

隧洞掘进机掘进数据预处理方法及工程案例验证

聂其坦¹, 肖浩汉², 刘飞¹, 刘立鹏², 牛瑞强¹

(1. 广东粤海粤西供水有限公司, 广东 湛江 524000; 2. 中国水利水电科学研究院, 北京 100048)

摘要: 全断面硬岩隧洞掘进机(TBM)在施工过程中自动获取了海量的掘进数据,适当筛选与清理掘进数据能够提高数据质量,对隧洞工程的智能化施工具有重要意义。因此,根据引绰济辽工程 TBM 掘进数据的特征,提出了集成的 TBM 掘进数据预处理方法,包括完整掘进段提取、内部掘进段分割以及掘进参数降噪。此外,采用时序预测更强的长短期记忆网络(LSTM)算法建立了扭矩切深指数(TPI)预测模型,以验证数据预处理方法的有效性。结果表明:所提出的数据预处理方法能够显著改善数据质量,提高深度学习模型的预测精度,验证集的拟合优度 R^2 由 0.503 提升至 0.721,相关系数 R' 由 0.809 提升至 0.900,平均相对误差 MRE 由 3.107 下降至 0.096。研究成果对于提高 TBM 掘进数据的准确性和可靠性具有重要意义,也可为相关领域的研究提供参考。

关键词: 隧洞掘进机(TBM);数据预处理;核密度估计;巴特沃斯滤波;长短期记忆网络(LSTM)

中图分类号:U455

文献标识码:A

文章编号:1672-643X(2024)05-0191-10

TBM excavation data preprocessing method and engineering case verification

NIE Qitan¹, XIAO Haohan², LIU Fei¹, LIU Lipeng², NIU Ruiqiang¹

(1. Guangdong Yuehai Yuexi Water Supply Co., Ltd., Zhanjiang 524000, China;

2. China Institute of Water Resources and Hydropower Research, Beijing 100048, China)

Abstract: The full face hard rock tunnel boring machine (TBM) automatically generates massive excavation data during the tunnel construction process. Proper screening and cleansing of excavation data is crucial to data quality, which also has great guiding significance for the intelligent construction of tunnel engineering. Therefore, based on the characteristics of TBM excavation data in the Yinchuo Project, an integrated TBM excavation data preprocessing method is proposed, which includes complete excavation segment extraction, internal excavation segmentation, and excavation parameter noise reduction. To verify the effectiveness of the proposed data preprocessing method, a torque cut depth index (TPI) prediction model is developed by the long short-term memory (LSTM) algorithm, which has strong temporal prediction capabilities. The results demonstrate that the proposed data preprocessing method can significantly improve data quality and enhance the prediction accuracy of deep learning models. For the validation dataset, R^2 increases from 0.503 to 0.721, R' ascends from 0.809 to 0.900, and MRE plummets from 3.107 to 0.096. These research achievements bear profound implications for enhancing the precision and reliability of TBM tunneling data, thereby offering invaluable insights for further exploration in the related domain.

Key words: tunnel boring machine(TBM); data preprocessing; kernel density estimation; Butterworth filtering; long short-term memory (LSTM)

1 研究背景

隧洞掘进机(tunnel boring machine, TBM)是一种

用于围岩切削、出渣、注浆、支护并实现连续隧洞挖掘作业的综合机械设备^[1]。与传统隧洞掘进方法相比, TBM 具有施工速度快、开挖质量高、经济效益好、掘进

收稿日期:2023-11-26; 修回日期:2024-05-13

基金项目:中国水利水电科学研究院基本科研业务费专项项目(GE0145B012021);国家自然科学基金项目(52179121);流域水循环模拟与调控国家重点实验室自主研究课题(SK12022ZD05)

作者简介:聂其坦(1987—),男,山东济宁人,学士,工程师,主要从事水利工程建设管理工作。

通讯作者:肖浩汉(1993—),男,河北衡水人,博士,工程师,主要从事水利工程科研工作。

过程安全、利于环境保护和降低劳动强度等优点,是地铁隧洞、输水隧洞和公路隧洞等工程施工过程中的重要工法之一,具有广阔的市场应用前景^[2]。

TBM 掘进过程可以看作是持续的大型岩-土现场实验,其中涉及到复杂的“人-机-岩”交互关系^[3]。由于受到地层不均匀和地质水文条件复杂等因素的综合影响,各个物理参数之间的关系十分复杂且多变,传统的物理和数学方法很难构建泛化能力强的模型来描述这种关系。随着计算机算力的提升和人工智能的快速发展,深度学习成为了城市智能规划、结构智能设计、建筑智能建造和隧洞智能维护等领域的研究和应用热点^[4-5]。许多学者开始利用 TBM 掘进过程中获得的数据进行掘进参数优化^[6]、地质风险识别^[7]、姿态智能调整^[8]等研究工作,已涌现出许多与隧洞掘进相关的人工智能研究成果,例如,结合机器学习、数值建模方法(如有限元法)^[9]和数字建模方法(如建筑信息建模)^[10-11],实现对隧洞全过程的实时建模分析;将人工智能技术和物联网技术相结合,实现对大量数据的实时处理,监测隧洞开挖引起的地面响应^[12]和掘进机性能^[13]。

“人-机-岩”交互关系的准确描述基于对数据的挖掘和解译。然而,TBM 掘进过程中采集到的数据规模巨大,具有时序性、高维多元性和低价值密度性等特点^[14],直接使用未经处理的原始数据进行智能化研究很难达到良好的应用效果。此外,由于定义

不统一、数据源不规范、处理技术和评估方法的缺乏,庞大的 TBM 掘进数据给数据挖掘和分析带来了巨大的困难。因此,对海量的 TBM 掘进数据进行适当的筛选与清理以提高数据质量是十分必要的。

本研究基于引绰济辽工程 2 号隧洞的 TBM 历史掘进大数据,提出了适合长输水隧洞的 TBM 掘进数据预处理方法,并通过掘进效率指标进行预测验证,对比了数据预处理前后深度学习模型的预测效果。研究成果对隧洞工程智能化建设具有重要的指导作用和战略意义。

2 数据来源与研究方法

2.1 研究区概况

引绰济辽工程位于内蒙古自治区东北部,水路起自嫩江支流绰尔河,经过绰尔河和洮儿河两个流域,将水引至西辽河,设计最大年调水量为 $4.88 \times 10^8 \text{ m}^3$,该工程将缓解西辽河流域缺水问题。本研究所用的 TBM 数据来自引绰济辽工程的输水工程施工二标(2 号隧洞),其全长为 58.51 km,洞线比降为 1/3 000,开挖直径为 5.2 m,断面呈圆形。隧洞的埋深主要集中在 70~120 m 之间,最大埋深达 250 m,引绰济辽工程地理位置及本文研究段概况见图 1。洞线岩性主要包括砂岩、砾岩、熔岩、凝灰角砾岩和凝灰岩等,围岩的种类主要分为 II、III、IV 和 V 类,各种类别占比见图 2。

