

DOI: 10.11779/CJGE20221139

# 考虑不确定性的 TBM 施工速度加权随机森林预测模型

闫长斌<sup>1</sup>, 高子昂<sup>1</sup>, 姚西桐<sup>2</sup>, 汪鹤健<sup>3</sup>, 杨风威<sup>4</sup>, 杨继华<sup>4</sup>, 卢高明<sup>5</sup>

(1. 郑州大学土木工程学院, 河南 郑州 450001; 2. 广西交通设计集团有限公司, 广西 南宁 530029; 3. 国网河南省电力公司开封供电公司, 河南 开封 475000; 4. 黄河勘测规划设计研究院有限公司, 河南 郑州 450003; 5. 盾构及掘进技术国家重点实验室, 河南 郑州 450001)

**摘要:** TBM 施工速度影响因素众多, 具有显著的不确定性。对于地质参数的模糊性, 采用岩体分级系统 RMR、岩石耐磨性 CAI 和岩石硬度  $H$  衡量地质条件; 对于施工过程中机械参数的随机性, 利用 TBM 刀盘推力  $TF$  与转速  $RPM$  等主动控制参数进行分析; 同时提出以其他因素停机时间占比来量化人为因素的不确定性。依托兰州水源地建设工程输水隧洞双护盾 TBM 施工实测数据, 建立了考虑不确定性的 TBM 施工速度预测数据库和加权随机森林算法模型, 并与随机森林、支持向量回归、BP 神经网络等预测方法进行了对比分析。研究表明, 加权随机森林模型中测试集的预测结果均方根误差和决定系数分别为 1.59, 0.97, 预测精度及可靠性均优于其他 3 种模型。该模型采用不同权重赋值的方法优化超参数, 具有高精度、不易过拟合等优点, 表现出更好的泛化能力和鲁棒性。

**关键词:** 隧道掘进机 (TBM); 施工速度; 不确定性; 加权随机森林; 支持向量回归; 神经网络; 预测模型

**中图分类号:** U455; TV554 **文献标识码:** A **文章编号:** 1000-4548(2023)12-2575-09

**作者简介:** 闫长斌(1980—), 男, 博士, 教授, 主要从事岩土与地下工程等方面的教学和科研工作。E-mail: yanchangbin\_2001@163.com。

## Weighted random forest prediction model for TBM advance rate considering uncertainty

YAN Changbin<sup>1</sup>, GAO Ziang<sup>1</sup>, YAO Xitong<sup>2</sup>, WANG Hejian<sup>3</sup>, YANG Fengwei<sup>4</sup>, YANG Jihua<sup>4</sup>, LU Gaoming<sup>5</sup>

(1. School of Civil Engineering, Zhengzhou University, Zhengzhou 450001, China; 2. Guangxi Communications Design Group Co., Ltd., Nanning 530029, China; 3. Kaifeng Power Supply Company, State Grid Henan Electric Power Company, Kaifeng 475000, China; 4. Yellow River Engineering Consulting Co., Ltd., Zhengzhou 450003, China; 5. State Key Laboratory of Shield Machine and Boring Technology, Zhengzhou 450001, China)

**Abstract:** There are many factors affecting TBM construction speed, and they have significant uncertainty. In view of the ambiguity of geological parameters, the rock classification system RMR, rock abrasiveness CAI and rock hardness  $H$  are used to measure the geological conditions. For the randomness of mechanical parameters in the construction process, the active control parameters such as TBM cutter head thrust  $TF$  and rotational speed  $RPM$  are used for analysis. At the same time, it is proposed to quantify the uncertainty of human factors by the proportion of downtime of other factors. Based on the measured data of the double-shield TBM construction in the water conveyance tunnel of the Lanzhou water source construction project, the prediction database of the TBM advance rate and the weighted random forest algorithm model considering uncertainty are established. In addition, other models such as random forest, support vector regression and BP neural network are used to verify the prediction accuracy of the proposed model. The results show that the error of the root mean square and the determination coefficient of the predicted results of the test set in the weighted random forest model are 1.59 and 0.97, respectively, and the prediction accuracy and reliability of the weighted random forest prediction model are better than those of the other three models. The model adopts the method of optimizing hyperparameters by assigning different weights, which has the advantages of high accuracy, difficult overfitting and better generalization capability and robustness.

**Key words:** tunnel boring machine; advance rate; uncertainty; weighted random forest; support vector regression; neural network; prediction model

## 0 引言

随着“一带一路”倡议的深入推进、交通强国战

基金项目: 国家自然科学基金项目 (41972270, 42002281); 盾构及掘进技术国家重点实验室开放基金项目 (SKLST-2019-K06); 水利部黄河流域水治理与水安全重点实验室 (筹) 研究基金项目 (2022-SYSJJ-06)  
收稿日期: 2022-09-16

略目标的明确提出以及水资源优化配置的相继展开,隧道掘进机(TBM)施工方法已成为长距离隧道(洞)施工的首选方法。鉴于施工过程的风险性、地质参数的不确定性和人为因素的不可控性,TBM施工性能预测是实现安全高效破岩的前提与保证。如何充分考虑地质条件的不确定性和施工过程的风险性,基于现场实测得到的围岩参数,机械参数以及掘进参数等信息,建立科学有效的TBM施工速度(AR)预测模型,一直是TBM施工领域研究的重点与难点问题之一。

