

融合多种时空自注意力机制的 Transformer 交通流预测模型^①



曹 威^{1,2}, 王 兴^{1,2}, 邹复民³, 金 虬^{1,2}, 王小军^{1,2}

¹(福建师范大学 计算机与网络空间安全学院, 福州 350117)

²(福建师范大学 数字福建大数据安全技术研究所, 福州 350117)

³(福建理工大学 福建省汽车电子与电驱动技术重点实验室, 福州 350118)

通信作者: 王 兴, E-mail: wangxing@fjnu.edu.cn

摘要: 交通流预测是智能交通系统中实现城市交通优化的一种重要方法, 准确的交通流量预测对交通管理和诱导具有重要意义。然而, 因交通流本身存在高度时空依赖性而表现出复杂的非线性特征, 现有的方法主要考虑路网中节点的局部时空特征, 忽略了路网中所有节点的长期时空特征。为了充分挖掘交通流数据复杂的时空依赖, 提出一种融合多种时空自注意力机制的 Transformer 交通流预测模型 (MSTTF)。该模型在嵌入层通过位置编码嵌入时间和空间信息, 并在注意力机制层融合邻接空间自注意力机制, 相似空间自注意力机制, 时间自注意力机制, 时间-空间自注意力机制等多种自注意力机制挖掘数据中潜在的时空依赖关系, 最后在输出层进行预测。结果表明, MSTTF 模型与传统时空 Transformer 相比, MAE 平均降低了 10.36%。特别地, 相比于目前最先进的 PDFormer 模型, MAE 平均降低了 1.24%, 能取得更好的预测效果。

关键词: 交通流预测; 智能交通; 时空依赖性; Transformer; 自注意力机制

引用格式: 曹威, 王兴, 邹复民, 金彪, 王小军. 融合多种时空自注意力机制的 Transformer 交通流预测模型. 计算机系统应用, 2024, 33(4): 82–92.
<http://www.c-s-a.org.cn/1003-3254/9451.html>

Transformer Traffic Flow Prediction Model Integrating Multiple Spatiotemporal Self-attention Mechanisms

CAO Wei^{1,2}, WANG Xing^{1,2}, ZOU Fu-Min³, JIN Biao^{1,2}, WANG Xiao-Jun^{1,2}

¹(College of Computer and Cyber Security, Fujian Normal University, Fuzhou 350117, China)

²(Digital Fujian Institute of Big Data Security Technology, Fujian Normal University, Fuzhou 350117, China)

³(Fujian Key Laboratory of Automotive Electronics and Electric Drive, Fujian University of Technology, Fuzhou 350118, China)

Abstract: Traffic flow prediction is an important method for achieving urban traffic optimization in intelligent transportation systems. Accurate traffic flow prediction holds significant importance for traffic management and guidance. However, due to the high spatiotemporal dependence, the traffic flow exhibits complex nonlinear characteristics. Existing methods mainly consider the local spatiotemporal features of nodes in the road network, overlooking the long-term spatiotemporal characteristics of all nodes in the network. To fully explore the complex spatiotemporal dependencies in traffic flow data, this study proposes a Transformer-based traffic flow prediction model called multi-spatiotemporal self-attention Transformer (MSTTF). This model embeds temporal and spatial information through position encoding in the embedding layer and integrates various self-attention mechanisms, including adjacent spatial self-attention, similar spatial self-attention, temporal self-attention, and spatiotemporal self-attention, to uncover potential spatiotemporal dependencies in the data. The predictions are made in the output layer. The results demonstrate that the MSTTF model achieves an

① 基金项目: 福建省科技厅对外合作项目 (2020I0014)

收稿时间: 2023-10-08; 修改时间: 2023-11-09; 采用时间: 2023-11-15; csa 在线出版时间: 2024-01-30

CNKI 网络首发时间: 2024-02-01

average reduction of 10.36% in *MAE* compared to the traditional spatiotemporal Transformer model. Particularly, when compared to the state-of-the-art PDFormer model, the MSTTF model achieves an average *MAE* reduction of 1.24%, indicating superior predictive performance.

Key words: traffic flow prediction; intelligent transportation; spatial-temporal dependence; Transformer; self-attention mechanism

1 引言

交通流预测是智能交通系统的重要研究课题之一,利用大数据、人工智能等技术对未来的交通流状况进行预测,提供准确的交通信息,对优化交通管理,提高城市交通运输效率和改善出行体验等具有重要意义。当前,交通流预测的研究方法主要分为3类:统计学方法^[1,2],基于机器学习方法^[3-6]及基于深度学习的方法^[7,8]。

统计学方法通常会根据交通数据的概率分布情况进行建模。其中,时序模型是进行交通流预测的一种有效方法,如Zhao等^[9]为解决整合移动平均自回归模型(autoregressive integrated moving average, ARIMA)训练残差必须服从高斯分布的问题,在ARIMA的基础上提出了一种与GARCH模型(generalized autoregressive conditional Heteroskedasticity)相结合的交通流预测方法。基于隐马尔可夫模型的方法也被广泛使用,如Li等^[10]提出了一种全局PageRank算法的动态隐马尔可夫模型(discrete hidden Markov model, DHMM),该模型通过计算PageRank指标以衡量交叉口处交通异常的累积情况,从而直观地描述交通异常程度;Raskar等^[11]使用均值适应度导向的蜻蜓算法改进隐马尔可夫模型,对特定的输入特征进行预测,并通过优化状态数量提升了预测性能。

基于机器学习进行交通流预测的方法可以分为两类:一类是无监督学习方法,聚类算法是交通流预测中最常用的无监督算法之一。如Sun等^[12]考虑到不同的交通流模式会影响短期交通流预测结果,设计了一种K均值聚类和门控循环单元(gate recurrent unit, GRU)的组合模型,该模型对历史交通流数据进行聚类,充分考虑了交通流模式的多样性,提高了预测的准确率。另一类是监督学习的方法,使用历史数据来训练模型,并使用已知输入和输出之间的关系来预测未来的交通流量。如Xing等^[13]提出了一种S-ELM-Cluster(symmetric extreme learning machine cluster)的高效学习方法,能够将大规模数据转化为小规模数据上的不同问题,提

