

doi: 10.3969/j.issn.1672-6073.2025.03.012

# 基于机器学习的近接工程 隧道沉降预测方法

韩玉珍<sup>1</sup>, 聂小凡<sup>1</sup>, 华福才<sup>1</sup>, 何纪忠<sup>1</sup>, 张连卫<sup>1</sup>, 潘 毫<sup>1</sup>, 张 雷<sup>2</sup>, 王旭阳<sup>1</sup>

(1. 北京城建设计发展集团股份有限公司, 北京 100037; 2. 北京城建集团有限责任公司, 北京 100088)

**摘要:** 针对既有地铁隧道受新建隧道近接施工引发沉降的预测问题, 采用自动化机器学习方法与多源数据加权思路, 建立一个以隧道属性、地层属性、相对位置关系为输入, 以既有隧道沉降值为输出的预测模型, 并以万级别仿真数据及现实数据进行算例分析。结果表明: 自动化机器学习算法能基于仿真数据得到精度较高的预测模型, 但在多源数据集上表现不佳; 通过对多源数据加权的方式可以提升模型泛化能力, 达到以仿真数据为基础, 专精现实数据的效果, 当加权后的现实数据数量与仿真数据比在同一数量级时, 模型误差较低; 根据表现最佳模型的特征重要度, 地层损失率是预测的核心输入参数, 地质、空间关系、施工属性三者的重要程度近乎一致。

**关键词:** 城市轨道交通; 近接工程; 隧道沉降预测; 机器学习; 多源数据; 加权方法

中图分类号: U231

文献标志码: A

文章编号: 1672-6073(2025)03-0084-07

## Machine Learning-Based Prediction Method for Tunnel Settlement in Proximity Engineering

HAN Yuzhen<sup>1</sup>, NIE Xiaofan<sup>1</sup>, HUA Fucui<sup>1</sup>, HE Jizhong<sup>1</sup>, ZHANG Lianwei<sup>1</sup>,  
PAN Hao<sup>1</sup>, ZHANG Lei<sup>2</sup>, WANG Xuyang<sup>1</sup>

(1. Beijing Urban Construction Design and Development Group Co., Ltd., Beijing 100037;  
2. Beijing Urban Construction Group Co., Ltd., Beijing 100088)

**Abstract:** To address the issue of settlement in existing subway tunnels due to the construction of new adjacent tunnels, an automated machine learning approach and a strategy for weighting multi-source data were employed. A predictive model was developed, taking into account tunnel characteristics, stratum properties, and relative positional relationships as input parameters, with the settlement values of the existing tunnels as the output. The model was tested using tens of thousands of simulation data points and real-world data. The findings suggest that while the automated machine learning algorithm can produce a highly accurate predictive model based on simulation data, it may not perform as well with multi-source data sets. By assigning weights to multi-source data, the model's ability to generalize can be improved, leading to an optimized model that specializes in real-world data, based on simulation data. When the quantity of weighted real-world data is comparable to the simulation data, the model's error rate is reduced. Additionally, according to the feature importance of the best-performing model, the stratum loss rate emerges as a critical input parameter for prediction, with the significance of geological conditions, spatial relationships, and construction attributes being nearly equivalent.

**Keywords:** urban rail transit; proximity engineering; tunnel settlement prediction; machine learning; multi-source data; weighting method

收稿日期: 2024-08-22 修回日期: 2024-12-30

第一作者: 韩玉珍, 女, 博士, 教授级高级工程师, 主要从事地下工程方面的研究工作, hanyuzhen@bjucd.com

通信作者: 聂小凡, 男, 硕士, 工程师, 主要从事地下工程方面的研究工作, niexiaofan@bjucd.com

基金项目: 雄安新区建设科学技术计划项目(No.2021-03, No.2021-07); 广西重点研发计划(2023AB29062)

引用格式: 韩玉珍, 聂小凡, 华福才, 等. 基于机器学习的近接工程隧道沉降预测方法[J]. 都市轨道交通, 2025, 38(3): 84-90.

HAN Yuzhen, NIE Xiaofan, HUA Fucui, et al. Machine learning-based prediction method for tunnel settlement in proximity engineering[J]. Urban rapid rail transit, 2025, 38(3): 84-90.

