

# 基于 BLSTM-AM 模型的 TBM 稳定段掘进参数预测

周小雄, 龚秋明, 殷丽君, 许弘毅, 班超

(北京工业大学 城市防灾与减灾教育部重点实验室, 北京 100024)

**摘要:** 隧道掘进机(TBM)掘进数据的上升段为实时岩体条件感知和掘进性能参数预测提供了丰富的信息。提出一种融合注意力机制的双向长短时记忆网络, 来实现 TBM 掘进稳定段的性能参数预测。在所提出的模型中, 4 个主要参数的时间序列数据作为主要输入来提取岩机作用关系, 稳定段的推进速度和刀盘转速作为辅助输入来考虑主司机的控制行为, 模型输出推力和扭矩预测值。不同于传统的预测模型, 所提出的模型不依赖于地质参数, 通过自动学习上升段的特征来建立控制参数与预测性能参数之间的映射关系。模型建立过程, 采用多项数据处理技术来修正异常值、过滤噪声及归一化等, 并提出了基于扭矩时序曲线来识别 TBM 上升段和稳定段的方法。依托于吉林引松供水隧洞工程, 验证了该模型的有效性和准确性。结果表明, 所建模型有较好的预测效果, 可辅助于类似地质条件的 TBM 智能化施工。

**关键词:** 地下工程; 硬岩掘进机; 掘进参数预测; 双向长短时记忆网络; 注意力机制; 智能化施工

**中图分类号:** TU 94

**文献标识码:** A

**文章编号:** 1000 - 6915(2020)增 2 - 3505 - 11

## Predicting boring parameters of TBM stable stage based on BLSTM networks combined with attention mechanism

ZHOU Xiaoxiong, GONG Qiuming, YIN Lijun, XU Hongyi, BAN Chao

(Key Laboratory of Urban Security and Disaster Engineering of Ministry of Education,

Beijing University of Technology, Beijing 100024, China)

**Abstract:** The ascending branch of tunnel boring machine(TBM) tunnelling data provides rich information for real-time rock mass condition perception and the prediction of boring performance parameter. This paper proposed a bidirectional long short-term memory network combined with attention mechanism to predict the performance parameters in the stable phase of TBM tunneling. In our model, time series data of four main parameters are taken as the main input to extract the rock-machine interaction relationship, and the advance speed and RPM given in the stable phase are taken as auxiliary inputs to consider the human control behavior, and the output of the model is the predicted values of thrust and torque. Different from the traditional prediction model, the proposed model does not require geological parameters, and establishes the mapping relationship between control parameters and performance parameters by automatically learning the characteristics of the ascending branch data. In the process of model establishment, some data preprocessing techniques are used to correct the outlier data, filter noise and normalize data, etc., and a method for segmenting the ascending and stable branch based on the torque-time curve is proposed. Relying on the Jilin Yinsong water supply tunnel project, the effectiveness and accuracy of the model are verified. The results show that the overall prediction effect of the proposed model is good, which can assist the

**收稿日期:** 2019 - 12 - 02; **修回日期:** 2020 - 05 - 26

**作者简介:** 周小雄(1991 - ), 男, 2013 年毕业于甘肃农业大学土木工程专业, 现为博士研究生, 主要从事 TBM 智能化施工方面的研究工作。E-mail: zhouxiaoxiong@emails.bjut.edu.cn。通讯作者: 龚秋明(1969 - ), 男, 现任教授、博士生导师。E-mail: gongqiuming@bjut.edu.cn

**DOI:** 10.13722/j.cnki.jrme.2019.1158

intelligent construction of TBM with similar geological conditions.

**Key words:** underground engineering; hard rock TBM; boring parameters prediction; bidirectional long short-term memory (BLSTM); attention mechanism; intelligent construction

## 1 引言

近十年来,随着我国掘进机生产技术的发展以及基础建设的加快,隧道掘进机(tunnel boring machine, TBM)已经成为铁路隧道、水利隧洞以及采矿工程的首选<sup>[1]</sup>。然而,我国是一个工程地质条件复杂的国家,随着隧道工程逐渐向大埋深和长距离方向发展,未来我国的隧道工程将面临更大的挑战。相比于钻爆法, TBM 工法虽然具有众多优势,但是其对地质条件的适应性差,以及对施工人员的经验要求高,制约着 TBM 在复杂地质条件下的应用。实时掘进性能预测是 TBM 辅助智能化施工中的一个重要内容,对于保障 TBM 安全、高效掘进具有重要意义<sup>[2]</sup>。

传统的 TBM 掘进性能预测都是从岩体条件的角度出发,通过研究岩体参数与掘进参数之间的关系,并采用多元回归分析或者人工智能方法建立掘进性能的预测模型。其中,基于多元回归方法的预测模型主要包括 CSM 模型<sup>[3]</sup>, Q-TBM<sup>[4]</sup>, NTNU 模型<sup>[5]</sup>和岩体特征可掘性模型<sup>[6]</sup>等。此外,还有一些学者基于人工智能方法预测掘进性能,包括人工神经网络方法<sup>[7]</sup>、支持向量回归方法<sup>[8]</sup>、粒子群算法<sup>[9]</sup>和模糊逻辑算法<sup>[10]</sup>等。TBM 的掘进性能影响因素众多,其机理非常复杂,大多数模型一般基于单条隧道或者特定岩性条件所提出,无法考虑完全影响 TBM 掘进性能的地质因素,其应用存在局限性。此外,由于 TBM 施工中掌子面前方的实时地质参数并不明确,基于岩体参数的模型难以指导 TBM 实时掘进性能预测。

随着物联网技术的快速发展,不同 TBM 工程的掘进数据能够被实时采集并传输到云平台系统进行快速分析<sup>[11]</sup>。TBM 掘进数据蕴含了丰富的岩机相互作用信息,基于人工智能技术对海量掘进数据挖掘,实现 TBM 的岩体条件感知和掘进性能预测具有可行性。W. Sun 等<sup>[12]</sup>采用随机森林方法的方法对盾构的动态荷载进行预测,其同时使用了掘进参数和工程地质参数。T. Yamamoto 等<sup>[13]</sup>基于锚杆钻机的数据和 TBM 掘进参数对掌子面的岩体信息进行预测,实现 TBM 的掘进控制。O. Zhang 等<sup>[14]</sup>采用聚类分析方法,通过对实时掘进参数进行聚

类,来预测实时地质条件。X. Gao 等<sup>[15]</sup>采用了循环神经网络模型分别对 TBM 推力、扭矩和贯入度等主要参数进行实时预测,结果表明循环神经网络对 TBM 的掘进参数的序列建模有很好的效果。虽然这些研究在掘进参数的挖掘上取得了一些成果,但是没有同时考虑掘进参数的时间变化特征以及各参数间相关性。掘进数据本质上属于多元时间序列,其特征提取和分析尤为困难,需要非常强健的算法。