高了模型的计算效率。由于神经网络强大的非线性建模能力和自适应学习能力,在交通流预测中被广泛应用^[14]。如BiLSTM(bidirectional long short-term memory)^[15,16],RNN(recurrent neural network)^[17],GNN(graph neural network)^[18,19]等。GCN(graph convolutional network)^[20]是GNN的一种具体实现,它引入了卷积操作来进行节点特征的聚合。如Song等^[21]提出了一个时空同步建模模型STGCN(spatial temporal synchronous graph convolutional network),通过构造局部时空图来同时捕获时空依赖关系;Guo等^[22]考虑到目前基于GCN的方法忽略了交通系统的自然层次结构,因此提出了一种新的层次图卷积网络模型HGCN(hierarchical graph convolution network),该模型在交通流预测中通过对微观和宏观交通图进行操作,在短期预测方面取得了良好的效果,但在长期预测方面表现不佳。部分学者还引入了注意力机制^[23],使得模型能够有选择性地关注输入中的重要部分,从而提高处理效果和性能。

虽然GCN在处理交通网络这种复杂的图结构方面有着不错的优点,但也存在一些缺陷。首先,GCN的信息传播是通过邻接节点局部聚合实现,这导致信息在传播过程中逐渐衰减,因而无法适用于长期预测。Transformer是一种基于自注意力机制的神经网络模型,交通流数据具有长期时序依赖性,Transformer通过自注意力机制,能够在序列中建模长期依赖关系,从而更好地预测未来交通流状况。如文献[24]中提出了一种基于多头注意力机制的Transformer交通流预测模型,并对分析了GRU和LSTM模型,证明了它在处理时序数据长期相关性上的优势。Yan等^[25]设计了一种层次Transformer模型,使得不同层可以学习不同的特征,然后分层融合,显著提高了模型的预测精度。Wen等^[26]采用相对位置编码对多头注意力进行线性映射,以数据驱动的方式获得序列节点之间的相对位置关系。Xu等^[27]提出了一种新的时空Transformer框架STTN(spatial-temporal Transformer network, STTN),采用多

头注意力机制对不同空间依赖模式进行联合建模，实验表明其在交通流预测方面有不错的效果。此外，文献[28]中还考虑车祸会在一定时间后影响到周围路段，因此作者通过对局部时空图进行时空建模，在提取时空特征方面有一定的优势。尽管 Transformer 在交通流预测中取得了一定的效果，但在长序列数据和较长时间跨度的交通流预测中仍然会导致预测准确率大幅下降，并且 Transformer 未明确地建模交通流数据的时空依赖关系，这可能导致其无法充分挖掘交通流时空规律。

为了解决上述问题，本文提出了一种基于 Transformer 的多种时空自注意力机制融合的交通流预测模型 MSTTF (multiple spatio-temporal self-attention Transformer)。该模型可以同时学习时空信息和路网结构信息，充分捕获短期时间和空间特征的同时，还能针对交通流的长期规律进行学习。在数据嵌入层，模型对过往相同时间段的交通流进行融合，将上周同一时间段，昨天相同时间段，以及当前时刻临近时间段的交通流数据嵌入到输入数据中，在注意力机制层，除了传统的时间自注意力机制，空间自注意力机制外，还引入相似空间自注意力机制，并新设计了一种时间-空间自注意力机制，挖掘输入数据的长期时空相关性，最后对 4 种不同的自注意力机制进行融合，进而充分捕获交通流数据中的时空依赖关系。

2 本文方法

交通预测可以被建模为一个函数，它将观测到的路网交通数据映射到未来的交通状态，如流量、速度等，同时也是一个序列建模问题，通过滑动窗口技术构建历史数据与未来数据之间的关系^[29]。在本文中，将交通网络定义为图 $G = (V, E, A)$ ，其中 $V = \{v_1, v_2, \dots, v_N\}$ 是路网中传感器的 N 个顶点的集合 ($|V| = N$)， E 是表示图中顶点之间的连通性边的集合， $A \in \mathbb{R}^{N \times N}$ 是图 G 的邻接矩阵。本文将图 G 中观察到的交通状态记为 $X_G^{(t)} \in \mathbb{R}^{N \times 1}$ ，表示道路网络中 N 个传感器在时间 t 的交通状态，即观测到的交通流。交通流预测问题可以表示为给定历史 T 步的交通流量，来学习一个函数 f ，该函数可以预测下一个 T 步的交通流量，可以将这个过程描述为：

$$[X_G^{(t-T+1)}, \dots, X_G^{(t-1)}, X_G^{(t)}] \rightarrow [X_G^{(t+1)}, \dots, X_G^{(t+T-1)}, X_G^{(t+T)}] \quad (1)$$

2.1 模型介绍

交通流数据不仅具有单节点的时间相关性和多节

点间的短期空间相关性外，路网中所有节点间还存在长期的时空依赖关系。为了充分挖掘交通流的时空依赖性，本文提出了一种多种时空自注意力机制融合的交通流预测模型 MSTTF，总体网络架构如图 1 所示。主要由数据嵌入层、注意力机制层及输出层组成。数据嵌入层除了嵌入时间信息外，还嵌入了空间信息。注意力机制层引入了 4 种不同的自注意力机制，并通过特征融合层将 4 种自注意力机制进行融合，识别交通流数据中长期动态时空相关性，最后通过输出层进行预测。

