持向量回归方法建立了钻孔效率预测模型,结合混合蝙蝠算法

和非支配排序遗传算法 II 得到支持向量机 (support vector machine, SVM) 方法的超参数. Mehrad 等<sup>[11]</sup>提出了一种基于最小二乘支持向量机的钻孔效率预测模型,并采用布谷鸟优化算法用于模型超参数优化. Zhao 等<sup>[12]</sup>提出了一种基于人工神经网络的钻孔效率预测模型,并采用人工蜂群算法用于模型超参数优化. Elkatatny<sup>[13]</sup>使用 3 种机器学习算法 (包括人工神经网络、支持向量机和自适应神经模糊推理系统) 来预测钻孔效率. Gan 等<sup>[14]</sup>利用互信息法获取与钻孔效率强相关的参数,基于改进粒子群优化径向基函数的神经网络建立了钻孔效率预测模型. Zhou 等<sup>[15]</sup>通过遗传算法构建了不同的极限学习机模型以预测不同钻孔工况下的钻孔效率.然而,单一的机器学习算法容易低估预测的不确定性,导致其预测结果精度较低.因此,一些学者提出基于集成模型的钻孔效率预测模型来提高整体预测精度<sup>[16-17]</sup>.

超级学习器 (super learner, SL)<sup>[18]</sup>是一种基于堆叠泛化的集成模型,并通过  $k$  折交叉验证从基学习器中构建预测模型的最佳加权组合,从而最大限度地避免过拟合或欠拟合.由于超级学习器的鲁棒性较强,其在各领域得到了广泛的应用<sup>[19-25]</sup>.然而,超级学习器所代表的复杂机器学习算法往往是“黑盒”模型<sup>[26]</sup>,由于缺乏预测过程中的透明度,学者们仅通过准确率的高低来衡量预测模型的优劣,无法分析输入特征对输出结果的相关关系及影响程度,可解释性较差,限制了其在工程中的可靠性<sup>[27]</sup>.沙普利加性解释 (Shapley additive explanation, SHAP)<sup>[28]</sup>基于 SHAP 值概念,通过量化输入特征对模型的重要程度,从全局和局部两方面揭示模型内部关系的复杂性,提供了一种直观准确的方式来理解预测模型的重要特征以

及它们如何影响模型的输出结果<sup>[29]</sup>,为机器学习的可解释分析提供了新的途径.因此,本研究引入可解释机器学习 SHAP 致因分析方法,在超级学习器算法实现钻孔效率准确预测的同时,采用可解释 SHAP 理论解释不同影响因素对钻孔效率预测结果的影响程度,从而为料场运行管理人员提供准确和可靠的开挖进度管控依据.此外,超级学习器算法中各基学习器的超参数各不相同,超参数值的设定会对模型预测结果产生较大影响<sup>[30]</sup>.目前大部分研究采用人工搜索或网格搜索方法确定超级学习器算法的最佳超参数组合,超参数寻优效率低下.采用群智能优化算法优化模型超参数是一种高效准确获取最优超参数组合的有效途径<sup>[31-32]</sup>.猎人猎物优化(hunter-prey optimization, HPO)算法是 Naruei 等<sup>[33]</sup>提出的一种新型群智能算法,具有较强的全局搜索能力和较强的稳定性.然而,HPO 算法在局部搜索过程中忽略了历史搜索中的全局信息,导致局部搜索过程较为缓慢,且容易陷入局部最优.

综上所述,为了弥补机器学习预测模型超参数寻优效率低、预测精度低、现有钻孔效率预测难以分析影响因素与模型输出结果的关系等不足,本研究提出土石方开挖钻孔效率预测可解释超级学习器集成学习模型,其中强化 Q 学习改进的猎人猎物优化(Q-learning improved hunter-prey optimization, QIHPO)算法用于优化超级学习器中的 `n_estimators`、`learning_rate`、`max_depth` 等超参数,进而通过具有互补特征的异构基学习器提高模型预测精度,并通过可解释机器学习 SHAP 理论解释不同因素对钻孔效率的影响程度.

### 1 研究框架

本研究所提模型的研究框架如图 1 所示,主要包括数据采集及数据集制作、土石方开挖钻孔效率预测超级学习器集成学习模型与致因解释、案例分析 3 部分.

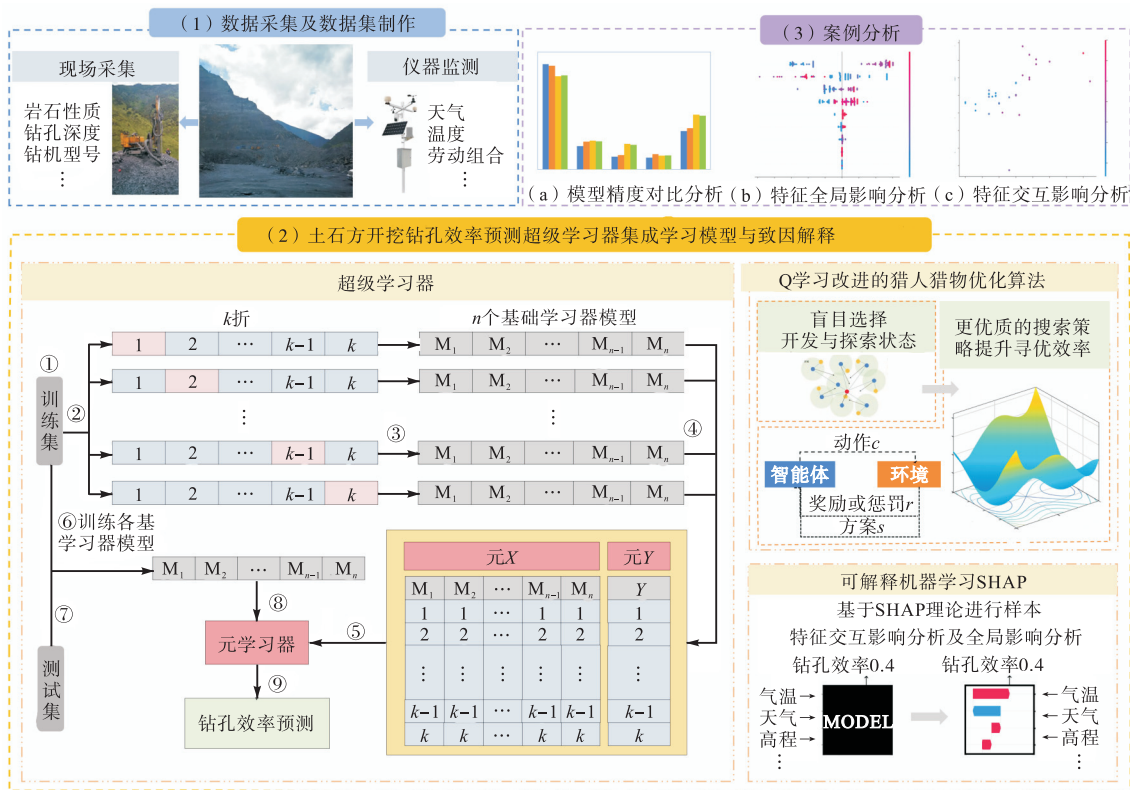


图 1 土石方开挖钻孔效率预测可解释超级学习器集成学习模型流程

Fig.1 Flow chart of interpretable super learner ensemble learning model of earth-rock excavation drilling efficiency prediction

(1)数据采集及数据集制作. 钻孔效率的主要影响因素包括地质、作业、环境和机械特性 4 个方面.为构建高精度的钻孔效率预测模型,选取岩性、高程、劳动组合情况、作业时段、温度、钻机型号、加钻杆数量、钻孔深度作为钻孔效率施工参数. 钻孔效率

影响因素及对应的钻孔效率构成钻孔效率超级学习器算法的数据集,将该数据集按一定比例划分为训练集和测试集.