影响TBM施工性能的不确定性因素较多:①围岩地质信息的模糊性,有限的前期勘察和超前地质预报系统无法全面揭示复杂隧道沿线复杂的地质信息,这就难免导致TBM施工过程中遭遇各类复杂地质问题;②施工的管理水平、工人的操作能力、TBM机械设备的运行性能的不确定性,其差异性和随机性同样制约着TBM施工性能有效发挥。近年来,许多学者相继开展了关于TBM施工速度预测的有益研究:①基于全尺寸线性切割实验数据开发了半理论半经验的CSM模型<sup>[1]</sup>,通过对大量工程的岩体参数与掘进参数进行回归分析得到了NTNU和 $Q_{TBM}$ 经验模型<sup>[2-3]</sup>;②随着人工智能与工程建设的交叉融合,国内外掀起基于机器学习算法的TBM施工预测模型研究浪潮。Alvarez等<sup>[4]</sup>收集了640座TBM掘进隧洞的岩体参数和掘进参数,提出基于神经模糊网络的TBM施工速度预测模型;Armaghani等<sup>[5]</sup>综合利用岩石单轴抗压强度(UCS)、岩石单轴抗拉强度(BTS)、刀盘推力(TF)、刀盘转速(RPM)以及其他岩体指标等8个参数,建立了基于PSO-ANN算法的TBM施工速度预测模型。Zhou等<sup>[6]</sup>通过ANN和GP算法,以岩石质量指标(RQD)、UCS、岩体分级系统(RMR)、BTS、TF和RPM为输入参数,对TBM施工速度进行了预测。然而,经验模型考虑的影响因素缺乏合理性,忽略了关键的岩体参数且输入项指标冗余会对TBM施工速度预测产生较大影响;基于智能算法的预测模型需要大量的数据集支撑,训练模型时超参数调节与优化十分繁琐,易出现过拟合现象。此外,现有的TBM施工速度预测模型尚未考虑施工过程中的不确定性与风险性,预测精度较低且适用范围具有一定的局限性。

针对现有机器学习算法模型的不足,引入一种简单可靠的超参数优化方法,即随机森林(RF)算法。该方法具有稳定性好,不易产生过拟合,预测精度高等优点。邓军等<sup>[7]</sup>提出了一种基于随机森林方法的采空区煤自燃预测模型,通过在多个矿井煤自燃预测中

应用,验证了RF方法的稳定性和普适性。王仁超等<sup>[8]</sup>基于随机森林回归方法建立了爆破块度预测模型,结合某工程实际验证了该模型的可行性,为堆石坝的爆破施工管理与控制提供了科学指导。李明超等<sup>[9]</sup>采用改进的随机森林方法提出了一种新的多尺度岩体完整性评价指标和相应分析方法,该指标与勘探数据吻合良好且能够对岩体完整性进行不同尺度的精细化评价。然而,在TBM施工性能预测领域尚未有研究引入随机森林算法,本文采用不同权重赋值的方法优化超参数,提出一种基于加权随机森林方法的TBM施工速度预测模型。

## 1 加权随机森林算法模型

随机森林(RF)是一种用于分类和回归的监督式集成学习方法,RF由许多决策树组成,通过聚合单个决策树的不同输出结果来降低导致决策树错误的方差,对于回归问题则通过取不同预测器的平均值作为结果。加权随机森林(WRF)是一种改进的RF算法,通过分配不同的权重,惩罚节点的错分从而提高预测的精确度<sup>[10]</sup>。

WRF算法模型的构建流程,如图1所示。具体步骤如下:

(1)形成训练输入集 $D=\{X, Y\}$ , $X$ 由具有 $M$ 个属性的 $N$ 个样本组成, $Y$ 为唯一的向量。

(2)采用Bootstrap重抽样方法对输入训练样本集合 $D$ 进行 $k$ 次有放回重复抽样,得到 $k$ 个伪样本。每个伪样本中的集合称为袋内(IB)数据,而每次未被抽中的数据称为袋外(OOB)数据。WRF通过袋外数据进行内部误差估计,以提升模型的泛化能力。

(3)构建CART树进行训练,随机取 $m$ 个属性作为当前节点的子集,根据加权不纯度 $G(u_i, v_i)$ 来寻找最优的节点分离值, $G(u_i, v_i)$ 值越小分离效果越好。每棵树在分裂生长时不进行剪枝, $m$ 值在整个训练过程中也保持不变,待其满足终止条件后停止生长。

(4)将 $k$ 棵回归树组合生成随机森林,将预测集数据输入到训练的随机森林中,对单棵回归树的预测值取平均作为模型的输出,其最终预测结果为

$$f(x_i) = \frac{1}{k} \sum_{i=1}^k h_i(x_i) \quad (1)$$

式中: $f(x_i)$ 为最终模型预测结果; $h_i(x_i)$ 为第 $i$ 棵回归树得出的结果。

针对决策树中子节点分离的好坏,一般以切分后节点的加权不纯度 $G(u_i, v_i)$ 来衡量:

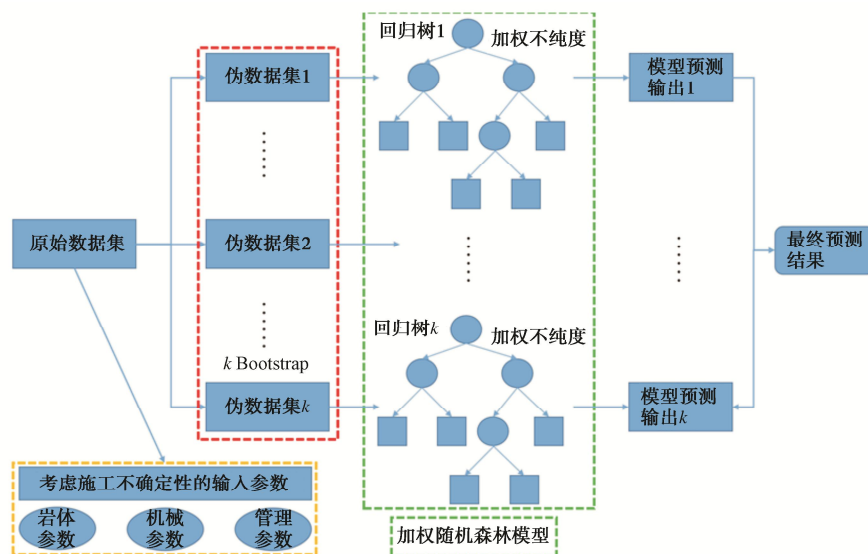


图 1 加权随机森林算法模型构建示意图

Fig. 1 Schematic graph of model for weighted random forest algorithm

$$G(u_i, v_i) = \frac{N_L}{N_S} H(X_L) + \frac{N_R}{N_S} H(X_R) \quad (2)$$

式中:  $u_i, v_i$  分别为某一个切分变量及其对应的切分值;  $N_S$  为当前节点所有训练样本个数;  $N_L, N_R$  分别为切分后左子节点和右子节点的训练样本个数;  $X_L, X_R$  分别为切分后左、右子节点的训练样本集合;  $H(X)$  为衡量节点不纯度的函数。回归任务常采用平方平均误差 (MSE) 表达。