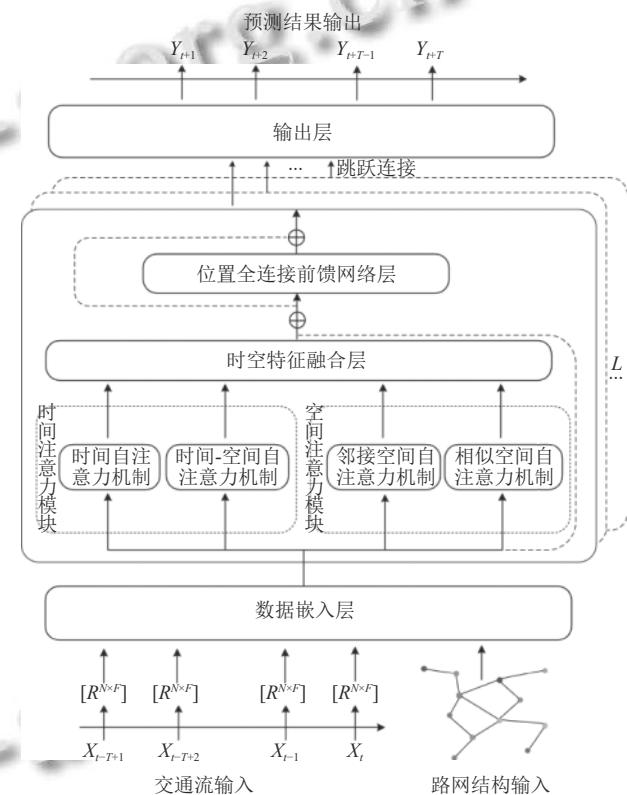


图 1 本文方法的整体框架图

2.2 数据嵌入层

如图 2 所示，考虑到交通流的时间相关性，当前时刻的交通流往往和当前时刻的前一段时间有着极强的关联性，和前一天相同时刻以及上一周同一天的相同时刻有很大的相似性，因此本文对输入数据进行了融合处理，将当前时间点的前 T 步数据 X_h ，前一天相同时段的交通流数据 X_d ，以及上周同一时间段的交通流数据 X_w 进行融合，用如下公式表示：

$$X_h = (X_G^{t-T+1}, \dots, X_G^{t-1}, X_G^t) \quad (2)$$

$$X_d = (X_G^{t-\tau+1}, X_G^{t-\tau+2}, \dots, X_G^{t-\tau+T}) \quad (3)$$

$$X_w = (X_G^{t-7 \times \tau+1}, X_G^{t-7 \times \tau+2}, \dots, X_G^{t-7 \times \tau+T}) \quad (4)$$

$$X_{\text{input}} = \text{Linear}(\text{Concat}(X_h, X_d, X_w)) \quad (5)$$

其中, τ 表示一天中检测器记录的总次数. 通过一个全连接层将原始输入数据 X 转换为 $X_{\text{input}} \in \mathbb{R}^{T \times N \times d}$, d 是嵌入维数.

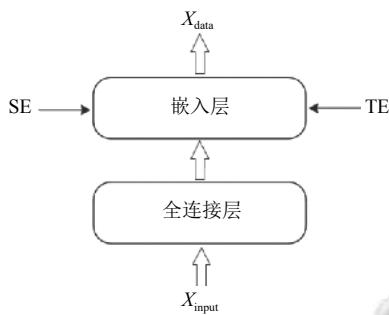


图 2 数据嵌入层结构

为了学习交通流数据中的时间模式和规律, 在嵌入层对数据进行时间嵌入编码, 将时间信息融入到模型中. 由于每周的天数在时间维度上是唯一的, 并且它们之间没有大小顺序的关系, 这里使用可以将离散类别数据转换为向量表示的独热编码来将一周的天数编码到 \mathbb{R}^7 的向量空间. 若对一天中的时间步也采用独热编码, 则会导致时间维度过于稀疏, 同时计算开销也会变大, 因此本文对时间步进行归一化处理, 通过拼接运算将两种编码变换到 \mathbb{R}^8 的向量空间, 最后用全连接层将时间编码转换成模型输入数据维度, 用 $X_{\text{TE}} \in \mathbb{R}^{T \times N \times d}$ 表示.

现实中的路网可以被建模为不规则的图结构, 常用对应的拉普拉斯矩阵将其转换为可分析的结构, 来捕捉其特征和连接关系. 标准化的拉普拉斯矩阵定义为:

$$L = I_N - D^{-\frac{1}{2}} A D^{-\frac{1}{2}} = U \Lambda U^T \quad (6)$$

其中, $A \in \mathbb{R}^{N \times N}$ 是邻接矩阵, I_N 是单位矩阵, 度矩阵 $D \in \mathbb{R}^{N \times N}$ 是一个对角矩阵, 由节点度组成. Λ 为 L 矩阵的特征值组成是对角矩阵, U 是傅里叶基. 这里使用最小 k 个非平凡特征向量上的线性投影来生成空间图拉普拉斯嵌入, 并通过全连接层将其扩展为输入数据维度, 用 $X_{\text{SE}} \in \mathbb{R}^{T \times N \times d}$ 表示. 之后, 参考传统 Transformer 模型中的位置编码方法对空间信息进行编码, 如下所示:

$$X_{\text{SE}} = \begin{cases} \sin\left(\frac{pos}{10000^{\frac{2i}{d_{\text{model}}}}}\right), & i = 2k \\ \cos\left(\frac{pos}{10000^{\frac{2i}{d_{\text{model}}}}}\right), & i = 2k + 1 \end{cases} \quad (7)$$

其中, pos 表示第 pos 个位置的编码, d_{model} 表示编码长度. 最后, 将上述 3 个嵌入向量进行叠加, 得到数据嵌入层的输出:

$$X_{\text{data}} = \text{Linear}(X_{\text{input}} + X_{\text{SE}} + X_{\text{TE}}) \quad (8)$$

2.3 注意力机制层

在本文提出的 MSTTF 模型中, 使用了 4 种不同的自注意力机制, 分别为邻接空间自注意力机制 (adjacent spatial self-attention, *AdjSSA*), 相似空间自注意力机制 (similar spatial self-attention, *SimSSA*), 时间自注意力机制 (temporal self-attention, *TSA*) 以及时间-空间自注意力机制 (temporal-spatial self-attention, *TSSA*). 其关系图如图 3 所示, 后文会进行详细说明.