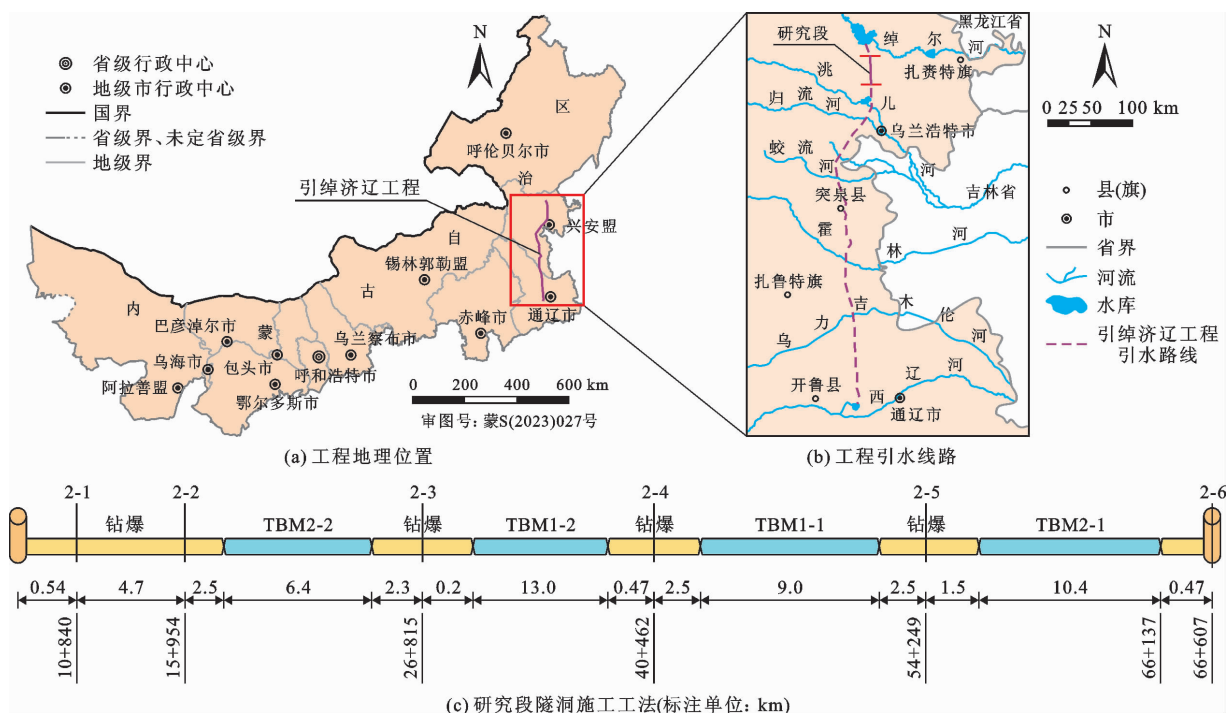


图1 引绰济辽工程地理位置及本文研究段概况

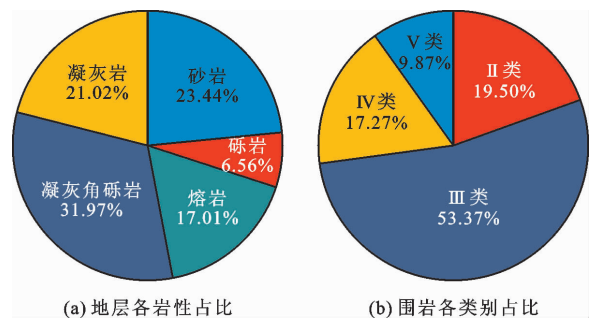


图 2 研究区域的地层岩性和围岩类别占比

该研究区域采用敞开式 TBM 进行长距离掘进,掘进总长度为 38.17 km。本研究收集了桩号 66 + 124 ~ 60 + 799 和桩号 53 + 583 ~ 49 + 786,共 9.122 km 的掘进数据,总数据量高达 100 GB,数据详细记录了总推进力、刀盘扭矩和掘进速度等 401 项参数,完整地保存了设备在不同地层和不同工况下的数据。TBM 的主要设备参数见表 1。

表 1 引绰济辽工程 TBM 设备参数

设备参数	单位	规格
刀盘直径	mm	5 200
刀具数量	把	34
刀盘功率	kW	2 100
整机长度	m	300
最大油缸行程	mm	1 850
额定刀盘扭矩	kN · m	3 340
额定刀盘推力	kN	11 340
最大刀盘转速	r/min	11.5
最大出渣速度	m ³ /h	560
数据采集频率	Hz	1
数据采集种类	个	403

2.2 数据来源

2.2.1 掘进性能参数 TBM 设备采集的参数按照属性可分为 3 类:标记参数、设定参数和运行参数^[15]。标记参数反映掘进机在时间和空间上的运行信息;设定参数是指在掘进过程中驾驶者常常需要调节的参数,反映出驾驶者的操作状态;运行参数则是一种被动响应参数,其数值与地质条件和驾驶者操作密切相关。

(1)标记参数。代表性的标记参数为掘进时间和桩号。掘进时间(Time)记录了掘进机的运行时间信息,在掘进过程中,以 1 Hz 的频率进行全天 24 h 的连续记录,以便通过时间节点快速找到对应的掘进信息。桩号(Chainage)记录了掘进机的空间位置信息,从隧洞起点开始测量,依次通过编号表示,单位为 m,例如桩号 98 + 750 代表距离隧洞起点

98 750 m 的位置。

(2)设定参数。在引绰济辽工程中,设定参数包括刀盘转速设定值(n_s)和掘进速度设定值(v_s)。刀盘转速(运行参数,缩写为 n)由设定值 n_s 控制。驾驶者首先通过控制刀盘转速电位器的输出电压信号来设定刀盘转速,然后通过模数转换器将电压信号转换为数字信号,在上位机上记录 n_s 的数值。随后, n_s 的数字信号进一步转换成频率信号,用于控制电机变频器的输出频率,从而控制电机的转速。掘进速度(运行参数,缩写为 v)由设定值 v_s 控制,但同时会受到负载压力的影响,掘进速度的控制方式与刀盘转速类似。 v_s 的数字信号被传递给推进油缸的电磁阀,从而控制推进油缸的流量,当单位时间内推进油缸的流量增大时,则 v 随之增大。但负载压力过大会导致溢流,即使不断增加 v_s 来改变推进油缸的流量, v 也难以提高。