## 0 引言

随着地下空间利用强度的增加,既有地铁隧道周边的近接施工活动日益频繁,此类施工活动将不可避免地对既有隧道产生干扰<sup>[1]</sup>,因此预先评估施工影响程度至关重要。目前,常用的评估手段包括理论计算、数值模拟、模型试验等,其中,数值模拟相比其他方法,能够灵活应对不同问题的边界条件,较贴合实际工程的复杂性。然而数值模拟的运用,高度依赖于操作者的经验,其计算效率与模型采用的本构关系及建模的精细化程度密切相关<sup>[2]</sup>。因此,数值模拟的合理应用与深入研究存在一定门槛,在处理愈发多样化与普遍化的地下设施影响评估问题时,即使针对同一场景下的类似任务,仍需要耗费大量的时间与人力,难以利用现有成果合理推算。鉴于此,评估分析中迫切需要一种能够高效、准确预测影响程度,并尽量减少人为操作和引导的方法。

而随着计算机算法与硬件水平的不断提升,近年来机器学习为解决传统预测问题提供了新方向。该方法与传统基于物理模型的正向推导不同,机器学习关注数据之间的潜在相关性,直接构建输入参数与输出参数之间的函数关系,实现输入与输出之间“端对端”的联系。已有学者在地下工程<sup>[3-4]</sup>、病害识别<sup>[5]</sup>等领域使用机器学习方法开展研究。ZHANG等<sup>[6]</sup>综合比较了5种常用的机器学习算法的预测性能,认为随机森林算法在识别隧道沉降方面最为出色;陈仁朋等<sup>[7]</sup>综述了机器学习方法预测盾构掘进地表沉降的相关研究,围绕预测模型的输入参数、预测目标、预测算法和沉降控制4个关键环节进行系统性的阐述;陈泽彬<sup>[8]</sup>基于大量有限元案例,提出一种用深度卷积神经网络预测悬臂结构二维弹性应力场的方法,预测结果与数值模拟的平均误差小于3%;ZHANG等<sup>[9]</sup>以7个涵盖土层层性质与开挖空间特征的输入参数构建可预测隧道水平位移的模型,结果显示基于机器学习算法构建的模型对于未参与训练的工程监测数据亦具有一定泛化能力;WANG等<sup>[10]</sup>提出一个通用的非线性结构深度学习框架DeepSNA,针对钢板剪力墙结构受力性能的案例,该框架显示出强大的预测能力且计算效率超越数值方法千倍以上。

众多学者的研究为自动化和智能化解决工程问题提供了有效路径,然而,各类方法与数值模拟一样,仍有部分环节高度依赖于经验或人工操作。一方面,预测模型的构建者需要根据物理力学准则和工程实践经

验来筛选关键因素,并将其作为预测模型的输入变量;另一方面,机器学习的诸多算法需预先设定若干个超参数,这些超参数在很大程度上决定了预测模型的表现,而超参数的选取往往依赖于经验判定。

鉴于上述问题,本文针对新建隧道近距离施工引发既有地铁隧道沉降这一典型预测任务,采用自动化机器学习算法的思路,建立基于决策树及其衍生算法的预测模型;模型训练集的构建纳入了工程实测数据和数值模拟等多源数据,通过权重分配来区分不同数据来源的贡献度,进而开发了一套针对多源数据的加权训练算法;经包含实际工程数据的检验,计算模型满足工程预测精度,并分析了权重与数据量对于模型表现的影响。

## 1 研究对象与数据

### 1.1 研究对象

既有地铁隧道周边新建隧道是典型的近接施工工程,其中既有隧道附近新建隧道是这一问题的重要分析场景。以两条单洞隧道为例,既有隧道受到附近新建隧道的施工扰动时,会出现竖向沉降变形,影响既有隧道的受力性能与服役状态。对于该研究对象,预测任务通常以隧道属性、地层参数、隧道空间位置关系等作为输入信息,以既有隧道竖向变形作为输出信息,一个工况下的输入与输出共同组合成一个有效数据。有效数据可通过数值模拟、理论分析、现场监测以及模型试验等方法得到。数值模拟与理论分析能对多种工况进行分析,且通常计算时间较为稳定,可批量提供较多数据;而模型试验与现场监测均依赖于实物世界中的操作,因此每一个工况的实施成本均较高,所能提供的数据量较少。对于输入信息特征数量较多的近接工程,通常需要数倍特征数量的数据才可支撑算法训练,目前尚无公开的、可直接满足机器学习模型训练数据规模的数据集。

