

杨溢,王淑贤,沈亚玺,等. 基于可解释 GA-VR 集成学习的爆破质点峰值振动速度预测[J]. 安全与环境学报, 2026, 26(6): 2084-2093.

YANG Y, WANG S X, SHEN Y X, et al. Predicting peak vibration velocity of blasting particles using interpretable GA-VR integrated learning[J]. Journal of Safety and Environment, 2026, 26(6): 2084-2093.

基于可解释 GA-VR 集成学习的爆破质点峰值振动速度预测*

杨 溢¹,王淑贤¹,沈亚玺²,石玉莲³,丁秋月^{1,4}

(1 昆明理工大学公共安全与应急管理学院,昆明 650032;2 天津大学建筑工程学院,天津 300072;

3 昆明理工大学国土资源工程学院,昆明 650032;4 云南智科安全咨询有限公司,昆明 650032)

摘 要:为提升爆破质点峰值振动速度(Peak Particle Velocity, PPV)的预测精度,提出了一种基于遗传算法(Genetic Algorithm, GA)优化的投票回归(Voting Regressor, VR)集成模型(GA-VR)。基于某露天矿实测数据,采用 GA 优化梯度提升回归、随机森林回归和决策树回归的超参数,并通过投票机制集成模型;同时利用夏普利加性解释(SHapley Additive exPlanations, SHAP)从全局和局部角度分别解释影响因素预测 PPV 的机制。结果表明:GA-VR 模型在测试数据上 R^2 为 0.948,预测性能优于 6 种对比模型及 5 类异质优化算法。通过 SHAP 分析可知影响因素中测点与爆破位置水平距离占主导因素,其次为最大段药量、测点与最小抵抗线夹角。提出的模型在保障高精度预测的同时提供了具有可解释的预测依据,可用于指导爆破参数的智能优化与安全控制。

关键词:安全工程;爆破振动预测;遗传算法;集成学习;可解释分析;特征工程

中图分类号:X932;X936

文献标志码:A

文章编号:1009-6094(2026)06-2084-10

DOI: 10.13637/j.issn.1009-6094.2025.0770

0 引 言

爆破工程技术在矿山开采、隧道建设等领域广泛应用,但爆破过程中仅有 20%~30% 的炸药能量用于破碎岩石,其余能量主要通过地震波、地面振动及飞石等形式耗散。这些能量耗散现象不仅降低效率,更会威胁周边环境与建筑安全,尤其爆破振动是影响施工安全和建(构)筑物结构稳定性的主要因素^[1]。爆破质点峰值振动速度(Peak Particle Velocity, PPV)是评估地面振动影响的关键指标,建立精准的 PPV 预测模型对保障爆破作业安全高效意义重大,是爆破工程研究的关键问题和难点问题^[2]。

早期 PPV 预测研究主要依托经验公式构建预测模型,通过对爆破参数与地质条件进行线性简化处理实现预测,然而在复杂地质环境及多变量耦合作用场景下存在显著局限性,常引发预测结果与实测数据间的系统性偏差^[3-4]。相较而言,机器学习方法通过集成多维特征参数并构建非线性映射关系,显著增强了模型对复杂系统关联特征的解析能力,进而实现预测精度的突破性提升。

随着计算技术的快速发展,机器学习方法在 PPV 预测中得到广泛应用^[5-8]。2020 年,郭钦鹏等^[6]运用遗传算法(Genetic Algorithm, GA)对反向传播(Back Propagation, BP)神经网络的权值与阈值进行优化,构建了 GA-BP 神经网络预测模型,使用具体爆破工程监测数据作为数据集进行模型预测验证,并与 BP 神经网络模型的预测结果进行对比。范勇等^[2]将 BP 神经网络解决复杂非线性函数逼近能力和粒子群优化(Particle Swarm Optimization, PSO)算法全局寻优能力相结合,建立了改进的 PSO-BP 神经网络预测模型,将提出的模型与萨道夫斯基公式的检验结果进行对比可知,改进的 PSO-BP 神经网络预测模型的预测值与实测值吻合更好,预测的结果更为可靠。Wang 等^[9]基于现场监测与数值模拟,对经验方程的标度距离进行了改进,提出了考虑爆破引起的振动在相邻隧道周围的衍射和反射放大效应的新方程,在预测紧密相邻岩石隧道中的 PPV 时取得了更加准确的结果。2023 年,张西良等^[10]首次将浅部地下矿山爆破数据中有用的知识迁移至深部矿山爆破振动速度预测模型中,提出一种逻辑回归(Logistic Regression, LR)迁移学习算法(LR-

* 收稿日期:2025-05-08

作者简介:杨溢,教授,博士生导师,从事爆破工程、安全工程研究,2919847230@qq.com。

基金项目:国家自然科学基金项目(41861134008)

TrAdaBoost), 同时将该算法与支持向量回归机 (Support Vector Regression, SVR)、回归迁移学习 (TrAdaboost-R2) 算法分别进行预测和对比, 结果可知, 所提出模型的预测误差更低, 性能更好。李振阳等^[11] 使用 PSO 算法优化极端梯度提升树 (Extreme Gradient Boosting, XGBoost), 构建了一种参数自优化的 PSO-XGBoost 预测模型, 将该模型与传统 XGBoost 模型、SSA-XGBoost 优化模型及萨道夫斯基经验公式的 PPV 回归预测进行了对比分析。高宇璠等^[1] 建立了一种融合鲸鱼优化算法 (Whale Optimization Algorithm, WOA) 与信息准则 (Information Criterion, IC) 的人工神经网络 (Artificial Neural Network, ANN) 爆破振动预测模型, 该研究通过整合定性定量地质参数构建数据集, 系统对比了 WOA-IC-ANN 模型与传统经验公式等方法的预测性能, 结果表明, 该模型显著提升了预测精度与泛化能力。上述研究通过机器学习技术显著改善了传统方法的局限性并提升了预测精度。然而, 由于机器学习模型固有的“黑箱”特性, 其无法充分揭示变量之间的潜在交互机制, 导致模型内部预测逻辑和变量因果关系的阐释受限, 制约了模型在 PPV 预测及爆破控制等应用场景中的可解释性和可信度, 现有研究很少深入探讨模型与个体 PPV 预测案例中特征值具体参数之间的关联性。