(2)土石方开挖钻孔效率预测超级学习器集成学习模型与致因解释. 针对单一机器学习方法存在超

参数寻优效率低、预测模型精度低且可解释性较差的不足,利用超级学习器算法能够通过具有互补特征的异构基学习器捕捉样本特征差异性的优势,建立土石方开挖钻孔效率预测模型.进一步将 Q 学习引入传统猎人猎物优化算法的局部搜索过程进行改进,建立 QIHPO 算法用于自适应优化超级学习器算法的超参数.最后,将可解释机器学习 SHAP 理论与 QIHPO-SL 超级学习器算法结合,从而建立土石方开挖钻孔效率预测可解释超级学习器集成学习模型,实现钻孔效率的精确预测,并对影响因素进行全局解释及局部解释,分析钻孔效率影响因素对钻孔效率预测结果的正负影响、多个钻孔效率影响因素相互作用对钻孔效率预测结果的影响以及单个样本的钻孔效率影响因素对钻孔效率预测结果的影响,从而提高 QIHPO-SL 集成学习模型的可解释性.

(3) 案例分析. 将本研究提出的 QIHPO-SL 模型应用于中国西南某坝料土石方开挖工程并对钻孔效率影响因素进行致因分析,与其他模型进行对比以验证模型预测性能的优越性.

## 2 钻孔效率预测可解释超级学习器集成学习模型

### 2.1 钻孔效率预测 QIHPO-SL 集成学习模型

为了揭示地质、作业、环境和机械特性等众多因素与钻孔效率的复杂非线性关系,本文提出钻孔效率预测 QIHPO-SL 集成学习模型.首先,利用 Q 学习能够与全局信息进行交互的能力改进 HPO 算法局部搜索过程,弥补其因盲目搜索导致的收敛效率过慢的不足,用于超级学习器超参数优化.其次,利用超级学习器建立钻孔效率预测模型,采用具有互补特征的异构机器学习算法作为基学习器,有效提高模型的预测精度.

#### 2.1.1 超级学习器

超级学习器是一种集成学习算法,区别于简单平均集成与加权平均集成,其通过自适应融合训练策略从候选算法库中构建预测的最佳加权组合,从而生成一个更加强大而复杂的模型<sup>[34]</sup>.为了使超级学习器模型的预测效果最优,应选择学习能力强、差异度大的模型作为基学习器.本研究使用了以下 9 种机器学习算法作为基学习器,包括 SVM、K 近邻(K-nearest neighbors, KNN)、线性回归(linear regression, LR)、梯度提升(gradient boosting, GB)、随机森林(random forest, RF)、额外树(extra tree, ET)、极限

梯度提升(extreme gradient boosting, XGB)、最小绝对收缩和选择算子(least absolute shrinkage and selection operator, LASSO)回归和决策树(decision tree, DT),并选择泛化性强的 GB 作为元学习器以融合各个基学习器的预测结果.具体过程如下.

(1) 将获得的钻孔效率数据集  $(X, Y)$  按比例划分为训练集  $(X_{tr}, Y_{tr})$  和测试集  $(X_{te}, Y_{te})$ ;  $X = \{x_i, i = 1, 2, \dots, n\}$ . 对于钻孔效率数据集  $O_i = (X_{tr}, Y_{tr}), i = 1, 2, \dots, n$ , 其目标是估计回归  $\varphi_0(X) = E(Y|X)$ , 其中  $X(X \in \chi)$  和  $Y$  分别是钻孔效率影响因素和钻孔效率预测值,  $\chi$  为钻孔效率输入数据集. 回归问题可以定义为钻孔效率期望损失  $E([L(O, \varphi)])$ , 其中  $L$  是损失函数, 如式(1)所示.

$$\varphi_0(X) = \arg \min_x E([L(O, \varphi)]) \quad (1)$$

(2) 整个钻孔效率输入数据集  $\chi$  以  $k$ -折交叉验证的方式进行分割, 其中每折的验证和训练样本分别表示为  $V(v)$  和  $T(v)$ ,  $v = 1, 2, \dots, k$ . 设  $\psi_j (j = 1, 2, \dots, J)$  是来自一组默认算法的  $j$  个基学习器的集合. 对于第  $v$  层, 将每个基本模型  $\psi_j$  拟合到  $T(v)$  上, 得到相应钻孔效率验证样本上的钻孔效率预测如式(2)所示.

$$\psi_{j,T(v)}(V(v)) \quad j = 1, 2, \dots, J \quad (2)$$

(3) 将每个基学习器的钻孔效率预测结果堆叠在一起, 形成一个钻孔效率预测矩阵  $Z$ ,  $Z = \{\psi_{j,T(v)}(V(v))\}$ . 由权重向量  $a$  索引的候选基学习器的加权组合可以通过式(3)得到.

$$m(Z|a) = \sum_{j=1}^J a_j \psi_{j,T(v)}(V(v)), \sum_{j=1}^J a_j = 1 \quad (3)$$

(4) 在接下来的步骤中, 通过式(4)确定使所有允许的权重向量组合与钻孔效率预测值  $Y$  之间的交叉验证误差最小的权重向量  $a'$ .

$$a' = \arg \min_a \sum_{i=1}^n (Y_i - m(Z_i|a))^2 \quad (4)$$

(5) 根据  $m(Z|a)$  将最优权重向量  $a'$  与  $\psi_j(X)$  结合, 生成最终的钻孔效率预测结果  $\psi_{SL}(X)$ , 如式(5)所示.

$$\psi_{SL}(X) = \sum_{j=1}^J a'_j \psi_j(X) \quad (5)$$

#### 2.1.2 QIHPO-SL 集成学习模型

超级学习器算法中各基学习器的超参数各不相同, 超参数值的设定会对模型预测效果产生较大影响. 目前大部分研究采用人工搜索或网格搜索方法确定超级学习器算法的最佳超参数组合, 寻优效率低下. 采用群智能优化算法对超参数进行寻优能够快速准确获得最优参数组合.