### 1.2 数据获取与处理

为了使数据规模满足机器学习模型的基础要求,多数研究以数值模拟数据打造足量数据,因而训练模型可以视其为对数值模拟的代理计算。但数值模拟是对实际工程在一个角度上的数学假定,预测模型更是在数学假定之上的函数拟合,对实际工程的指导能力并不稳定。为获取多角度信息,提升模型的工程指导能力,本文在采用数值模拟方法取得了大量仿真数据的同时,还收集了现场监测与模型试验得到的现实数据,以此得到了一个数据总规模超万条的数据集。其

中现实数据主要来源于公开文献<sup>[11-15]</sup>以及各城市地铁隧道在近接施工期的监测报告；仿真数据主要来自于前序研究中基于 Python-有限元计算的自动化生成方案<sup>[16]</sup>，该方案可根据预设的工况参数自动化生成模型、计算模型、并提取既有隧道受周边不同位置单洞隧道施工的响应位移，主要的输入参数及对应取值范围如表 1 所示。

表 1 自动化生成方案的输入参数及取值范围  
Table 1 Input parameters and ranges of values for automated generation schemes

类型	参数	取值范围
隧道属性	隧道外径 $D_0/m$	5.9、6.0、6.2
	衬砌厚度 $d/m$	根据 $D_0$ 取 0.3 或 0.35
	隧道埋深 $h_t/m$	10~30
	衬砌弹性模量 $E/kPa$	$3.45 \times 10^7$ 或 $3.55 \times 10^7$
	地层损失率 $\beta/\%$	0.1~1.25
地层属性	地层厚度 $H/m$	0~100
	割线模量 $E_{50}/kPa$	$1 \times 10^3 \sim 3 \times 10^4$
	切线模量 $E_{oed}/kPa$	根据 $E_{50}$ 计算
	卸载再加载模量 $E_{ur}/kPa$	根据 $E_{50}$ 计算
	初始剪切模量 $G_0/kPa$	根据 $E_{50}$ 计算
	黏聚力 $c/kPa$	0~150
	内摩擦角 $\varphi/(\circ)$	5~45
	剪胀角 $\psi/(\circ)$	根据 $\varphi$ 计算
相对位置	地下水深度 $h_w/m$	0~10
	隧道间水平方位夹角 $\theta_a/(\circ)$	0~90
	隧道间竖直方位夹角 $\theta_b/(\circ)$	0~180
	隧道净距 $D/m$	1.5~24

以 3 层地层下两个近接的单洞隧道作为典型场景，则以上生成方案中的一个案例可由 38 个参数表征，具体为：新建隧道及既有隧道属性各 5 个、相对位置关系 3 个、每层地层属性 8 个以及一个统一的地下水深度。保持其他来源数据的输入参数与表 1 中各项仿真参数一致，同时以既有隧道在新建施工后竖向沉降位移作为输出数据，以此组成共计 10 024 个可供学习模型训练的结构化数据，每个数据均表征了一个独特的工况。

## 2 研究方法

### 2.1 决策树及其衍生算法

机器学习的核心目标是通过程序手段使机器更熟练和高效地完成工作。在机器学习分类中，隧道的沉降预测问题通常属于监督式学习(supervised learning)中的回归(regression)问题。这类问题的特点是利用已

标注的数据来建立输入变量与输出变量之间的映射关系模型，其中模型的输出为连续值。对于该类问题，目前常用的算法包括决策树、神经网络、支持向量机等。其中，决策树是一种树状结构的预测模型，其基本原理是将输入数据从各个结点排列到某个叶结点。最初的决策树模型已经演进出多种算法，其中分类与回归决策树(classification and regression tree, CART)模型采用了一种每次分裂产生两个子节点的二叉树结构，既可用于分类问题也可用于回归问题，因此得到了广泛应用。

CART 模型作为一种基学习器(base learner)，为提升其预测精度及性能，研究者们提出了集成学习(ensemble learning)的概念，即通过优化或组合多个基学习器以完成学习任务。在集成学习中，有两种主流的思路：Boosting 和 Bagging。Boosting 方法是在训练一个基学习器后，以残差为目标再次进行预测，以此连续地训练模型来关注前一个模型在预测中的错误，从而逐步提高整体的预测性能，得到强学习器。该思路下具有代表性的算法有梯度提升决策树(gradient boosting decision tree, GBDT)、Xgboost、LightGBM、Catboost 等。Bagging 方法则是通过从原始数据集中多次放回取样，并行地训练多个不同基学习器，最终将各基学习器组合形成一个强学习器，具有代表性的算法有随机森林(random forest, RF)、极度随机树(extra tree)等。集成学习得到的集成模型在预测性能上远优于基学习器，能够将预测精度仅比随机猜度略高的弱学习器增强为预测精度高的强学习器，因而已成为预测模型的首选方法。