深度学习是近年来兴起的一种表征学习方法,相比传统机器学习方法,由于它能逐层自动学习数据中的特征关系,具有不需要大量的特征工程、适应性强和易于迁移等优势。循环神经网络(recurrent neural network, RNN)<sup>[16]</sup>是神经网络中的一种特殊网络结构,主要应用于序列问题,如语音识别<sup>[17]</sup>、机器翻译<sup>[18]</sup>和时间序列预测<sup>[19]</sup>等。长短期记忆网络(long short-term memory, LSTM)<sup>[20]</sup>和双向循环神经网络(bidirectional RNN, BRNN)<sup>[21]</sup>,分别是对 RNN 的循环单元结构和传播方式进行了改进。在注意力机制被提出后,与循环神经网络的编解码框架相结合,进一步提高了循环神经网络对长序列的特征表达能力<sup>[18]</sup>。目前,融合注意力机制的双向长短期记忆网络(bidirectional LSTM, BLSTM)在时间序列的预测上取得了很好的效果<sup>[22-23]</sup>。

TBM 作为高度机械化的设备,其掘进过程中的众多传感器所记录数据包含了丰富的信息,反映了施工人员的操作行为和岩机作用关系等。TBM 的掘进数据可分解为时间维度和特征维度,时间维度反应了设备参数随人为操作的变化关系,特征维度反应了当前时刻 TBM 的各个参数之间的关系。本文提出了一种融合注意力机制的 BLSTM 模型,对 TBM 上升段的主要掘进参数的时序数据进行特征提取,抽象出当前岩体条件下的岩机作用关系,并考虑施工人员的掘进控制参数,对 TBM 的推力和扭矩进行预测,为 TBM 的辅助智能化施工提供支持。

## 2 问题描述

TBM 的一个掘进循环可以定义为从设备启动到停止的过程。每个循环按照加载过程可以分为上升

段和稳定段。上升段从刀盘接触掌子面开始, 持续到以稳定参数掘进。在该阶段, 推进速度和刀盘转速逐渐增大, 对应的推力和扭矩也随之增大, 可以理解为一个连续掘进试验<sup>[24]</sup>。施工人员通过观察上升段中主要参数的连续变化特征, 根据经验确定稳定段的刀盘转速和推进速度, 该控制参数的选择将很大程度决定 TBM 的掘进效率。

TBM 掘进上升段反映了不同加载状态的掘进参数之间的变化规律, 其本质是当前岩体条件下的岩机作用关系。通过上升段数据所提供的信息, 可以预测出给定的推进速度和刀盘转速条件下的推力和扭矩。该问题可以表示如下:

$$y = f(x_1, x_2, \dots, x_{T-1}, x_T, pr, rpm) \quad (1)$$

式中:  $x_i$  为  $i$  时刻的上升段输入特征向量, 包括该时刻的多个传感器值;  $rpm$  和  $pr$  分别为施工人员所给定的刀盘转速和推进速度;  $y$  为预测变量, 即总推力和刀盘扭矩;  $T$  为所使用的上升段数据的时长;  $f$  为模型所表示的函数关系。

该问题中, 上升段的时间序列数据作为主要输入, 为模型提供当前岩体条件下的掘进参数变化特征。稳定段的推进速度与刀盘转速以标量形式作为辅助输入, 代表主司机在稳定段的决策。模型最终输出的是决策控制参数所对应的预测性能参数。因此, 问题的关键在于构建的模型能够准确提取出上升段数据中的掘进参数变化特征。

### 3 注意力机制的 BLSTM 网络

#### 3.1 LSTM 神经网络

循环神经网络(RNN)是一种主要用于处理序列数据的神经网络结构, 沿序列传播方向依次对每个时刻输入的数据采用共享权重的 RNN 单元进行计算, 来提取出整个序列的高级特征。由于 RNN 在每次计算时都将上一个时刻隐藏层的状态加入到该时刻的输入中, 因此 RNN 可以利用序列的历史信息, 实现“记忆”功能。理论上, RNN 可以将整个序列的上下文信息映射到输出的目标向量中。然而, RNN 在采用沿时间反向传播算法(backpropagation through time, BPTT)<sup>[25]</sup>进行训练时, 存在梯度消失和膨胀的问题, 导致 RNN 无法捕获到长期的依赖关系<sup>[26]</sup>。长短时记忆(LSTM)和门控循环单元(gated recurrent unit, GRU)<sup>[27]</sup>都是对传统 RNN 的改进, 通过引入了门来控制信息的遗忘和新增, 从而可以更有效的学习到长期依赖关系, 是目前

RNN 的两种主流结构。如图 1 所示, 一个标准的 LSTM 单元包含 3 个控制门和 1 个 cell 状态向量, 其中,  $i$  为输入门,  $f$  为遗忘门,  $o$  为输出门。遗忘门控制了历史信息的保留和舍弃, 输入门控制了哪些信息被新增到 cell 中, 输出门决定计算输出时用到 cell 中的哪些信息。在每一个时间步  $t$ , 隐藏层输出  $h^t$  可以由前一个时刻的隐藏层输出  $h^{t-1}$  和当前时刻的输入  $x^t$  来更新, 即

$$i^t = \sigma(W^i x^t + V^i h^{t-1} + b^i) \quad (2)$$

$$f^t = \sigma(W^f x^t + V^f h^{t-1} + b^f) \quad (3)$$

$$o^t = \sigma(W^o x^t + V^o h^{t-1} + b^o) \quad (4)$$

$$c^t = f^t \odot c^{t-1} + i^t \odot \tanh(W^c x^t + V^c h^{t-1} + b^c) \quad (5)$$

$$h^t = o^t \odot \tanh(c^t) \quad (6)$$

式中:  $i^t$ ,  $f^t$  和  $o^t$  分别为  $t$  时刻的输入门、遗忘门和输出门;  $W^i$ ,  $W^f$ ,  $W^o$ ,  $V^i$ ,  $V^f$  和  $V^o$  分别为输入门、遗忘门和输出门对应  $x^t$  和  $h^{t-1}$  的权重向量;  $b^i$ ,  $b^f$ ,  $b^o$  分别为输入门、遗忘门和输出门的偏置向量;  $c^t$  为  $t$  时刻的 cell 状态;  $h^t$  为  $t$  时刻的隐藏层输出向量;  $\odot$  为向量点积运算符号;  $\sigma$  为 sigmoid 函数, 将门的值控制在 0 和 1 范围。