### 2.1.2.1 QIHPO 算法

HPO 算法<sup>[33]</sup>在局部搜索过程中是盲目的,从不在历史搜索中总结经验,导致局部搜索过程较为缓慢,且只能聚焦于当前利益而无法实现全局最优.针对上述问题,本研究将强化学习中的 Q 学习引入到 HPO 算法的局部搜索阶段, Q 学习方法可以帮助发现参数的最佳设计和更平衡的策略,允许算法在探索和开发阶段之间切换,使算法能够选择更合适的更新策略,使搜索过程更具适应性和效率,训练搜索代理做出更有利的动作以提高 HPO 算法的搜索能力,从而提出 QIHPO 算法,具体过程如下.

**步骤 1** 根据超级学习器超参数优化的个数对超级学习器超参数种群进行初始化,并初始化  $q\_table$  及  $r\_table$ .

**步骤 2** 计算各个搜索代理的适应度值,即超级学习器模型的预测误差,确定当前最佳搜索代理及适应度.

**步骤 3** 更新搜索代理位置的相关参数,并根据搜索代理的  $q\_table$  值选择猎人狩猎策略或猎物搜索策略,进而更新超级学习器超参数的取值.

**步骤 4** 在融合了 Q 学习的 QIHPO 算法中,  $q\_table$  在局部搜索过程中指导猎人智能体选择不同位置更新模式,摒弃了原始随意评估搜索策略的盲目搜索.  $q\_table$  可以精确引导搜索代理选择使预测精度最大化的搜索策略,从而能够返回更优质的搜索策略并提升寻优效率.搜索代理通过计算每一种可能搜索策略的效益程度来选择最佳状态,并采取  $q\_table$  值最高的模式进行下一步搜索.搜索代理在每一步之后会根据当前的行为得到奖励或惩罚.在每次迭代中,猎物智能体使用式(6)来计算和加权下一步的每个可能搜索策略,然后选择最有可能接近最优解的最佳搜索策略.

$$q\_table(s,c) = Q(s,c) + \beta(r + \gamma \times \max_{c'} Q(s',c') - Q(s,c)) \quad (6)$$

式中:  $s$  为当前猎人附近所有可能的解决方案;  $c$  为当前动作;  $\beta$  为学习率(控制每次更新的幅度);  $r$  为猎人因采取当前动作而获得的即时奖励或惩罚;  $\gamma$  为折扣因子(平衡当前搜索策略奖励和未来搜索策略奖励的重要性);  $s'$  为下一步猎人附近所有可能的解决方案;  $c'$  为下一步可能动作.

**步骤 5** 更新群体目标函数.若某个猎物的目标函数大于步骤(2)中的最优值,则将其设为新的最优解,对应将其目标函数设为当前最优值.

**步骤 6** 重复步骤 3~5,直到达到最大迭代次数,

输出最优超参数组合.

### 2.1.2.2 QIHPO-SL 集成学习模型预测流程

采用超级学习器算法建立土石方开挖钻孔效率预测模型,通过 QIHPO 算法对超级学习器的超参数进行优化,使超级学习器训练过程中预测结果和真实值之间误差最小,最后得到土石方开挖钻孔效率预测 QIHPO-SL 集成学习模型的预测结果. QIHPO-SL 集成学习模型的实现流程如图 2 所示,伪代码如下.

**Input:**  $N$ ;  $iter_{max}$ ;  $S$ ;  $D$ ;  $ub$ ;  $lb$ ; 钻孔效率数据集

**Output:** 最优超参数组合、钻孔效率预测值  
定义基学习器及元学习器

**For**  $train\_idx, test\_idx$  **in**  $KFold(X\_train)$ :

生成随机初始种群并初始化  $q\_table, r\_table$

**While**  $iter < iter_{max}$

计算种群适应度,评估全局最优位置

更新搜索代理参数  $C, Z$

**If**  $q\_table[0,0] \geq q\_table[0,1]$

更新猎人位置

**Else**

更新猎物位置

**End if**

评估适应度值及全局最优位置

更新  $q\_table$  及  $r\_table$

**End while**

训练基学习器

训练元学习器

预测钻孔效率

**Return** 最优超参数组合、钻孔效率预测值

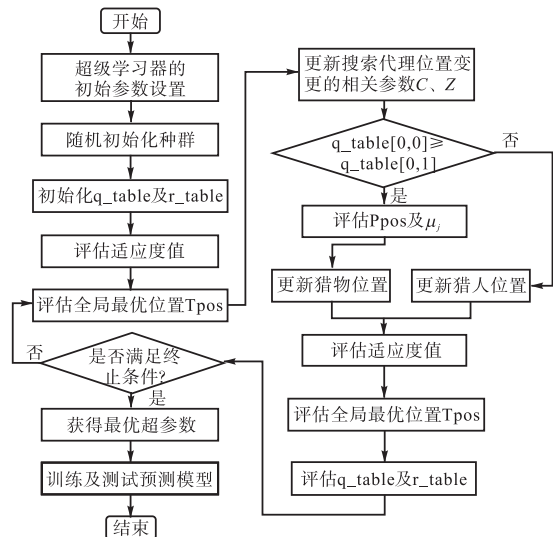


图 2 QIHPO-SL 集成学习模型流程

Fig.2 Flow chart of QIHPO-SL ensemble learning model

## 2.2 基于可解释机器学习SHAP的钻孔效率致因分析

可解释机器学习 SHAP 可以解释不同影响因素对钻孔效率预测结果的影响程度,从而为料场运行管理人员提供准确和可靠的钻孔效率分析结果. SHAP

是一种用于解释机器学习模型预测结果的方法。SHAP 的核心思想是通过计算特征值的不同组合对预测结果的影响来确定每个特征的重要性。具体而言, SHAP 使用了一个引入了噪声的数据集作为参考, 并生成了一个包含所有可能的特征组合的子集, 称为“博弈”。然后, 它计算每个特征在不同特征组合中的贡献, 并通过 SHAP 值来解决特征之间的“合理分配”问题。SHAP 可以帮助更好地理解机器学习模型的预测结果, 通过计算每个特征对预测结果的贡献, 以可视化的方式呈现每个特征对预测结果的贡献, 使解释结果更加直观和易于理解。

针对每一组钻孔效率影响因素及对应的钻孔效率预测结果  $\psi_{SL}(x_i)$ , 分别计算每个钻孔效率影响因素的 SHAP 值。设第  $i$  个钻孔效率样本为  $x_i$ , 第  $i$  个钻孔效率样本的第  $j$  个钻孔效率影响因素为  $x_{ij}$ , 模型对该样本的钻孔效率预测值为  $\psi_{SL}(x_i)$ , 整个模型的基线(模型没有任何输入特征时的预测结果均值)为  $y_{base}$ ,  $f(x_{ij})$  为第  $i$  个钻孔效率样本的第  $j$  个钻孔效率影响因素的 SHAP 值, 则 SHAP 值服从式 (7), 即

$$\psi_{SL}(x_i) = y_{base} + f(x_{i1}) + f(x_{i2}) + \dots + f(x_{ij}) \quad (7)$$