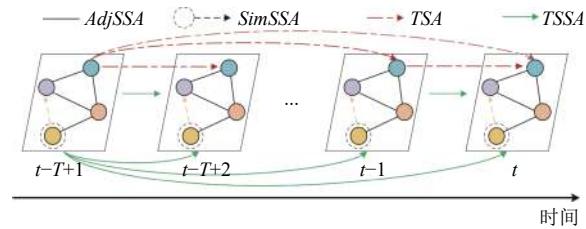


图 3 注意力关系图

2.3.1 空间注意力模块

(1) 邻接空间自注意力机制. 为了捕获路网节点间的空间依赖性, 本文设计了一种邻接空间自注意力机制. 形式上, 在时刻 t , 首先获得自关注操作的查询 $Q_t^{(S,F)}$, 键 $K_t^{(S,F)}$ 和值 $V_t^{(S,F)}$ 矩阵, (S,F) 表示空间和交通量维度. 这里选择使用缩放点积乘法来计算两个节点之间的相关性, 并获得全局时间注意力矩阵, 将这个过程公式化, 可以得到下面的方程:

$$Q_t^{(S,F)} = W_Q^{(S,F)} X^T, K_t^{(S,F)} = W_K^{(S,F)} X^T, V_t^{(S,F)} = W_v^{(S,F)} X^T \quad (9)$$

$$\text{Score}_t^{(S,F)} = \frac{Q_t^{(S,F)} (K_t^{(S,F)})^T}{\sqrt{d'}} \quad (10)$$

$$\text{SSA}(Q_t^{(S,F)}, K_t^{(S,F)}, V_t^{(S,F)}) = \frac{\exp(\text{Score}_t^{(S,F)})}{\sum_{k=1}^N \exp(\text{Score}_t^{(S,F)})} \cdot V_t^{(S,F)} \quad (11)$$

其中, $W_Q^{(S,F)}$ 、 $W_K^{(S,F)}$ 、 $W_v^{(S,F)} \in \mathbb{R}^{d' \times d}$ 是可学习参数, d' 是查询, 键和值矩阵的维数. 对于式 (11) 中的空间自注意力, 每个节点与所有节点交互, 相当于将空间图视为

完全连通的图。考虑到路网中一定范围内的道路交通状态存在相互影响的情况，使用了一种邻接空间矩阵 M_{adj} ，用来描述这种在空间上相近的道路节点，然后在计算邻接空间自注意力时，引入邻接空间矩阵 $M_{adj_{i,j}}$ ，用来捕获这种邻接空间中的相互关联，计算方式如式(12)和式(13)所示：

$$M_{adj_{i,j}} = \begin{cases} 1, & dis(i, j) < \lambda \\ 0, & \text{其他} \end{cases} \quad (12)$$

$$\begin{aligned} AdjSSA(Q_t^{(S,F)}, K_t^{(S,F)}, V_t^{(S,F)}) = \\ \frac{\exp(Score_t^{(S,F)})}{\sum_{k=1}^N \exp(Score_t^{(S,F)})} \odot M_{adj} \cdot V_t^{(S,F)} \end{aligned} \quad (13)$$

其中， $dis(i, j)$ 表示节点*i*和节点*j*之间的距离，也就是它们之间的跳数， λ 是一个阈值， \odot 表示 Hadamard 乘积。

(2) 相似空间自注意力机制。不同道路之间可能存在相似的交通流状态，为了捕获这种相似的交通状态，本文采用文献[28]的方法，使用 DTW 算法引入了相似空间矩阵 M_{res} ，然后利用 K-shape 聚类算法更新键值矩阵 $K_t^{(S,F)}$ ，用于捕捉相邻节点邻近时间段的短期空间相关性，从而扩展当前节点的影响范围。将其公式化为如下过程：

$$M_{res_{i,j}} = \begin{cases} 1, & j \in top-K(i) \\ 0, & \text{其他} \end{cases} \quad (14)$$

$$r_{t,n} = \sum_{i=1}^{N_p} \text{Softmax}((X_{(t-T:t),n} W_1) p_i W_2) (p_i W_3) \quad (15)$$

$$\tilde{K}_t^{(S,F)} = K_t^{(S,F)} + R_t \quad (16)$$

其中， $top-K(i)$ 表示通过 DTW 算法计算出的与节点*i*的历史交通流最相似的 K 个节点的集合。每个簇的质心 p_i 来代表该簇，这个质心也是一个长度为 T 的时间序列，以便更好地描述和分析交通模式。 W_1, W_2, W_3 为可学习参数， $r_{t,n}$ 表示节点*n*的相似的历史序列综合， R_t 表示所有的 N 个节点的历史相似矩阵，时间片 t 处的新 $\tilde{K}_t^{(S,F)}$ 矩阵整合了从时间片 $(t-T)$ 到 t 的所有节点的历史交通流信息。利用这种方法，能够找到那些相似的交通模式的远程节点对，当计算查询矩阵和新键值矩阵的乘积以获得式(4)中时间片 t 的空间相关得分 $Score_t^{(S,F)}$ 时，查询矩阵就可以考虑到与其相似节点的交通流状态，从而更加准确地获取空间上的相关性。由此得到相