为解决上述研究瓶颈, 本文提出一种融合 GA 优化的集成投票回归 (Voting Regressor, VR) 模型 (GA-VR)。首先, 通过构建梯度提升回归 (Gradient Boosting Regressor, GBR)、随机森林回归 (Random Forest Regressor, RFR) 与决策树回归 (Decision Tree Regressor, DTR) 的多算法融合框架, 实现超参数协同优化与异质模型优势互补, 显著提升 PPV 预测的精度与稳健性。其次, 基于夏普利加性解释 (SHapley Additive exPlanations, SHAP) 框架构建可解释性分析体系, 通过特征贡献度量化评估与交互效应可视化, 同步提供模型预测的全局解释与局部解释。最后, 结合实际爆破工程案例验证模型准确性, 为爆破振动控制提供一种兼具预测性能与可解释性的智能决策支持工具, 推进机器学习在岩土工程领域的进一步应用。

1 矿山爆破振动数据库构建

基于特征重要性排序与相关性分析对所收集的数据进行预处理, 并将处理后的数据划分为训练集和测试集, 用于 PPV 预测模型的构建。

1.1 数据来源与处理

数据来源依托大冶有色金属集团铜绿山矿露天采场爆破振动控制工程案例^[12], 通过采集爆破监测数据, 并剔除异常记录数据和信息记录不全的数据, 最终获取 137 组有效样本数据。结合前人研究与监测数据统计特性, 构建了包含 8 个爆破动力参数与 3 个地质构造参数的多维度特征空间, 如表 1 所示。

表 1 参数列表
Table 1 Parameter list

特征类别	参数名称
爆破动力特征	最大段药量 X_1
	总药量 X_2
	水平距离 X_3
	高差 X_4
	前排抵抗线 X_5
	微差时间 X_6
	预裂缝穿透率 X_7
	炸药爆速 X_8
地质构造特征	完整性系数 X_9
	坚固性系数 X_{10}
	测点与最小抵抗线方向夹角 X_{11}

首先基于 XGBoost 算法, 建立基于增益权重的特征重要性量化评估体系, 为 PPV 预测的特征筛选提供评估标准; 同时引入 Pearson 相关系数分析方法, 评估各特征指标间的相关性。

1.1.1 特征重要性排序

采用 Chen 等^[13] 于 2016 年提出的 XGBoost 算法对 PPV 预测特征进行重要性排序, 通过迭代优化可微损失函数并融合正则化策略增强模型泛化能力。在建模过程中, 构建爆破动力特征与地质构造特征组成的多维输入空间, 以 PPV 为预测目标, 设置关键超参数: 最大树深度 (max_depth) 为 8, 子节点最小样本权重 (min_child_weight) 为 1, 全局偏置项初始化值 (base_score) 为 0.5。特征重要性评估结果如图 1 所示。

1.1.2 特征相关性分析

根据上述特征重要性评估结果, 结合特征相关性分析以简化机器学习的输入特征维度, 进而提升模型预测性能。采用 Pearson 相关系数方法 [式 (1)], 评估各参数之间的相关性。具有强相关性 (Pearson 系数绝对值 > 0.6) 的两个特征将结合重要

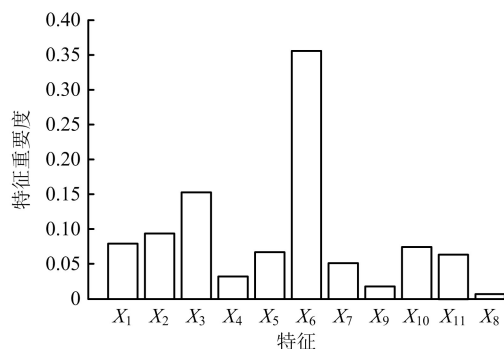


图1 特征重要性评估

Fig.1 Feature importance evaluation

性排序结果,删除重要度较低的特征。该筛选机制有效平衡了特征冗余消除与关键信息保留的工程需求。

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (1)$$

式中 r 表示皮尔逊相关系数; X_i 和 Y_i 分别表示变量 X 和 Y 的样本值; $\bar{X}$ 和 $\bar{Y}$ 是相应变量的平均值。

图2展示了特征参数间线性关联强度的定量表征结果。爆破动力参数中的水平距离与高差之间存在显著正相关,地质构造参数中的完整性系数与坚固性系数呈现显著线性关系。基于本文提出的特征工程,结合图1和2结果,高差与完整性系数被视为冗余特征并剔除,构建由7个爆破动力参数(最大段药量、总药量、水平距离、前排抵抗线、微差时间、预裂缝穿透率、炸药爆速)和2个地质构造参数(坚固性系数、测点与最小抵抗线方向夹角)组成的特征向量。