(3)运行参数。运行参数种类繁多,其中刀盘推力(F)和刀盘扭矩(T)是现场驾驶者最为关注的两个参数。刀盘推力是通过安装在推进油缸上的压力传感器测量得到的;刀盘扭矩是通过测量驱动电机的功率反算得到的。

为了消除不同设备之间的差异,考虑了平均载荷,也称为单刀推力 F 或单刀扭矩 T 。单刀推力 F 和单刀扭矩 T 分别为总推进力 F_t 和总扭矩 T_t 与滚刀数量 N 的比值,见公式(1)、(2)。

$$F = \frac{F_t}{N}$$
(1)

$$T = \frac{T_t}{N}$$
(2)

2.2.2 掘进效率指标 掘进性能参数是掘进机运行状态的直观反映。为了准确地衡量掘进机的破岩能力,一些学者对掘进性能参数进行了特征转换和重构。例如,Sanio^[15]通过大量室内岩石切削试验发现,刀盘推力(F)与贯入度(p)的 0.5 ~ 1.5 次方成正比,从而提出了基于掘进性能参数的场切深指数 FPI ,该指数表示单位贯入度下所需要的刀盘推力,可以反映机-岩相互作用的过程。

然而,目前对于 FPI 的现场计算仍存在争议。罗华^[16]认为 FPI 增长率与岩体力学参数呈现复杂的对数关系,而 Jing 等^[17]通过现场分级掘进试验证明了刀盘推力和贯入度的关系可通过一次线性函数表达。Wang 等^[18]则通过统计分析大量的掘进数据,提出了 F 和 p 的“双折线”形式。针对 FPI 计算方式的争议,Chen 等^[19]从能量守恒的角度提出了扭矩切深指数

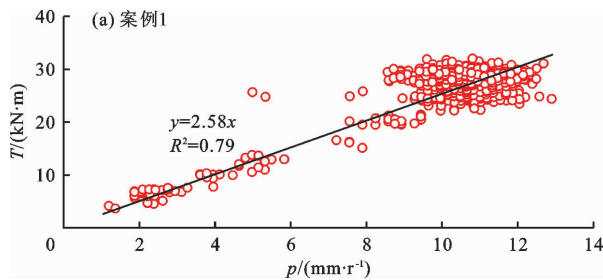
TPI , 该指数表示单位贯入度下所需要的刀盘扭矩, 能够反映 TBM 设备在不同地层条件下的运行状态。陈祖煜等^[20]认为掘进过程中破岩做功等于刀盘扭矩的能量消耗, 可以使用公式(3)表示。

$$W_T = T \cdot n = \theta \cdot E \cdot v \cdot A \quad (3)$$

式中: W_T 为单位时间内掘进机扭矩能量消耗, J; T 为扭矩, $\text{kN} \cdot \text{m}$; n 为刀盘转速, r/min ; θ 为掘进机能量输出效率, 同一台设备或同一工程中 θ 可认为是不变量; E 为单位体积岩体破碎所需能量, J; v 为掘进速度, mm/min ; A 为开挖掌子面面积, m^2 , 同一工程中 A 为常数。

在较短掘进过程中一般认为地质条件是均质的, 则公式(3)可转换为公式(4)。

$$P_T = \frac{W_T}{v} = \frac{T \cdot n}{v} = \theta \cdot E \cdot A \quad (4)$$



式中: P_T 为掘进机刀盘扭矩能量消耗速率, $\text{kN} \cdot \text{m}/(\text{mm}/\text{r})$ 。岩体性质基本一致时, 掘进机能量消耗速率 P_T 为定值。因此, 单组掘进段内扭矩切深指数 TPI 的定义可表达为公式(5)~(6)。

$$TPI = \frac{T \cdot n}{v} = \frac{T}{p} \quad (5)$$

$$p = \frac{v}{n} \quad (6)$$

图3展示了典型刀盘扭矩(T)与贯入度(p)关系的回归结果。由图3可以发现, 掘进过程中刀盘扭矩基本随着贯入度的增加而呈线性增大, 两者具有非常好的线性关系, 其比值(T/p)可以表示为反映破岩能力的 TPI 指标。与 FPI 相比, TPI 的定义更加明确且计算方式简单, 适用于 TBM 现场掘进过程, 该指标在一定程度上也反映了岩石力学基础理论^[20]。

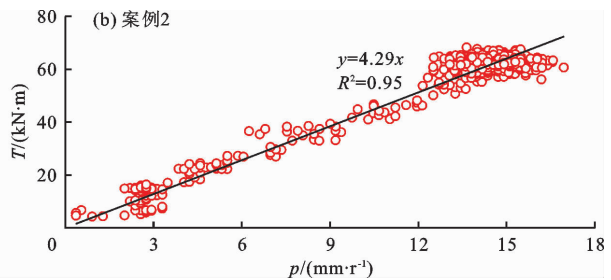


图3 典型刀盘扭矩(T)与贯入度(p)关系的回归结果

2.3 研究方法

本节旨在分析 TBM 掘进过程并总结掘进数据的特征, 进而归纳数据的发展规律, 在此基础上, 提出一种集成的 TBM 掘进数据预处理方法, 包括完整掘进段提取、内部掘进段分割和掘进参数降噪 3 个过程。这 3 个过程之间具有明显的先后顺序, 首先, 由完整掘进段提取出有效掘进过程数据; 其次, 由内部掘进段分割确定适用于机器学习的上升段和稳定段数据; 最后, 进行掘进参数降噪, 滤除掘进过程的噪音信息。集成的 TBM 掘进数据预处理方法见图 4。

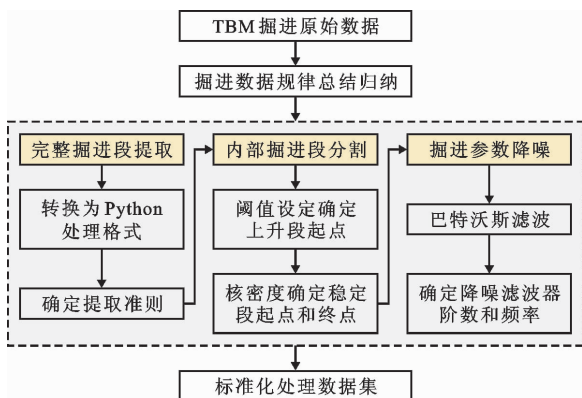


图4 集成的 TBM 掘进数据预处理方法

2.3.1 典型 TBM 掘进过程 TBM 工作模式遵循严格的先后顺序。首先, 驾驶员从零开始逐渐增加刀盘转速设定值, 同时观察刀盘扭矩与刀盘推力的数值。一旦达到指定转速, 驾驶员缓慢调节掘进速度设定值。当设备达到适当的掘进状态时, 驾驶员保持参数稳定, 直到掘进结束。最后, 驾驶员逐渐降低刀盘转速设定值至零, 完成单个掘进段的开挖工作, 然后进行其他工序作业。归纳掘进参数的发展规律, 整个掘进过程可分为空推、上升、稳定和下降 4 个阶段。