$f(x_{ij})$  的计算式为

$$f(x_{ij}) = \sum_{H \subseteq F \setminus \{j\}} \frac{|H|!(|F|-|H|-1)!}{|F|!} [d(H \cup \{j\}) - d(H)] \quad (8)$$

式中:  $F$  为钻孔效率样本  $x_i$  所有影响因素的全集;  $H$  为钻孔效率样本  $x_i$  中任意多个钻孔效率影响因素形成的子集;  $d(H)$  为子集  $H$  中所包括的钻孔效率影响因素共同作用所产生的贡献;  $d(H \cup \{j\}) - d(H)$  为钻孔效率影响因素  $j$  为该共同作用带来的贡献。

计算每个钻孔效率影响因素对于钻孔效率预测结果的影响程度。对所有钻孔效率样本的钻孔效率影响因素  $j$  累加求均值, 即为钻孔效率影响因素  $j$  的 SHAP 值  $f(x_j)$ , 如式 (9) 所示。

$$f(x_j) = \sum_{i=1}^M f(x_{ij}) \quad (9)$$

式中  $M$  为钻孔效率样本  $x_i$  的样本总数。

### 3 工程实例

#### 3.1 数据采集与预处理

本研究以中国西南某坝料土石方开挖工程为研究对象, 收集了土石方开挖的钻孔效率及其影响因素数据集, 如图 3 所示。通过现场调查、搜集专家意见以及对实测数据的分析, 综合考虑了地质、作业、环境和机械特性对钻孔效率的影响, 选择了 8 个特征作

为钻孔效率的影响因素, 如表 1 所示。土石方开挖钻孔效率预测数据集以高程、钻孔深度、气温、劳动组合情况、加钻杆数量、岩石性质、钻机型号、作业时段为输入特征, 以钻孔效率为输出结果, 最终样本量为 150 组。

在上述输入特征中, 高程指钻孔工作面的高度, 其变化可能会影响气压、机械性能和人员状态; 钻机的压力会随着钻孔深度的变化而变化, 从而改变钻孔效率; 不同的气温和岩石性质可能直接影响钻机的效率; 不同类型的钻机型号在性能和效率上各不相同; 不同的作业时段会影响照明, 从而影响钻杆的装配和操作; 钻孔效率与换杆时间有关, 加钻杆数量越少, 钻孔效率越高; 对于劳动组合情况, 在原有人员的基础上增加辅助人员, 可以显著减少增加钻杆所需的时间, 从而提高钻孔效率。



图 3 某工程料场开挖作业

Fig.3 Excavation operations in one construction yard

表 1 钻孔效率影响因素

Tab.1 Influencing factors of drilling efficiency

类别	影响因素	特征描述
地质	岩石性质	0 代表板岩, 1 代表砂岩
作业	高程/m	工作面高度
	劳动组合情况/人	每个钻孔操作机器的工人人数
	作业时段	晴天或多云; 0 表示多云, 1 表示晴天
	加钻杆数量/根	每个钻孔安装钻杆的数量
	钻孔深度/m	每个钻孔的深度
环境	气温/°C	钻孔时的大气温度
机械特性	钻机型号	(0,0)、(0,1)、(1,1)、(1,0) 代表不同类型的钻机

#### 3.2 土石方开挖钻孔效率预测结果分析

将钻孔样本数据集按 7 : 3 随机划分为训练集和测试集, 训练集用于训练预测模型, 测试集用于检测预测模型的预测性能。采用 3 折交叉验证方法训练 QIHPO-SL 集成学习模型, 其中, QIHPO 的种群数量设置为 200, 最大迭代次数设置为 100。QIHPO-SL 集成学习模型训练完成后, 对测试集钻孔效率进行预测, 预测结果如图 4 所示。从图 4 中可以看出, 钻孔

效率预测值与真实值拟合较好。

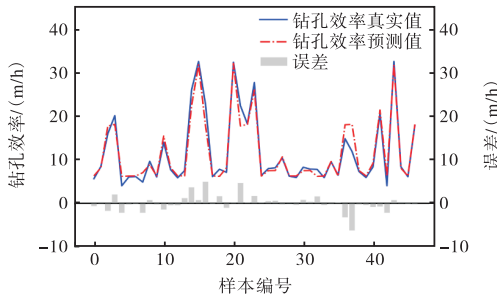


图4 QIHPO-SL集成学习模型的钻孔效率预测效果

Fig.4 Drilling efficiency prediction effect based on QIHPO-SL ensemble learning model

采用决定系数  $R^2$ 、均方误差 MSE、平均绝对百分比误差 MAPE、均方根误差 RMSE 4 种误差指标评价本研究提出的 QIHPO-SL 模型预测性能, 具体计算式如式(10)~(13)所示。

$$R^2 = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (\bar{y}_i - y_i)^2} \quad (10)$$

$$E_{MS} = \frac{1}{n} \sum_i (y_i - \hat{y}_i)^2 \quad (11)$$

$$E_{MAP} = \frac{100\%}{n} \sum_i \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (12)$$

$$E_{RMS} = \sqrt{\frac{1}{n} \sum_i (y_i - \hat{y}_i)^2} \quad (13)$$

式中:  $\hat{y}_i$  为钻孔效率预测值;  $y_i$  为钻孔效率真实值;  $\bar{y}_i$  为真实值  $y_i$  的平均值;  $n$  为样本的数量。

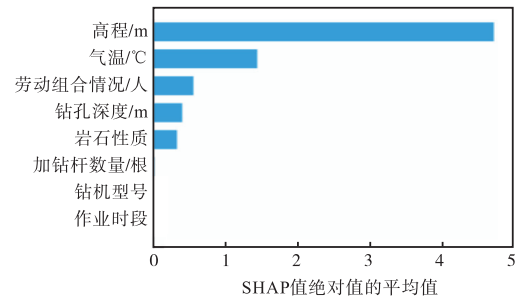
分析预测结果的误差指标可知,  $R^2$ 、 $E_{RMS}$ 、 $E_{MS}$  和  $E_{MAP}$  分别为 0.8982、0.3282、0.1077 和 0.1010, 因此 QIHPO-SL 集成学习模型的预测结果较为准确。

### 3.3 钻孔效率致因分析

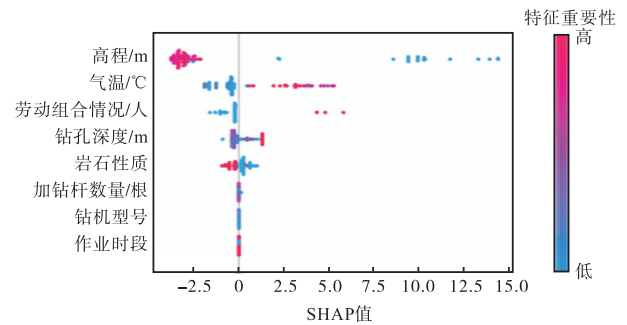
为进一步探究钻孔效率致因机理以挖掘影响模型预测结果的关键特征, 采用可解释机器学习 SHAP 对钻孔效率预测模型进行全局解释及局部解释, 提高 QIHPO-SL 集成学习模型的可解释性。