本文通过特征重要性排序和特征相关性分析,总结了一种面向PPV预测数据处理的特征工程流

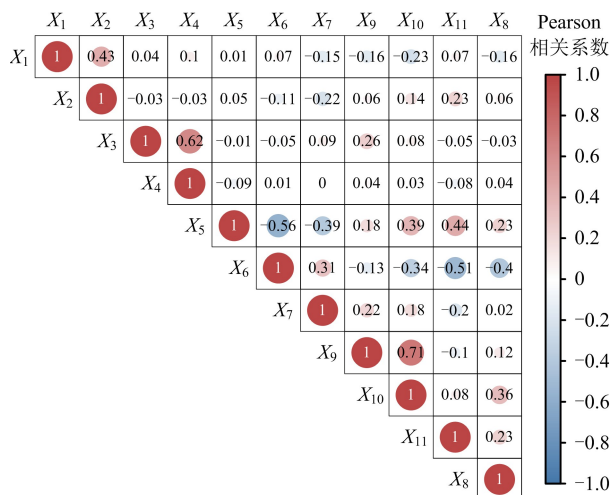


图2 相关矩阵图

Fig.2 Correlation matrix plot

程,通过多维度特征降维实现数据集的优化。首先,对所收集数据进行特征重要性排序;其次,计算皮尔逊相关系数分析特征间的相关性,基于预设的相关系数绝对值 >0.6 的判定标准,剔除高相关条件下的冗余特征;最终保留其余特征,构成优化后的数据集。该流程框架具备跨场景迁移能力,其模块化设计不仅适用于矿山爆破PPV预测数据集的处理,也为复杂高维数据提供了可复用的数据降维方法。

1.2 数据集划分

将处理后的数据集按照8:2的比例划分为训练集(109组样本)和测试集(28组样本)。训练集用于训练预测模型,测试集用于检验模型的预测性能。为了获得具有鲁棒性的预测模型,选用5折交叉验证方法对训练集进一步拆分,拆分为训练子集和验证子集,其基本思想如图3所示:将数据集均分为5个子集,每次迭代以1个子集作为验证集,其余4个作为训练集,通过5次迭代使每个子集均参与验证,最终平均5次评估结果获得综合性能指标^[14]。通

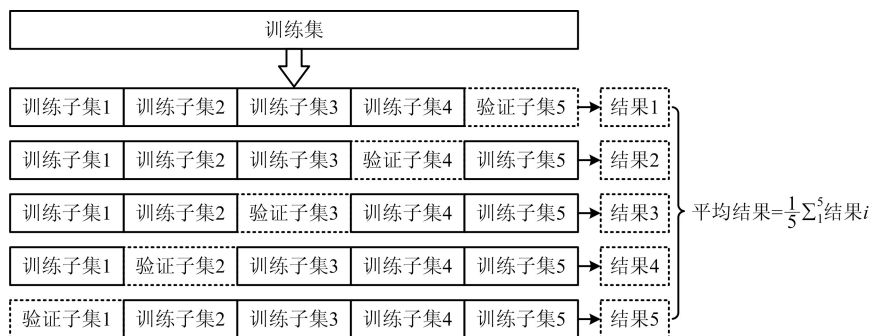


图3 5折交叉验证法

Fig.3 5-fold cross validation method

过 5 折交叉验证可在小样本数据集中有效利用训练集,从而获取模型最佳超参数。由于篇幅限制,仅列出部分测试集的样本,如表 2 所示。

2 模型构建

2.1 机器学习算法

2.1.1 投票回归(VR)模型

集成学习通过协同整合多个异质基估计器构建高可靠性识别系统,其预测精度与稳健性通常优于单一模型^[15]。集成投票模型作为该领域的通用框架,采用多数表决或均值融合策略聚合各基模型的预测输出,从而增强模型泛化能力,适用于分类与回归任务。集成模型的理论错误率可通过下式进行概率估计。

$$P(E) = \sum_{k=n/2}^n \binom{n}{k} p^k (1-p)^{n-k} \quad (2)$$

式中 $P(E)$ 表示集成模型的理论错误率; n 表示试验的总次数; k 表示在 n 次试验中发生“错误”的次数; p 表示每次试验发生“错误”的概率。可知,当 $p < 0.5$ 时, $P(E)$ 随 n 的增大单调递减。换言之,只要每个基分类的准确率大于 0.5,随着参与集成的基分类器增多, $P(E)$ 将单调递减,最终趋向于 0,集成性能不断提升。

集成学习方法依据基估计器类型可分为异质集成(异构算法组合)与同质集成(同构算法参数差异化调整)两类,实证研究证实异质集成架构在预测性能方面具有显著优势^[16]。基于此理论优势,遴选出 GBR^[17]、RFR^[18] 及 DTR^[19] 作为基估计器。本文构建的 VR 基于上述基回归器的独立训练结果,通过加权平均机制生成集成预测结果。该架构可有效克

服单一模型局限性,通过算法优势互补提升复杂数据集的预测效能。

2.1.2 GA-VR 模型

GA 作为 Holland^[20] 提出的仿生智能优化算法,通过生物进化算子(选择、交叉、变异)实现全局最优解搜索,具备全局收敛性与参数自适应性优势。本文首先对投票集成模型中的基估计器(GBR、RFR、DTR)分别进行参数调优,然后通过集成构建 VR,并基于 GA 进行权值优化获得 GA-VR 模型,并采用 SHAP 解析模型决策机制。

2.2 性能评价标准

采用决定系数 R^2 和均方误差 (Mean Squared Error, MSE) S_{ME} 评估模型的预测性能,计算公式如下。

$$R^2 = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (\bar{y}_i - y_i)^2} \quad (3)$$

$$S_{ME,i} = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2 \quad (4)$$

式中 y_i 表示样本点的真实值; $\hat{y}_i$ 表示模型预测值; $\bar{y}_i$ 则是所有样本点的真实平均值; m 是测试集中样本点的数量。MSE 越小, R^2 越接近 1,表示预测性能越好。