(1) 空推段: 在一个完整的掘进过程结束后, 掘进机会向后退一段距离 (2 ~ 3 cm), 然后再次向前推进, 形成空推段。在此阶段, 驾驶员根据地质勘探结果设定初始的 n_s 和 v_s 数值。由于刀盘尚未接触掌子面, 空推段的 v 数值较大, 而 T 和 F 的数值保持较小水平。

(2) 上升段: 空推段结束后, 刀盘与隧洞掌子面岩体接触, T 和 F 数值呈现上升趋势, 驾驶员根据出渣情况再次调整 n_s 和 v_s 数值, 以达到最佳的掘进效果。

(3) 稳定段: 在均质岩体中, 驾驶员基本不再主动调整 n_s 和 v_s 数值, 掘进机基于上升段设定的参数

开始稳定地进行掘进,各项掘进参数保持稳定。

(4)下降段:完成掘进工作后,驾驶员关闭推进系统,此时机器不再向前推进。 v 和 F 瞬间降为零,而 T 和 n 则逐步下降至零。

2.3.2 掘进数据预处理

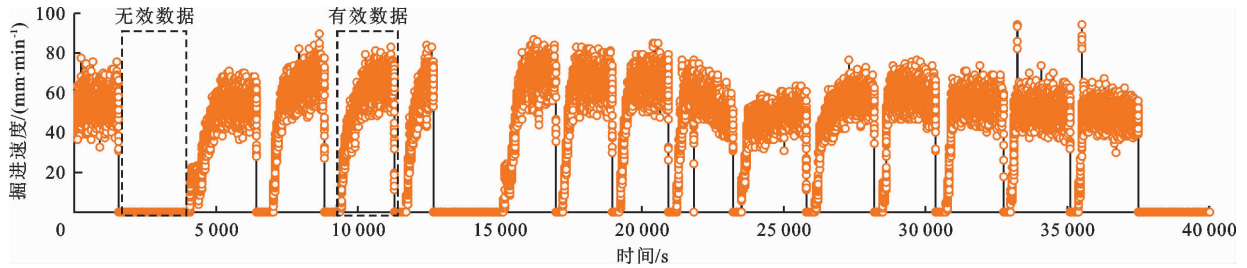


图5 引绰济辽工程TBM掘进速度有效数据与无效数据示例

本文在总结引绰济辽工程整个TBM掘进过程数据的变化特征基础上,提出了一种新的有效数据提取方法,该方法第 i 组掘进段的提取表达式如(7)~(10)。

$$\text{起点 } n_s^i > 0 \quad (7)$$

$$\text{终点 } n_e^i \leq 0 \quad (8)$$

$$0 < s_e^i - s_s^i < l_{\text{油缸}} \quad (9)$$

$$T_{p-\min} < T_p(n_s^{i+1} - n_e^i) \quad (10)$$

式中: n_s^i 为掘进段刀盘转速第1个大于零的索引值,r/min; n_e^i 为掘进段刀盘转速第1个小于等于零的索引值,r/min; s_e^i 、 s_s^i 分别为掘进段 i 的终点位移、起点位移,m; $l_{\text{油缸}}$ 为推进油缸的理论伸长量,m,可通过设备出厂默认设置获得; $T_{p-\min}$ 为相邻掘进段的最小时间间隔,s,可通过统计分析获得。条件式(7)~(10)同时成立时,则认为第 i 组掘进段是一个有效的掘进过程。

上述提取方法简化了起始点判别参数,并限定了正常掘进数据范围,与二值函数判别方法^[17]相比,该方法考虑了现场地质条件的复杂性和驾驶员行为的难预料性,规避了部分掘进参数短暂异常变为零值,并将一组完整掘进段拆分为多个的风险。

(2)内部掘进段分割。完整掘进段中上升段和稳定段数据用于TBM破岩,驾驶员可以根据反馈结果调整掘进参数。目前,利用上升段的数据推演稳定段的掘进参数是一种常见的做法^[21],有助于为驾驶员提供掘进指导。上升段起点可以通过阈值设定来确定,但由于操作者习惯和地层条件差异等因素,上升段终点即稳定段起点,难以通过常规统计方法确定。因此,本研究根据稳定段数据的特点,提出了一种基于核密度估计法的掘进段内部分割方案。核

(1)完整掘进段提取。在TBM掘进过程中,由于设备的日常维护、刀具更换以及非正常停机等原因,会产生大量无效数据,有效数据与无效数据示例如图5所示。因此,在提取有效数据之前,首先需要排除这些无效数据。

密度估计是直方图的一种自然拓展,属于非参数检验方法^[22],该方法不依赖有关数据分布的先验知识,也不做任何分布假设,在统计学理论和实际应用领域得到广泛应用。本文核密度估计法确定内部掘进段分割的思路如下。

给定掘进参数样本 $(x_1, x_2, \dots, x_n)$,通过经验分布函数(函数式11)来计算估计分布函数。

$$F_n(t) = \frac{\text{样本中小于等于 } t \text{ 的元素数}}{n} = \frac{1}{n} \sum_{i=1}^n I(x_i \leq t) \quad (11)$$

式中: $F_n(t)$ 为小于等于 t 的样本的概率; n 为样本数量; $I(\cdot)$ 为指示函数,当 $x_i \leq t$ 时, $I(x_i \leq t) = 1$,否则 $I(x_i \leq t) = 0$ 。

当求解 $X = x$ 处的密度函数值时,可采用公式(12)进行计算。

$$f(x) = \lim_{h \rightarrow 0} \frac{F(x+h) - F(x-h)}{2h} \quad (12)$$

式中: h 为点 x 附近的小区间,也称为带宽。

将分布函数用公式(11)替代,公式(12)中分子项为落在 $[x-h, x+h]$ 区间的点个数。则 $f(x)$ 的估计表达为公式(13)。

$$\begin{aligned} \hat{f}_h(x) &= \frac{1}{2h} \frac{\sum_{i=1}^N I(x-h \leq x_i \leq x+h)}{N} \\ &= \frac{1}{Nh} \sum_{i=1}^N \frac{I}{2} \times \left(\frac{|x - x_i|}{h} \leq 1 \right) \end{aligned} \quad (13)$$

