

doi: 10.3969/j.issn.1002-0268.2025.05.002

基于 BERT 改进模型的交通运行态势预测方法

胡立伟*, 余先林, 赵雪亭, 杨志莹, 王兴中, 胡飞宇, 武加宝
(昆明理工大学 交通工程学院, 云南 昆明 650500)

摘要:【目标】为应对交通网络复杂的动态时空特性和提升路网的承载能力, 改变路网不确定性导致的交通拥堵问题, 提升特征提取过程中的精度及速度, 基于 BERT 模型提出用于交通态势预测的 TCPBERT 模型。【方法】首先将基于 Transformer 编码器框架的标准 BERT 模型引入交通预测中, 以利用其强大的上下文信息学习能力和时间序列建模能力; 随后在标准 BERT 模型基础上对模型参数及其嵌入方式进行改进, 提出用于交通运行态势预测的 TCPBERT 模型, 模型在堆叠的 Transformer 编码器中使用纯注意力机制对时间序列相关性进行捕获, 通过将归一化处理后的时间序列数据集对模型进行训练和性能检验, 与基线模型进行误差对比后验证 TCPBERT 模型在交通运行态势预测任务上可行性和有效性, 最后为进一步评估模型的稳定性和鲁棒性, 对 TCPBERT 模型进行敏感性分析, 验证模型参数设置的最优性。【结果】TCPBERT 模型在多个评价指标上均优于基线模型, 相较于 LSTM, STGCN, ST-ANet 模型, TCPBERT 模型的平均绝对百分比误差分别降低了 8.08%, 6.59%, 4.4%。在 6 h 的长时预测中, TCPBERT 模型的预测结果与原始交通数据的趋势最为接近, 显示出较强的时序建模能力和预测稳定性。【结论】通过引入 BERT 模型、Transformer 双向编码器架构和多头自注意力机制, 采用 TCPBERT 模型加强了对时间序列信息的捕获能力, 有效提升了交通运行态势预测的精度和效率。

关键词: 智能交通; 交通态势预测; Transformer 双向编码器; 时间序列相关性; 注意力机制

中图分类号: U491

文献标识码: A

文章编号: 1002-0268 (2025) 05-0018-08

Traffic situation prediction method based on improved BERT model

HU Liwei*, YU Xianlin, ZHAO Xueting, YANG Zhiying, WANG Xingzhong, HU Feiyu, WU Jiabao
(Faculty of Transportation Engineering, Kunming University of Science and Technology, Kunming, Yunnan 650500, China)

Abstract: [Objective] To address the complex dynamic spatio-temporal features of transportation networks, improve the traffic congestion caused by network uncertainty, and upgrade the accuracy and efficiency of feature extraction, this study proposed TCPBERT model for traffic situation prediction based on BERT model. [Method] First, the standard BERT model based on Transformer encoder framework was introduced in traffic prediction by using its powerful contextual learning capability and time series modeling ability. Subsequently, the model parameters and embedding modes were improved based on the standard BERT model. The TCPBERT model for traffic situation prediction was proposed. The model employed a pure attention mechanism within stacked Transformer encoders to capture time series correlations. The normalized time series dataset was used to train the model, followed by performance evaluation. Comparing prediction errors with baseline model, the feasibility and effectiveness of TCPBERT model in traffic situation prediction

收稿日期: 2022-12-08 修改日期: 2024-04-27

基金项目: 国家自然科学基金项目 (42277476); 云南省基础研究专项-重点项目 (202401AS070065)

作者简介 (* 通讯作者): 胡立伟 (1978-), 男, 山东潍坊人, 博士, 教授, 研究方向为交通运输规划与管理。(liweihukm@kust.edu.cn)

引用格式: 胡立伟, 余先林, 赵雪亭, 等. 基于 BERT 改进模型的交通运行态势预测方法 [J]. 公路交通科技, 2025, 42 (5): 18-25.

HU Liwei, YU Xianlin, ZHAO Xueting, et al. Traffic situation prediction method based on improved BERT model [J]. Journal of Highway and Transportation Research and Development, 2025, 42 (5): 18-25.

© The Author(s) 2025. This is an open access article under the CC-BY 4.0 License. (<https://creativecommons.org/licenses/by/4.0/>)

were verified. Finally, to further assess the stability and robustness of model, a sensitivity analysis was conducted to verify the optimality of model parameters setting. [**Result**] TCPBERT model outperforms baseline models in terms of multiple evaluation indicators. Compared with LSTM, STGCN and ST-ANet models, the MAPEs with TCPBERT model are 8.08%, 6.59%, 4.40% lower respectively. During the long-term 6-hour prediction, the prediction result with TCPBERT model is closest to the original traffic data trend, demonstrating strong temporal modeling capability and prediction stability. [**Conclusion**] By incorporating BERT model, Transformer bidirectional encoder framework and multi-head self-attention mechanism, the TCPBERT model enhances the capability of capturing temporal sequence information, significantly improving the accuracy and efficiency of traffic situation prediction.