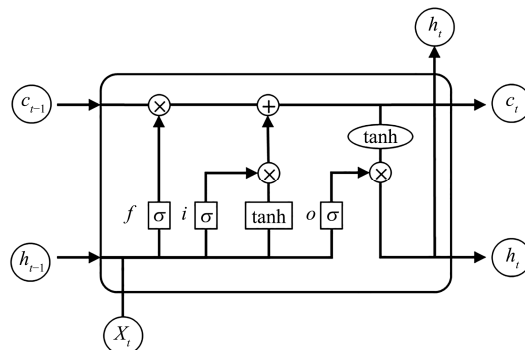


图 1 LSTM 单元结构组成

Fig.1 Architecture of LSTM units

LSTM 网络在计算每个时刻的隐藏层输出时, 共享所有的权重和偏置。当目标输出为单一变量时, 沿序列顺序依次计算完整个输入序列后, 将最后一个时刻的隐藏层向量与输出层建立线性连接。通过 BPTT 算法, LSTM 的网络的权重参数可以由反向误差进行更新, 以最小化损失函数。对于回归问题, 本文采用 MSE 作为损失函数:

$$loss = \frac{1}{N} \sum_{i=1}^N (y_i - \bar{y}_i)^2 \quad (7)$$

式中:  $N$  为一个训练批次的样本数量,  $y$  为实际值,

$\bar{y}$  为预测值。

图 2 为单层 LSTM 网络结构的展开形式, 在每个时刻, 上一时刻的隐藏层向量  $h^{t-1}$  和该时刻的数据  $x^t$  合并成新的输入向量, 经过 LSTM 单元计算处理后得到该时刻的隐藏层输出向量, 并继续把该值传递给下一个时刻。同时, cell 状态向量在沿序列传播过程中, 其信息也得到更新。LSTM 网络只沿序列向前传播, 因此在计算该时刻的输出时, 仅能利用历史信息, 而无法利用后续时刻的信息。

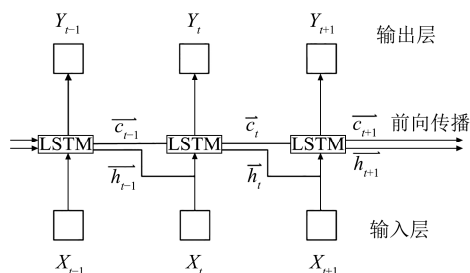


图 2 LSTM 网络结构

Fig.2 Illustration of LSTM network

### 3.2 BLSTM 网络

对于 TBM 掘进参数而言, 存在较强的时间依赖性, 数据的变化关系不仅体现在之前时刻上, 还反应在后续时刻上, 因此需要考虑信息传播的双向性。M. Schuster 和 K. K. Paliwal<sup>[21]</sup>于 1997 年提出了双向循环神经网络(BRNN)的结构, 在传统 RNN 网络结构上, 新增了一个反向传播层, 双向结构可以同时利用上下文信息。基于该结构, 将 RNN 单元替换成 LSTM 单元, 形成了双向长短时间记忆网络(bidirectional LSTM, BLSTM)。如图 3 所示, 该结构每层含有 2 个并行层, 分别控制向前和向后传播, 每个传播层为一个独立的 LSTM 结构, 因此 BLSTM 的隐藏层输出由 2 个部分组成:

$$h_f^t = o_f^t \odot \tanh(c_f^t) \quad (8)$$

$$h_b^t = o_b^t \odot \tanh(c_b^t) \quad (9)$$

式中:  $h_f^t$ ,  $h_b^t$  分别为  $t$  时刻的前向和后向传播的隐藏层输出, 对于整个网络的隐藏层输出为

$$h^t = h_f^t \oplus h_b^t \quad (10)$$

式中:  $\oplus$  为向量拼接运算,  $h^t$  为双向网络的最终隐藏层输出向量。BLSTM 与 LSTM 一样, 可以通过堆叠构建更深的网络模型。深层循环神经网络相比浅层循环神经网络, 可以抽象出更加高级的特征, 学习到更加复杂的关系。

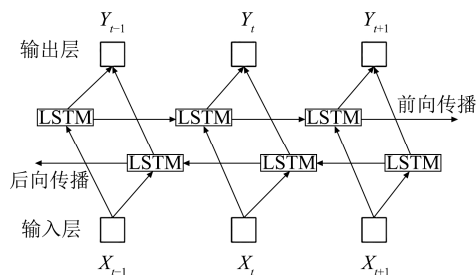


图 3 BLSTM 网络结构

Fig.3 Illustration of a BLSTM network

### 3.3 注意力机制

虽然循环神经网络很擅长处理时间序列任务, 但是其仍然存在局限性。循环神经网络可以理解为一个编码-解码过程, 输入序列的全部信息被压缩到特定长度的特征向量中, 再将特征向量解码到输出目标向量。当输入信息过长时, 固定长度的特征向量无法完全储存全部上下文信息, 早期历史信息被覆盖, 造成部分信息丢失, 从而模型的性能变差。

注意力机制是受人类视觉选择性的启发而提出, 可以让神经网络集中学习数据中的重要特征, 已经在很多学习任务中取得了良好的表现。机制核心是给每个元素分配不同的权重系数, 权重系数的大小表示注意力的分配比例。注意力机制的计算可以分为 2 个过程, 权重计算和加权求和, 其中权重的计算如下所示:

$$u_i = \tanh(wh^i + b) \quad (11)$$

$$a_i = \text{soft max}(u_i^T, u_w) \quad (12)$$

式中:  $h^i$  为 BLSTM 的隐藏层输出向量;  $a_i$  为分配的权重值;  $u_w$  通过随机初始化后, 在训练过程中更新;  $u_i$  为计算过程中  $h^i$  的激活向量。

采用注意力权重进行加权求和之后的得到的输出向量  $c$  可以由下式计算:

$$c = \sum_{j=1}^T \alpha_j h_j \quad (13)$$

### 3.4 模型构建

图 4 为本文所建立的模型网络结构。输入由 2 个部分组成, 上升段时间序列数据和给定的刀盘转速与推进速度。其中, 上升段的时间序列作为 BLSTM 网络的输入, 经过 2 层 BLSTM 后, 得到 BLSTM 网络的隐藏层输出向量。隐藏层向量经过注意力机制处理后, 得到提取出的上升段特征向量。该特征向量与给定的刀盘转速和推进速度拼接

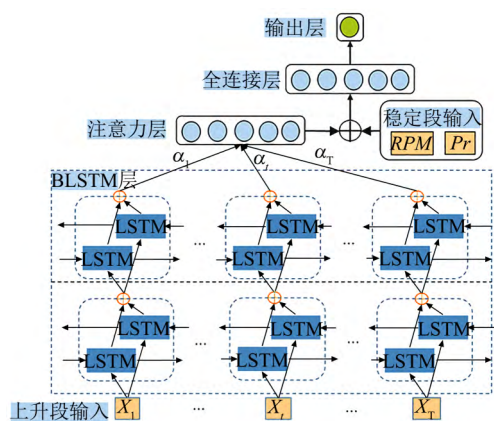


图4 注意力机制的 BLSTM 模型

Fig.4 BLSTM model with attentional mechanisms

合并, 构成全连接网络的输入。经过 1 个全连接层后, 输出最后预测值, 即总推力或者刀盘扭矩。

## 4 数据处理与模型训练

### 4.1 掘进循环提取

本文的数据源自吉林引松供水隧洞工程 TBM3 标段。TBM 掘进参数由中铁工程装备集团的云平台系统收集提供, 工程地质资料由吉林省水利勘察设计院提供。该隧道 TBM 施工段长 17.488 km, 围岩类别从 II 类到 V 类均有分布, 其中以 III 类和 IV 类为主。TBM 掘进参数的获取频率为 1 Hz, 每天有 86 400 条数据被记录进数据库中, 每条数据都包含 TBM 各个系统的 198 个指标。