2.3 模型的可解释性

解释预测过程是解决机器学习黑箱问题的有效途径,能够帮助深入理解影响模型输出的因素,如特征重要性等^[21]。其中,SHAP 方法^[22] 基于合作博弈理论中的 Shapley 值,通过计算所有特征组合中单个特征的平均边际贡献,量化单个特征对模型预测的贡献度,从而实现全局可解释性。具体地,SHAP 根据确定的 9 个特征组合计算影响特定样本 x 预测结

表 2 爆破质点峰值振动速度样本测试集

Table 2 Sample test set of PPV of blasting particles

序号	X_1/kg	X_2/kg	X_3/m	X_5/m	X_6/ms	$X_7/\%$	X_{10}	$X_{11}/(^{\circ})$	$X_8/(\text{m} \cdot \text{s}^{-1})$	PPV/ $(\text{cm} \cdot \text{s}^{-1})$
1	468	936	75.5	6	50	0	6	180	2 800	3.350
2	494	3 952	183.2	5	100	0	5	120	2 800	0.105
3	1 180	7 080	91.2	6	25	70	8	180	4 200	3.093
4	1 660	8 300	175.6	4	50	0	6	60	2 800	0.445
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
25	1 790	5 370	363.8	4	100	60	6	50	2 800	0.681
26	350	1 050	143.5	5	100	50	6	180	2 800	0.609
27	1 012	3 036	85.3	4	100	0	6	0	2 800	1.715
28	494	1 482	132.5	5	50	0	6	130	2 800	0.413

果的边际贡献。给定第 i 个样本 x_i 的第 j 个特征为 x_{ij} , 模型对 PPV 预测问题中第 i 个样本的预测值为 y_i , 所有训练样本的平均预测值为 y_{base} , 即响应变量的均值, 则 x_{ij} 的 SHAP 值可表示为

$$y_i = y_{\text{base}} + \sum_{j=1}^j \varphi_j \quad (5)$$

式中 y_i 表示第 i 个样本的预测值; y_{base} 表示所有训练样本的平均预测值; φ_j 的符号决定特征向量对 PPV 预测目标的影响性正负, 大小表明该特征对预测目标影响程度的高低, 其贡献值之和等于最终的 PPV 预测输出。特征 j 的 SHAP 值 φ_j 计算方法为

$$\varphi_j = \sum_{T \subseteq G \setminus \{j\}} |T|! \frac{(G - |T| - 1)!}{G!} \times [f_k(T) - f_k(T - \{j\})] \quad (6)$$

式中 G 表示原始特征集; T 为 G 中任意特征子集; $G \setminus \{j\}$ 为除特征 j 外所有特征序列的子集; $f_k(T)$ 为特征特征子集 T 的累积贡献值; $f_k(T - \{j\})$ 为除特征 j 之外的剩余特征子集的累积贡献值。

3 结果与讨论

为了验证模型的预测性能和模型解释合理性, 讨论预测模型和优化算法的选择对 PPV 预测结果的影响, 并通过 SHAP 分别从全局和局部的角度解释模型的预测过程。

3.1 模型对比

为检验投票集成机制的有效性, 将提出的 GA-VR 模型与其他基模型——SVR^[23]、RFR、GBR、极度随机树回归 (Extra-Trees Regression, ETR)^[24]、DTR 及 K 近邻 (K -Nearest Neighbors, KNN)^[25] 回归展开对比分析, 为排除优化算法的差异化影响, 所有模型统一使用 GA 算法优化超参数。表 3 为本文方法与其他基模型在 GA 优化下的预测性能对比结果, 结果表明, 在基模型中, GA-GBR、GA-RFR 与 GA-DTR 在 PPV 预测任务上表现最优。因此, 本文提出

的 GA-VR 通过集成这三个性能最佳的基模型是合理的, GA-VR 在 PPV 预测任务中预测性能最优, 其决定系数 R^2 达到 94.8%, 验证了方法可靠性。该模型通过投票集成架构实现特征空间协同优化与权值自适应分配, 更能有效拟合 PPV 特征参数的动态变化规律。

由万策同效定理可知^[26], 优化算法并不具备普适性的优化性能。因此, 本文通过优化算法性能的对比确立最佳方案。VR 由三个异质基模型构成, 其差异化的特征解译机制导致对同源样本的预测置信度分布呈现异质性, 需通过权值优化实现集成效能最大化。针对基模型共存的参数敏感性与异常值脆弱性问题, 建立差异化优化方案: GBR 须具备泛化性能与参数自适应性; RFR 需强化噪声鲁棒性与多样性保持能力; DTR 需平衡全局搜索与局部求精能力。为此, 比较了以下六种异质优化算法: 梯度下降 (Gradient Descent, GD) 法^[27]、PSO^[28]、GA、KNN、萤火虫算法 (Firefly Algorithm, FA)^[29] 和蚁群算法 (Ant Colony Optimization, ACO)^[30] 对 VR 模型的优化效果。表 4 为 GD-VR、PSO-VR、KNN-VR、FA-VR、ACO-VR 模型预测性能的对比结果。结果表明, GA 优化 VR 的预测精度显著优于其他优化算法, 验证了该方法在多维度参数优化的技术优势。