似空间自注意力机制计算公式如下所示：

$$(Q_t^{(S,F)}, K_t^{(S,F)}, V_t^{(S,F)}) = \frac{\exp(Score_t^{(S,F)})}{\sum_{k=1}^N \exp(Score_t^{(S,F)})} \odot M_{res} \cdot V_t^{(S,F)} \quad (17)$$

其中， \odot 表示 Hadamard 乘积。通过这种方式，空间注意力模块同时结合了短距离邻接空间信息和远距离相似空间信息，从而能够更加准确全面地捕获道路节点间的空间相关性。

2.3.2 时间注意力模块

(1) 时间自注意力机制。时间自注意力机制首先需要将时间序列中的每个时间步的交通流表示为一个 d 维的向量，从而形成一个 $N \times d$ 的输入表示矩阵。接着针对每个时间步，该模型会将其表示为 3 个向量，分别是 $Q_t^{(T,F)}$ 、 $K_t^{(T,F)}$ 以及 $V_t^{(T,F)}$ 向量， (T, F) 表示时间-交通流维度。在计算注意力分布时，时间自注意力机制通常使用点积乘法的方式，即将 $Q_t^{(T,F)}$ 向量和 $K_t^{(T,F)}$ 向量的点积作为相似度得分，用 $Score_t^{(T,F)}$ 矩阵表示，然后将得分通过 $Softmax$ 函数进行归一化，以得到注意力分布。这些注意力分布会被用于对 $V_t^{(T,F)}$ 向量进行加权平均，得到当前时间步的输出向量。参考空间自注意力机制，将这个过程总结为如下公式：

$$TSA = Softmax \left(\frac{\left(Q_t^{(T,F)} (K_t^{(T,F)})^T \right)}{\sqrt{d'}} \right) \cdot V_t^{(T,F)} \quad (18)$$

(2) 时间-空间自注意力机制。传统的自注意力机制通常只考虑时间或者空间维度的相关性，而忽略了节点间跨时空性的相互关联。例如，时间自注意力机制只能计算某个节点在不同时间片上的相关性，而无法考虑不同节点之间的相互作用。空间自注意力机制只能计算某个时刻路网节点之间的相关性，而无法考虑不同时刻之间的相互作用。这些传统的自注意力机制忽略了节点之间跨时空的相互关联，因此未能从路网整体上捕获这种复杂的动态时空特征。具体来说，在现实生活中，道路节点的交通流量不仅受到当前时刻的影响，还受到过去和未来时刻周围节点的影响。如图 4 所示：以道路节点 *B* 为例，假设需要预测 $t+n$ 时刻的交通流量，时间-空间自注意力机制会考虑 $t+n$ 时刻 *B* 和 *C* 的交通流量， t 时刻节点 *A* 的交通流量以及 t 时刻节点 *D* 的交通流量之间的关系，从而更准确地预测 $t+n$ 时刻道路节点 *B* 的交通流量。

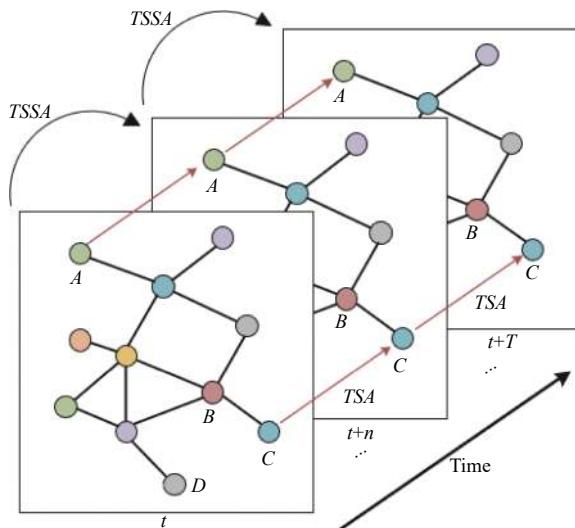


图4 时间注意力关系图

时间-空间自注意力机制是一种有效的方法,可以更好地处理交通流在不同时间和位置之间的跨时空相互关联,从而提高交通流预测和交通规划的准确性和效果。其基本思想是将不同时间和位置的交通流量进行比较,并根据它们的相对重要性给出加权值,以此来捕捉交通流在不同时间和位置之间的复杂关系。在这种自注意力机制中,模型会计算不同时间和位置之间节点的相对重要性,并将其应用于加权求和操作中,从而更好地捕捉交通流在不同时间和位置之间的相互关联。具体来说,对于每个时刻和位置的交通量,模型会计算该位置与其他时刻和位置之间的相对重要性,并根据这些重要性给出加权值。计算公式如下:

$$Score_t^{(T,S)} = \frac{W_Q^{(T,S)}(X_{TFS})^T W_K^{(T,S)}(X_{TFS})}{\sqrt{d'}} \quad (19)$$

$$\begin{aligned} TSSA(Q_t^{(T,S)}, K_t^{(T,S)}, V_t^{(T,S)}) = \\ \frac{\exp(Score_t^{(T,S)})}{\sum_{k=1}^N \exp(Score_t^{(T,S)})} \cdot W_V^{(T,S)}(X_{TFS})^T \end{aligned} \quad (20)$$

其中, $W_Q^{(T,S)}$, $W_K^{(T,S)}$, $W_V^{(T,S)}$ 为可学习的参数,用于计算时间-空间自注意力机制中的查询,键,值向量。时间-空间自注意力机制可以将时间和空间信息结合起来,能更准确地建模交通流量的时空分布和变化规律。

经过4个自注意力机制的处理后,需要进行特征融合。考虑到模型的计算复杂度,本文采用了拼接方法,将各个自注意力机制的输出拼接在一起,并通过一个全连接层将数据从原始空间映射到输出空间。该过程可表示为以下公式:

$$X_o = Linear(Concat(AdjSSA, ResSSA, TSA, TSSA)W_{ST}) \quad (21)$$

其中, $AdjSSA$, $ResSSA$, TSA , $TSSA$ 分别为4个自注意力机制的输出, $Concat$ 表示维度拼接操作, W_{ST} 表示全连接层的权重矩阵,最终得到注意力模块的输出 $X_o \in \mathbb{R}^{T \times N \times d}$ 。此外为了优化训练效率,在每个组件中还采用了残差学习框架。