为得到更光滑的密度函数,取

$$K_0(t) = \frac{1}{2} \times I(t < 1) \quad (14)$$

则 $f(x)$ 的估计变为

$$\int \hat{f}_h(x) = \frac{1}{Nh} \sum_{i=1}^N K_0\left(\frac{x-x_i}{h}\right) \quad (15)$$

$f(x)$ 的估计积分表示为

$$\begin{aligned} \int \hat{f}_h(x) dx &= \frac{1}{Nh} \sum_{i=1}^N \int K_0\left(\frac{x-x_i}{h}\right) dx = \frac{1}{N} \sum_{i=1}^N \int K_0(t) dt \\ &= \int K_0(t) dt \end{aligned} \quad (16)$$

当 K_0 的积分等于 1 时, 则 $f(x)$ 估计积分等于 1。如果采用正态分布的密度函数作为核函数, 则 $\phi(x)$ 估计为

$$\hat{f}_h(x) = \frac{1}{Nh} \sum_{i=1}^N \phi\left(\frac{x-x_i}{h}\right) \quad (17)$$

核密度估计中的带宽 h 和核函数 K_0 是决定 TBM 掘进参数的密度分布平滑程度的两个重要参数^[23]。如果 h 选择过大, 则不符合 h 趋向于零的要求, h 选择过小, 则用于估计 $f(x)$ 的点过少。

(3) 掘进参数降噪。在采集传感器信号时, 由于受到刀盘不规则频率振动、系统电磁辐射和设备电路等因素的影响, 会产生“噪声”干扰。这些噪声具有较大波动性, 对数据挖掘过程造成一定影响。巴特沃斯滤波器 (Butterworth filter) 是一种常见的电子滤波器, 常用于模拟电路中对信号进行频率选择和滤波^[24], 它具有最大平坦的幅度响应, 在通带内具有最小的幅度波纹, 并且能够在阻带内实现最大的衰减。因此, 本文使用巴特沃斯滤波器对掘进参数进行降噪处理。当给定滤波器的阶数为 m 、截止频率为 ω_c 时, 巴特沃斯滤波器的平方幅度可采用公式 (18) 进行计算^[25]。

$$|H(\omega)|^2 = \frac{1}{1 + (\omega/\omega_c)^{2m}} \quad (18)$$

通过选择适宜的滤波器参数 (滤波器阶数和截止频率), 可以实现对 TBM 掘进数据准确和有效的频率选择和滤波。

3 结果与分析

3.1 预处理结果

将集成完整掘进段提取、内部掘进段分割和掘进参数降噪 3 个过程的 TBM 掘进数据预处理方法转化为 Python 语言, 开发 TBM-Processing 数据预处理软件。通过该软件, 共从引绰济辽工程原始掘进数据中提取出 3 851 组有效数据, 平均掘进长度为 1.38 m, 平均掘进时间为 1 498 s, 单位长度平均掘进时间为 1 085 s/m。

3.1.1 内部掘进段分割 以掘进速度参数为例, 说明通过所述内部掘进段分割方法确定稳定段起点和

终点索引值的过程, 如图 6 所示。

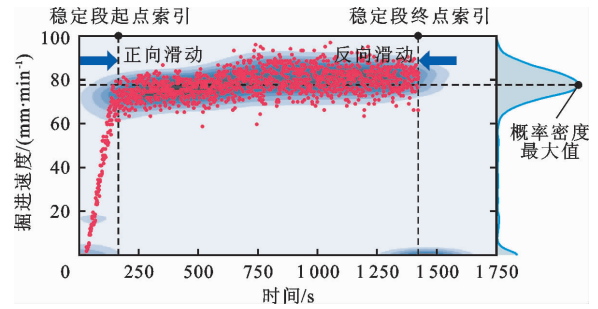


图 6 掘进速度的核密度估计图

首先, 采用高斯核函数作为核函数 K_0 。带宽 h 根据样本分布的经验公式进行评估^[26]。

$$h = 1.06 \min\left\{\sigma, \frac{IQR}{1.34}\right\} n^{-1/5} \quad (19)$$

式中: σ 为样本的标准差; IQR 为箱线图的四分位距离, 即第 3 四分位数与第 1 四分位数之差。

然后, 得到掘进速度的核密度估计并将掘进速度密度的最大值作为稳定段的代表值。进而, 从左向右正向遍历掘进段的数据, 在误差小于设定误差 (本文设定为 0.2 mm/min) 的情况下, 将对应点的索引作为稳定段起点的索引值。最后, 从右向左反向遍历掘进段的数据, 以同样的方式得到稳定段终点的索引值。通过上述这些步骤, 完成内部掘进段的分割工作。

3.1.2 掘进参数降噪 巴特沃斯滤波器阶数与阻带振幅衰减速度成反比^[26]。较高阶数的滤波器可以提供更陡的衰减, 但同时也增加了计算和实现的复杂性。为满足工程应用的要求, 本研究选择了较低的阶数 2 作为滤波器的参数。

截止频率决定了信号通过滤波器的频率范围, 可以通过调整截止频率来选择需要保留或者消除的信号频率范围。图 7 展示了截止频率为 0.02、0.04、0.08、0.10、0.20、0.40 时掘进速度或刀盘扭矩参数的原始值和滤波值。分析图 7 可知, 当截止频率设置为 0.02 时, 滤波曲线过于平滑; 而当截止频率为 0.40 时, 滤波曲线与原始曲线之间的差异不显著, 无法实现有效降噪。在综合考虑平滑性和准确性基础上, 本研究将滤波器的截止频率设置为 0.10。

3.1.3 掘进性能参数变化 表 2 统计了数据预处理前后掘进性能参数的变化情况。表 2 显示, 数据预处理后, 运行参数刀盘扭矩、刀盘推力和设定参数刀盘转速的平均值均有所增大, 标准差和变异系数均相应降低, 数据的离散程度大幅减弱。同时, 掘进速度的极端值 (如最小值和最大值) 得以平抑, 参数

回到正常范围。

表 3 统计了数据预处理前后刀盘扭矩 (T) 与贯入度 (p) 关系线性拟合优度 R^2 的变化情况。由表 3 可以看出,数据预处理前,工程中拟合优度 $R^2 > 0.6$ 的掘进段占比为 31.40%,而数据预处理后,拟合优度 $R^2 > 0.6$ 的掘进段占比为 79.39%。可见经过数据预处理后 $T-p$ 拟合优度提升效果明显,数据质量大为改善。