表 5 展示了模型超参数优化结果。对 GBR 而言, GA 对 max_depth、learning_rate 和 min_samples_split 三个参数进行了优化, 从预测效果来看, 优化后的 GBR 的适应度函数值比未优化时小 0.003; 对 RFR 而言, GA 对 min_samples_split 参数进行了优化, 从预测效果来看, 优化后的 RFR 的适应度函数值比未优化时小 0.020; 对 DTR 而言, GA 对 min_samples_split 参数进行了优化, 从预测效果来看, 优化后的 DTR 的适应度函数值比未优化时小 0.172。由此可知 GA 优化过的单个模型较未优化时性能均

表 3 不同模型的 R^2

Table 3 R^2 for different models

模型	GA-DTR	GA-ETR	GA-GBR	GA-KNN	GA-RFR	GA-SVR	GA-VR
R^2	0.843	0.797	0.881	0.648	0.878	0.696	0.948

表 4 不同优化算法的 R^2

Table 4 R^2 for different optimisation algorithms

优化算法	GD-VR	PSO-VR	GA-VR	KNN-VR	FA-VR	ACO-VR
R^2	0.903	0.893	0.948	0.912	0.909	0.890

有提升。

表 5 三个基模型超参数

Table 5 Model hyper parameters of three bases

模型	超参数	超参数范围	最佳超参数
GBR	$n_estimators$	[75 , 105]	94
	max_depth	[3 , 10]	6
	$learning_rate$	[0.020 , 0.030]	0.027
	$min_samples_split$	[5 , 12]	11
RFR	max_depth	[5 , 20]	10
	$n_estimators$	[150 , 300]	240
	$min_samples_split$	[2 , 10]	3
DTR	$min_samples_split$	[2 , 20]	2
	max_leaf_nodes	[10 , 50]	20
	max_depth	[2 , 30]	10

注: $n_estimators$ 表示要构建的弱学习器(决策树)的数量; max_depth 定义了决策树的最大深度,即决策树可以生成的最长路径的长度; $learning_rate$ 代表在每一次迭代中梯度向损失函数最优解移动的步长; $min_samples_split$ 用于控制内部节点(非叶子节点)分裂所需的最小样本数; max_leaf_nodes 表示决策树能够生成的最大叶子节点数。

3.2 预测结果评估

图 4 通过综合评价体系直观对比了各模型的预测表现:各模型预测向量到坐标原点的欧氏距离表征预测结果与观测值间的标准差(Standard Deviation, SD);方位角余弦值映射为预测与观测序列的皮尔逊相关系数 r ;至参考原点的径向距离量化中心化均方根误差(Centered Root Mean Square Error, CRMSE)。由图 4 可知,GA-VR 模型展现出最优预测性能,验证了该模型在多元评价维度上的综合优势。为进一步展示本文 GA-VR 的具体预测性能,表 6 列出了 GA-VR 在部分代表性测试样本

上的预测结果。上述结果表明,所提出的模型具有良好的预测性能,在预测精度上取得了较好的效果。

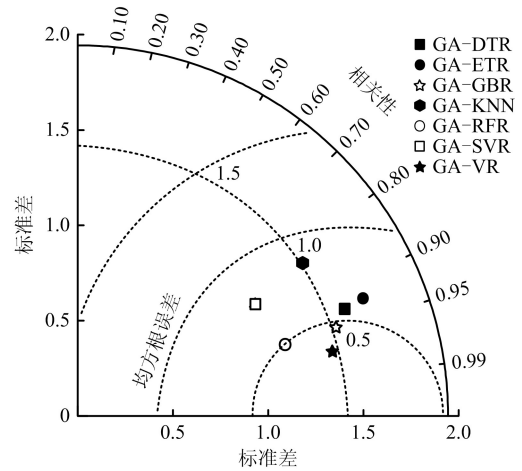


图 4 可视化预测准确度的泰勒图

Fig. 4 Taylor diagram for visualizing prediction accuracy

3.3 模型预测解释分析

不同特征对 PPV 预测的 SHAP 值影响如图 5 所示,图中通过颜色梯度(浅色表示特征值大,深色表示特征值小)表征各个特征与对应的 SHAP 值的关系。以 SHAP 值为 0 作为分界线,左边的点表示对 PPV 预测的值产生负面影响(即减小 PPV 的预测值),右边的点表示对 PPV 预测的值产生积极影响(即增大 PPV 的预测值)。由图 5 可知,在水平距离(X_3)较小的情况下,GA-VR 模型对 PPV 的预测值显著增加,在最大段药量(X_1)、总药量(X_2)、前排抵抗线(X_5)、测点与最小抵抗线方向夹角(X_{11})偏低时,GA-VR 模型倾向于将这些条件视为降低 PPV 预测值的影响因素。上述结果与文献[31]中发表的发现一致。

表 6 本文模型在部分测试集上的性能

Table 6 Performance of the model in this paper on a subset of the test set

序号	X_1/kg	X_2/kg	X_3/m	X_5/m	X_6/ms	$X_7/\%$	X_{10}	$X_{11}/(^{\circ})$	$X_8/(\text{m} \cdot \text{s}^{-1})$	真实 PPV/ ($\text{cm} \cdot \text{s}^{-1}$)	预测 PPV/ ($\text{cm} \cdot \text{s}^{-1}$)
1	1 090	5 450	253.1	5	75	100	8	90	4 200	0.395	0.388
2	370	2 150	101.2	7	25	0	6	115	4 200	1.554	1.548
3	400	3 600	31.5	6	50	0	6	160	4 200	5.436	5.470
4	1 410	6 780	79.4	5	50	0	7	180	4 200	2.690	2.737
5	1 090	5 450	391.5	5	75	100	8	90	4 200	0.375	0.321