图 5(a) 为钻孔效率影响因素的特征重要性排序。图 5(b) 所示的散点图中, 每个点表示钻孔效率样本点对应的每个钻孔效率影响因素的 SHAP 值, 并用色彩梯度表达各影响因素数值间的关系。图 5(b) 中的 x 轴为 SHAP 值, 用以衡量影响因素对模型预测值的贡献程度和影响; y 轴为影响因素, 按照重要性从上至下排列。每个点的颜色越红, 表示影响因素 SHAP 值越高, 对模型预测结果贡献越大; 反之, 每个点的颜色越蓝, 表示影响因素 SHAP 值越低, 对模型预测结果贡献越小。

结合图 5(b), 高程、气温及劳动组合情况对钻孔效率预测结果影响较大。高程对于钻孔效率预测结果影响最大; 当高程逐渐增大时, 其对应的 SHAP 值小于 0, 当高程逐渐减小时, 其对应的 SHAP 值大于 0, 高程对钻孔效率预测结果的正影响效果较为显著; 在料场开挖钻孔过程中, 高程越大, 作业面越狭窄, 随着钻孔作业的高程逐渐降低, 作业面越宽阔, 钻孔效率会随之提升。对于气温, 当其特征取值增大时, 其对应的 SHAP 值大于 0; 当气温的特征取值减小时, 其对应的 SHAP 值小于 0, 气温对钻孔效率预测结果的正影响效果大于其对钻孔效率预测结果的负影响效果; 在料场开挖钻孔过程中, 低温对机械效率有一定影响, 钻孔效率随气温降低而下降, 气温升高, 料场开挖强度增大, 钻孔效率随之增大。对于劳动组合情况, 当其特征取值偏大时, 对应的 SHAP 值大于 0, 当其特征取值偏小时, 对应的 SHAP 值小于 0, 表明钻孔效率会随钻机操作人员数量增加而增大。



(a) 特征重要性排序



(b) 特征重要性散点图

图5 钻孔效率影响因素特征重要性排序及特征重要性散点图

Fig.5 Feature importance ranking and feature importance scatterplot of influencing factors of drilling efficiency

事实上, 钻孔效率受到多种影响因素的影响, 仅关注单一特征无法全面解析钻孔效率改变的致因结果, 因此有必要进一步探究多因素耦合作用下对钻孔效率的影响。由于交互特征组合较多, 且高程是影响模型预测结果最重要的特征, 因此研究重点以高程、气温、劳动组合情况与钻孔深度作为研究范例。如图

6 所示, SHAP 依赖图可以揭示 2 个特征的交互作用, 提供了特征变量对钻孔效率预测值的边际效应, 其中  $x$  轴表示某一输入特征的特征值,  $y$  轴为该特征的 SHAP 值, 颜色梯度表示另一输入特征的特征值.

以图 6(a) 与 6(b) 为例, 这两个分图展示了高程与气温相互作用对钻孔效率预测值的总体效果. 综合分析气温与高程可知, 气温升高时气温的 SHAP

值增大, 对钻孔效率预测结果有正影响; 高程增大时高程的 SHAP 值减小, 对钻孔效率预测结果有负影响. 然而, 在气温与高程共同作用下, 高程升高至 2850 m 以上时气温较低, 此时 SHAP 值小于 0, 高程与气温共同作用导致钻孔效率降低; 高程降低至 2850 m 以下时气温升高, 此时 SHAP 值大于 0, 气温与高程共同作用导致钻孔效率升高.

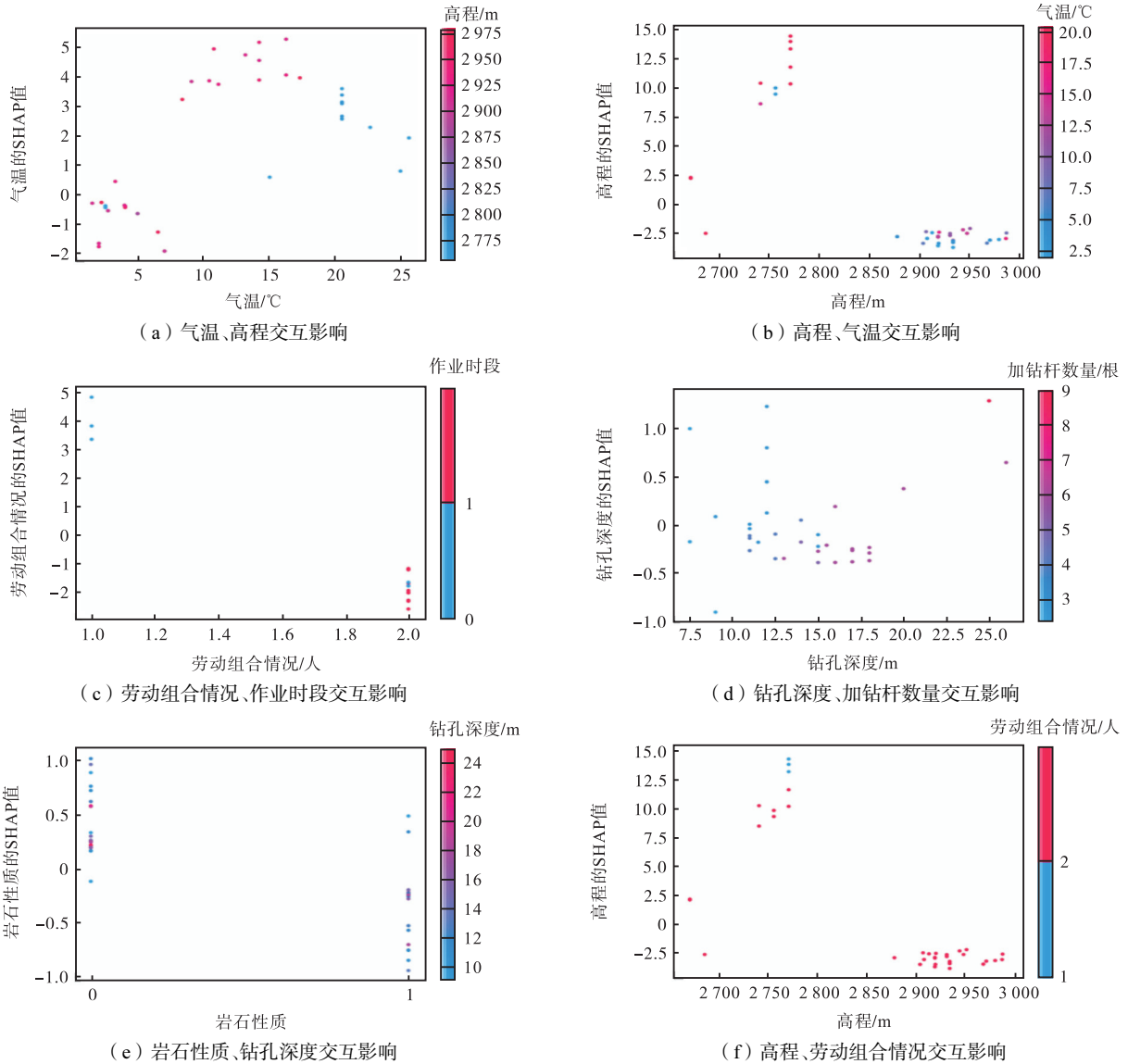


图 6 钻孔效率影响因素对模型预测结果的交互影响分析

Fig.6 Analysis of interaction effect of influencing factors of drilling efficiency on model prediction results