2.4 输出层

本文在每个时空编码器层之后使用由 1×1 卷积组成的跳跃连接来将输出 X_o 转换为跳跃维度 $X_{sk} \in \mathbb{R}^{T \times N \times d_{sk}}$ 。这里 d_{sk} 是跳过维度。然后,通过对每个跳跃连接层的输出求和,获得最终的隐藏状态 $X_{hid} \in \mathbb{R}^{T \times N \times d_{sk}}$ 。为了进行多步预测,直接使用输出层将最终隐藏状态 X_{hid} 转换为所需维度,如下所示:

$$X_{output} = Conv_2(Conv_1(X_{hid})) \quad (22)$$

其中, $X_{output} \in \mathbb{R}^{T' \times N \times C}$ 是 T' 步的预测结果, $Conv_1$ 和 $Conv_2$ 是 1×1 卷积操作。通过最小化预测值和真实值之间的平均绝对误差 (MAE),并利用优化器进行端到端的训练。

$$Loss = \frac{1}{n} \sum_{i=1}^n |Y_i - X_{output}| \quad (23)$$

3 实验与结果分析

3.1 实验数据集

为了检验MSTTF的实际有效性,本文分别在3个真实交通数据集PeMS04, PeMS07和PeMS08上进行实验。PeMS数据集是加利福尼亚州交通管理中心(Caltrans)发布的交通数据集,由Caltrans性能测量系统(California Transportation Agencies (Caltrans) Performance Measurement System (PeMS))通过部署在加利福尼亚州主要城市的高速公路上的探测器实时采集。数据集信息如表1所示。其中,PeMS04主要由307个探测器探测而来,PeMS07来自883个探测器,PeMS08来自170个探测器。为了与大多数现代方法保持一致,本文分别将3个数据集按6:2:2的比例划分为训练集,验证集和测试集。

表1 数据描述

数据集	节点数	步数	间隔(min)	时间范围
PeMS04	307	16992	5	2018/01/01–2018/02/28
PeMS07	883	28224	5	2017/05/01–2017/08/31
PeMS08	170	17856	5	2016/07/01–2016/08/31

3.2 评价指标

在实验中,本文通过3个广泛使用的指标,即 MAE (平均绝对误差), $RMSE$ (均方根误差)和 $MAPE$ (平均绝对百分比误差),来衡量模型的准确性,公式如下:

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (24)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|^2} \quad (25)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (26)$$

3.3 参数设置

所有实验都在配备NVIDIA GeForce 3060 GPU和32 GB内存的机器上进行,使用PyTorch 1.13.1和Python 3.9.16实验环境。时间步长设置为5 min,使用过去1 h(12步)的数据来预测下1 h(12步)的交通流量,即多步预测。本文对最小非平凡特征向量的个数 k ,