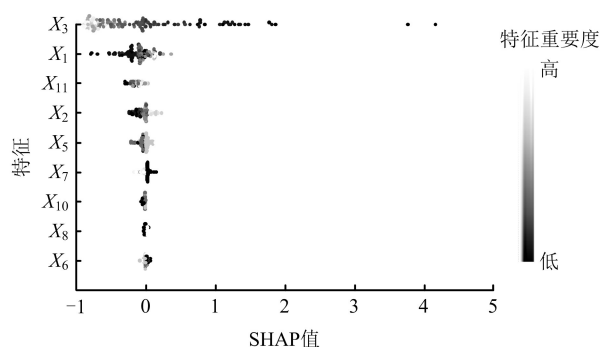


图5 GA-VR模型全局解释

Fig. 5 Global explanation of the GA-VR model

此外,SHAP可分析GA-VR模型对单个样本预测过程的决策机制,为每个样本建立特征贡献度可加性模型(即单个样本所有特征的SHAP值之和等于模型预测值与该样本基线期望值的差值),以93#测试样本为例。

图6为GA-VR模型根据93#样本特征预测PPV时各特征指标的SHAP值:灰色箭头表征积极影响,黑色箭头指示负面影响,基线值1.3685为训练集总体样本的均值。图7为93#样本各特征具体的影响程度, X_6 、 X_7 、 X_8 和 X_{11} 呈现积极影响(分别提升了0.022、0.020、0.030、0.141); X_1 、 X_2 、 X_3 、 X_5 和 X_{10} 则表现负面影响(分别降低了0.020、0.008、0.047、0.050、0.061)。综上累积各特征的SHAP贡献及惩罚,得到93#样本的PPV预测值为1.39 cm/s。通过SHAP全局解释和局部解释协同的模式有效提升了模型决策透明度,为复杂工程场景下的机器学习可信应用提供方法支持。

4 结论

本文提出了一种融合GA优化的投票回归集成(GA-VR)与SHAP解释的PPV预测模型。主要结论如下。

(1)通过包括特征重要性排序和特征相关性分析的特征工程程序,可实现PPV数据集的优化,该

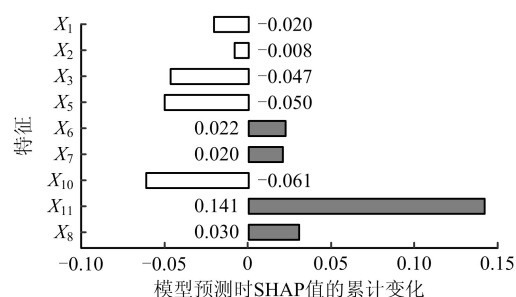


图7 93#样本各指标的SHAP贡献值或惩罚值

Fig. 7 SHAP contribution or penalty values for each indicator for the 93# sample

程序为复杂高维数据提供了一种有效的数据降维方法。

(2)通过对比分析研究,相较于单一基模型及其他优化算法,GA-VR集成模型表现出更优的性能($R^2 = 0.948$),说明了异质集成架构在解析复杂非线性的PPV预测方面更具优势。

(3)通过SHAP分析可知影响因素中测点与爆破位置水平距离占主导因素,其次为最大段药量、测点与最小抵抗线夹角。SHAP解释有效提升了模型预测过程的透明度,有助于工程人员进行控制和分析。

(4)当前研究受限于样本规模,对特殊情况的泛化能力有待提升,未来可通过引入更多有效数据与建立动态特征更新机制,增强模型对其他特殊情况的适应性。

参考文献 (References):

- [1] 高宇璠,傅洪贤. 基于WOA-IC优化神经网络的隧道爆破振动预测研究[J]. 振动与冲击, 2025, 44(4): 229-237.
GAO Y F, FU H X. Research on tunnel blasting vibration prediction based on WOA-IC optimized neural network [J]. Journal of Vibration and Shock, 2025, 44(4): 229-237.
- [2] 范勇,裴勇,杨广栋,等. 基于改进PSO-BP神经网络