### 2.2 多源加权算法

数据来源与数据价值息息相关，模型试验、现场监测等方法得到的输出数据纳入了现实空间各项因素的影响，但数量稀少、获取代价高；而数值模拟、理论计算等方法基于各类数学假定，输入与输出具有复杂但稳定的映射关系。不同规则下，各类数据来源在输入输出过程中呈现出各异的映射特征，若能构建一个整合这些映射特征的复合模型，即可从多个角度数据观察该问题的数据空间，得到与总规则更接近的预测模型。由此，模型在数据设置上，应充分挖掘现实数据在模型中的指导能力，又加入仿真数据稳定的力学框架，使模型兼顾多来源数据进行综合学习。为了表现这种数据差异，本文提出一种多源加权算法，通过权重凸显不同来源数据的差异，完成对多源数据的

选择性学习过程。经过自动化生成,大量的仿真数据可以驱动模型训练,而多源数据加权算法可以突出少量现实数据的作用,从而达成以仿真数据为基础,挖掘现实数据作用的效果。

权重主要运用在模型训练的误差计算中,权重的直接含义是一个现实数据的误差相比于仿真数据放大多少倍,也即认定在训练过程中一个现实数据的影响力等于多少个仿真数据,是对不同来源数据价值的主观认定。比如,对于现实数据的均方根误差(RMSE),通过加权方式可调整为:

$$E_H = \omega \cdot \text{RMSE} = \omega \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - p_i)^2} \quad (1)$$

式中,  $E_H$  为现实数据的误差,在本文场景中具体指近接施工后隧道的预测沉降值与实际沉降值的差值,误差越小表示模型表现越好;  $\omega$  为现实数据的权重;  $n$  为数据个数;  $y_i$  和  $p_i$  分别为第  $i$  个数据的真实值与预测值。

### 2.3 自动化机器学习算法

在确立了学习算法并准备构建预测模型之前,必须对算法中的一系列超参数进行预设,超参数的选择对模型的预测效果具有决定性影响,不恰当的超参数将导致过拟合或欠拟合现象。因此,超参数的优化是模型构建的必要流程,优化的复杂程度取决于超参数的个数及对应的取值空间。验证一组超参数的效果便需要完整地训练一次模型,超参数较少时,可以使用穷举思路的网络搜索法寻求最优超参数,但对于集成学习的复杂度,穷举思路会导致计算成本剧增。而这类优化又进一步考验了使用者对相关算法的熟悉程度,这些调试内容与岩土工程所对应物理问题的本身含义相距甚远,从而提高了工程人员的使用门槛。为了解决这一问题,近年来学者们提出了自动化机器学习的解决思路。

自动化机器学习的目标是自动设计和开发机器学习系统,以此减少人为交互的工作量。其可以在没有相关领域知识的情况下,完成数据特征提取、算法选择、超参数优化、多模型集成等学习任务中的一项或全部。自动化机器学习大幅度减轻了在数据科学侧的经验需求,成为各专业领域建立预测模型的重要助力。根据不同的任务类型及设计原则,学者们提出了诸多自动化思路,如 FLAML<sup>[17]</sup>(fast library for automated machine learning)框架中综合分析学习模型、超参数、样本数量以及重采样策略间的关系,运用调节机制定量了计算开销,以此完成多模型比选及超参数优化。该框架

的流程主要如图 1 所示。

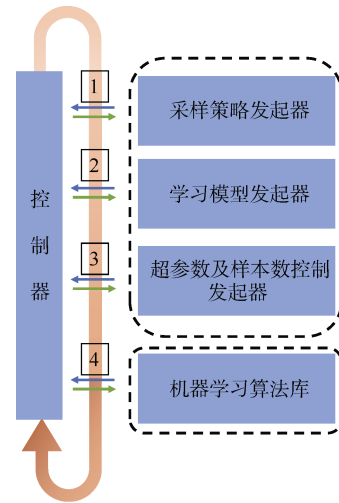


图 1 FLAML 的运算框架

Figure 1 FLAML framework

基于数据集的特征个数与数据规模,本研究采用决策树及其衍生算法这一类匹配性较好的机器学习算法。在选择具体决策树算法类型以及超参数时,采用自动化机器学习的 FLAML 框架进行运算。该框架可在 Python 平台中实现,运行时,仅需要定义拟花费的总运算时间。在时间限制内,算法将按照图 1 中的顺序逐步完善模型细节,并通过迭代寻找最优模型。该框架根据数据规模及运行时限,首先通过采样策略发起器确定验证集,之后通过学习模型发起器调用某个决策树模型进行训练,然后根据训练结果,对模型的超参数以及每次训练的样本个数进行调整。以上三步后,通过机器学习算法库对不同的决策树算法进行选择,重复以上过程直至得到最优模型。使用该框架可大幅减轻模型调试阶段的工作,同时采用统一的优化思路有助于客观对比多源数据的相关权重设置。