$$H(X) = \frac{1}{N_S} \sum_{i=1}^{N_S} (y_i - \bar{h})^2 \quad (3)$$

式中:  $X$  为当前节点的训练样本集合;  $y_i, \bar{h}$  分别为当前节点样本目标变量的真实值和 WRF 对样本数据预测值的平均值。

由于 WRF 采用的是有放回重抽样方法, 训练集样本的数据不尽相同且输入的属性随机选取,  $k$  个伪样本之间相互独立, 进而整个集成模型的性能可以得到有效提高, 其训练速度较快, 不容易出现过拟合且无需复杂的调参处理。此外, 由于 OOB 数据未参与模型回归树的构建, OOB 数据可充当测试集。

在检验加权随机森林的回归效果时, 本文采用均方根误差 (RMSE) 和决定系数  $R^2$  两个指标进行评估。

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (f_i - y_i)^2} \quad (4)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - f_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (5)$$

式中:  $f_i$  为样本预测值;  $y_i$  为样本实际值;  $n$  为预测样本数量;  $\bar{y}_i$  为样本实测值的平均值。

## 2 工程概况

兰州水源地建设工程从刘家峡水库取水后通过输水隧洞、输水支线和水厂、市政管线向兰州市区供水。输水隧洞是该项目的重点控制性工程, 全长 31.57 km (其中 TBM 施工段约 25.48 km), 采用以两台双护盾 TBM 为主和钻爆法为辅的联合施工形式。输水隧洞设计开挖洞径为 5.46 m, 衬砌后 4.60 m, 为压力隧洞<sup>[1]</sup>。输水隧洞沿线穿越的地层岩性依次为加里东中期石英闪长岩、前震旦系马衔山群黑云石英片岩和角闪石英片岩、加里东中期花岗岩、奥陶系上中统雾宿山群变质安山岩以及白垩系下统河口群砂岩与泥岩互层、砂砾岩等。其中, 石英片岩以片状构造为主, 片理较发育; 闪长岩抗风化能力强, 节理发育较少, 岩体完整性好; 花岗岩抗风化能力强, 多呈岩脉状侵入, 与围岩接合较好, 完整性好; 泥质砂岩受构造影响, 局部产状变化较大, 节理较发育, 岩体完整性一般, 遇水易崩解; 变质安山岩抗风化能力强, 岩体完整性较好。隧洞线路工程地质剖面情况见图 2 所示。隧洞沿线发育多条断层, 断层带宽度为 1~80 m, 多与洞线斜交。隧洞施工中断层破碎带会显著降低围岩稳定性, 增加施工风险。隧洞沿线不仅发育有石英片岩、花岗岩等硬脆性岩质, 且存在以砂岩、黏土岩和砂砾岩为主, 单轴饱和抗压强度小于 15 MPa 的软岩地层, 地质不确定性较为显著。隧洞多处洞段埋深较大, 工程区构造应力较大, 产生岩爆和软岩大变形的可能性较大。此外, 隧洞施工期间的正常涌水量约 15911.5 m<sup>3</sup>/d, 最大涌水量约 31823 m<sup>3</sup>/d, 易发生涌水、突泥

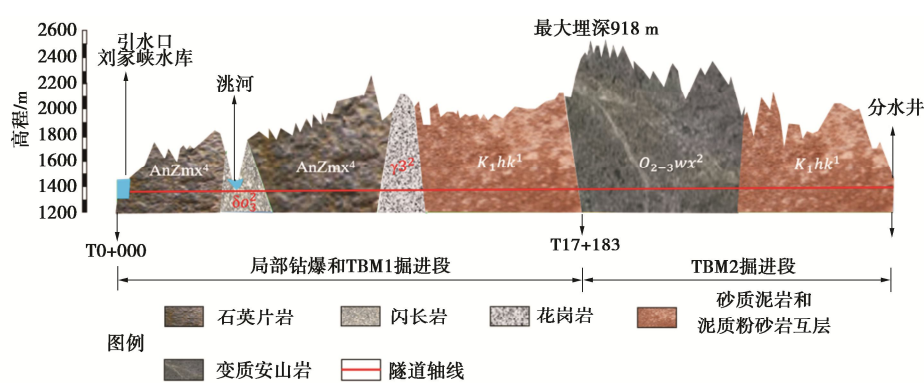


图2 输水隧洞线路区工程地质剖面图

Fig. 2 Engineering geological section of water conveyance tunnel

等风险较大的地质灾害。

3 TBM 施工速度预测

3.1 模型输入参数

掘进过程中，TBM 施工速度与地质条件、设备参数以及管理操作的不确定性密切相关。其中，地质条件的不确定性主要受岩石强度、岩体完整性、岩石硬度与耐磨性等因素影响。而 TBM 设备参数的不确定性与许多因素有关，但是其中能够独立量化控制的只有刀盘转速和刀盘推力，而其它机械参数相对来说是随动的。此外，管理操作水平的不确定性取决于施工组织是否合理，材料供应是否及时等众多因素。由于 TBM 对地质条件极其敏感，穿越不同地质构造单元的风险性和不确定性较大。基于此，针对上述涉及的 TBM 施工速度影响参数，提出考虑施工不确定性的量化分析指标。