Key words: intelligent transport; traffic situation prediction; BERT; time series correlation; attention mechanism

0 引言

在各国家和地区的智能交通系统发展建设过程中道路交通条件变得越发复杂, 路网不确定性更加突出, 但在此过程中累积了大量的交通数据, 这些数据对道路交通运行态势预测变得更加重要, 其对于缓解交通拥堵, 对道路进行管控至关重要。在早期交通预测中多使用时间序列模型参数模型, 常用的模型有自回归 (AR) 模型、移动平滑 (MA) 模型、自回归求和滑动平均 (ARIMA) 模型等, 这些时间序列模型可以从给定的一组数据中学习时间序列的特征进行预测, 其中应用最广泛的是 ARIMA 模型。Ahmed^[1]最早使用 ARIMA 模型用于预测极短的路段, 随后 Williams^[2]考虑到交通流量的周期性等特点提出了季节性 ARIMA 模型和具有外生变量的 ARIMA 模型^[3], 但是这类模型无法解释交通流的非线性和随机性, 准确率较低, 误差大。为了提升准确率, 开始使用早期机器学习非参数模型进行预测, 常见的模型有支持向量机模型、K 最近邻模型、贝叶斯组合模型等, 这类模型较为复杂, 算法收敛较慢, 且容易产生过拟合问题, 容易陷入局部最优解。

随着交通大数据的累积和计算机计算性能的提高, 深度学习被越来越多地应用于交通流预测当中, 不论模型精度还是计算速度都已经远超了传统统计学模型和早期机器学习模型, 比如卷积神经网络 (CNN)^[4]、循环神经网络 (RNN)、长短期记忆 (LSTM)、门控制循环单元 (GRU)、图卷积网络 (GCN)、时空图卷积网络 (STGCN)^[5]等。Zhang^[6]使用基于 CNN 的模型通过相邻道路的交通状况数据和完全卷积网络 (FCN) 来捕获时间和空间相关性; Zhao^[7]使用 LSTM 来识别时空相关性; Wang^[8]使用

双向 LSTM 来提高性能; Wu^[9]通过结合 CNN 和 RNN 来捕获时空信息进行预测。

除了上述的模型以外, 近年来机器翻译和自然语言处理任务广泛地使用基于注意力的模型来建模序列到序列的预测。学者们也将这类模型引入到交通预测当中, Guo^[10]提出基于注意力的时空卷积网络用于捕获空间维度上节点之间的动态相关性; 闻川和冯凤江^[11-12]基于注意力机制和图卷积网络获取路网节点和道路上下游的空间特征; Pan^[13]将注意力机制与学习单元相结合来对空间特征进行提取; Zhang^[14]在门控制注意网络 (GaAN) 中加入了注意力机制对节点特征进行聚合; Yao^[15]设计了周期性转移的注意力机制来处理长期性时序偏移, 对时空特征的提取效果显著。

近些年来无论是在图像识别领域还是在自然语言处理 (NLP) 领域, 大量的研究人员都致力于开发出精度更高、速度更快的模型, 其中自然语言处理领域的 Transformer^[16]摒弃了以往 CNN 模型和 RNN 模型的思路, 采用纯注意力机制且模型并行效果更好, 紧随其后的 BERT^[17]模型更是在一系列的 NLP 任务, 包括文本摘要和句子嵌入任务中都取得了优异成绩。目前 BERT 已经被用于其他科学应用中, BioBERT^[18]和 ClinicalBERT^[19]的表现证明了 BERT 在生物研究中的潜在适用性。本研究基于 BERT 模型提出用于交通态势预测的 TCPBERT 模型, 将交通数据与 BERT 集成进而对交通运行态势进行预测, 将 BERT 模型引入到交通态势预测当中能够进一步提高模型的预测精度和效率。

1 预测方法描述

1.1 道路运行状态量化指标界定

根据《城市道路交通管理评价指标体系》, 城市

主干道交通拥堵情况根据其平均车速分为4个等级,其中,平均车速 $V \geq 30$ km/h 为畅通, 20 km/h $\leq V < 30$ km/h 为轻度拥堵, 10 km/h $\leq V < 20$ km/h 为拥堵, $V < 10$ km/h 为严重拥堵。据此引入拥堵评价因子^[20] C_f , 车速越低表明越拥堵,即 C_f 值越大表示越拥堵,因此进行归一化处理见式(1):

$$C_f = \begin{cases} 0, & V \geq 30 \\ 1 - \frac{V - 10}{30 - 10}, & 10 < V < 30 \\ 1, & V \leq 10 \end{cases} \quad (1)$$

在进行归一化处理之后可以得到拥堵评价因子的各个评价区间为: $C_f = 0$, 非常畅通; $C_f \in (0, 0.5]$, 轻度拥堵; $C_f \in (0.5, 1)$, 拥堵; $C_f = 1$, 严重拥堵。