## 3 算例分析

### 3.1 训练时间及误差分析

使用自动化机器学习算法时,虽然延长运算时间通常能获得更优的模型性能,但性能提升的边际效益会随着时间转移显著下降。通常情况下,学习过程存在一个临界优化时间,当超过这个时间时,模型的表现不会再显著提升。为节约资源,在未加入现实数据及加权操作前,使用充足的时间(21 600 s)供 FLAML 进行模型调试。其中预先随机抽样 5%的仿真数据作为独立的测试集,其余仿真数据作为训练集,以测试集的 RMSE 作为误差标准,衡量不同预测模型的表现,结果如图 2 所示。

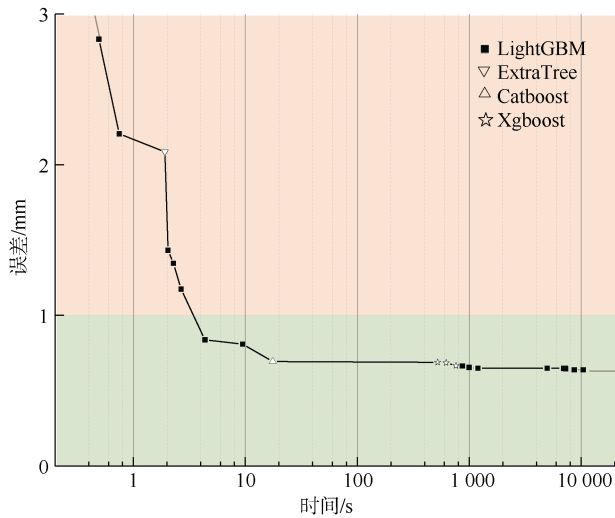


图 2 误差与计算时间的关系

Figure 2 Relationship between error and calculation time

从图 2 中可以看出,模型误差在 10 s 内迅速降至 1 mm 以下(图 2 中绿色区域),在 10~1 000 s 内出现两次明显的下降,而超过 1 000 s 后,误差几乎不再变化,达到 0.63 mm。而在自动优化过程中,共使用过 4 种基于决策树的算法模型,其中 LightGBM 等 Boosting 算法在误差较大时(图 2 中红色区域)完成了显著的误差降低,最优模型亦为该算法调试得到,体现出这一算法对近接工程隧道沉降预测的适用性。根据误差的下降趋势,后续模型采用 1 200 s 作为最大训练时间,以在节约运算成本的条件下完成多工况对比。

### 3.2 加权数据表现

数据集中大量数据来源于自动化生成的有限元仿真数据,而为了体现量少但含有指导价值的现实数据,采用了前文所述的多源加权思路。设定仿真数据权重为 1,对现实数据分别设置权重为 0、1、10、100、1 000、10 000 共 6 个水平进行测试,得到各个模型的 RMSE 误差如图 3 所示。值得说明的是,权重仅在训练过程中计算误差使用,而在训练后的调用过程中不使用,即图 3 中所列测试集误差未进行加权。

由图 3 可以看出,在现实数据权重为 0,即在训练中不考虑现实数据时,模型训练集误差较小,但测试集误差很大,相比图 2 中最优的模型,测试集误差从 0.63 mm 升为了 20.54 mm,说明仅由仿真数据训练出的单源模型对现实数据预测性能很差,不同来源的数据不具备相同的映射关系,不经过处理则会拉低模型的整体表现。而当权重为 1,即认为现实数据与仿真数据相同时,测试集与训练集误差均有所下降,但仍存在较大误差。当权重在 10~1 000,即现实数据

加权后的数据量与仿真数据逐渐达到同一个量级时,误差均有显著降低,但测试集与训练集的误差仍存在较大差异,模型存在一定过拟合现象。在权重超过 1 000,即加权后现实数据的数据量级超过仿真数据时,误差水平再次升高,训练集误差开始显著上升,算法已难以调整误差水平过高的现实数据。可以看出,当现实数据加权后的数量与仿真数据在一个数量级,即两者比在 1/10~10 以内时,模型的表现较好,可以同时学习到多源数据中的规律,并综合得到一个满足两者数据输入的映射关系,在该范围内进一步分析,误差结果如图 4 所示。