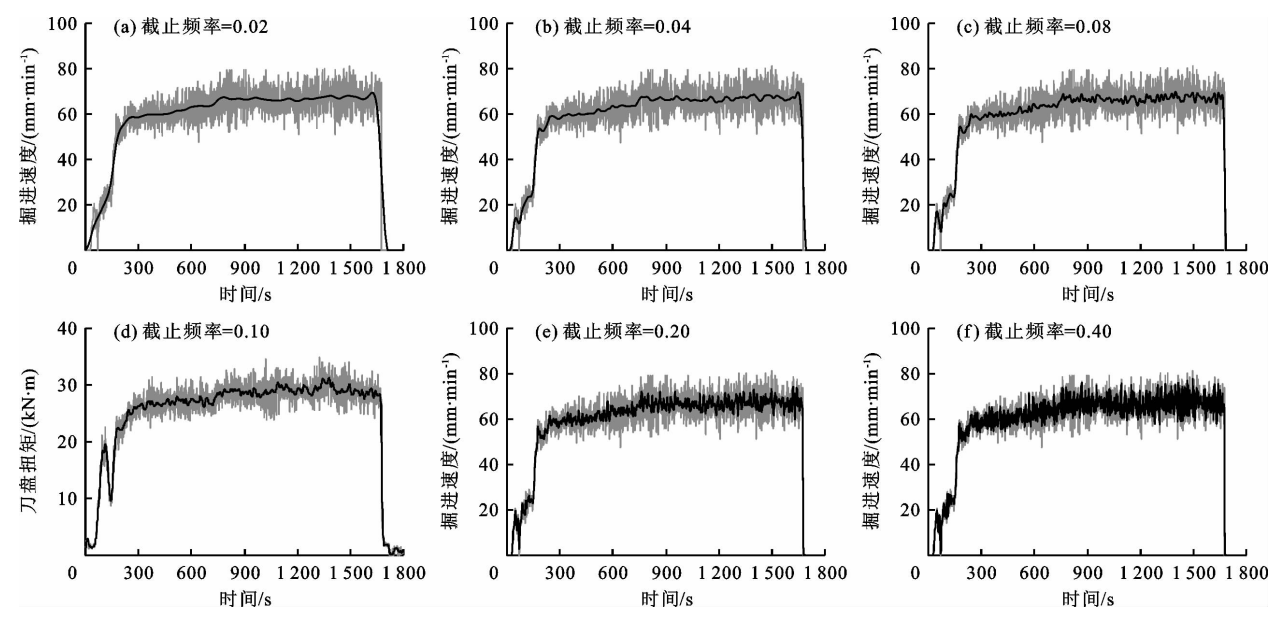


图 7 不同截止频率下掘进速度或刀盘扭矩的原始值和滤波值

表 2 数据预处理前后掘进性能参数统计指标的变化

性能参数	预处理前					预处理后				
	平均值	标准差	最小值	最大值	变异系数	平均值	标准差	最小值	最大值	变异系数
刀盘扭矩 $T/(\text{kN} \cdot \text{m})$	25.98	5.64	0	92.45	0.22	26.36	4.90	0.28	91.87	0.19
刀盘推力 F/kN	216.46	20.13	0	384.14	0.10	218.79	16.33	34.36	372.23	0.07
刀盘转速 $n/(\text{r} \cdot \text{min}^{-1})$	8.26	0.30	0	10.25	0.04	8.29	0.24	1.41	10.25	0.03
掘进速度 $v/(\text{mm} \cdot \text{min}^{-1})$	58.79	24.3	0	138.76	0.38	57.88	9.16	0.69	119.53	0.16

表 3 数据预处理前后 $T-p$ 关系线性拟合优度的变化

拟合优度数据段	掘进段占比/%	
	预处理前	预处理后
$R^2 < 0.6$	68.60	20.61
$0.6 \leq R^2 < 0.8$	23.40	14.78
$0.8 \leq R^2 < 1$	8.00	64.61

3.2 预处理方法工程验证及应用

为了验证本文提出的数据预处理方法的有效性,在 TPI 指标预测场景中,采用所提出的集成数据预处理方法来处理引绰济辽工程 TBM 掘进数据,并建立模型 1。同时,采用二值函数判别方法^[16]来提取完整掘进段数据,但没有进行滤波降噪处理,然后建立模型 2。模型 1 和模型 2 均采用擅长时序处理的长短期记忆网络(long short-term memory, LSTM)

进行构建,通过比较模型 1 和模型 2 对于 TPI 预测效果情况,验证所提数据预处理方法的有效性。

3.2.1 LSTM 算法简介 LSTM 是一种常用的循环神经网络(recurrent neural network, RNN)模型,特别适合处理具有长期时间依赖性的时序数据^[27]。LSTM 在 RNN 的基础上增加了存储单元以及输入门、遗忘门和输出门 3 个门控单元,用来控制不同时间的信息传输,以提高模型学习长期依赖关系的能力,其代表性的计算公式如公式(20)~(23)。

$$f_t = \sigma(W_{fh}h_{t-1} + U_{fx}x_t + b_f)$$
 (20)

$$i_t = \sigma(W_{ih}h_{t-1} + U_{ix}x_t + b_i)$$
 (21)

$$a_t = \tanh(W_{ah}h_{t-1} + U_{ax}x_t + b_a)$$
 (22)

$$C_t = C_{t-1} \odot f_t + i_t \odot a_t$$
 (23)

式中: f_t 为 t 时刻遗忘门输出; x_t 为 t 时刻的输入; W 和 U 为模型参数矩阵; b 为偏置向量; h_{t-1} 为 $t-1$

时刻隐藏层状态; i_t 和 a_t 分别为 t 时刻输入门第 1 部分和第 2 部分的输出; $\sigma(\cdot)$ 为 sigmoid 函数, 即 $\sigma(x) = 1/(1 + e^{-x})$; C_t 为 t 时刻的单元状态; $\odot$ 表示 Hadamard 乘积, 即两向量对应元素相乘, 且二者维数必须相同。

LSTM 算法特别适用于时间序列数据或具有时间依赖性的数据, 它通过在模型中引入门控单元的机制来有效捕捉长期依赖关系, 从而在序列模型建立任务中取得了很好的效果。