1.2 交通运行态势预测描述

交通运行态势预测是近年来交通领域的研究热点,一般预测方法分为两大类。第1类为模型驱动,是通过构建交通流中各类参数的数学模型对交通运行状态的复杂变化进行描述,并在此基础上实现交通状态预测;第2类为数据驱动,是通过智能计算挖掘历史数据隐含的信息,实现交通流状态的迭代估计^[21]。

本研究在进行交通态势预测时是通过对给定的道路空间结构和 T 个时间步长的历史交通状态进行学习,进而得到一个映射函数 $f(\cdot)$ 来预测出未来 T 个时间步长的交通状态,即:

$$[\hat{y}_1, \dots, \hat{y}_T] = f([s_1, \dots, s_T], G), \quad (2)$$

式中, s_1, \dots, s_T 为历史交通数据的时间序列; G 为路网拓扑图; $\hat{y}_1, \dots, \hat{y}_T$ 为预测出的未来 T 个时间步的交通运行状态,本研究所提出的 TCPBERT 模型中 T 取 10, 单个时间步长为 6 min, 即预测未来 10 个时间步长 60 min 的交通运行态势,为长时预测。

2 模型构建

2.1 TCPBERT 整体架构

TCPBERT 的整体架构是基于多头注意力的 Transformer 双向编码器,相较于基于 CNN 和 RNN 的交通预测深度学习模型,在嵌入和训练方法上均不相同。TCPBERT 保留了标准 BERT 模型的特点,能够用更大、更多样化的数据集对模型进行训练,将模型的预测性能进一步提高。TCPBERT 模型的整体架构图如图 1 所示。

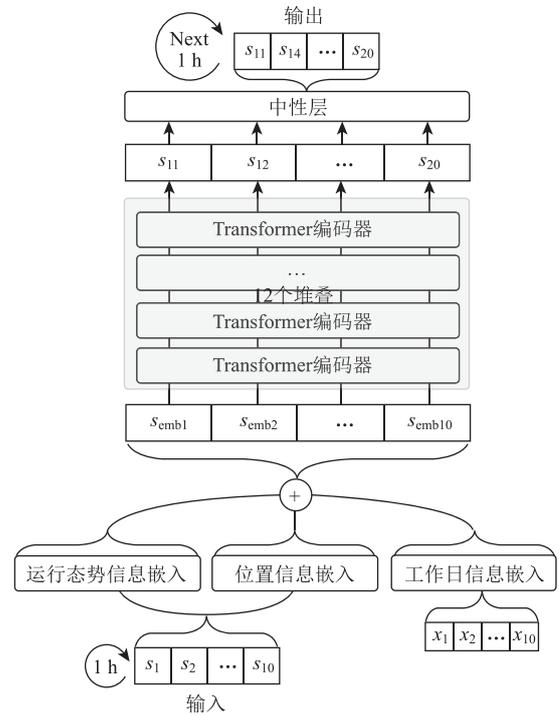


图 1 TCPBERT 模型的整体架构

Fig. 1 Overall framework of TCPBERT model

本研究使用数据集中包含的数据有旅行时间 $t = \{t_1, \dots, t_N\}$, 其中原始数据为传感器每隔 2 min 采集一次,将其汇总成时间间隔为 6 min 的数据样本用于记录道路路段上车辆的行驶时间,通过计算出平均行驶速度后根据上文方法对其进行归一化处理,称其为交通运行态势信息 $S = \{s_1, \dots, s_N\}$ 。在每个训练批次中,随机选择一个时间步长 t 作为数据输入的第一个时间步长,因此输入数据为 s_t, \dots, s_{t+T} , 由于 TCPBERT 与自编码器具有相同的结构,所以输出大小与输入大小相同,因此输入和输出数据分别是 s_t, \dots, s_{t+10} 和 $s_{t+11}, \dots, s_{t+20}$ 。

2.2 模型参数嵌入

标准 BERT 模型的嵌入方式有 3 类,包括词元嵌入、分段嵌入和位置嵌入,并且会将这些嵌入在后续进行求和。鉴于标准 BERT 模型是针对自然语言处理任务的,在本研究中对嵌入方法进行以下改进以适用于交通运行态势预测:第一,词元嵌入涉及到 Word2Vec 嵌入技术,但交通流速度数据是连续变量,所以使用一个线性函数将其进行矢量化,称为交通运行态势信息嵌入;第二,采用工作日作为与日期相关的信息进行嵌入,使用线性函数拟合一周中每天的数据,称为工作日信息嵌入;第三,同样使用位置嵌入来确定输入数据时间节点间的关系。

在标准 BERT 模型中,词元嵌入层的大小等于

隐藏层的大小, 这是 Lan 在 ALBERT 中提出的最佳解决方案, 而 BERT 的高性能主要归因于隐藏层嵌入, 因此本研究将隐藏层 H 的尺寸进行放大, 用于专注相邻节点间的识别。