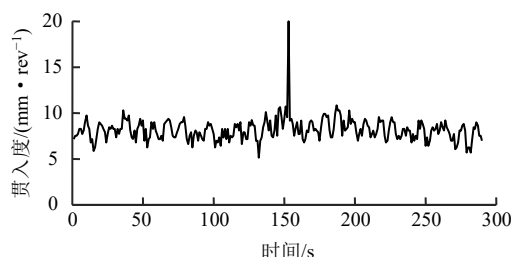
TBM 施工一般以油缸行程作为一个掘进循环, 每个掘进循环之间由换步或者停机状态分隔, 需要从记录文件中提取出每个步进的掘进数据。本文通过推力、扭矩和贯入度 3 个指标, 对每个时刻 TBM 是否处于掘进状态进行判定。只有同时满足推力和扭矩都大于 0, 且贯入度不超过 20 mm/rev, 则认为该时刻 TBM 处于掘进状态, 否则判定为非掘进状态。

判定完每个时刻点的工作状态之后, 一个掘进循环则对应一组连续掘进的序列, 每个循环之间的数据代表换步或临时停机状态。为了消除设备异常的数据, 只有掘进时长超过 600 s 时, 才视为有效步进。经过以上判定, 总共提取了 10 421 个有效掘进进步。

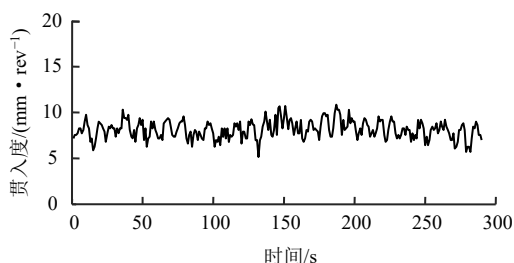
### 4.2 异常值判定和处理

TBM 掘进过程中, 受施工环境或者传感器性能的影响, 数据中会存在异常值, 从而影响模型的

预测性能, 图 5 为贯入度时序曲线中的异常值。对于掘进过程中的极大异常点, 采用  $3\sigma$  准则判定。通过计算整个掘进贯入度和掘进速度的均值  $\mu$  和标准差  $\sigma$ , 当掘进速度和贯入度大于  $\mu + 3\sigma$  时, 则按照异常点处理。对识别出来的异常点, 取 5 个临近数据点的平均值替代。



(a) 处理前



(b) 处理后

图5 异常值处理结果

Fig.5 Results of outlier processing

### 4.3 数据平滑处理

本文数据集的采样频率为 1 Hz, 采集的数据中受到 TBM 动态破岩和设备振动的影响, 存在一些噪声, 这些噪声会影响模型的训练速度以及预测性能。为了消除掘进参数中的白噪声, 本文采用滑动均值滤波来处理。取窗口长度为  $n$ , 从第一个数据点开始, 计算相邻的  $n$  个数据点的算术平均值并作为该点滤波之后的新值, 即

$$\bar{x}_t = \frac{1}{n} \sum_{i=t-n+1}^t x_i \quad (14)$$

均值滤波的窗口长度决定影响该点数值的数据范围, 当选择较大的窗口长度时, 可得到更加平滑曲线, 但是忽视了很多数据的变化细节, 如果选择的窗口长度过小, 则噪声消除的效果不够理想。本文采用的滤波窗口长度为 15 s, 图 6 为贯入度的均值滤波处理结果。

### 4.4 上升段与稳定段分割

上升段和稳定段的划分是构建高质量数据集的关键, 上升段识别的准确度直接影响训练数据的信息丰富程度和模型性能。其基本原则是保证所识别

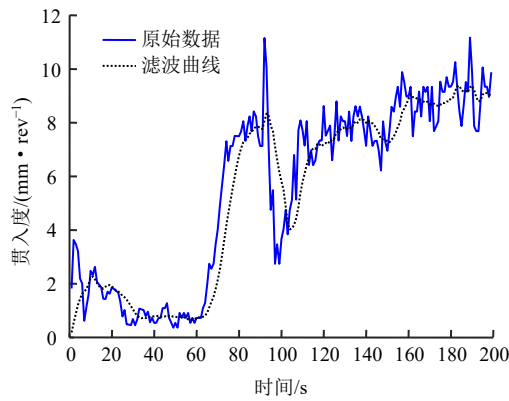


图 6 均值平滑滤波的结果  
Fig.6 Results of average value filtering

的上升段起点与实际 TBM 刀盘接触掌子面越接近越好。本文以刀盘扭矩作为识别上升段起点的依据。在刀盘启动和空推时, 扭矩值稳定在一个较低的水平, 当刀盘与掌子面产生了相互作用, 扭矩开始显著增大, 直至扭矩曲线逐渐平稳并进入稳定段。而贯入度和推力在这个过程中受支撑盾与围岩摩擦力的影响, 数值波动明显且在上升段开始时没有显著的变化特征。

为了进一步从扭矩曲线中自动识别上升段和稳定段, 本文采用了 Bottom-Up 算法进行识别。Bottom-Up 算法是一种基于分段线性拟合的时间序列分割无监督算法<sup>[28]</sup>。该方法通过输入指定的分段数量, 对扭矩曲线进行分段线性拟合, 其中总 RMSE 最小的分割结果则为最优的分段结果, 此处不再详细阐述其理论。图 7 给出了本文使用该算法的分割过程, 基于该算法识别上升段和稳定段的过程可以分为 3 步。首先, 设定了 10 个分段点, 使用 Bottom-Up 算法寻找最优分段线性拟合的分段点。然后, 根据线性分段后的结果, 从斜率最大的一段开始遍历, 连续相邻的多段斜率为正的部分则共同构成了上升段, 稳定段为上升段之后的连续多段斜率较小的部分。最后, 将上升段和稳定段的起始和结束时刻对应到原始数据上, 得到最终的分割结果。

4.5 特征选择与数据集生成

TBM 设备数据包含 198 个参数指标, 记录了各子系统的运行数据。在这些数据中, 刀盘驱动系统和推进系统的运行参数是岩机相互作用最主要的部分, 其他参数受人为主观因素影响较大或者与岩机作用的关系较弱。此外, 在刀盘驱动系统中的部分参数, 如电机电流、刀盘扭矩和刀盘功率等, 存在很强的线性相关性。因此本文只选择了其中与岩机关系最密切的总推力、刀盘扭矩、贯入度和刀盘转