3.2.2 时序预测模型 TPI 预测的时序模型结构包括 1 个输入层、1 个 LSTM 层、1 个全连接层和 1 个输出层, 如图 8 所示。

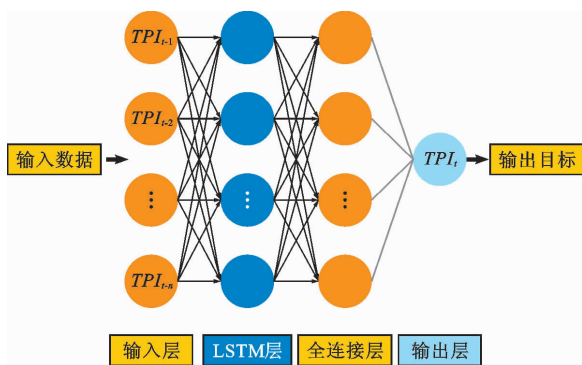


图 8 TPI 预测的时序模型结构示意图

首先, 将过去 5 组掘进段的 TPI 值输入到 LSTM 层, LSTM 层从原始数据中提取特征后传递至全连

接层, 最后结果通过全连接层传递至输出层。考虑到 LSTM 层具有强大的序列记忆能力, 本研究中设置了 1 层 LSTM 层, 其权重和偏差通过自适应矩估计 (adaptive moment estimation, Adam) 方法确定^[28], 其他的模型结构参数通过网格搜索确定。最终, LSTM 隐藏层神经元个数为 8, 全连接层神经元个数为 8, 优化器的学习率为 0.001、批处理大小为 32、迭代次数为 200。

以引绰济辽工程桩号 66 + 097 ~ 60 + 799 对应的掘进数据为对象, 分析了模型 1 和模型 2 对于 TPI 的预测效果。训练集和验证集部分的分割比例根据掘进距离确定为 4:1, 即桩号 66 + 097 ~ 61 + 858 作为 LSTM 模型的训练集, 桩号 61 + 858 ~ 60 + 799 作为 LSTM 模型的验证集。模型预测精度通过拟合优度 R^2 、平均相对误差 MRE (mean relative error) 和相关系数 R' 指标综合评价。

3.2.3 TPI 预测结果 图 9 显示了模型 1 和模型 2 对于 TPI 的预测值与初始值对比。分析图 9 可知, 在模型 2 中 (图 9(b)), LSTM 模型能够较好地学习到 0 ~ 400 数据编号范围内 TPI 数值的变化规律。然而, 从数据编号 600 开始, 出现了大量的 TPI 极低值和突变点, 这导致了模型学习效果的显著下降, 进而使得预测效果变差。采用所提出的数据预处理方法处理后, 极低值和突变点的数量大为减少, 模型的预测效果得到了显著改善。

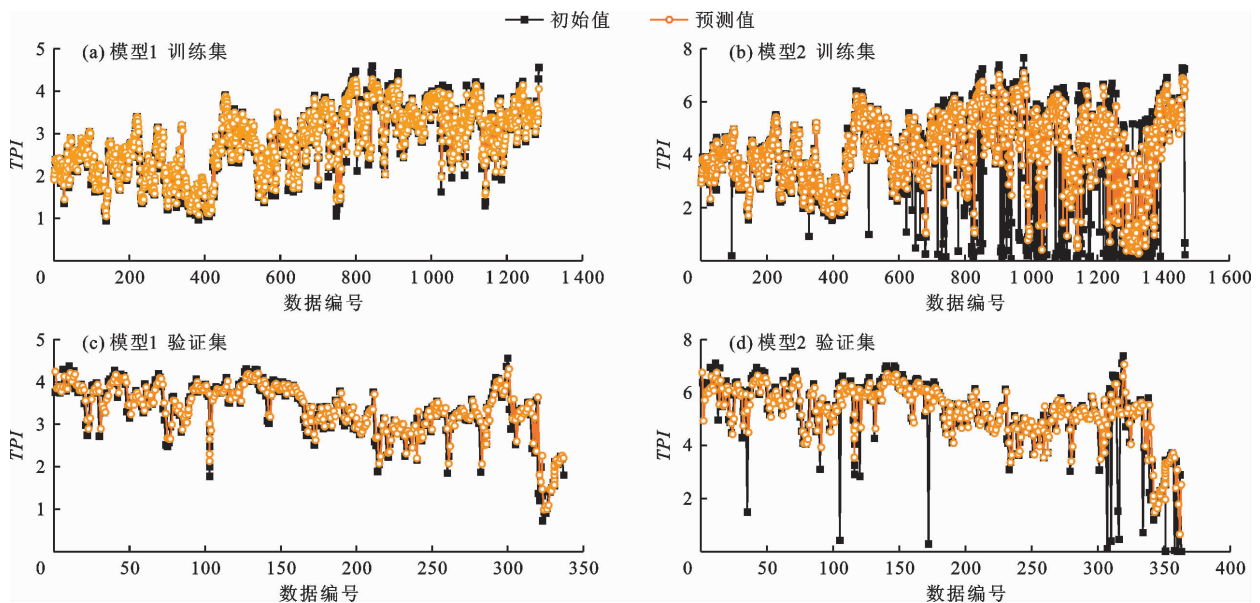


图 9 模型 1 和模型 2 对于 TPI 的预测值与初始值对比

表 4 为模型 1 和模型 2 对于 TPI 预测结果的评价指标值。由表 4 可知, 采用所提出的数据预处理方法在一定程度上会降低数据总量, 训练集的数据

量从 1 507 组减少至 1 317 组, 验证集的数据量从 406 组减少至 370 组。尽管训练集数据量有所减少, 但是在相同原始数据的基础上, 模型 1 的预测精

度显著提高,具体而言,训练集的 R^2 从 0.310 提升至 0.830,验证集的 R^2 从 0.503 提升至 0.721;训练集的 MRE 从 1.381 降低至 0.084,验证集的 MRE 从 3.107 降低至 0.096;训练集的 R' 从 0.659 提升至 0.894,验证集的 R' 从 0.809 提升至 0.900。

综上所述,本文提出的 TBM 掘进数据预处理方

法能够提取有效数据并去除噪声信息。尽管预处理后机器学习模型的总训练数据量有所减少,但却取得了更高的 TPI 指标预测精度。此外,上述结果也表明,数据预处理在数据挖掘过程中起着至关重要的作用,并且对机器学习模型预测性能的显著提升具有帮助。

表 4 模型 1 和模型 2 对于 TPI 预测结果的评价指标值

模型	掘进数据	桩号	数据量	R^2	MRE	R'
模型 1	训练集	66 + 097 ~ 61 + 858	1 317	0.830	0.084	0.894
	验证集	61 + 858 ~ 60 + 799	370	0.721	0.096	0.900
模型 2	训练集	66 + 097 ~ 61 + 858	1 507	0.310	1.381	0.659
	验证集	61 + 858 ~ 60 + 799	406	0.503	3.107	0.809

4 结 论

针对目前缺少标准但适用性强的 TBM 掘进数据清理和预处理体系的研究现状,本文基于引绰济辽工程长输水隧洞 TBM 掘进数据的普遍特征,提出了集成的 TBM 掘进数据预处理方法,包括完整掘进段提取、内部掘进段分割和掘进参数降噪 3 个过程。然后,采用时序预测提取能力更强 LSTM 算法建立了 TPI 预测模型,验证了所提数据预处理方法的有效性。得出的主要结论分述如下。