模型的输入序列 $S = \{s_1, \dots, s_N\}$ 和工作日信息 $X = \{x_1, \dots, x_i\}$ 的嵌入方法表示如下:

$$E_{\text{traffic}}(S) = W^M(W^T S), \quad (3)$$

$$E_{\text{position}}(S) = W^M(W^P S), \quad (4)$$

$$E_{\text{weekday}}(X) = W^M(W^W X), \quad (5)$$

$$E_{\text{total}} = E_{\text{traffic}} + E_{\text{position}} + E_{\text{weekday}}, \quad (6)$$

式中, $E_{\text{traffic}}(S)$ 为交通运行态势信息嵌入函数; $W^T \in \mathbb{R}^{1 \times d_e}$ 为嵌入维度为 d_e 的运行态势嵌入权重; $E_{\text{position}}(S)$ 为位置嵌入函数; $W^P \in \mathbb{R}^{1 \times d_{\text{max}}}$ 为最大位置长度为 d_{max} 的位置嵌入权重; $E_{\text{weekday}}(X)$ 为工作日嵌入函数, $W^W \in \mathbb{R}^{7 \times d_e}$ 为工作日嵌入权重, 因为一周 7 天, 所以 W^W 的初始维度为 7; $W^M \in \mathbb{R}^{d_e \times d_{\text{model}}}$ 是将每个嵌入值拟合到模型 d_{model} 维度的权重; E_{total} 为权重和。

2.3 多头自注意力机制

目前用于时间序列预测的模型多是基于 RNN 的深度学习模型, 而基于 RNN 的深度学习模型有一个重大的缺点是对时间序列的依赖性过高并且难以并行处理, 这就会导致训练的时间成本非常高。为了解决这一问题, 进一步提高模型在大规模数据训练中的捕获信息特征的能力, TCPBERT 采用多头自注意力机制来提取数据在时间序列中的依赖关系。

多头自注意力机制的本质是将多个自注意力机制结构进行结合, 计算如下所示:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{Head}_1, \dots, \text{Head}_n)W^O, \quad (7)$$

$$\text{Head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V), \quad (8)$$

式中, $W_i^Q, W_i^K, W_i^V \in \mathbb{R}^{d_{\text{model}} \times d_k}$ 均为线性变换参数, 多头就是对 Q, K, V 的不同参数的线性变换, $W^O \in \mathbb{R}^{hd_i \times d_{\text{model}}}$ 为输出权重矩阵, 用于连接 n 个并行的头 $\text{Head}_1, \dots, \text{Head}_n$ 。

自注意力机制使用缩放点积来计算注意力输出, 计算流程如图 2 所示, 计算公式如下:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V. \quad (9)$$

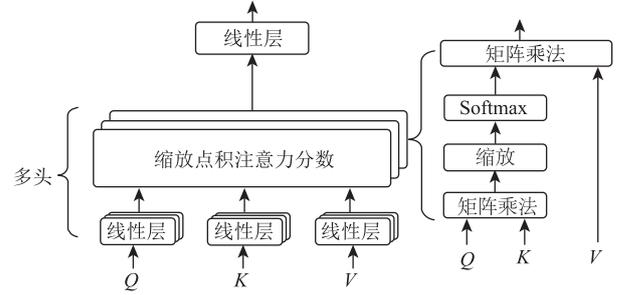


图 2 多头注意力机制计算流程

Fig. 2 Multi-head attention mechanism calculation process

3 模型应用

3.1 数据来源

本研究所用数据集来源于 2017 年阿里天池智慧交通预测挑战赛, 是由贵州省大数据发展管理局与贵州省交通运输厅提供的真实数据, 采用 2017 年 4 月至 2017 年 6 月 3 个月 132 条路段上检测器所采集的数据, 数据集信息如表 1 所示, 2 min 粒度, 覆盖全天。数据集中包括有路段 ID、路段长度、路段宽度、道路等级、上游路段 ID、下游路段 ID、日期、时间段、车辆在路段上的平均行驶时间。

表 1 数据集描述

Table 1 Data set description

数据集	阿里天池智慧交通预测挑战赛数据集
路段数量/段	132
路段等级	城市主干道
时间段	2017 年 4 月至 2017 年 6 月
训练集/个	1 900 800
测试集/个	982 080

3.2 数据处理和模型设置

根据路段长度和平均行驶时间首先计算出平均行驶速度 V , 提取出数据集中的交通速度信息和工作日信息, 并将交通速度信息通过式 (1) 归一化处理转化为交通运行态势信息。数据集中交通速度信息为每 2 min 一个时间片, 本研究将 3 个时间片取平均值, 将数据设置为 6 min 一个时间片。取数据集中 4 月和 5 月的数据作为训练集, 6 月的数据作为测试集。

本研究中数据输入为 10 个时间步长, 每个时间步长 6 min, 因此当模型接收到上一个小时 (60 min) 的输入时, 它会预测出下一个小时 (60 min) 的交通运行态势, 如将 12:00—13:00 的