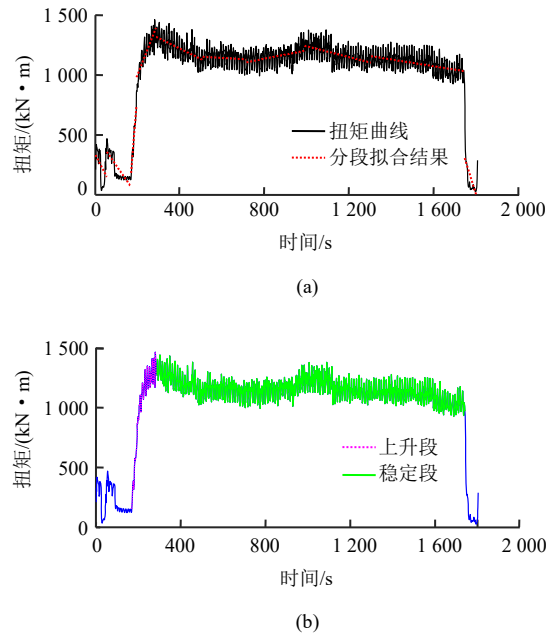


图 7 上升段和稳定段的分割过程  
Fig.7 The segmentation process of ascent phase and stable phase

速共 4 个指标。除了上升段的时间序列数据外, 还需要输入刀盘转速和推进速度, 这 2 个值为人为给定的控制参数。根据上升段的时间序列数据和 2 个给定的控制参数分别建立总推力和刀盘扭矩的预测模型, 其参数表如表 1 所示。

表 1 模型的输入和输出

Table 1 The inputs and outputs of model		
输入		输出
上升段序列数据	给定参数	预测参数
刀盘转速	刀盘转速	总推力
总推力		
刀盘扭矩	推进速度	刀盘扭矩
贯入度		

上升段数据作为模型中岩机关系感知的部分, 一方面, 其输入的数据时长影响了所包含的信息量丰富程度, 另一方面, 输入的数据时长也影响了模型在设备启动之后的决策速度。图 8 为所有掘进循环的上升段总时长的频数分布图, 从图中可以看出大部分的上升段总时长分布在 60~120 s 范围, 均值在 100 s 左右。为了探索最合理的输入上升段长度, 本文从上升段起点开始, 分别选择了 10, 20, 30, 40, 50 和 60 s 共 6 种时长的上升段数据, 构建不同的数据集。当该掘进循环的上升段总时长小于所选择的模型输入时长时, 则从稳定段中的数据进行补充。



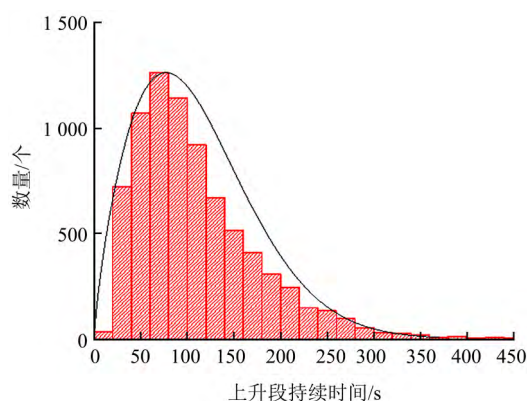


图 8 上升段持续时间分布

Fig.8 Distribution of ascending branch durations

模型的给定控制参数和输出参数从稳定段中获取, 以一个掘进循环为单位, 分别计算稳定段的推进速度、刀盘转速、总推力和刀盘扭矩的平均值。稳定段推进速度和刀盘转速的平均值则作为给定参数输入到模型, 稳定段刀盘扭矩和总推力的平均值则作为模型的预测输出。最终, 一个步进生成了 6 组不同上升段时长的样本, 每种时长的数据集均包含 10 421 个样本。

#### 4.6 模型训练和评估

本文采用平均绝对百分误差(mean absolute percentage error, *MAPE*)和平均绝对误差(mean absolute error, *MAE*)作为评价模型的性能指标, 其计算表达式如下:

$$MAPE = \frac{100}{n} \sum_{i=1}^n \frac{|\bar{y}_i - y_i|}{y_i} \% \quad (15)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\bar{y}_i - y_i| \quad (16)$$

式中:  $\bar{y}_i$  为模型预测值,  $y_i$  为真实值。

数据集按照 6:1:3 的比例划分成为训练集、验证集和测试集。对于不同上升段时长的数据集, BLSTM 网络的输入序列长度分别设置为对应的上升段时长。BLSTM 的隐藏层神经元个数和全连接层的个数是需要设定的超参数, 将在结果部分进行讨论。输出层只有 1 个神经元, 对推力和扭矩分别训练两个网络。训练的批次大小设置为 256 个。为了防止过拟合, 在每次参数的更新中, 采用 0.5 的比例来随机激活 BLSTM 的神经元。模型的训练采用 Adam 算法进行权重更新, 其初始学习率设定为 0.01。

对于输入特征数据, 均按下式进行归一化, 可以消除输入数据间的量纲差异:

$$X_{\text{norm}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (17)$$

式中:  $X_{\text{norm}}$  为归一化后的向量;  $X_{\max}$ ,  $X_{\min}$  分别为向量的最大值和最小值。

每当模型在所有训练集样本上完成一次遍历后, 随机打乱训练样本顺序, 并在验证集上进行一次测试。当测试集的平均损失连续 50 次迭代都没有继续降低, 则停止训练。

本文的实验是基于 python 语言环境下的 Tensorflow 1.11.0 深度学习框架, 硬件环境采用的是 Intel Xeon E5 2678V3 CPU, 64GB RAM 以及 NVIDIA TITAN RTX 显卡。

## 5 结果分析与讨论

### 5.1 模型结构分析

本文模型有 2 个部分参数需要设置, 即 BLSTM 部分的隐藏层神经元数量和层数, 以及全连接层的神经元数量。为了找到最优的网络结构, 在上升段时长为 40 s 的数据集上, 分别对不同的模型结构进行训练并在测试集上测试。由于推力和扭矩的模型在上升段的特征提取采用相同的结构, 因此只使用推力的预测结果来评价模型结构的性能。为了方便表述, 将模型的参数组合记作  $B(n_1, n_1)F(n_2)$ 。其中,  $B(n_1, n_1)$  代表 BLSTM 层的参数, 括号中数字的个数代表其循环神经网络的层数, 该示例代表了采用了 2 层的 BLSTM 结构, 每层隐藏层神经元个数都设置为  $n_1$  个,  $F(n_2)$  代表全连接层神经元个数为  $n_2$ 。全连接层的神经元数量与 BLSTM 的隐藏层神经元数量设置成为相同, 这在保证模型性能的同时, 减小模型的调参工作<sup>[29]</sup>。网络结构中的隐藏层神经元个数分别为 16, 32, 64, 128。为了使结果更加稳定, 分别对每种模型结构, 在随机初始化权重之后训练 10 次, 并取 10 次预测的 *MAE* 和 *MAPE* 的平均值作为最终的模型性能评价指标。