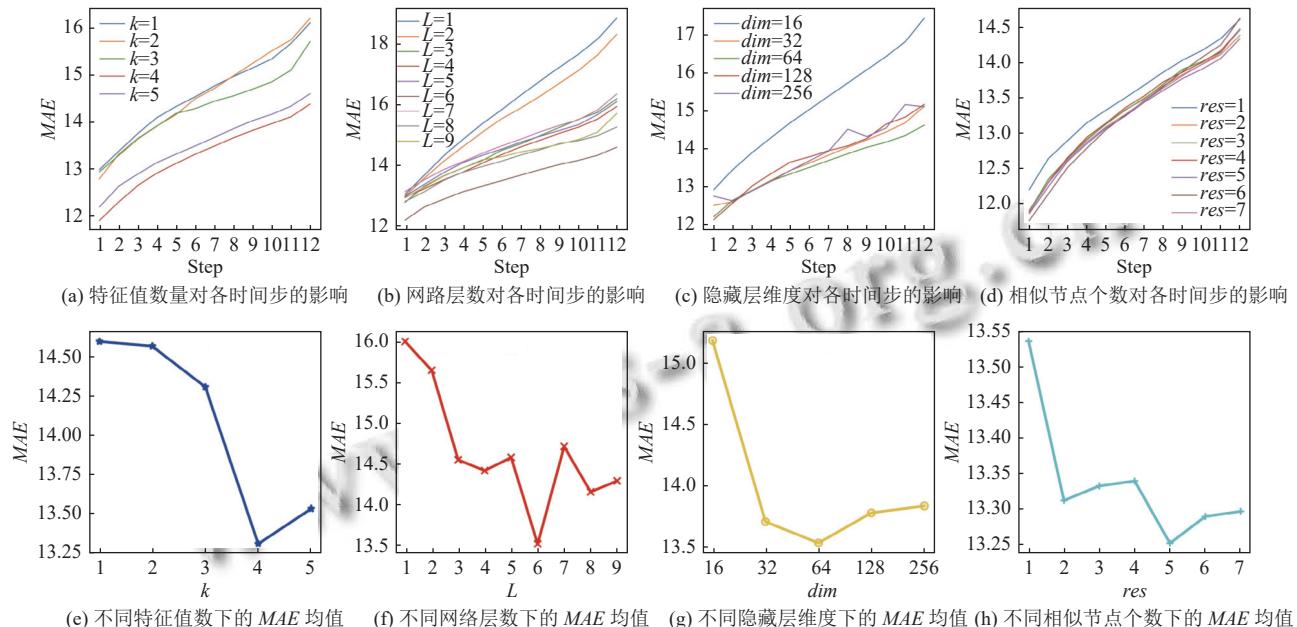


图5 不同参数设置下的 MAE 的变化趋势

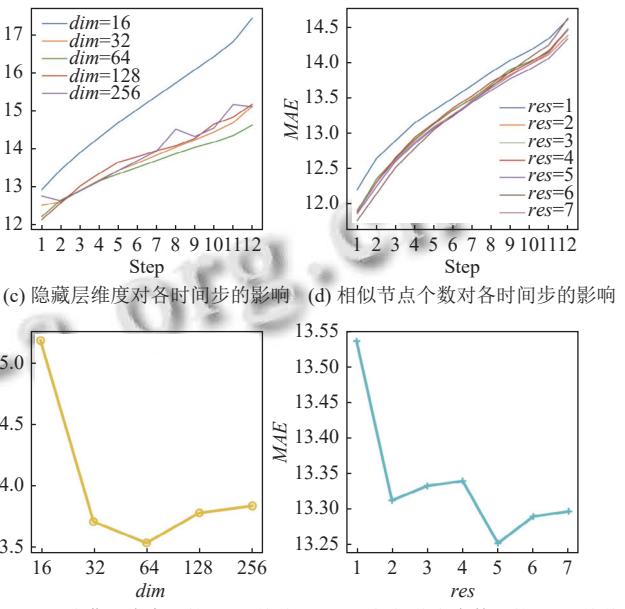
3.5 实验结果分析

表2展示了在不同数据集下,各个模型对未来1 h交通流的平均预测性能。结果表明,本文模型MSTTF取得了更好的实验效果。STGCN, STGCN和GWNET这些基于图结构的神经网络虽然能够有效处理非规则数据,灵活适应复杂的交通网络拓扑结构,捕捉交通流

编码器的层数 L ,隐藏层维度 dim 以及相似空间的大小 res 进行探索,如图5所示,当 $k=4$, $L=5$, $dim=64$, $res=5$ 时,实验的平均绝对误差最小,效果最优。优化器采用AdamW优化器。学习率为0.001,批量大小为16(由于PeMS07路网节点过多,本研究将批量大小设置为4),epoch为300。模型参数拟合过程中的误差变化曲线如图6所示,本文将前100次进行可视化展示,随着迭代次数epoch的次数增加,预测精度逐渐升高,误差值逐渐减小,模型很快收敛。

3.4 模型对照基线

为了验证MSTTF在大规模区域交通流长期预测中的有效性,将其与5种当前较为先进的交通流预测模型进行对比分析。这5个基线模型分别为STGCN,STGCN, GWNET, STTN, PDFormer。其中STGCN,STGCN, GWNET为当前较为先进的时空图卷积网络模型;STTN和PDFormer是目前最先进的Transformer模型。在实验过程中,每个基准模型的参数都与MSTTF模型的参数设置对标,保证实验数据的真实可靠性。



的空间依赖性,但对交通流的时间特征无法充分挖掘,无法准确应对交通流的突变场景。STTN模型在传统Transformer中加入了时间自注意力机制和空间自注意力机制,有效地挖掘了交通流在时间和空间上的规律,提高了预测准确率。而PDFormer通过加入延迟注意力机制挖掘了邻接节点的历史交通流特征,扩大了节点

间相互影响的范围,取得了不错的预测效果.在此基础上, MSTTF 通过设计一种时间-空间自注意力机制, 进

一步学习交通流在路网中所有节点间的长期规律, 使得模型能够更好地拟合交通流的变化.

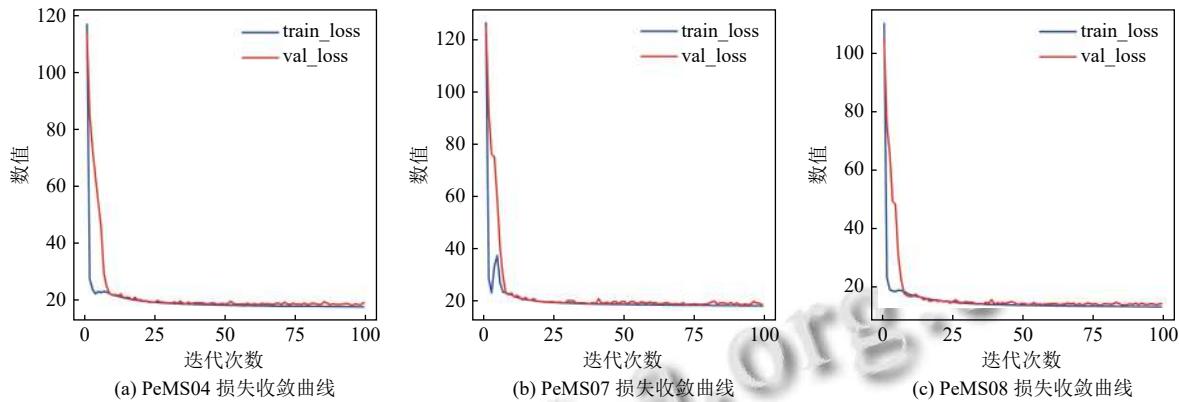


图 6 损失收敛曲线

表 2 各模型在不同数据集下的预测性能

模型	PeMS04			PeMS07			PeMS08		
	MAE	MAPE (%)	RMSE	MAE	MAPE (%)	RMSE	MAE	MAPE (%)	RMSE
STGCN	21.276	14.812	32.909	22.018	12.243	33.683	16.9889	12.224	26.542
STSGCN	21.145	14.823	33.017	23.864	11.369	36.542	16.649	11.140	25.978
GWNET	19.409	13.464	30.695	20.791	9.546	33.865	15.068	10.311	24.056
STTN	20.023	14.071	31.463	21.283	10.546	34.284	15.559	11.293	24.468
PDFormer	18.364	12.115	29.998	19.854	8.526	32.889	13.584	9.128	23.501
MSTTF	18.121	12.264	29.847	19.544	8.57	32.796	13.470	9.243	22.801