(1) 基于地质条件不确定性的参数选取

一旦隧洞线路确定，地质条件就客观存在，不会以施工人员的意志而改变，TBM 施工能否达到安全高效掘进目标，主要取决于隧洞围岩地质条件。然而隧洞地质条件具有很大不确定性，对于 TBM 施工而言，影响掘进性能的地质因素有多种，主要包含岩石的强度、硬度、耐磨性以及岩体的完整性、节理走向与掘进方向之间的关系等。岩体分级系统（RMR）综合考虑了岩体的单轴抗压强度、岩石质量指标、节理间距、节理状况、地下水状况以及修正系数（根据节理面的方位与洞室轴线的方向关系确定）等因素。这 6 个指标的岩体评分值总和可有效表征部分岩体地质条件不确定性对 TBM 施工的适宜性，因此选取 RMR、岩石磨蚀系数 CAI 和岩石矿物加权硬度  $H$  值作为影响 AR 的地质条件输入参数，以此量化不同地质环境下岩体条件的不确定性。

(2) 基于掘进条件不确定性的参数选取

在 TBM 掘进过程中，能反映 TBM 掘进性能的机械参数主要包含刀盘转速、刀盘扭矩、刀盘推力和贯入度等，以上 4 个掘进参数并非是完全相互独立的，其中能够独立控制的掘进参数只有刀盘转速和刀盘推力，而其它掘进参数则是刀盘转速和刀盘推力共同作用的结果。鉴于岩-机相互作用的复杂性，选取能够主动控制的刀盘推力  $TF$  和转速  $RPM$  来量化施工过程中机械设备的不确定性。

(3) 基于管理操作不确定性的参数选取

除地质因素和机械因素外，TBM 施工速度受人为因素影响较大，包括施工团队的技术经验水平、不同环境下的工作状态、现场施工组织水平以及对施工过程中突发事件处理能力等。然而，机器故障和组织管理等影响因素难以量化，Frough 等<sup>[12]</sup>以 Karaj-Tehren 输水隧洞为例，研究发现地质和岩体条件相关停机时间（GRRD）约占总时间的 20%，且 GRRD 和 RMR 有很好的相关性，提出利用停机因素和停机时间对利用率进行分析预测。龚秋明等通过对 TBM 停机时间分类统计，将其分为岩体条件相关停机（RRD）与其他因素停机（ORD）两大类<sup>[13-14]</sup>。ORD 的影响因素主要包括 TBM 设备故障、皮带机出渣及通风设备维护以及正常换步、管片拼装和班组管理等，能够在一定程度上表征设备和操作管理 2 个方面的不确定性，因此将该指标引入 TBM 施工速度预测研究中。

利用兰州水源地建设工程输水隧洞双护盾 TBM 施工地质日报查阅得到的有关掘进段围岩地质条件信息，结合室内外岩石物理力学试验数据，可计算得到岩体分级系统 RMR 值；针对不同岩石开展 CERCHAR 耐磨性试验得到岩石耐磨性 CAI 值；将岩石中每种矿物的含量乘以摩氏硬度，得出岩石中矿物的加权硬度值  $H$ 。此外，收集并整理 TBM 施工记录，分析 TBM



各施工工序占用时间及停机原因,进而计算得到 ORD 等数据。通过计算得到以上这些数据在不同地质单元中的平均值,可建立不同地质条件下 TBM 运行参数数据库,其中训练样本输入、输出参数的描述性统计如表 1 所示。

表 1 模型训练样本输入、输出参数描述性统计

Table 1 Descriptive statistics of input and output parameters of training dataset				
数据类型	参数	范围	平均值	标准差
输入参数	RMR	10.00~80.00	51.98	17.71
	CAI	0.75~3.49	2.18	0.99
	$H$	4.01~5.97	5.24	0.85
	TF/kN	3250~9600	6216.00	2180.00
	$\text{RPM}/(\text{r}\cdot\text{min}^{-1})$	4.20~8.00	6.04	0.90
	ORD/%	3.12~65.77	34.65	17.15
输出参数	$\text{AR}/(\text{m}\cdot\text{d}^{-1})$	2.88~40.80	19.30	8.16

图 3 给出了输入数据集的散点矩阵图。由图 3 可见,除 RMR, TF 和 RPM 等 3 者之间存在较高的相关性外,其他输入参数之间并不具有明显的相关性,说明选取的参数之间具有较强的独立性,适用于建立基于加权随机森林算法的 TBM 施工速度预测模型。建立模型之前将样本中的自变量矩阵  $X$  和因变量矩阵  $y$  进行归一化和标准化处理,然后分别得到归一化和标准化后的矩阵  $E_0$  和  $F_0$ 。

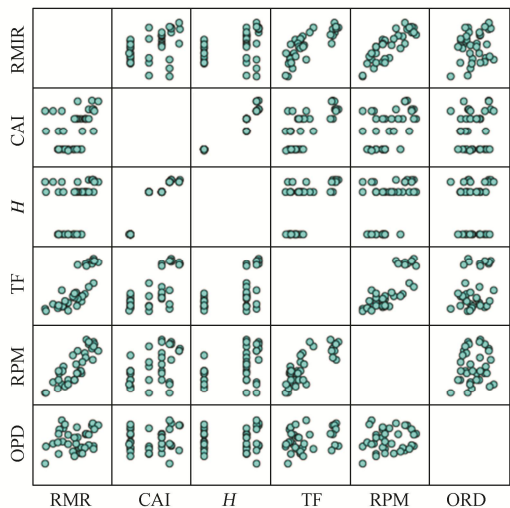


图 3 输入参数散点矩阵图

Fig. 3 Scatter plot matrix of input dataset

3.2 模型参数选择

对于随机森林,主要的超参数包含回归树棵数  $k$  和最大特征数  $m$ 。 $k$  值太小模型容易欠拟合,太大时计算繁琐且不能显著提升模型的性能。这里采用十折交叉检验<sup>[15]</sup>选取最优回归树棵数,图 4 为十折交叉检验的流程图,将原始数据集  $A$  划分为互斥的 10 个子

集,按顺序选取其中一个子集作为模型验证集,其他 9 个子集合并作为训练集,进而实现对模型的 10 次交叉验证,通过比较 10 个泛化正确率的平均值选取最优模型参数。