表 2 给出了不同结构模型的性能比较结果。其中, 模型 I~IV 都只采用了单层的 BLSTM 的结构。从结果来看, 当隐藏层神经元数量太少或者太多时, 模型的预测性能都较差。当神经元的数量设置为 64 时, 对于单层 BLSTM 的网络结构得到了最好的性能, 因此最终模型中各部分隐藏层神经元的数量均设置为 64 个。

为了探究更复杂的结构, 进一步对 BLSTM 的层数进行了分析。模型 III, V 和 VI 分别包含 1, 2

表 2 不同模型参数的网络结构平均性能比较  
Table 2 Average performance comparison of network structures with different model parameters

模型结构	MAE/kN	MAPE/%
I: B(16)L(16)	848.6	7.07
II: B(32)L(32)	826.2	6.88
III: B(64)L(64)	806.9	6.74
IV: B(128)L(128)	810.4	6.78
V: B(64, 64)L(64)	804.9	6.73
VI: B(64, 64, 64)L(64)	816.1	6.82

和 3 层的 BLSTM，所有隐藏层的数量均设置为 64 个。结果表明，当 BLSTM 的层数为 2 层时，相比单层的模型有更好的性能。但是当层数增加到 3 层后，模型的性能并没有继续提升，主要是因为复杂的模型容易增加过拟合风险。

不同模型参数的网络结构 10 次训练得到的预测误差箱型图如图 9 所示。图 9 中，模型 II 和 III 的多次重复训练的结果最不稳定，模型结构 VI 的预测结果最稳定，但是其平均误差不是最优的。2 种深度循环神经网络模型 V 和 VI 的预测性能都比较稳定。在平衡模型结构的稳定性和预测性能之后，由于模型结构 V 在所有模型中 MAE 和 MAPE 均表现最好，同时其多次训练的预测结果较稳定，则作为最终的模型结构。

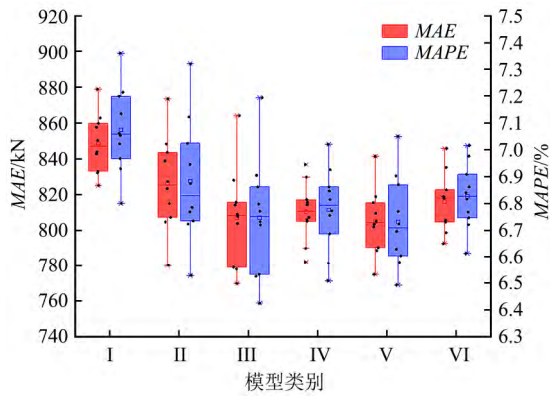


图 9 不同网络模型的预测 MAE 和 MAPE 的箱型图  
Fig.9 Box plot of prediction MAE and MAEP obtained by different network structure

5.2 模型成分的影响

本文提出的模型中包括循环神经网络和注意力机制层 2 个主要成分。为了评价模型的各个成分的影响，分别从循环单元类型，网络传播形式和注意力机制对模型性能的影响进行了分析。选择了 40 s 上升段的数据集，分别建立推力和扭矩的预测模型。每个模型分别训练 10 次，取 10 次测试集上预

测结果的 MAE 均值作为最终的评价指标。各个模型的预测性能如表 3 所示。表中的模型名称中，“RNN”，“LSTM”和“GRU”分别代表循环神经网络中 3 种单元类型，“B”代表采用了双向的网络结构，“AM”表示使用了注意力机制。每种模型对推力和扭矩预测性能的箱型图如图 10 和 11 所示。由表 3 可知，本文模型对推力和扭矩预测的 MAE 分别为 826.7 kN 和 196.4 kN · m，在所有模型中最优，说明模型中每个成分都对提升该网络的性能起到了作用。在各个模型中，模型 I 和 III 分别采用 RNN 和 GRU 作为循环单元。结果表明，在使用 RNN 作为循环单元时，其预测误差要显著高于 GRU 和 LSTM，说明 RNN 在对长序列的特征提取上有明显的局限，其主要原因是训练过程中的梯度消失导致其无法学习到长期依赖关系。LSTM 和 GRU 都通过门控制单元改善这个问题，但是在本网络结构中 LSTM 的预测性能要明显好于 GRU。

表 3 模型成分对性能影响

Table 3 The effect of model components for performance

模型	推力		扭矩	
	MAE/kN	MAPE/%	MAE/(kN · m)	MAPE/%
I: BRNN+AM	1 163.9	9.8	267.5	13.3
II: LSTM+AM	834.4	7.0	202.4	9.5
III: BGRU+AM	966.2	8.1	211.1	9.8
IV: BLSTM	829.4	6.9	205.7	9.5
V: BLSTM+AM	826.7	6.9	196.4	9.1

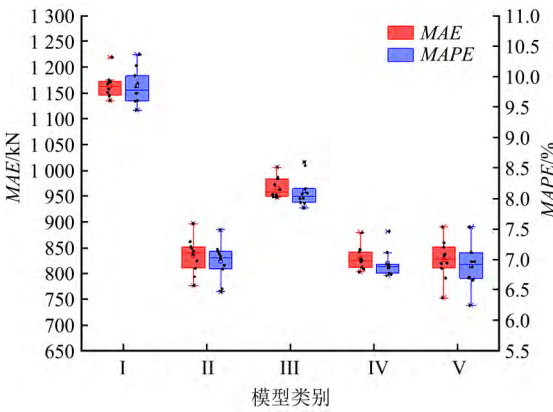


图 10 不同模型的推力预测性能箱型图

Fig.10 Boxplot of thrust performance prediction obtained by different model

循环神经网络的传播方式也影响了网络结构的预测性能。模型 II 只考虑了沿上升段序列的单向传播，其对推力和扭矩的预测误差都略高于双向传播的模型 V。结果说明双向传播形式的网络结构相比



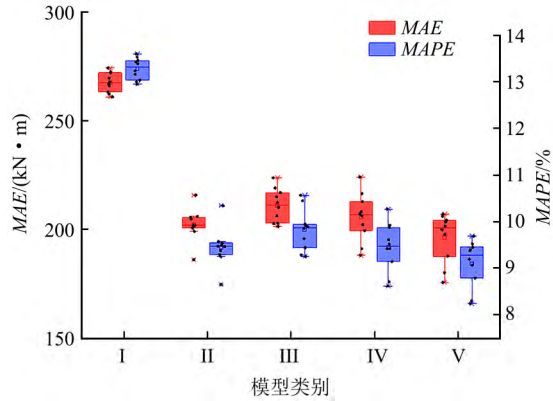


图 11 不同模型的扭矩预测性能箱型图

Fig.11 Boxplot of torque performance prediction obtained by different model

单向传播形式的网络结构一定程度上更好的利用上下文信息，从而提取出更丰富的 TBM 上升段数据特征。但是从箱型图上看，单向传播形式的多次训练结果比双向传播形式更稳定。