数据作为输入, 预测 13:00—14:00 的交通运行态势。

本研究 TCPBERT 模型使用 Adam 优化器, 初始学习率设为 1×10^{-5} , 在第 3、第 6 和第 9 个训练批次中分别下降 0.5 倍。一方面为了防止梯度爆炸, 使用梯度裁剪, 最大范数设置为 5; 另一方面, 为了防止过拟合, 以 0.2 的速率使用 dropout。同时使用高斯误差线性单元 (GELU) 作为非线性激活函数。

3.3 模型误差评估方法

利用交通流预测中常用的均方根误差 (RMSE)、平均绝对误差 (MAE)、平均绝对比例误差 (MASE) 和平均绝对百分比误差 (MAPE) 对预测结果进行评价, 预测值和实际值分别用 \hat{y} 和 y 表示。

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}。 \quad (11)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|。 \quad (12)$$

$$MASE = \frac{MAE}{\frac{N}{N-1} \sum_{n=2}^N |y_n - y_{n-1}|}。 \quad (13)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\%。 \quad (14)$$

3.4 预测结果分析

首先, 将本研究提出的 TCPBERT 模型与统计学模型 ARIMA 模型、深度学习模型 LSTM 模型、SAE (堆叠式自动编码器) 模型以及近年来主流的 STGCN 模型和 ST-ANet 模型^[22] 在同样的数据集下进行比较, 验证 TCPBERT 的性能。在测试集中随机选取前 60 min 数据作为数据输入进行预测, 对后 60 min 的预测结果与实际值进行可视化对比。各个模型在测试集中的 60 min 预测结果如图 3 所示, 图中黑色实线部分为真实原始交通数据时间序列, 虚线为模型预测值。

从预测结果来看, TCPBERT 模型的预测效果最佳, 预测出的总体趋势与原数据最为接近。加入部分注意力机制的 ST-ANet 模型预测结果也较好, STGCN 模型次之, 预测效果最差的是 SAE 模型和 ARIMA 模型。模型误差评价指标为连续 60 min 的预测结果平均损失, 如表 2 所示, 各预测模型性能指标对比如图 4 所示。

从表 2 可以看出, TCPBERT 模型在 60 min 预测的性能表现优于其他模型, 相较于 ARIMA, SAE,

LSTM, STGCN 和 ST-ANet; RMSE 值分别降低了 10.72, 9.16, 7.34, 4.68, 2.97; MAE 值分别降低了 8.38, 9.14, 3.01, 2.40, 0.92; MASE 值分别降低了 4.11, 0.83, 1.91, 0.45, 0.35; MAPE 值分别降低了 18.64%, 15.29%, 8.08%, 6.59%, 4.4%。

为了验证 TCPBERT 模型在更长时预测上的预测性能表现优于其他模型, 在测试集中随机选取前 6 h 数据作为数据输入进行预测, 对后 6 h 的预测结果与实际值进行可视化对比, 各个模型在数据集中 6 h 预测结果, 图 5 为图中黑色实线部分为真实原始交通数据时间序列, 虚线为模型预测值。

从预测结果来看, 将预测时间增大至 6 h 后 TCPBERT 模型的预测精度虽有下降, 但在与其他模型对比中效果仍为最佳, 预测的总体表现与原数据最为接近; ST-ANet 模型、STGCN 模型和 LSTM 模型在加长预测时间之后预测精度出现明显下降; 由于 SAE 模型和 ARIMA 模型已无法满足更长时间的预测, 故未作出可视化。不难看出, 在长时预测方面, 本研究所提出的 TCPBERT 模型相较于其他基线模型性能提升较大, 从图 3 的 60 min 和图 5 的 6 h 预测可视化图及图 4 的 60 min 模型误差对比能看出 TCPBERT 模型的预测值和实际值显示出类似的趋势, 表明模型预测的分布特征与实际数据的分布特征相似, 且在与其他 5 种模型的对比中 TCPBERT 模型的误差最小, 性能最佳。

3.5 模型敏感性分析

为了保证模型的层数以及参数的设置为最优方案, 本研究还对模型参数进行了比较, 最终确定模型的设置, 修改包括 BERT 模型的层数、模型中 Transformer 编码器单向/双向、是否使用工作日信息嵌入方法、嵌入层大小、隐藏层大小, 对比结果如表 3 所示。

表 3 显示了模型灵敏性分析的结果, 第 1~4 行表明, 随着层数的增加性能会提高, 且与 6 层相比使用 12 层性能显著提高, 但当层数增加到 24 层时性能不会进一步提高, 表明层数不应超过 12; 第 3 行和第 5 行表明, 双向模型比单向模型性能更好; 第 3 行和第 6 行表明增加嵌入层大小, 减少隐藏层的大小 RMSE 值反而会增大; 第 3 行和第 7 行表明, 当不使用工作日信息嵌入时, 性能会下降。