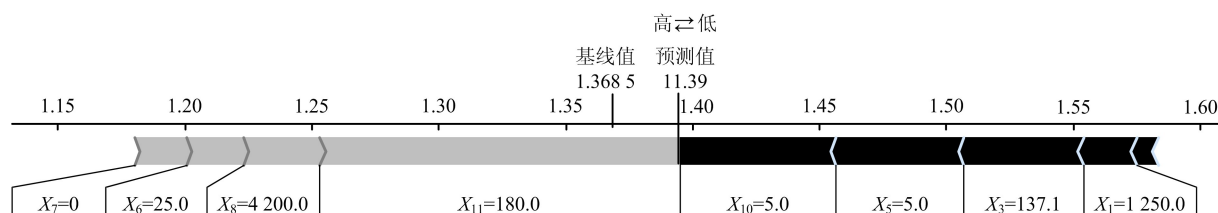


图6 93#样本的PPV预测中各指标的SHAP值

Fig. 6 SHAP values for each indicator in the PPV prediction for the 93# sample

- 络的爆破振动速度峰值预测[J]. 振动与冲击, 2022, 41(16): 194-203, 302.
- FAN Y, PEI Y, YANG G D, et al. Prediction of blasting vibration velocity peak based on an improved PSO-BP neural network[J]. Journal of Vibration and Shock, 2022, 41(16): 194-203, 302.
- [3] 朱大鹏, 谢昌建, 许红波. 隧道进口岩堆斜坡爆破振动动力响应研究[J]. 安全与环境学报, 2022, 22(3): 1275-1283.
- ZHU D P, XIE C J, XU H B. Dynamic response of talus slope at entrance of tunnel under blasting vibration[J]. Journal of Safety and Environment, 2022, 22(3): 1275-1283.
- [4] 言志信, 言湮, 江平, 等. 爆破振动峰值速度预报方法探讨[J]. 振动与冲击, 2010, 29(5): 179-182, 249.
- YAN Z X, YAN L, JIANG P, et al. Prediction methods for blasting-induced ground vibration velocity[J]. Journal of Vibration and Shock, 2010, 29(5): 179-182, 249.
- [5] 刘亚群, 李海波, 裴启涛, 等. 基于灰色关联分析的遗传神经网络在水下爆破中质点峰值振动速度预测研究[J]. 岩土力学, 2013, 34(增刊1): 259-264.
- LIU Y Q, LI H B, PEI Q T, et al. Prediction of peak particle velocity induced by underwater blasting based on the combination of grey relational analysis and genetic neural network[J]. Rock and Soil Mechanics, 2013, 34(Suppl. 1): 259-264.
- [6] 郭钦鹏, 杨仕教, 朱忠华, 等. 运用 GA-BP 神经网络对爆破振动速度预测[J]. 爆破, 2020, 37(3): 148-152.
- GUO Q P, YANG S J, ZHU Z H, et al. Prediction of blasting vibration velocity using GA-BP neural network[J]. Blasting, 2020, 37(3): 148-152.
- [7] 刘敦文, 崔朋波. 基于 BP 神经网络的新浇筑混凝土爆破安全震动速度预测[J]. 安全与环境学报, 2014, 14(1): 43-46.
- LIU D W, CUI P B. Forecasting for safety vibration velocity of freshly-made concrete based on BP neural network[J]. Journal of Safety and Environment, 2014, 14(1): 43-46.
- [8] 唐海, 石永强, 李海波, 等. 基于神经网络的爆破振动速度峰值预报[J]. 岩石力学与工程学报, 2007(增刊1): 3533-3539.
- TANG H, SHI Y Q, LI H B, et al. Prediction of peak velocity of blasting vibration based on neural network[J]. Chinese Journal of Rock Mechanics and Engineering, 2007(Suppl. 1): 3533-3539.
- [9] WANG X, LI J C, ZHAO X B, et al. Propagation characteristics and prediction of blast-induced vibration on closely spaced rock tunnels[J]. Tunnelling and Underground Space Technology, 2022, 123: 104416.
- [10] 张西良, 焦灏恺, 李二宝. 基于迁移学习算法的深部爆破振动速度预测[J]. 中国安全科学学报, 2023, 33(6): 64-72.
- ZHANG X L, JIAO H K, LI E B. Prediction of vibration velocity of deep blasting based on transfer learning[J]. China Safety Science Journal, 2023, 33(6): 64-72.
- [11] 李振阳, 张宝岗, 熊信, 等. 基于 PSO-XGBoost 的露天矿山 PPV 预测模型研究[J]. 黄金科学技术, 2024, 32(4): 620-630.
- LI Z Y, ZHANG B G, XIONG X, et al. Research on PPV prediction model of open-pit mine based on PSO-XGBoost[J]. Gold Science and Technology, 2024, 32(4): 620-630.
- [12] 史秀志. 爆破振动信号时频分析与爆破振动特征参量和危害预测研究[D]. 长沙: 中南大学, 2008.
- SHI X Z. Study of time and frequency analysis of blasting vibration signal and the prediction of blasting vibration characteristic parameters and damage[D]. Changsha: Central South University, 2008.
- [13] CHEN T, GUESTRIN C. XGboost: a scalable tree boosting system[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 13-17, 2016, San Francisco, California, USA. New York: Association for Computing Machinery, 2016: 785-794.
- [14] WITTEN I H, FRANK E, HALL M A, et al. Data Mining: Practical machine learning tools and techniques[M]. San Francisco, CA, USA: Elsevier, Morgan Kaufmann Publishers, 2025.
- [15] KUNCHEVA L I, WHITAKER C J. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy[J]. Machine Learning, 2003, 51(2): 181-207.
- [16] QI C C, TANG X L. A hybrid ensemble method for improved prediction of slope stability[J]. International Journal for Numerical and Analytical Methods in Geomechanics, 2018, 42(15): 1823-1839.
- [17] FRIEDMAN J H. Greedy function approximation: a gradient boosting machine[J]. Annals of Statistics, 2001, 29(5): 1189-1232.