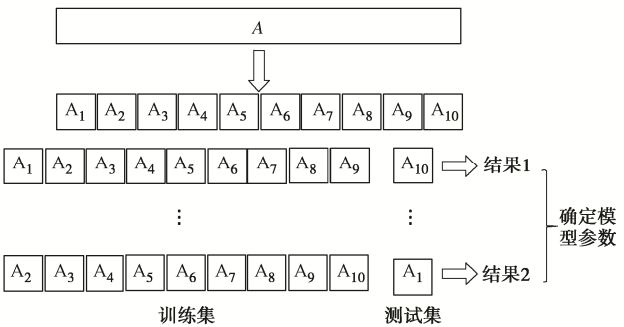


图 4 十折交叉验证

Fig. 4 10-fold cross-validation method

设 RF 模型中  $k$  取值范围为 $[1,800]$ ,迭代计算不同回归树棵数下测试集的均方误差,如图 5 所示。当  $k$  值大于 500 时, MSE 降至趋于稳定,为减小模型的复杂度  $k$  取为 500。 $m$  值决定着模型的扰动程度,一般根据经验公式确定。

$$m = \lceil \log_2 M \rceil \quad (6)$$

式中:  $M$  为模型输入参数的总特征数;  $\lceil \cdot \rceil$  表示取整运算。本研究建立的模型中输入参数具有 6 个属性,则根据经验公式估算  $m$  取 2。

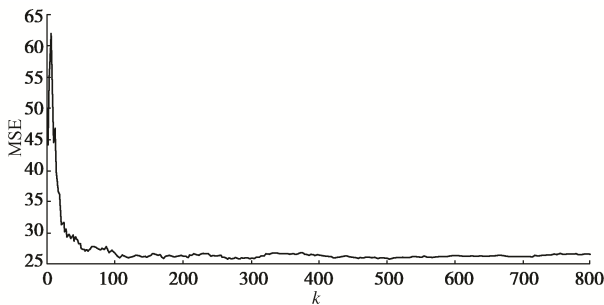


图 5 CART 数量  $k$  对 WRF 模型性能的影响

Fig. 5 Effects of number of CARTs ( $k$ ) on performance of WRF model

模型训练过程中, WRF 预测模型中的超参数根据十折交叉验证法估算取值,回归树棵数取 500,最大特征数取 2。对于其他预测模型中的超参数, BPNN 模型中的隐含层数目采用试算法进行寻优,当隐含层个数为 10 时模型性能最佳,迭代次数以及学习率根据默认参数标准分别设置为 1000, 0.01。SVR 预测模型中的超参数经网格搜索法确定,该函数的参数  $g$  为 0.32,惩罚因子系数  $c$  为 1。

3.3 模型训练与验证结果

为检验 WRF 模型在 TBM 施工速度预测中的合理性与有效性,基于兰州水源地输水隧洞工程的现场实测数据,选取 52 组数据集作为训练集,剩下 13 组作

为验证集。以能够量化岩体参数、机械参数、管理参数不确定性的 6 个指标 (表 1) 作为输入参数, 采用提出的 WRF 模型对相应地质区段的 TBM 施工速度进行预测, 基于加权随机森林算法回归得到的 AR 预测值与实测 AR 值的对比情况如图 6 所示。由图 6 可以看出, AR 预测值与真实记录值的散点分布相距不大, 二者数值较为接近。经统计其平均绝对误差为 1.16 m/d, 验证集决定系数达到 0.95, 表明模型具有良好的预测精度。而且预测值反映的整体趋势与真实值的走势类似, 表明加权随机森林算法对于施工过程中的地质变异能够合理识别, 不会出现与掘进区段内的现实地质条件相矛盾的情况。为了进一步验证 WRF 模型的有效性, 利用同样的训练集和验证集分别建立了随机森林回归模型 RF、反向传播神经网络模型 BPNN 以及支持向量回归模型 SVR。

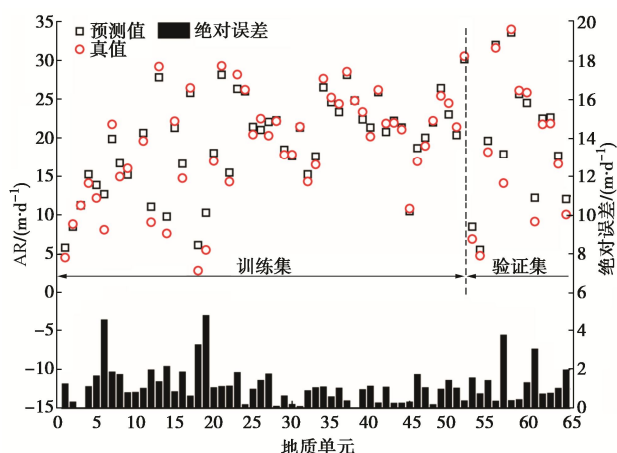
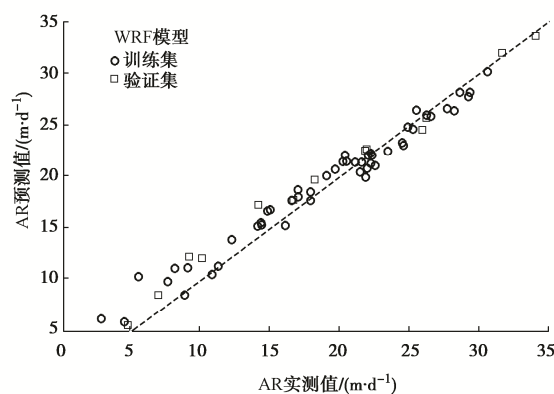


图 6 基于 WRF 模型的施工速度 AR 预测

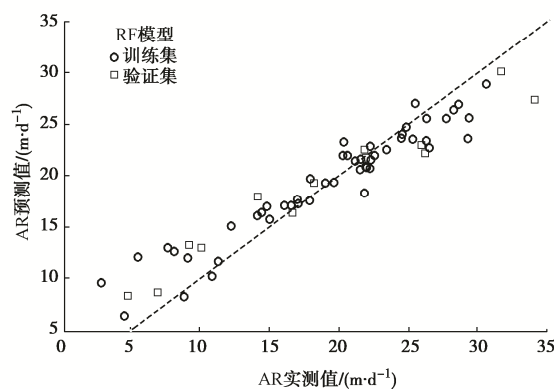
Fig. 6 Predicted values of AR by WRF model