4 结论

在交通态势预测领域许多学者试图通过融合不同类型的模型来提高预测性能, 然而这种预测方法

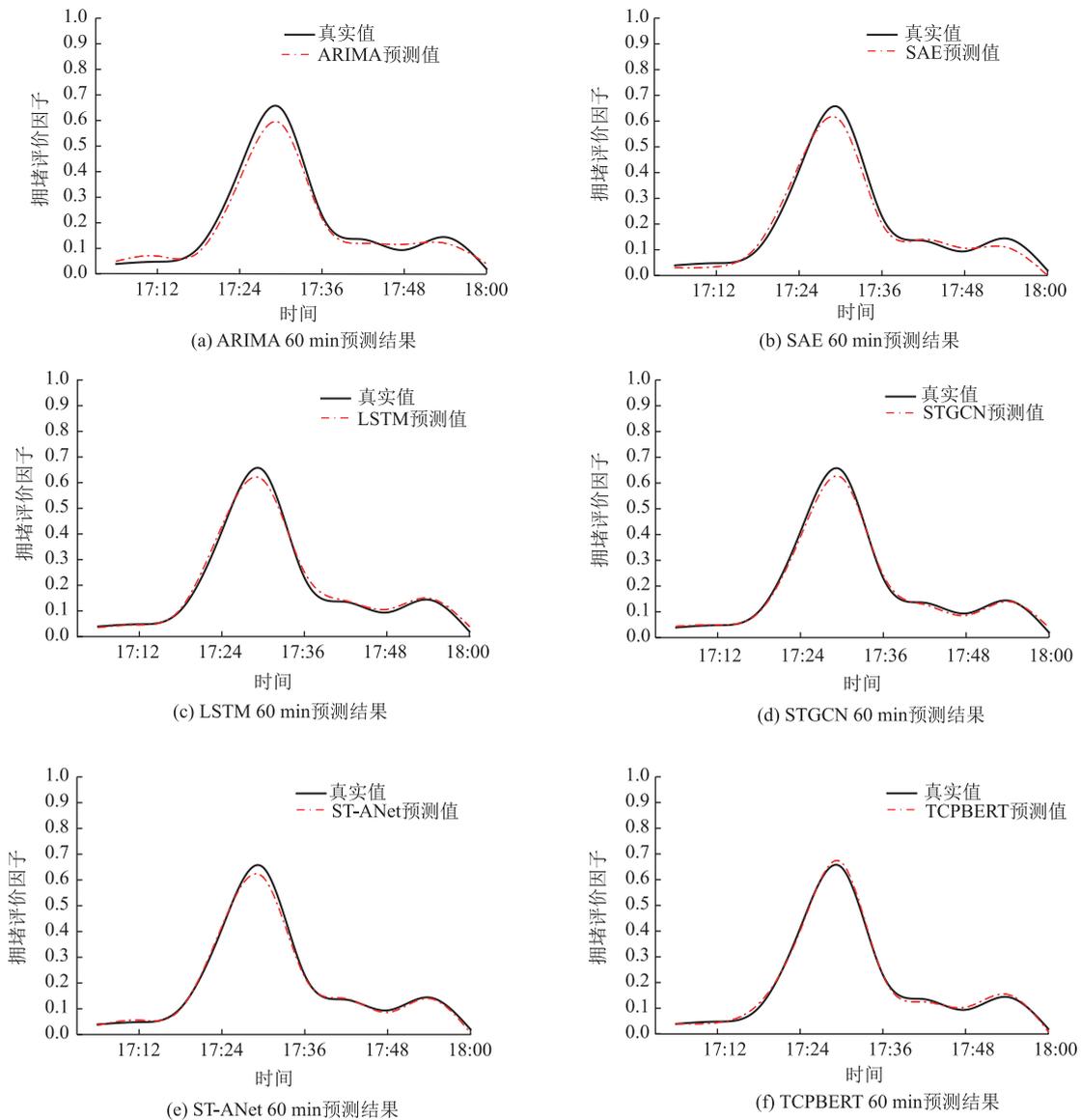


图 3 各模型 60 min 预测结果

Fig. 3 Prediction results with various models for 60 minutes

表 2 各模型 60 min 的 RMSE, MAE, MASE 和 MAPE

Table 2 RMSE, MAE, MASE and MAPE of each model for

60 minutes

模型	RMSE 值	MAE 值	MASE 值	MAPE 值/%
ARIMA	17.10	11.81	4.62	26.65
SAE	15.54	12.57	1.34	23.30
LSTM	13.72	6.44	2.42	16.09
STGCN	11.06	5.83	0.96	14.60
ST-ANet	9.35	4.35	0.86	12.45
TCPBERT	6.38	3.43	0.51	8.01

降低了模型的效率。本研究通过对 NLP 领域提出的基于 Transformer 双向编码器的 BERT 模型进行改进, 提出适用于交通领域的 TCPBERT 模型来对道路交通态势进行长时预测。在本研究中, 模型 (TCPBERT)