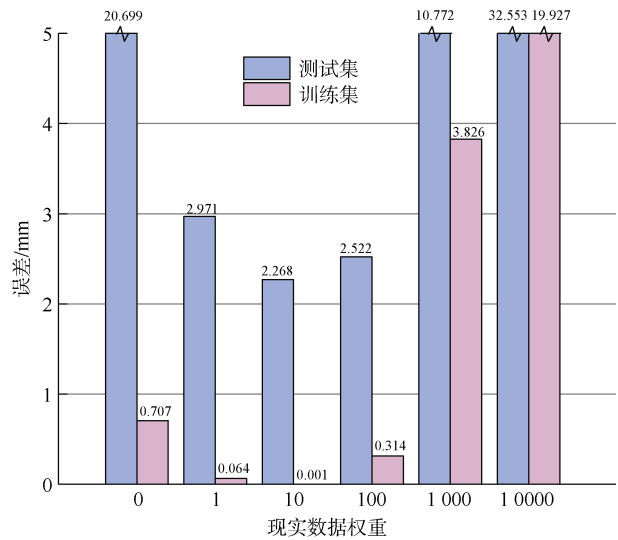


图 3 不同权重下的模型误差

Figure 3 Model error under different weights

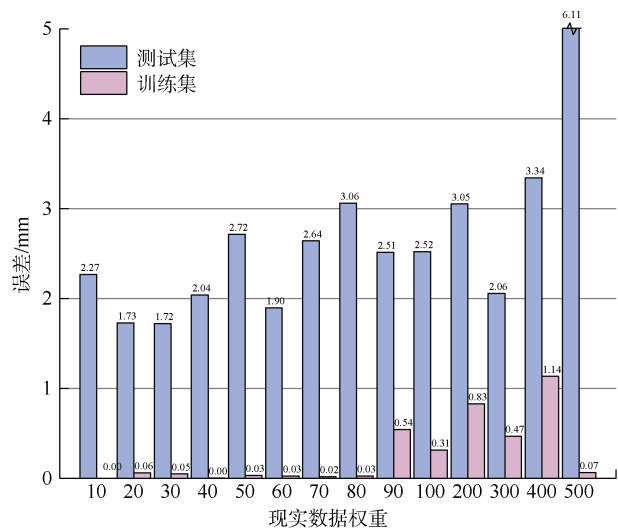


图 4 细分区间内的模型误差分布

Figure 4 Model error distribution

当加权后现实数据与仿真数据的数量在同一量级时,模型表现存在一定波动,大部分模型的误差介于1~3 mm之间。同时,权重小于100时的误差总体低于权重大于100时,说明现实数据占比略小于仿真数据时,多源加权算法能较好平衡少量现实数据与海量仿真数据的指导性,使模型预测表现较好。其中最优模型的权重取值为30,该模型在测试集中,均方根误差1.72 mm、决定系数0.96。根据误差案例具体分析,发现对于隧道隆起的若干个工况,模型的预测误差相对较大,可能因为训练样本中各量值工况分布不均,该类工况总数较少,在小样本量条件下难以完成该类预测规律的学习。

### 3.3 特征重要度分析

根据测试集表现,误差最低的最优模型通过XgBoost算法生成,是一个经457次迭代的11层决策树。对于决策树类的算法,可以根据树上结点的排列及出现次数,计算每个特征参数对于预测的重要程度,即特征重要度,其反映了各个参数对于预测输出值时的贡献。根据这一算法特点,权重为0的单源模型(训练中仅使用仿真数据)、权重为1的多源同权模型以及权重为30的多源加权模型的特征重要度如表2所示。表2中特征重要度经过归一化处理,代表了各特征对计算结果重要程度的百分比,同时,地层属性中4个模量被统一归纳为刚度参数,摩擦角、黏聚力、剪胀角被统一归纳为强度参数。表中未列出重要度低于1%的次要特征。