随机抽取两个样本进行 SHAP 局部可解释分析, 如图 7 所示, 图中  $x$  轴为钻孔效率预测结果,  $y$  轴为钻孔效率影响因素的特征值. 条状图若为红色, 则该钻孔效率影响因素的 SHAP 值为正; 条状图若为蓝色, 则该钻孔效率影响因素的 SHAP 值为负. 以图 7(b) 为例, 钻孔效率影响因素对样本的预测结果影响从大到小分别为高程、气温、钻孔深度、岩石性质、劳动组合情况、加钻杆数量、作业时段及钻机型号. 样

本 6 中, 高程的特征值为 2877 m, SHAP 值为-2.82, 高程使样本 6 的钻孔效率预测值减小; 气温的特征值为 4.95 °C, SHAP 值为-0.63, 气温使样本 6 的钻孔效率预测值减小; 钻孔深度的样本特征值为 18 m, SHAP 值为-0.28, 钻孔深度使样本 6 的钻孔效率预测值减小; 岩石性质的特征取值为 0, SHAP 值为 0.24, 岩石性质使样本 6 的钻孔效率预测值增加.

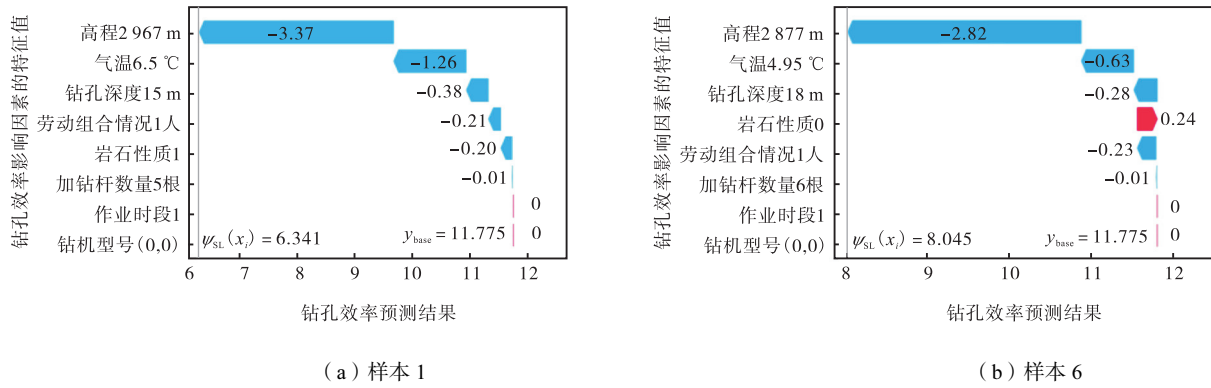


图7 单个样本预测  
Fig.7 Prediction based on single samples

### 3.4 对比分析与讨论

#### 3.4.1 QIHPO 性能对比分析

本节以典型单峰标准函数和多峰标准函数为测试函数,将本文提出的 QIHPO 算法与杜鹃搜索 (cuckoo search, CS)算法、鼠群优化 (rat swarm opti-

mizer, RSO)算法、原子搜索优化 (atom search optimization, ASO)算法、粒子群优化 (particle swarm optimization, PSO)算法、HPO 算法等算法进行比较.各算法的种群设置为 50,最大迭代次数设置为 500,对计算的结果进行对比分析,如图 8 所示.

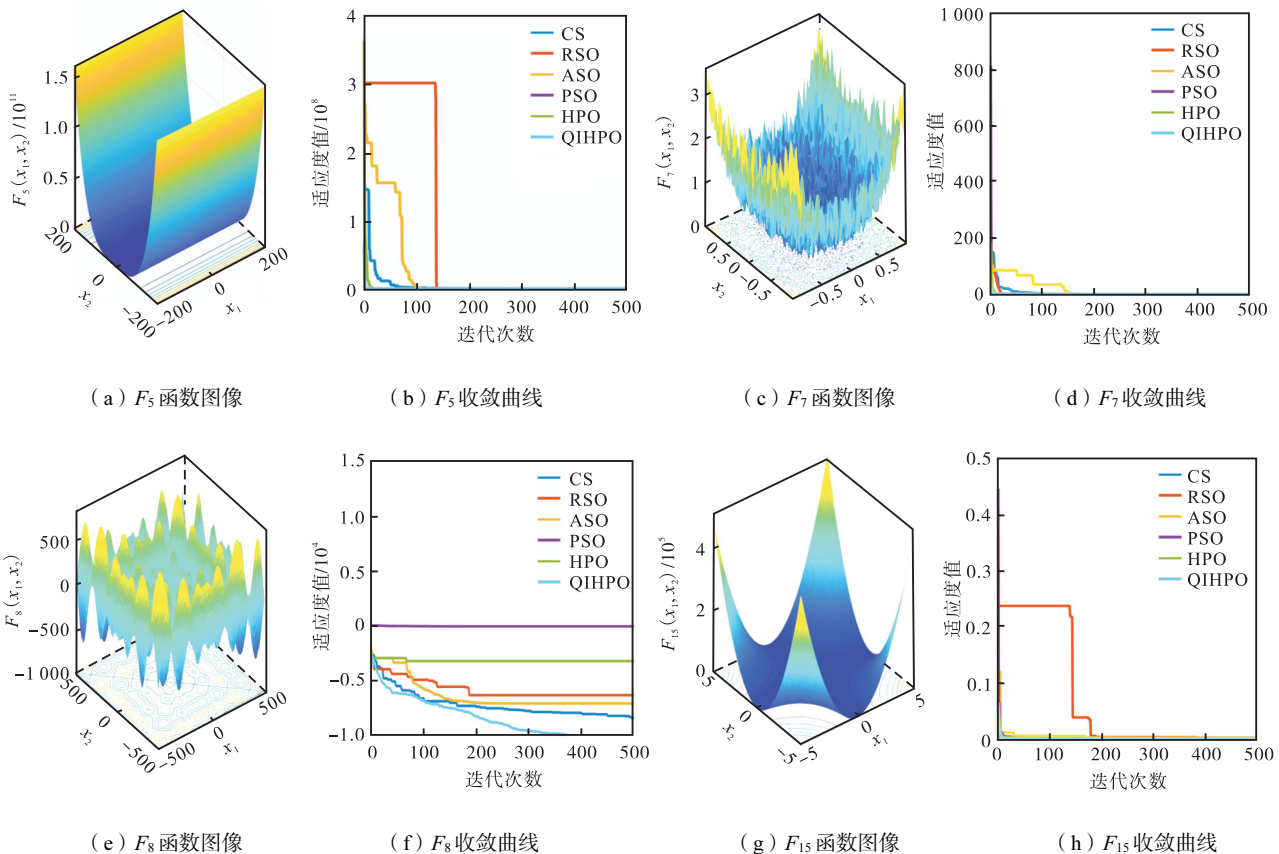


图8 各算法收敛曲线  
Fig.8 Convergence curves for different algorithms

从图 8 可以看出,在所选择的单峰测试函数、多峰测试函数上, QIHPO 算法较其他算法具备更好的收敛性.由于引入了 Q 学习策略改进 HPO 算法局部搜索过程,猎物能够迅速靠近猎物的最优位置附近,

且猎物能够跳出局部最优,快速找到全局最优解,因此 QIHPO 具备更快的收敛速度和更低的适应度极值.综上所述, QIHPO 算法相比于其他算法,具备较快的收敛速度、较强的搜索能力和较高的寻优精度.