各模型在训练集和验证集下的预测值与实测值散点结果, 如图 7 所示。

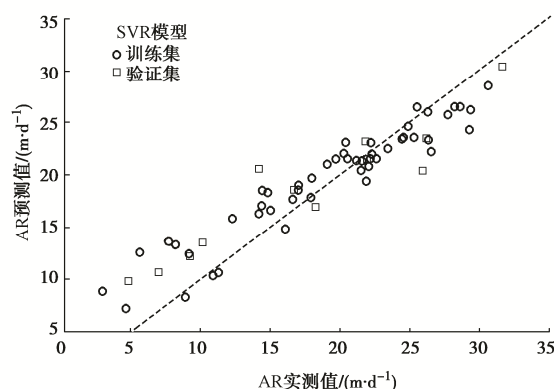
根据统计结果, 所有模型在训练集下的预测情况良好,  $R^2$  均大于 0.8, 除个别围岩条件极差的地质区段偏离零误差线较远外, 大部分数据的训练精度符合工程要求。从图 7 可以看出, 无论是训练集还是验证集数据, WRF 模型实测值和预测值的拟合结果与 1:1 线最接近, WRF 模型的预测性能最佳, 训练和验证阶段的  $R^2$  分别为 0.96, 0.97, 这说明改进的 WRF 模型通过加权的方法搜寻最优的超参数, 对默认参数情况下的 RF 预测效果优化作用显著, 基于赋权的随机森林模型对于输入参数的改变可以及时捕捉, 模型的冗余性低, 泛化性能好。RF 模型的拟合效果明显弱于 WRF 模型, 散点位置与 1:1 线的偏离程度相较于 WRF 存在明显差异, 训练集和验证集下的 RMSE 值分别为 2.62, 3.31, 与 WRF 模型相比差距也较大, 表明默认的超参数无法根据地质变异而进行自适应调整, 所训练的模型无法体现地质条件的不确定性。



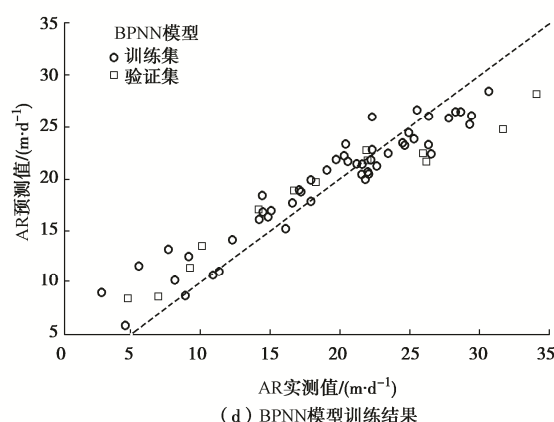
(a) WRF 模型训练结果



(b) RF 模型训练结果



(c) SVR 模型训练结果



(d) BPNN 模型训练结果

图 7 不同模型训练和测试样本预测结果散点图

Fig. 7 Scatter plot of predicted results for different models at training and testing stages

由图 7 中散点位置与 1:1 线的偏离程度可以看出, SVR 模型在训练集和验证集的结果误差均较大,

这是因为该模型对于超参数的选取十分敏感, 仅用小范围的随机寻优参数不能达到较高的预测精度, 模型泛化能力弱, 针对非线性回归预测问题需优化超参数才能够提高精度。值得注意的是, 对于 BPNN 模型, 训练集上的预测结果与实测值误差较小, 而在验证集上拟合结果偏离 1:1 线程度较大, 说明该模型出现过拟合现象。

此外, 4 个预测模型的性能评价指标  $R^2$ , RMSE 和 MAPE 进一步量化了上述特征, 如表 2 所示。由表 2 可见, 加权随机森林模型预测结果精度较高, 描述验证样本 AR 预测值和实测值之间偏差的 RMSE 值仅为 1.59, MAPE 值仅为 0.11, 说明 WRF 模型的预测性能最佳。训练和验证阶段的  $R^2$  分别为 0.96, 0.97, 数值并未发生很大偏差, 说明模型不存在欠拟合和过拟合现象。RF 模型的预测性能略逊于 WRF 模型, 但较其他两种模型的误差偏小。然而, 不论是训练集还是验证集, SVR 模型的 RMSE 值与 MAPE 值均较大,  $R^2$  值相较于其他模型明显偏低, 说明超参数的敏感性较大, 导致该模型的预测精度稳定性差。此外, BPNN 模型在训练集和验证集下的 RMSE 和  $R^2$  值差异较大, 表明模型的泛化能力及鲁棒性差。

表 2 不同模型的预测性能指标

Table 2 Predictive performance indice for different models								
评价 指标	WRF		RF		SVR		BPNN	
	训 练 集	验 证 集	训 练 集	验 证 集	训 练 集	验 证 集	训 练 集	验 证 集
RMSE	1.48	1.59	2.62	3.31	3.15	3.13	2.01	3.82
$R^2$	0.96	0.97	0.89	0.88	0.83	0.85	0.90	0.83
MAPE	0.11	0.11	0.19	0.21	0.21	0.28	0.18	0.22

4 TBM 掘进风险性印证

从现场实际掘进结果分析, TBM 在部分地质区段施工过程中遭遇过多次重大地质风险, 上述 4 种模型在对应地质区段的预测结果如表 3 所示。

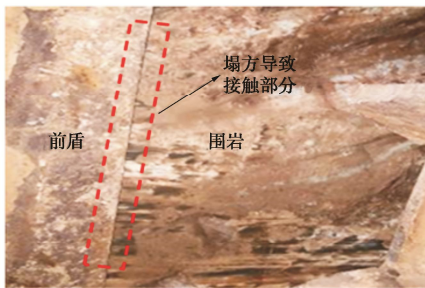
表 3 典型地质区段内不同模型的预测结果

Table 3 Comparison of prediction performance among different models						
桩号	地质 区段	AR 结果/(m·d <sup>-1</sup> )				
		真实值	WRF	RF	SVR	BPNN
T4+550— T8+737	1	12.24	13.92	15.02	15.89	14.05
	2	8.16	11.13	12.58	13.28	11.74
	3	21.84	19.95	18.27	19.37	20.03
T14+622— T15+100	4	7.68	9.86	12.98	13.57	13.20
T9+199— T9+331	5	2.88	6.17	9.62	8.87	9.00
T14+080— T14+100	6	5.54	10.34	12.11	12.60	11.60
	7	14.16	17.25	17.92	20.55	17.99
T19+752— T19+647	8	9.21	12.28	13.20	13.17	12.38