在表 3 中，模型 IV 在未使用注意力机制后，相比本文的模型对推力预测误差变化不显著，MAE 仅相差 2.7 kN 左右。但是对扭矩的预测性能影响较大，MAE 和 MAPE 分别增加了 9.3 kN·m 和 0.4%。结果说明注意力机制一定程度上提升了模型性能，主要得益于其对信息的权重调整。在提取 TBM 上升段数据的特征时，其中与岩机作用关系相关的重要信息能够被注意力机制更好的捕捉，从而得到更好的预测效果。

5.3 上升段数据时长的影响

对于 6 种不同时长的上升段数据集分别使用本文提出的 BLSTM-AM 模型结构训练 10 次，取 10 次测试的 MAE 和 MAPE 的平均值作为最终的性能评价指标。

表 4 列出了使用不同时长上升段数据的性能比较。图 12 和 13 分别为对推力和扭矩预测误差的箱型图。由表 4 可知，随着上升段数据时长的增加，模型对推力和扭矩的预测性能越来越好。特别是上升段时长从 10 s 增加到 20 s 时，性能提升的比较明显。但是，性能提升并不是线性递增的，从箱型图中可以看出，随着上升段数据时长的增加，性能的提升越来越缓慢。上升段数据时长较短时，预测性能较差，其原因可以归结为 2 个方面：一方面，其数据量少，所包含的信息量有限，无法准确表示出岩机作用关系；另一方面，所提取的上升段数据是上升段起点之后的掘进数据，而上升段起点的识别准确程度对短时长的上升段数据影响更大。若提取

表 4 不同上升段时长模型的平均性能比较

Table 4 Average performance comparison of models with different ascending branch durations

上升段时长/s	推力		扭矩	
	MAE/kN	MAPE/%	MAE/(kN·m)	MAPE/%
10	1 148.7	9.8	258.0	12.3
20	953.5	7.9	228.5	10.7
30	901.4	7.5	215.3	10.2
40	826.7	6.9	196.4	9.1
50	789.8	6.7	192.4	9.0
60	765.9	6.5	172.8	7.9

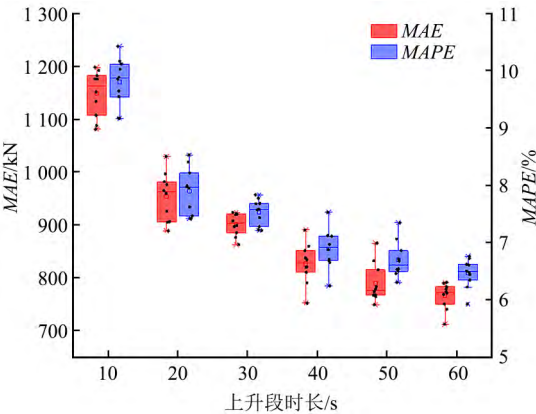


图 12 不同上升段时长模型的推力预测性能箱型图

Fig.12 Boxplot of thrust prediction performance of models with different ascending branch durations

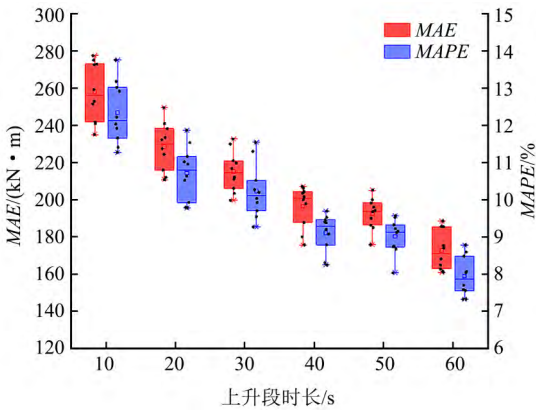


图 13 不同上升段时长的扭矩预测性能箱型图

Fig.13 Boxplot of torque prediction performance of models with different ascending branch durations

的上升段中包含了刀盘未接触掌子面的无效数据，则短时长的上升段中的有效信息将更少，而选择长时长的上升段作预测可以较好的避免这个问题。对于长时长的上升段输入，尽管其更多的数据可以提供更充足的信息，但是训练样本有限导致模型无法充分的学习到这些特征，相反容易造成模型的过拟合问题。

通过比较扭矩和推力的 *MAPE* 时,发现使用相同时长的上升段数据时,对扭矩的预测性能要比推力的预测性能差 1~2 个百分点。分析其原因,主要是由于 TBM 施工中主要采用的推力控制,在岩体条件变化不是很大时,推力控制在比较稳定的水平,而扭矩则很容易受到岩体条件的微弱变化所影响,导致同一个步进中的扭矩变化更大。当同时使用平均值评价整个稳定段的推力和扭矩时,平均值对于推力相比扭矩更接近于实际值,从而导致模型在使用上升段提取的岩机作用关系去预测稳定段的平均值时,在预测扭矩时误差更大。

在图 12 和 13 的箱型图中,当上升段时长选择 50 s 时,推力预测模型的性能已经达到一个比较好的结果。对于扭矩,上升段时长为 60 s 时,其预测性能相比 50 s 又有一次比较明显的提升,但是其多次训练的方差更大。但是,采用长时长的上升段数据进行预测时,模型需要较长的时间才能做出决策。而实际施工中,操作人员需要在掘进启动阶段预留尽可能多的时间用于参数调整,要求模型能够尽早给出决策建议。因此在平衡了模型的预测性能和决策效率之后,本文认为 50 s 时长的上升段数据更适合于实际 TBM 施工。

## 6 结 论

实时掘进性能预测是 TBM 智能辅助施工的重要内容,是保障 TBM 安全、高效掘进的关键。本文提出了一种融合注意力机制的 BLSTM 网络,以解决对 TBM 稳定段掘进性能参数的预测问题。该模型以归一化的总推力、刀盘扭矩、贯入度和刀盘转速 4 个参数的上升段时序数据为主要输入,以归一化的稳定段推进速度和刀盘转速作为控制参数输入,对稳定段的总推力和扭矩进行预测,并基于吉林引松供水隧洞工程,对本文的模型进行验证,主要的成果和结论如下:

(1) 本文模型能够通过自动学习 TBM 上升段掘进数据的变化特征,实现实时岩体条件下的岩机作用关系的提取,从而对给定控制参数下的掘进性能预测,可避免传统方法过于依赖地质参数的不足。

(2) 通过对模型的结构和成分进行分析,本文模型对推力和扭矩的预测都有较高准确率。在使用 40 s 的上升段时,对推力和扭矩的预测误差分别为 826.7 kN 和 196.4 kN·m,表明该模型对于指导