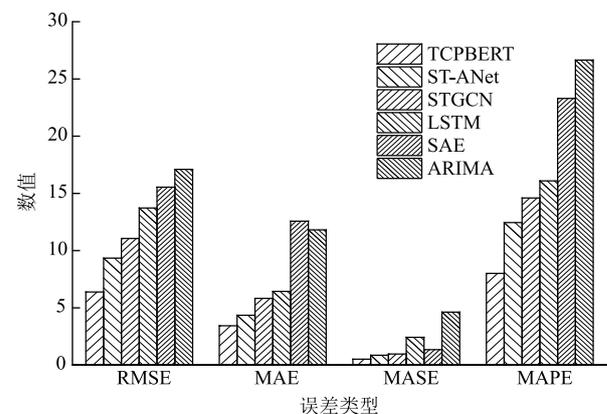


图 4 各预测模型性能指标对比

Fig. 4 Comparison of performance indicators of various models

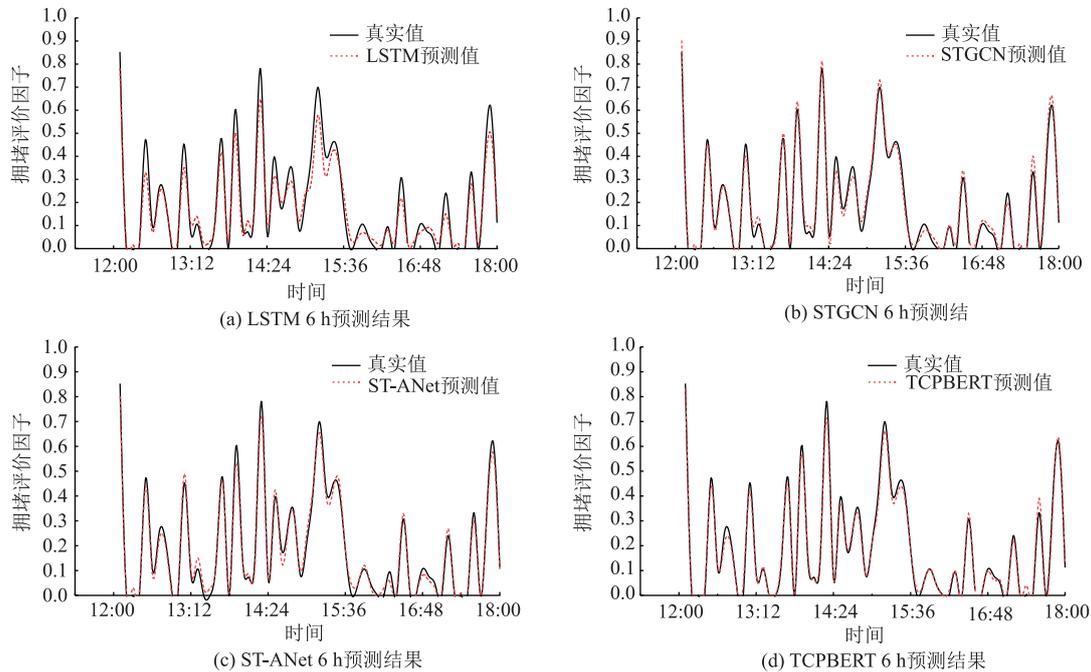


图 5 各模型 6 h 预测结果

Fig. 5 Prediction results with various models for 6 hours

表 3 模型不同参数设置预测结果对比

Table 3 Comparison of prediction results with different parameter settings for model

编号	模型层数	单向/双向	嵌入层大小	隐藏层大小	使用工作日信息嵌入	RMSE 值
1	3	双向	256	768	是	8.43
2	6	双向	256	768	是	8.24
3	12	双向	256	768	是	6.38
4	24	双向	256	768	是	6.46
5	12	单向	256	768	是	7.23
6	12	双向	512	512	是	7.37
7	12	双向	256	768	否	8.03

架构沿用了 BERT 模型，基本架构为堆叠 12 层的 Transformer 双向编码器，在参数嵌入方面将交通运行态势信息、工作日信息与位置信息通过编码嵌入到模型，数据输入后在多头自注意力机制的作用下自动捕获时间序列信息对参数进行学习，通过试验结果可以得到以下结论：

(1) 将 BERT 模型引入到交通态势预测中可以有效地提高预测精度，将 Transformer 双向编码器架构、多头自注意力机制与大规模交通数据相结合能够在交通态势长时预测上取得不错的效果。

(2) 相较于目前主流的 CNN 和 RNN 及其变体模型，多头注意力机制对提升模型精度起到了关键作用。

(3) 使用工作日信息嵌入能够加强模型对时间序

列信息的捕获能力，进一步提升模型预测精度。

本研究尚未考虑到道路交通系统中的动态突发因素，如交通事故和极端天气情况等对交通态势及模型预测精度的影响，这类问题将是下一步的研究内容。

参考文献：

References:

- [1] AHMED M S, COOK A R. Analysis of freeway traffic time-series data by using box-jenkins techniques [J]. Transportation Research Record, 1979, 722: 1-9.
- [2] WILLIAMS B M. Multivariate vehicular traffic flow prediction: Evaluation of arimax modeling [J]. Transportation Research Record, 2001, 1776 (1): 194-200.
- [3] WILLIAMS B M, HOEL L A. Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results [J]. Journal of Transportation Engineering, 2003, 129 (6): 664-672.
- [4] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks [C] // Advances in Neural Information Processing Systems. Nevada: MIT Press, 2012: 1097-1105.
- [5] 雷斌, 李佳璐, 张鹏, 等. 基于多图时空图卷积模型的城市交通流长时预测 [J]. 公路交通科技, 2024, 41 (4): 204-213.