### 3.4.2 QHPO-SL 集成学习模型预测性能对比分析

为了验证 QIHPO-SL 集成学习模型的钻孔效率预测效果,采用  $R^2$ 、 $E_{MS}$ 、 $E_{MAP}$  和  $E_{RMS}$  这 4 种误差指标,将本文模型与 QIHPO 优化的传统机器学习方法构建的单一预测模型以及未使用 QIHPO 优化的超级学习器模型(即 QIHPO-XGB、QIHPO-RF 和 SL)进行对比,结果如图 9 所示。

由图 9 可知,QIHPO-XGB、QIHPO-RF 和 SL 这 3 种机器学习模型的误差评价指标值均处于较低水平,具备良好的预测性能。然而,QIHPO-SL 超级学习器模型的  $R^2$ 、 $E_{MS}$ 、 $E_{MAP}$  和  $E_{RMS}$  值分别为 0.898 2、0.198 7、0.101 0 和 0.328 2;相比于 QIHPO-XGB 模型、QIHPO-RF 模型和 SL 模型,其  $R^2$  值分别提高了 12.94%、12.02% 和 1.58%, $E_{MS}$  值分别降低了 50.21%、48.60% 和 12.15%, $E_{MAP}$  值分别降低了 13.75%、15.55% 和 22.78%, $E_{RMS}$  值分别降低了 29.44%、28.31% 和 6.28%。由此可知,QIHPO-SL 集成学习模型通过具有互补特征的异构基学习器捕捉样本特征差异性,其预测性能明显优于单一机器学习模型预测性能,钻孔效率预测结果更加精确。

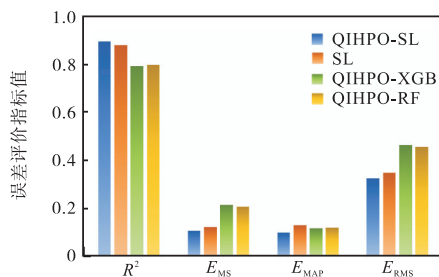


图 9 钻孔效率预测值误差分析结果

Fig.9 Results of error analysis of predicted values of drilling efficiency

## 4 结 论

土石方开挖钻孔效率直接影响料场开挖进度,受到地质、作业、环境和机械特性等众多因素影响,精确预测钻孔效率并分析影响因素对钻孔效率影响程度,对料场开挖进度管理至关重要。本文建立了土石方开挖钻孔效率预测可解释超级学习器集成学习模型,弥补了现有钻孔效率模型超参数寻优效率低、预测模型精度低、可解释性差的不足,并通过工程实例及对比分析,验证了所提方法的优越性。本文得到的主要结论如下。

(1) 本文提出基于 GB 元学习器集成 SVR、LASSO、XGB、RF、DT、GB、ET、KNN 和 LR 共 9 种机器学习算法的超级学习器模型,该模型实现了不同

算法优势互补,弥补了单一机器学习模型精度低的缺点,将 Q 学习引入 HPO 算法的局部搜索过程进行改进,并优化超级学习器算法的超参数,实现了钻孔效率的高精度预测。

(2) 将可解释机器学习 SHAP 理论与 QIHPO-SL 集成学习模型结合对钻孔效率进行致因分析,挖掘影响钻孔效率的关键特征,并解释不同因素对钻孔效率的影响程度,增强了 QIHPO-SL 集成学习模型的可解释性。

(3) 案例研究表明,QIHPO-SL 集成学习模型能够快速精准地预测钻孔效率,其预测精度相比于 QIHPO-XGB、QIHPO-RF 和 SL 算法,分别提高了 12.94%、12.02% 和 1.58%;通过可解释机器学习 SHAP 致因分析,得出高程、气温、劳动组合情况和钻孔深度是影响钻孔效率预测结果的关键因素;SHAP 方法不仅可以挖掘关键特征对钻孔效率预测结果的影响,还能够在全局解释中揭示输入特征与预测结果之间的正负相关趋势以及特征之间的相互作用。SHAP 方法从全局及局部对模型预测结果进行解释,提高了超级学习器模型的可解释性。

## 参考文献:

- [1] 马洪琪, 钟登华, 张宗亮, 等. 重大水利水电工程施工实时控制关键技术及其工程应用[J]. 中国工程科学, 2011, 13(12): 20-27.  
Ma Hongqi, Zhong Denghua, Zhang Zongliang, et al. Key technologies of real-time construction control for major hydraulic and hydroelectric projects[J]. Engineering Sciences, 2011, 13(12): 20-27(in Chinese).
- [2] 赖永明, 吴学雷, 申明亮, 等. 基于土石方平衡的面板堆石坝料场开采建模研究[J]. 中国农村水利水电, 2011(9): 75-77, 81.  
Lai Yongming, Wu Xuelei, Shen Mingliang, et al. Research on quarry excavation modeling of concrete face rockfill dams based on earthwork balance[J]. China Rural Water and Hydropower, 2011(9): 75-77, 81(in Chinese).
- [3] Kahraman S, Bilgin N, Feridunoglu C. Dominant rock properties affecting the penetration rate of percussive drills[J]. International Journal of Rock Mechanics and Mining Sciences, 2003, 40(5): 711-723.
- [4] 欧江萍. 分析影响煤田钻探效率的原因及其解决方法[J]. 民营科技, 2010(8): 3.  
Ou Jiangping. Analyzing the reasons affecting the efficiency of coalfield drilling and their solutions[J]. Sci-