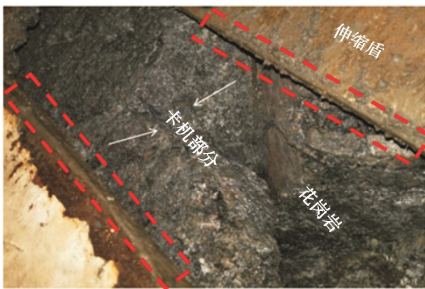
对于地质区段 1—区段 3, 该桩号范围内隧洞穿越  $f_2 \sim f_3$  小断层 (图 8 (a))。其中断层  $f_2$  产状  $100^\circ \angle 54^\circ$ , 断层带宽度约 0.3~1.0 m, 断层带物质为断层泥, 岩屑及碎裂岩体; 断层  $f_3$  上盘产状  $130^\circ \angle 60^\circ$ , 下盘产状  $140^\circ \angle 73^\circ$ , 断层带上部宽度 2~3 m, 下部宽度 8~9 m, 断层带物质为断层泥, 碎裂石英片岩岩体及糜棱岩。围岩类别为 III 类—IV 类, 掘进过程中会发生突水和塌方, 而且多组陡倾裂隙组易在两侧墙形成楔形体, 对施工有一定影响, 开挖过程中有沿侧墙小的掉块发生。局部洞段埋深大于 500 m, 硬质石英片岩在高应力作用下存在岩爆现象。基于该区段内收集的地质、机械以及施工管理参数建立的 WRF 模型预测结果分别为 13.92, 11.13, 19.95 m/d, 其中地质区段 2 的 AR 预测值的绝对误差达到 2.97 m/d, 表现为异常的掘进状况, 能够从侧面反映模型的输入参数具有表征 TBM 掘进过程中地质与机械异常信息的特点, 模型的训练可以提取各项有效特征, 对于不同掘进区段的风险划分起到提高精度与稳定性的作用。而且 WRF 模型的预测精度普遍高于其他模型, 表明基于加权改进的方法可以应对地质变异带来的参数不确定性问题, 具有良好的泛化性能。



(a) T4+550—T8+737段断层断裂带



(b) T14+622—T15+100段TBM前盾与围岩接触

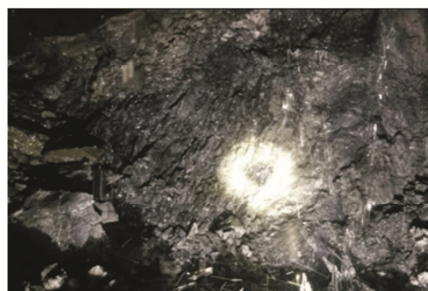


(c) T9+199—T9+331段TBM卡机情况





(d) T14+080—T14+100段TBM卡机情况



(e) T19+752—T19+647段围岩塌方

图8 典型地质区段风险

Fig. 8 Risks of typical geological section

地质区段4在掘进过程中隧洞拱顶出现塌方风险,造成TBM前盾与围岩接触(图8(b)),对应掘进桩号位于T14+622—T15+100 m, WRF模型的预测结果为9.86 m/d,该值与真实值之间的误差达到2.18 m/d,波动范围较正常掘进段表现出异常,可以反映当前地质区段面临一定的施工风险。而对比其他3种模型发现WRF模型预测精度最优,说明模型训练时能够充分挖掘输入参数的信息,对各种特征向量的权重实现了合理的反馈。

地质区段5,6掘进过程中突发卡机事故(图8(c)),掘进桩号为T9+199—T9+331。根据WRF模型预测得到的TBM施工速度结果分析发现,施工速度的输出目标分别为6.17, 10.34 m/d,该值大小处于较低的范围,表明TBM在此地质区段内掘进性能不佳,而且AR预测值与实际值的绝对误差分别达到3.3, 4.8 m/d,明显高于其他掘进段,即面临的施工不确定性较大,从而表明模型的预测结果可在一定程度上与掘进过程中的风险性相互印证。同时对比WRF与RF模型之间的绝对预测误差发现,经过自适应赋权处理后的模型能够较好地识别地质变异,该处地质风险区段的预测结果精确度明显提高,表明WRF的加权行为具有合理性。

此外,地质区段7掘进过程中发生软弱围岩的挤出变形,其掘进桩号位于T14+080—T14+100,其中顶拱的变形量超过了洞壁与护盾之间的间隙(约8 cm),围岩与护盾发生挤压接触(图8(d))。针对

WRF模型的预测结果来说,此掘进段内AR的预测值达到17.25 m/d,单纯的点预测值并未隐含较强的风险信息,然而观察TBM施工速度的绝对误差可发现明显异常,绝对误差达到3.09 m/d,相对稳定的模型出现较大误差波动,能够反映掘进过程中的变化。而通过对比WRF与其他3种模型的预测结果发现,基于加权方法改进的模型精度最高,进一步反映了权重的赋值具有衡量地质参数不确定性的功能,对于不同输入参数的训练起到至关重要的作用。

根据现场掘进反馈,地质区段8(桩号T19+752—T19+647)掌子面和顶拱发生严重塌方,TBM缓慢掘进通过了V类围岩破碎带(图8(e))。在该掘进段内基于WRF模型的TBM施工速度预测值为12.28 m/d,绝对误差达到3.07 m/d,其波动范围明显超过正常掘进区段,根据预测的绝对误差能够反馈前方掌子面面临较大的施工不确定性,TBM掘进受阻或发生卡机的风险较大。通过AR预测结果发现,WRF预测目标值相较于RF模型更接近真实值,说明模型训练时赋予输入参数的权值具有合理性,在保证模型稳定性的同时可以提高预测精度。