TBM 施工具有可行性。

(3) 对不同时长的上升段数据建模的结果表明,上升段时长的增加能提供丰富的岩机作用信息,从而提高模型对推力和扭矩的预测性能。但考虑到模型的决策效率与预测性能的平衡,在实际 TBM 施工中,50 s 的上升段时长是一个较好的选择。

**致谢:**感谢 973 计划(2015CB058100)“TBM 安全高效掘进全过程信息化智能控制与支撑软件基础研究”项目组提供 TBM 掘进数据,感谢吉林省水利水电勘测设计院提供地质资料。

## 参考文献(References):

- [1] 荆留杰,张娜,杨晨. TBM 及其施工技术在中国的发展与趋势[J]. 隧道建设, 2016, 36(3): 331 - 337.(JING Liujie, ZHANG Na, YANG Chen. Development of TBM and Its construction technologies in China[J]. Tunnel Construction, 2016, 36(3): 331 - 337.(in Chinese))
- [2] JING L J, LI J B, YANG C, et al. A case study of TBM performance prediction using field tunnelling tests in limestone strata[J]. Tunnelling and Underground Space Technology, 2019, 83(1): 364 - 372.
- [3] ROSTAMI J. Development of a force estimation model for rock fragmentation with disc cutters through theoretical modeling and physical measurement of crushed zone pressure[Ph. D. Thesis][D]. USA Colorado: Colorado School of Mines, 1997.
- [4] BRULAND, A. Hard rock tunnel boring[Ph. D. Thesis][D]. Norway Trondheim: Norwegian University of Science and Technology, 2000
- [5] BARTON N R. TBM tunnelling in jointed and faulted rock[M]. Balkema, Rotterdam: Crc Press, 2000: 3 - 38.
- [6] GONG Q M, ZHAO J. Development of a rock mass characteristics model for TBM penetration rate prediction[J]. International Journal of Rock Mechanics and Mining Sciences, 2009, 46(1): 8 - 18.
- [7] BENARDOS A G, KALIAMPAKOS D C. Modelling TBM performance with artificial neural networks[J]. Tunnelling and Underground Space Technology, 2004, 19(6): 597 - 605.
- [8] SALIMI A, ROSTAMI J, MOORMANN C, et al. Application of non-linear regression analysis and artificial intelligence algorithms for performance prediction of hard rock TBMs[J]. Tunnelling and Underground Space Technology, 2016, 58(9): 236 - 246.
- [9] MOHAMMADI S D, TORABI-KAVEH M, BAYATI M. Prediction of TBM penetration rate using intact and mass rock properties(case

- study: Zagros long tunnel, Iran[J]. Arabian Journal of Geosciences, 2015, 8(6): 3 893 – 3 904.
- [10] ACAROGLU O, OZDEMIR L, ASBURY B. A fuzzy logic model to predict specific energy requirement for TBM performance prediction[J]. Tunnelling and Underground Space Technology, 2008, 23(5): 600 – 608.
- [11] 张 娜, 李建斌, 荆留杰, 等. TBM 掘进参数智能控制系统的研究与应用[J]. 隧道建设, 2018, 38(10): 150 – 156.(ZHANG Na, LI Jianbin, JING Liujie, et al. Study and application of intelligent control system of TBM tunneling parameters[J]. Tunnel Construction, 2018, 38(10): 150 – 156.(in Chinese))
- [12] SUN W, SHI M, ZHANG C, et al. Dynamic load prediction of tunnel boring machine (TBM) based on heterogeneous in-situ data[J]. Automation in Construction, 2018, 92(8): 23 – 34.
- [13] YAMAMOTO T, SHIRASAGI S, YAMAMOTO S, et al. Evaluation of the geological condition ahead of the tunnel face by geostatistical techniques using TBM driving data[J]. Tunnelling and Underground Space Technology, 2003, 18(2): 213 – 221.
- [14] ZHANG Q, LIU Z, TAN J. Prediction of geological conditions for a tunnel boring machine using big operational data[J]. Automation in Construction, 2019, 100(4): 73 – 83.
- [15] GAO X, SHI M, SONG X, et al. Recurrent neural networks for realtime prediction of TBM operating parameters[J]. Automation in Construction, 2019, 98: 225 – 235.
- [16] PINEDA F J. Generalization of back-propagation to recurrent neural networks[J]. Physical Review Letters, 1987, 59(19): 2 229 – 2 232.
- [17] HINTON G, DENG L, YU D, et al. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups[J]. IEEE Signal Processing Magazine, 2012, 29(6): 82 – 97.
- [18] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[R]. [S. l.]: [s. n.], 2014.
- [19] MALHOTRA P, VIG L, SHROFF G, et al. Long short term memory networks for anomaly detection in time series[M]. [S. l.]: Presses universitaires de Louvain, 2015: 89 – 94.
- [20] GERS F A, SCHMIDHUBER J, CUMMINS F, et al. Learning to forget: continual prediction with LSTM[J]. Neural Computation, 2000, 12(10): 2 451 – 2 471.
- [21] SCHUSTER M, PALIWAL K K. Bidirectional recurrent neural networks[J]. IEEE Transactions on Signal Processing, 1997, 45(11): 2 673 – 2 681.
- [22] SONG S, LAN C, XING J, et al. Spatio-temporal attention-based LSTM networks for 3D action recognition and detection[J]. IEEE Transactions on Image Processing, 2018, 27(7): 3 459 – 3 471.
- [23] ZHANG R, MENG F, YONG Z, et al. Relation classification via recurrent neural network with attention and tensor layers[J]. Big Data Mining and Analytics, 2018, 1(3): 234 – 244.
- [24] 龚秋明. 掘进机隧道掘进概论[M]. 北京: 科学出版社, 2014: 283 – 300.(GONG Qiuming. TBM tunnelling: An overview[M]. Beijing: Science Press, 2014: 283 – 300.(in Chinese))
- [25] HINTON G E, OSINDERO S, TEH Y W. A fast learning algorithm for deep belief nets[J]. Neural computation, 2006, 18(7): 1 527 – 1 554.
- [26] PASCANU R, MIKOLOV T, BENGIO Y. On the difficulty of training recurrent neural networks[C]// International Conference on Machine Learning. [S. l.]: [s. n.], 2013: 1 310 – 1 318.
- [27] CHO K, VAN MERRIËNBOER B, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[R]. [S. l.]: [s. n.], 2014.
- [28] KEOGH E, CHU S, HART D, et al. Segmenting time series: A survey and novel approach[J]. Data Mining in Time Series Databases, 2003, 57(3): 1 – 21.
- [29] ZHANG C, LIM P, QIN A K, et al. Multiobjective deep belief networks ensemble for remaining useful life estimation in prognostics[J]. IEEE Transactions on Neural Networks, 2017, 28(10): 2 306 – 2 318.