表2 不同权重模型的特征重要度  
Table 2 Feature importance of models with different weights %

类型	特征	单源模型 ( $\omega=0$ )	多源同权 模型( $\omega=1$ )	多源加权 模型( $\omega=30$ )
隧道位置	水平方向角	2.83	2.69	1.16
	竖向方向角	30.80	31.39	21.33
	净距	4.62	4.92	1.22
既有隧道	隧道埋深	2.31	2.46	1.40
	隧道埋深	11.02	10.49	4.51
新建隧道	地层损失率	11.61	12.82	32.50
	地层厚度	4.24	4.01	1.23
第一层土	刚度参数	17.43	16.49	14.76
	强度参数	3.14	3.06	2.20
	刚度参数	4.48	4.35	6.76
第二层土	强度参数	1.18	1.18	0.73
	刚度参数	3.22	2.28	3.11
地下水	水位深度	0.53	0.51	6.01

从表2中可以看出,单源模型与多源同权模型在各个特征的重要程度上较为相似,两个模型重要特征前三位均是竖向方向角、第一层土刚度参数、新建隧道地层损失率,同时各特征重要度的变化幅度在 $\pm 1.5\%$ 之间,说明模型在仿真数据的训练中学习到的规律较为通用,经少量调整后就能大幅提升泛化能力,更适应现实数据。而多源加权模型的特征重要度相比于前两个模型则发生大幅变化。在多源加权模型中,新建隧道的地层损失率成为了该模型预测的核心参数,占比近1/3;与之相对的,隧道位置关系及新旧隧道埋深等与方位相关特征的预测权重有所下降;对于各地层参数,3个模型的特征重要度均未发生大幅变化;而地下水深度的重要程度显著提升,至6%左右。可以认为,当强化现实数据的作用时,预测模型的预测模式产生了本质变化,其牺牲了部分仿真数据的规律性,而强化了现实数据中更为敏感重要的特征,比如决定沉降量值的地层损失率、地下水深度,但总体误差仍在工程精度要求内,满足了“以仿真数据为基础,专精现实数据”的预期。另一方面,对于最优模型,地质条件属性的重要度合计约35%,与新建隧道施工相关参数、两隧道位置属性的重要度接近,即地质、空间关系、施工属性为3个在支撑预测任务上同等重要的信息,较符合人工判别时的经验性共识。

## 4 结论

本文结合机器学习算法,对新建隧道近距离施工引发既有地铁隧道沉降这一典型场景进行了深入分析,得出以下结论。

1) 基于有限元仿真的数据经自动化机器学习算法训练后的预测表现较好,测试集均方根误差小于1 mm;而加入基于物理法则的监测、模型数据等多源数据后,该模型的预测误差显著上升,表现出对新数据的不适用性,反映出数据的来源将显著改变模型,应注重不同来源数据的差异,比如部分特征在不同来源中分布等问题。

2) 对多源进行加权可以有效改变预测模型的表现,当现实数据加权后的数量与仿真数据比在1/10~10之间时,模型效果相对较好。经平行试验得出现实数据相当于30个仿真数据时,可得到测试集均方根误差小于2 mm,决定系数0.96的最佳预测模型。

3) 单源模型与多源同权模型在特征重要度上较为相似,说明模型可以从仿真数据学习到较为通用的规律,并经少量调整后适应现实数据;而多源加权数

据强化了隧道施工与地下水的相关特征,使地质、空间关系、施工属性成为3个同等重要的信息。

4) 本文预测模型以仿真数据为基础,专精现实数据,但仍存在一些局限性。在仿真数据中,均为非同期施工的单洞隧道静力响应,有待在提升模型复杂度的情况下研究同期施工或施工时间差等长期效应。同时,数据权重作为一项依赖经验设置的参数,其与数据量之间的最优关系仍需要更多案例验证。