(1)提出了适用性更强的完整掘进段提取标准,确定了基于核密度估计法的掘进段内部分割方案。同时,建立了基于巴特沃斯滤波器的掘进参数滤波降噪方法,通过设定合理的滤波器阶数和截止频率,实现了掘进参数的滤波降噪。

(2)数据预处理前后掘进性能参数和 $T-p$ 拟合优度变化情况的对比结果表明,经过预处理后,数据的离散程度大幅减弱,掘进参数中的极端值(如最小值和最大值)得以平抑,数据质量大为改善。

(3)采用 TPI 预测验证了所提出的数据预处理方法的有效性,结果显示,数据经过预处理后,验证集的拟合优度 R^2 由 0.503 提升至 0.721;相关系数 R' 由 0.809 提升至 0.900;平均相对误差 MRE 由 3.107 下降至 0.096,时序预测模型的预测误差明显降低,预测精度显著提高。

参考文献:

[1] 洪开荣. 我国隧道及地下工程近两年的发展与展望[J]. 隧道建设,2017,37(2):123-134.
[2]《中国公路学报》编辑部. 中国交通隧道工程学术研究综

述·2022[J]. 中国公路学报,2022,35(4):1-40.
[3] GONG Qiuming, LIU Quansheng, ZHANG Qianbing. Tunnel boring machines (TBMs) in difficult grounds[J]. Tunneling and Underground Space Technology, 2016, 57: 1-3.
[4] HOU Shaokang, LIU Yaoru, YANG Qiang. Real-time prediction of rock mass classification based on TBM operation big data and stacking technique of ensemble learning[J]. Journal of Rock Mechanics and Geotechnical Engineering, 2022, 14(1):123-143.
[5] 鲍跃全,李惠. 人工智能时代的土木工程[J]. 土木工程学报,2019,52(5):1-11.
[6] 肖浩汉,陈祖煜,徐国鑫,等. 基于 GRU 算法的盾构掘进参数预测:以成都地铁 19 号线为例[J]. 长江科学院院报,2023,40(1):123-131.
[7] 郑永光,张娜,刘扬扬,等. 基于多模态控制策略的 TBM 掘进参数预测模型研究[J]. 隧道建设(中英文),2023, 43(4):583-591.
[8] 吴秉键,吴惠明,胡珉,等. 基于数据驱动的盾构自主姿态控制体系设计 and 应用[J]. 隧道建设(中英文),2023, 43(3):478-485.
[9] FREITAG S, CAO B T, NINIĆ J, et al. Recurrent neural networks and proper orthogonal decomposition with interval data for real-time predictions of mechanised tunnelling processes[J]. Computers and Structures, 2018, 207: 258-273.
[10] ALSAHLY A, HEGEMANN F, KÖNIG M, et al. Integrated BIM-to-FEM approach in mechanised tunnelling[J]. Geomechanics and Tunneling, 2020, 13(2): 212-220.
[11] NINIĆ J, BUI H G, MESCHKE G. BIM-to-IGA: a fully automatic design-through-analysis workflow for segmented tunnel linings [J]. Advanced Engineering Informatics, 2020, 46: 101137.
[12] CHEN Renpeng, ZHANG Pin, KANG Xin, et al. Prediction of maximum surface settlement caused by earth pres-

- sure balance (EPB) shield tunneling with ANN methods [J]. *Soils and Foundations*, 2019, 59(2): 284-295.
- [13] ZHANG Pin, CHEN Renpeng, DAI Tian, et al. An AloT-based system for real-time monitoring of tunnel construction[J]. *Tunnelling and Underground Space Technology*, 2021, 109: 103766.
- [14] XIAO Haohan, YANG Wenkun, HU Jing, et al. Significance and methodology: preprocessing the big data for machine learning on TBM performance [J]. *Underground Space*, 2022, 7(4): 680-701.
- [15] SANIO H P. Prediction of the performance of disc cutters in anisotropic rock[J]. *International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts*, 1985, 22(3): 153-161.
- [16] 罗 华. 基于线性回归和深度置信网络的 TBM 性能预测研究[D]. 杭州:浙江大学, 2018.
- [17] JING Liujie, LI Jianbin, YANG Chen, et al. A case study of TBM performance prediction using field tunneling tests in limestone strata[J]. *Tunnelling and Underground Space Technology*, 2019, 83, 364-372.
- [18] WANG Shuangjing, WANG Yujie, LI Xu, et al. Big data-based boring indexes and their application during TBM tunneling [J]. *Advances in Civil Engineering*, 2021, 2021: 2621931.
- [19] CHEN Zuyu, ZHANG Yunpei, LI Jianbin, et al. Diagnosing tunnel collapse sections based on TBM tunneling big data and deep learning: a case study on the Yinsong Project, China [J]. *Tunnelling and Underground Space Technology*, 2021, 108: 103700.
- [20] 陈祖煜, 范立涛, 张云旆, 等. 知识驱动和数据驱动在 TBM 智能施工机器学习中的应用[J]. *土木工程学报*, 2024, 1-12 [2024-05-26]. <https://doi.org/10.15951/j.tmgcxb.23090744>.
- [21] 姚 敏, 李 旭, 原继东, 等. 基于 TBM 破岩数据的岩体条件深度学习表征方法[J]. *地球科学*, 2023, 48(5): 1908-1922.
- [22] SILVA I, FLEMING C H, NOONAN M J, et al. Autocorrelation-informed home range estimation: a review and practical guide[J]. *Methods in Ecology and Evolution*, 2022(3): 13(3): 534-544.
- [23] 张博文, 刘 智, 桑国明. 基于核密度波动的异常检测算法[J]. *计算机工程与应用*, 2021, 57(12): 132-136.
- [24] GUSTAFSSON F. Determining the initial states in forward backward filtering[J]. *IEEE Transactions on Signal Processing*, 1996, 44(4): 988-992.
- [25] 肖锦龙. 巴特沃斯低通滤波器在图像增强中的应用研究[J]. *智能计算机与应用*, 2021, 11(5): 179-182.
- [26] REYES M, FRANCISCO-FERNÁNDEZ M, CAO R. Nonparametric kernel density estimation for general grouped data[J]. *Journal of Nonparametric Statistics*, 2016, 28(2): 235-249.
- [27] SHERSTINSKY A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network[J]. *Physica D: Nonlinear Phenomena*, 2020, 404: 132306.
- [28] KINGMA D P, JIMMY B. Adam: a method for stochastic optimization[J/OL]. *arXiv*, 2014, 1412. 6980. <https://api.semanticscholar.org/CorpusID:6628106>.