- ence and Technology of Non-State-Run Enterprises, 2010(8): 3(in Chinese).
- [5] 张 伟. 金刚石绳索取心钻进施工效率影响因素分析[J]. 探矿工程(岩土钻掘工程), 2007(10): 22-24, 34.  
Zhang Wei. Analysis on factors influencing drilling efficiency of diamond wire-line drilling[J]. Exploration Engineering(Rock & Soil Drilling and Tunneling), 2007(10): 22-24, 34(in Chinese).
- [6] Bourgoynne A T, Young F S. A multiple regression approach to optimal drilling and abnormal pressure detection[J]. Society of Petroleum Engineers Journal, 1974, 14(4): 371-384.
- [7] Hareland G, Rampersad P R. Drag-bit model including wear[C]//Proceedings of the SPE Latin America/Caribbean Petroleum Engineering Conference. Buenos Aires, Argentina: SPE, 1994.
- [8] Winters W J, Warren T M, Onyia E C. Roller bit model with rock ductility and cone offset[C]//Proceedings of the SPE Annual Technical Conference and Exhibition. Dallas, USA: SPE, 1987.
- [9] Wallace S P, Hegde C M, Gray K E. A system for real-time drilling performance optimization and automation based on statistical learning methods[C]//Proceedings of the SPE Middle East Intelligent Oil and Gas Conference and Exhibition. Abu Dhabi, UAE: SPE, 2015.
- [10] Zhou Y, Lu C D, Zhang M L, et al. A novel rate of penetration model based on support vector regression and modified bat algorithm[J]. IEEE Transactions on Industrial Informatics, 2023, 19(5): 6659-6668.
- [11] Mehrad M, Bajolvand M, Ramezanzadeh A, et al. Developing a new rigorous drilling rate prediction model using a machine learning technique[J]. Journal of Petroleum Science and Engineering, 2020, 192: 107338.
- [12] Zhao Y R, Noorbakhsh A, Koopialipoor M, et al. A new methodology for optimization and prediction of rate of penetration during drilling operations[J]. Engineering with Computers, 2020, 36: 587-595.
- [13] Elkatatny S. Real-time prediction of rate of penetration while drilling complex lithologies using artificial intelligence techniques[J]. Ain Shams Engineering Journal, 2021, 12(1): 917-926.
- [14] Gan C, Cao W H, Wu M, et al. Two-level intelligent modeling method for the rate of penetration in complex geological drilling process[J]. Applied Soft Computing, 2019, 80: 592-602.
- [15] Zhou Y, Chen X, Zhao H B, et al. A novel rate of penetration prediction model with identified condition for the complex geological drilling process[J]. Journal of Process Control, 2021, 100: 30-40.
- [16] Alsaihati A, Elkatatny S, Gamal H. Rate of penetration prediction while drilling vertical complex lithology using an ensemble learning model[J]. Journal of Petroleum Science and Engineering, 2022, 208: 109335.
- [17] Lü F, Yu J, Zhang J, et al. A novel stacking-based ensemble learning model for drilling efficiency prediction in earth-rock excavation[J]. Journal of Zhejiang University-Science A, 2022, 23(12): 1027-1046.
- [18] Gada M, Haria Z, Mankad A, et al. Super learner: Stack generalization algorithm for AutoML[C]//Proceedings of the 2021 12th International Conference on Computing Communication and Networking Technologies(ICCCNT). Kharagpur, India: IEEE, 2021.
- [19] Mohan B, Badra J. A novel automated SuperLearner using a genetic algorithm-based hyperparameter optimization[J]. Advances in Engineering Software, 2023, 175: 103358.
- [20] Neto H N C, Lopez M A, Fernandes N C, et al. MineCap: Super incremental learning for detecting and blocking cryptocurrency mining on software-defined networking[J]. Annals of Telecommunications, 2020, 75: 121-131.
- [21] Srivani I, Sridhar M, Swamy K C T, et al. Multi-class classification of ionospheric scintillations using SMOTE-Super Learner ensemble technique[J]. Advances in Space Research, 2024, 73(7): 3845-3854.
- [22] Tyrallis H, Papacharalampous G, Langousis A. Super ensemble learning for daily streamflow forecasting: Large-scale demonstration and comparison with multiple machine learning algorithms[J]. Neural Computing and Applications, 2021, 33: 3053-3068.
- [23] 陈晨鑫, 郭晓晶, 许金芳, 等. 基于 Super Learner 的边缘结构模型研究及发展[J]. 药物流行病学杂志, 2022, 31(2): 101-106.  
Chen Chenxin, Guo Xiaojing, Xu Jinfang, et al. Introduction about researches and development of marginal structural models based on Super Learner[J]. Chinese Journal of Pharmacoepidemiology, 2022, 31(2): 101-106(in Chinese).
- [24] 孙浩浩, 章剑林, 张子莹, 等. 基于 Super Learner 集成学习模型对脑卒中的预测研究[J]. 杭州师范大学学报(自然科学版), 2023, 22(6): 590-597, 610.

- Sun Haohao, Zhang Jianlin, Zhang Ziyang, et al. A study of stroke prediction based on Super Learner integrated learning model[J]. Journal of Hangzhou Normal University(Natural Science Edition), 2023, 22(6): 590-597, 610(in Chinese).
- [25] 尹传洪. 基于 MFS-SUPL 模型的支付欺诈检测研究[D]. 武汉: 中南财经政法大学统计与数学学院, 2022.
- Yin Chuanhong. Research on Payment Fraud Detection Based on MFS-SUPL Model[D]. Wuhan: School of Statistics and Mathematics, Zhongnan University of Economics and Law, 2022(in Chinese).
- [26] Wen X, Xie Y C, Wu L T, et al. Quantifying and comparing the effects of key risk factors on various types of roadway segment crashes with LightGBM and SHAP[J]. Accident Analysis & Prevention, 2021, 159: 106261.
- [27] Mangalathu S, Hwang S H, Jeon J S. Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations(SHAP) approach[J]. Engineering Structures, 2020, 219: 110927.
- [28] Lundberg S M, Lee S-I. A unified approach to interpreting model predictions[C]//31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc., 2017.
- [29] Parsa A B, Movahedi A, Taghipour H, et al. Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis[J]. Accident Analysis & Prevention, 2020, 136: 105405.
- [30] Wong J, Manderson T, Abrahamowicz M, et al. Can hyperparameter tuning improve the performance of a Super Learner? A case study[J]. Epidemiology, 2019, 30(4): 521-531.
- [31] Parameswari A, Ganeshan R, Ragavi V, et al. Hybrid rat swarm hunter prey optimization trained deep learning for network intrusion detection using CNN features[J]. Computers & Security, 2024, 139: 103656.
- [32] 余红玲, 王晓玲, 任炳昱, 等. 土石坝渗流性态分析的 IAO-XGBoost 集成学习模型与预测结果解释[J]. 水利学报, 2023, 54(10): 1195-1209.
- Yu Hongling, Wang Xiaoling, Ren Bingyu, et al. IAO-XGBoost ensemble learning model for seepage behavior analysis of earth-rock dam and interpretation of prediction results[J]. Journal of Hydraulic Engineering, 2023, 54(10): 1195-1209(in Chinese).
- [33] Naruei I, Keynia F, Molahosseini A S. Hunter-prey optimization: Algorithm and applications[J]. Soft Computing, 2021, 26: 1279-1314.
- [34] Zhao J, Hosseini S, Chen Q Y, et al. Super learner ensemble model: A novel approach for predicting monthly copper price in future[J]. Resources Policy, 2023, 85: 103903.

(责任编辑: 武立有)