- [18] BREIMAN L. Random forests[J]. Machine Learning, 2001, 45(1): 5–32.
- [19] MOLNAR C. Interpretable machine learning: a guide for making black box models explainable[EB/OL]. (2025–03–12)[2025–05–01]. <https://christophm.github.io/interpretable-ml-book>.
- [20] HOLLAND J H. Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence [M]. Cambridge, MA, USA: MIT Press, 1992.
- [21] 银星, 李庆华. 基于机器学习的超高韧性水泥基复合材料板在近场爆炸荷载作用下的损伤预测[J/OL]. 工程力学: 1–13(2025–04–03)[2025–05–01]. <https://doi.org/10.6052/j.issn.1000-4750>. 2024. 10.0791.
- YIN X, LI Q H. Machine learning based damage prediction of ultra-high toughness cementitious composite panels under near range explosion[J/OL]. Engineering Mechanics: 1–13(2025–04–03)[2025–05–01]. <https://doi.org/10.6052/j.issn.1000-4750>. 2024. 10.0791.
- [22] LUNDBERG S M, LEE S I. A unified approach to interpreting model predictions[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems, December 4–9, 2017, Long Beach, California, USA. New York: Association for Computing Machinery, 2017: 4768–4777.
- [23] AWAD M, KHANNA R. Support vector regression [M]//AWAD M, KHANNA R. Efficient learning machines: theories, concepts, and applications for engineers and system designers. Berkeley, CA: Apress, 2015: 67–80.
- [24] WANG Z, MU L, MIAO H, et al. An innovative application of machine learning in prediction of the syngas properties of biomass chemical looping gasification based on extra trees regression algorithm[J]. Energy, 2023, 275: 127438.
- [25] SONG Y S, LIANG J Y, LU J, et al. An efficient instance selection algorithm for k nearest neighbor regression[J]. Neurocomputing, 2017, 251: 26–34.
- [26] HO Y C, PEPYNE D L. Simple explanation of the no-free-lunch theorem and its implications[J]. Journal of Optimization Theory and Applications, 2002, 115(3): 549–570.
- [27] ANDRYCHOWICZ M, DENIL M, GÓMEZ S, et al. Learning to learn by gradient descent by gradient descent [C]//Advances in Neural Information Processing Systems 29. Red Hook: Curran Associates, Inc., 2016: 3988–3996.
- [28] KENNEDY J, EBERHART R. Particle swarm optimization [C]//Proceedings of the 1995 IEEE International Conference on Neural Networks (ICNN'95), November 27–December 1, 1995, Perth, WA, Australia. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers, 1995: 1942–1948.
- [29] FISTER I, FISTER I, YANG X S, et al. A comprehensive review of firefly algorithms[J]. Swarm and Evolutionary Computation, 2013, 13: 34–46.
- [30] DORIGO M, BIRATTARI M, STUTZLE T. Ant colony optimization [J]. IEEE Computational Intelligence Magazine, 2006, 1(4): 28–39.
- [31] KAHRIMAN A. Analysis of parameters of ground vibration produced from bench blasting at a limestone quarry[J]. Soil Dynamics and Earthquake Engineering, 2004, 24(11): 887–892.

Predicting peak vibration velocity of blasting particles using interpretable GA–VR integrated learning

YANG Yi¹, WANG Shuxian¹, SHEN Yaxi², SHI Yulian³, DING Qiuyue^{1,4}

(1 Faculty of Public Safety and Emergency Management, Kunming University of Science and Technology, Kunming 650032, China;

2 School of Civil Engineering, Tianjin University, Tianjin 300072, China;

3 School of Land and Resource Engineering, Kunming University of Science and Technology, Kunming 650032, China;

4 Yunnan Zhike Safety Consulting Co., Ltd., Kunming 650032, China)

Abstract: Accurately predicting Peak Particle Velocity (PPV) is essential for ensuring operational safety and minimizing environmental impact. While machine learning models present advantages over traditional empirical formulas, their “black-box” nature and limited

2092

interpretability can hinder practical applications in optimizing blast parameters. This study introduces a novel, interpretable prediction framework that combines Genetic Algorithm (GA) optimization with a Voting Regressor (GA - VR) ensemble. The model integrates three distinct base learners: the Gradient Boosting Regressor (GBR), Random Forest Regressor (RFR), and Decision Tree Regressor (DTR). The GA is utilized to globally optimize critical hyperparameters for each base learner: for GBR, $n_estimators$ (range: 75 - 105), max_depth (3 - 10), $learning_rate$ (0.020 - 0.030), and $min_samples_split$ (5 - 12); for RFR, max_depth (5 - 20), $n_estimators$ (150 - 300), and $min_samples_split$ (2 - 10); and for DTR, $min_samples_split$ (2 - 20), max_leaf_nodes (10 - 50), and max_depth (2 - 30). The optimized base models are then integrated through a weighted average voting mechanism within the VR framework. Additionally, SHapley Additive exPlanations (SHAP) analysis is employed to provide both global and local interpretability of the model's predictions. Experimental results based on field data from the Daye Nonferrous Metals Group Tonglvshan open-pit mine demonstrate the superior performance of the GA - VR model, which achieved a coefficient of determination (R^2) of 0.948 on the independent test set. This significantly outperformed six individual GA-optimized base models: GA - DTR ($R^2 = 0.843$), GA - ETR ($R^2 = 0.797$), GA - GBR ($R^2 = 0.881$), GA - KNN ($R^2 = 0.648$), GA - RFR ($R^2 = 0.878$), and GA - SVR ($R^2 = 0.696$). Furthermore, GA - VR surpassed VR models optimized by five alternative algorithms: Gradient Descent (GD - VR, $R^2 = 0.903$), Particle Swarm Optimization (PSO - VR, $R^2 = 0.893$), K -Nearest Neighbors Regression (KNN - VR, $R^2 = 0.912$), Firefly Algorithm (FA - VR, $R^2 = 0.909$), and Ant Colony Optimization (ACO - VR, $R^2 = 0.890$). SHAP analysis revealed that the horizontal distance between the measurement point and the blast location (X_3) was the most influential factor in PPV predictions, followed by the maximum charge per delay (X_1) and the angle between the measurement point and the direction of the minimum resistance line (X_{11}). Local interpretability analysis for specific samples (e.g., sample 93#) further illustrated the contribution of each feature to individual predictions. Overall, the GA - VR model not only delivers high-precision PPV forecasting but also provides a transparent and interpretable rationale for its predictions, making it a powerful tool for the intelligent design of blast parameters and enhanced safety control in mining and construction engineering.

Key words: safety engineering; blast vibration prediction; genetic algorithms; integrated learning; interpretable analysis; feature engineering