## 参考文献

- [1] 罗程, 夏雪勤, 曾云嵘, 等. 邻近基坑降水影响下既有隧道受力变形分析[J]. 都市轨道交通, 2024, 37(2): 99-106.  
LUO Cheng, XIA Xueqin, ZENG Yunrong, et al. Force and deformation of shield tunnel under the influence of pre-dewatering in adjacent foundation pit[J]. Urban rapid rail transit, 2024, 37(2): 99-106.
- [2] 樊健生, 王琛, 宋凌寒. 土木工程智能计算分析研究进展与应用[J]. 建筑结构学报, 2022, 43(9): 1-22.  
FAN Jiansheng, WANG Chen, SONG Linghan. Research and application of intelligent computation in civil engineering[J]. Journal of building structures, 2022, 43(9): 1-22.
- [3] MORGENROTH J, KHAN U T, PERRAS M A. An overview of opportunities for machine learning methods in underground rock engineering design[J]. Geosciences, 2019, 9(12): 504.
- [4] 张研沁, 狄宏规, 徐永刚. 基于机器学习的地铁盾构隧道长期沉降预测[J]. 都市轨道交通, 2022, 35(3): 89-94.  
ZHANG Yanqin, DI Honggui, XU Yonggang. Machine learning based prediction model for long-term settlement in a metro-shield tunnel[J]. Urban rapid rail transit, 2022, 35(3): 89-94.
- [5] 孟瑞锋, 梁楨, 贾超, 等. 钢轨表面缺陷检测 Mask R-CNN 算法研究与优化[J]. 都市轨道交通, 2024, 37(5): 68-77.  
MENG Ruifeng, LIANG Zhen, JIA Chao, et al. Study and optimization of mask R-CNN algorithm for surface defect detection on railway tracks[J]. Urban rapid rail transit, 2024, 37(5): 68-77.
- [6] ZHANG Pin, WU Huaina, CHEN Renpeng, et al. Hybrid meta-heuristic and machine learning algorithms for tunneling-induced settlement prediction: a comparative study[J]. Tunneling and underground space technology, 2020, 99: 103383.
- [7] 陈仁朋, 邹聂, 吴怀娜, 等. 盾构掘进地表沉降机器学习预测与控制研究综述[J]. 华中科技大学学报(自然科学版), 2022, 50(8): 56-65.  
CHEN Renpeng, ZOU Nie, WU Huaina, et al. Review of prediction and control for surface settlement caused by shield tunneling based on machine learning[J]. Journal of Huazhong University of Science and Technology (Natural Science Edition), 2022, 50(8): 56-65.
- [8] 陈泽彬. 基于深度学习的有限元计算研究[D]. 哈尔滨: 哈尔滨工业大学, 2020.  
CHEN Zebin. Study of the finite element modeling and analysis based on deep learning[D]. Harbin: Harbin Institute of Technology, 2020.
- [9] ZHANG Dongmei, SHEN Yiming, HUANG Zhongkai, et al. Auto machine learning-based modelling and prediction of excavation-induced tunnel displacement[J]. Journal of rock mechanics and geotechnical engineering, 2022, 14(4): 1100-1114.
- [10] WANG Chen, SONG Linghan, FAN Jiansheng. End-to-End Structural analysis in civil engineering based on deep learning[J]. Automation in construction, 2022, 138: 104255.
- [11] 孔恒, 宋克志. 城市地下工程邻近施工关键技术与应用[M]. 北京: 人民交通出版社, 2013.  
KONG Heng, SONG Kezhi. Key technologies and application of close construction in city underground engineering[M]. Beijing: China Communications Press, 2013.
- [12] 汪洋, 何川, 曾东洋, 等. 盾构隧道正交下穿施工对既有隧道影响的模型试验与数值模拟[J]. 铁道学报, 2010, 32(2): 79-85.  
WANG Yang, HE Chuan, ZENG Dongyang, et al. Model test and numerical simulation of influence of perpendicular undercross shield tunnel construction on existing tunnel[J]. Journal of the China railway society, 2010, 32(2): 79-85.
- [13] LIANG Rongzhu, XIA Tangdai, HONG Yi, et al. Effects of above-crossing tunnelling on the existing shield tunnels[J]. Tunneling and underground space technology, 2016, 58: 159-176.
- [14] 金大龙, 袁大军, 韦家昕, 等. 小净距隧道群下穿既有运营隧道离心模型试验研究[J]. 岩土工程学报, 2018, 40(8): 1507-1514.  
JIN Dalong, YUAN Dajun, WEI Jiaxin, et al. Centrifugal model test of group tunneling with small spacing beneath existing tunnels[J]. Chinese journal of geotechnical engineering, 2018, 40(8): 1507-1514.
- [15] 韩玉珍, 潘毫, 张雷, 等. 城市轨道交通穿越风险工程案例集[M]. 北京: 中国建筑工业出版社, 2021.  
HAN Yuzhen, PAN Hao, ZHANG Lei, et al. Urban Rail Transit Penetration Risk Engineering Case Collection[M]. Beijing: China Architecture & Building Press, 2021.
- [16] WANG Binghua, et al. Machine learning-based prediction method of subway tunnel settlement under the influence of adjacent construction[C]//CICTP 2024. Shenzhen, China. American Society of Civil Engineers, 2024.
- [17] Wang C, Wu Q, Weimer M, et al. FLAML: A fast and lightweight automl library[J]. Proceedings of Machine Learning and Systems, 2021(3): 434-447.

(编辑: 傅依萱)