LEI Bin, LI Jialu, ZHANG Peng, et al. Long term

- prediction on urban traffic flow based on multi-source spatio-temporal graph convolutional neural network model [J]. *Journal of Highway and Transportation Research and Development*, 2024, 41 (4): 204–213.
- [6] ZHANG J B, ZHENG Y, SUN J K, et al. Flow prediction in spatio-temporal networks based on multitask deep learning [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2020, 32 (3): 468–478.
- [7] ZHAO Z, CHEN W H, WU X M, et al. LSTM network: A deep learning approach for short-term traffic forecast [J]. *IET Intelligent Transport Systems*, 2017, 11 (2): 68–75.
- [8] WANG J T, WANG H F. One-step fabrication of coating-free mesh with underwater superoleophobicity for highly efficient oil/water separation [J]. *Surface and Coatings Technology*, 2018, 340: 1–7.
- [9] WU Y K, TAN H C, QIN L Q, et al. A hybrid deep learning based traffic flow prediction method and its understanding [J]. *Transportation Research Part C: Emerging Technologies*, 2018, 90: 166–180.
- [10] GUO S N, LIN Y F, FENG N, et al. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting [J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, 33: 922–929.
- [11] 闻川, 成卫, 肖海承. 基于多维流量特征的短时交通流量预测模型 [J]. *公路交通科技*, 2023, 40 (7): 191–199.
- WEN Chuan, CHENG Wei, XIAO Haicheng, et al. A model for predicting short-term traffic volume based on multi-dimensional traffic volume characteristics [J]. *Journal of Highway and Transportation Research and Development*, 2023, 40 (7): 191–199.
- [12] 冯凤江, 杨增刊. 基于图卷积和注意力机制的高速公路交通流预测 [J]. *公路交通科技*, 2023, 40 (9): 215–223.
- FENG Fengjiang, YANG Zengkan. Expressway traffic flow forecast based on graph convolution and attention mechanism [J]. *Journal of Highway and Transportation Research and Development*, 2023, 40 (9): 215–223.
- [13] PAN Z Y, LIANG Y X, WANG W F, et al. Urban traffic prediction from spatio-temporal data using deep meta learning [C] // *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. New York: Association for Computing Machinery, 2019: 1720–1730.
- [14] ZHANG J N, SHI X J, XIE J Y, et al. Gaan: Gated attention networks for learning on large and spatiotemporal graphs [C] // *34th Conference on Uncertainty in Artificial Intelligence*. Monterey: AUAI, 2018: 339–349.
- [15] YAO H X, TANG X F, WEI H, et al. Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction [J]. *Proceedings of the AAAI conference on artificial intelligence*, 2019, 33 (1): 5668–5675.
- [16] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C] // *Proceedings of the 31st international Conference on Neural Information Processing Systems*. Long Beach: NIPS, 2017: 6000–6010.
- [17] DEVLIN J, CHANG M W, LEE K, et al. Bert: pre-training of deep bidirectional transformers for language understanding [C] // *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Minneapolis: Association for Computational Linguistics, 2019: 4171–4186.
- [18] LEE J, YOON W, KIM S, et al. BioBERT: A pre-trained biomedical language representation model for biomedical text mining [J]. *Bioinformatics*, 2020, 36 (4): 1234–1240.
- [19] HUANG K X, ALTOSAAR J, RANGANATH R. Clinicalbert: Modeling clinical notes and predicting hospital readmission [J/OL]. *arXiv preprint arXiv: 1904.05342*, 2019. <http://10.48550/arXiv.1904.05342>.
- [20] 谭娟, 王胜春. 基于深度学习的交通拥堵预测模型研究 [J]. *计算机应用研究*, 2015, 32 (10): 2951–2954.
- TAN Juan, WANG Shengchun. Research on prediction model for traffic congestion based on deep learning [J]. *Computer Research of Application*, 2015, 32 (10): 2951–2954.
- [21] 王祥雪, 许伦辉. 基于深度学习的短时交通流预测研究 [J]. *交通运输系统工程与信息*, 2018, 18 (1): 81–88.
- WANG Xiangxue, XU Lunhui. Short-term traffic flow prediction based on deep learning [J]. *Journal of Transportation Systems Engineering and Information Technology*, 2018, 18 (1): 81–88.
- [22] 邹国建, 赖子良, 李晔. 基于时空注意力网络的动态高速路网交通速度预测 [J]. *计算机工程*, 2023, 49 (2): 303–313.
- ZOU Guojian, LAI Ziliang, LI Ye. Traffic speed prediction based on spatio-temporal attention network for dynamic expressway network [J]. *Computer Engineering*, 2023, 49 (2): 303–313.