## 5 结 论

鉴于施工过程的随机性、地质参数的模糊性和人为因素的不确定性,选取可对岩体参数、机械参数及施工管理参数等参数的不确定性进行量化的指标,采用不同权重赋值的方法优化超参数,提出一种TBM施工速度加权随机森林预测模型。

(1)对于地质条件的不确定性,以岩石强度、岩体完整性、岩石硬度与耐磨性等因素量化;对于机械设备的不确定性,则以能够主动控制的刀盘推力和转速来衡量;同时考虑ORD能够在一定程度上表征设备和管理这两个方面的不确定性,引入该指标衡量人为因素的不确定性。

(2)通过对输入参数赋予不同的权重,提出加权随机森林优化超参数,该模型训练速度较快,不易出现过拟合且无需复杂的调参处理,具有高精度、不易过拟合等优点。

(3)建立的TBM施工速度加权随机森林预测模型,其训练集和验证集的预测效果良好,验证集的均方根误差和决定系数分别为1.59, 0.97,优于RF, SVR和BPNN模型,表现出更佳的泛化能力和鲁棒性。

(4)通过分析预测模型绝对误差异常值发现,预测模型的波动与超前地质预报结果或现场掘进过程中遭遇的地质风险相互印证,基于模型绝对误差提出的预警可在一定程度上降低地质灾害的影响程度,对于



保障 TBM 快速、安全施工具有重要的指导意义。

## 参考文献:

- [1] ROSTAMI J, OZDEMIR L. A new model for performance prediction of hard rock TBMs[C]// Proceedings of Rapid Excavation and Tunneling Conference (RETC). Boston, 1993: 793-809.
- [2] BRULAND A. Hard Rock Tunnel Boring[D]. Trondheim: Norwegian University of Science and Technology (NTNU), 1998.
- [3] BARTON N. TBM Tunnelling in Jointed and Faulted Rock[M]. Boca Raton: CRC Press, 2000:170-175.
- [4] ALVAREZ G M, BRUINES P A, VERHOEF P N W. Modeling tunnel boring machine performance by neuro-fuzzy methods[J]. Tunnelling and Underground Space Technology, 2000, **15**(3): 259-269.
- [5] ARMAGHANI D J, KOOPIALIPOOR M, MARTO A, et al. Application of several optimization techniques for estimating TBM advance rate in granitic rocks[J]. Journal of Rock Mechanics and Geotechnical Engineering, 2019, **11**(4): 779-789.
- [6] ZHOU J, YAZDANI B B, JAHED A D, et al. Forecasting of TBM advance rate in hard rock condition based on artificial neural network and genetic programming techniques[J]. Bulletin of Engineering Geology and the Environment, 2020, **79**(4): 2069-2084.
- [7] 邓 军, 雷昌奎, 曹 凯, 等. 采空区煤自燃预测的随机森林方法[J]. 煤炭学报, 2018, **43**(10): 2800-2808. (DENG Jun, LEI Changkui, CAO Kai, et al. Random forest method for predicting coal spontaneous combustion in gob[J]. Journal of China Coal Society, 2018, **43**(10): 2800-2808. (in Chinese))
- [8] 王仁超, 朱品光. 基于随机森林回归方法的爆破块度预测模型研究[J]. 水力发电学报, 2020, **39**(1): 89-101. (WANG Renchao, ZHU Pinguang. Study on blasting fragmentation prediction model based on random forest regression method[J]. Journal of Hydroelectric Engineering, 2020, **39**(1): 89-101. (in Chinese))
- [9] 李明超, 史博文, 韩 帅, 等. 基于对穿声波波速的岩体完整性多尺度评价新指标与分析方法[J]. 岩石力学与工程学报, 2020, **39**(10): 2060-2068. (LI Mingchao, SHI Bowen, HAN Shuai, et al. New index and analysis method for multi-scale rock mass integrity assessment based on P-wave velocity[J]. Chinese Journal of Rock Mechanics and Engineering, 2020, **39**(10): 2060-2068. (in Chinese))
- [10] BREIMAN L. Random forests[J]. Machine Learning, 2001, **45**(1): 5-32.
- [11] 闫长斌, 姜晓迪, 刘章恒, 等. 基于岩碴粒径分布规律的 TBM 破岩效率分析[J]. 岩土工程学报, 2019, **41**(3): 466-474. (YAN Changbin, JIANG Xiaodi, LIU Zhangheng, et al. Rock-breaking efficiency of TBM based on particle-size distribution of rock detritus[J]. Chinese Journal of Geotechnical Engineering, 2019, **41**(3): 466-474. (in Chinese))
- [12] FROUGH O, TORABI S R, YAGIZ S, et al. Effect of rock mass conditions on TBM utilization factor in Karaj-Tehran water conveyance tunnel[C]// World Tunnel Congress. Thailand, 2012.
- [13] FROUGH O, TORABI S R, YAGIZ S. Application of RMR for estimating rock-mass-related TBM utilization and performance parameters: a case study[J]. Rock Mechanics and Rock Engineering, 2015, **48**(3): 1305-1312.
- [14] 龚秋明, 卢建伟, 魏军政, 等. 基于岩体分级系统(RMR)评估预测 TBM 利用率研究[J]. 施工技术, 2018, **47**(5): 92-98, 127. (GONG Qiuming, LU Jianwei, WEI Junzheng, et al. Study on estimation and prediction of TBM utilization rate using rock mass rating(RMR)[J]. Construction Technology, 2018, **47**(5): 92-98, 127. (in Chinese))
- [15] KOHAVI R. A study of cross-validation and bootstrap for accuracy estimation and model selection[C]//Proceedings of the 14th International Joint Conference on Artificial Intelligence-Volume 2. New York, 1995.