为了更直观地展示 MSTTF 模型的有效性, 本文还随机选择了连续两天 (day1, day2) 的真实数值和各模型的预测结果进行可视化展示, 以验证模型的预测准确性. 如图 7 所示, 考虑到时间跨度太大会导致图像模糊, 不利于观察, 这里我们将两天的数据分开进行展示. 其中黑色曲线表示交通流的真实数值. 从图 7 可以观察到, 基线模型也都能够大致预测交通流的动态变化, 但总体而言, 基于图卷积神经网络的模型相对于 Transformer 模型预测误差大, 说明时空自注意力机制能够更好地捕捉交通流的时空特征. 同时, MSTTF 模型相对于 PDFormer 模型更好地拟合了交通流的突变, 这表明交通流不仅受到相邻节点之间的短期时空特征影响, 而且路网中所有节点之间的长期时空特征也会对交通流的动态变化产生影响.

交通流在不同时间段可能会表现出不同的特征和模式, 例如, 在高峰时段, 交通流量通常会更加拥堵, 而在低峰时段, 交通流量可能较为平稳. 因此本文还考虑了模型在不同时间段的预测准确性, 在 PeMS04 数据集中选取了 4 个有代表性的时间段进行实验, 来评估

每个模型的预测性能, 这 4 个时间段分别为工作日的高峰交通时间段 (2018 年 2 月 13 日星期二的 3:00–9:00, 16:00–22:00) 以及休息日对应的时间段 (2018 年 2 月 17 日星期六的 3:00–9:00, 16:00–22:00). 实验结果如图 8 所示, 从图 8 中可以观察到, 在工作中日期间, 所有模型基本都能拟合交通流的变化趋势. 然而, 在休息日期间, 各模型的预测性能相对于工作日存在显著差异. 这可能是因为在周末交通流会受到更多复杂因素的影响, 如休闲活动, 旅游, 购物等. 这些因素会导致交通流的突然增减, 从而表现出更加复杂和不稳定的特性. 而 MSTTF 在嵌入层中加入了上周同一时间段的交通流数据, 让模型能够周期性地学习这种不平稳数据的时空特征, 从而使得模型能更准确地预测交通流的动态变化, 因此表现得更加稳定.

为验证所提模型的准确性和稳定性, 将 MSTTF 模型与其他线模型在不同时间间隔 (5 min, 15 min, 30 min, 45 min, 60 min) 下的预测性能进行比较, 实验结果如表 3 所示. 由于 PeMS07 道路节点过多, PeMS08 道路节点又过少, 本文选择道路节点数量适中 PeMS04 数

据集进行实验。为了衡量预测性能,计算了 MAE 和 $MAPE$ 指标,并将实验结果表 3 中。从表 3 的数据可以看出,随着预测时间的增加,所有模型的实验误差均在增加,预测准确性普遍降低,有些模型甚至出现大幅波动的情况,这表明这些模型不适合做长期预测。而 MSTTF 在预测时长为 60 min 时, MAE 指标比其他模型分别降低 19.8%, 21.2%, 8.3%, 13.5%, 1.5%, 表现出了更好的准确性。而随着预测时长的增加,MSTTF 与 STTN, PDFormer 相比, MAE 变化的更小,充分说明 MSTTF 通过多种时空自注意力机制的融合能够更好地捕捉交通流的长期时空特征,更加适合用于交通流长期预测。

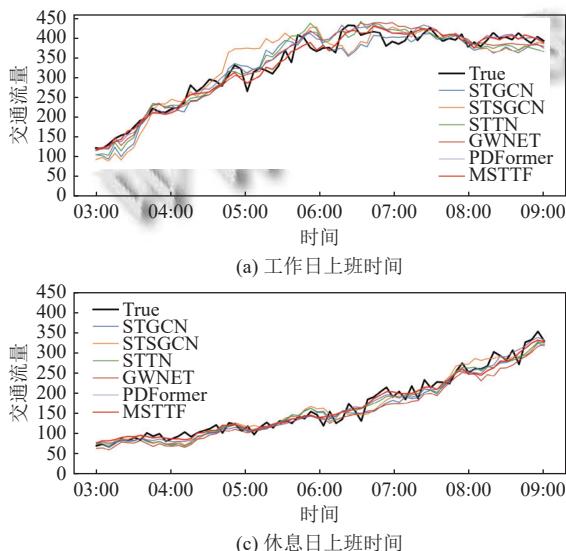


图 8 各模型在工作日和休息日的预测情况

表 3 各个模型在不同预测时间长度下的性能

模型	5 min		15 min		30 min		45 min		60 min	
	MAE	MAPE (%)								
STGCN	17.983	12.316	19.610	13.801	20.973	14.345	22.688	15.970	24.237	16.959
STSGCN	17.539	12.454	19.081	13.457	20.920	14.831	22.661	16.011	24.687	17.206
GWNET	17.536	12.481	18.385	12.740	19.415	13.490	20.271	14.111	21.205	14.787
STTN	17.375	12.042	18.702	12.906	19.819	13.796	20.969	14.742	22.483	16.156
PDFormer	16.370	11.252	17.688	11.874	18.517	12.590	18.940	13.042	19.732	13.522
MSTTF	16.576	11.100	17.517	11.721	18.267	12.235	18.810	12.664	19.443	13.229

3.6 注意力机制消融实验

为了评价不同自注意力机制对模型预测效果的影响,本文设置了 5 种模型的变体: TF 表示剔除了 SimSSA 和 TSSA; STF 表示只剔除了 SimSSA; TTF 表示只剔除了 TSSA; MSTF 表示剔除 AdjSSA 和 SimSSA; MTTF 表示剔除了 TSA 和 TSSA; 最后是本文模型 MSTTF。同样在 PeMS04 数据集上进行实验,标准步长为 5 min,

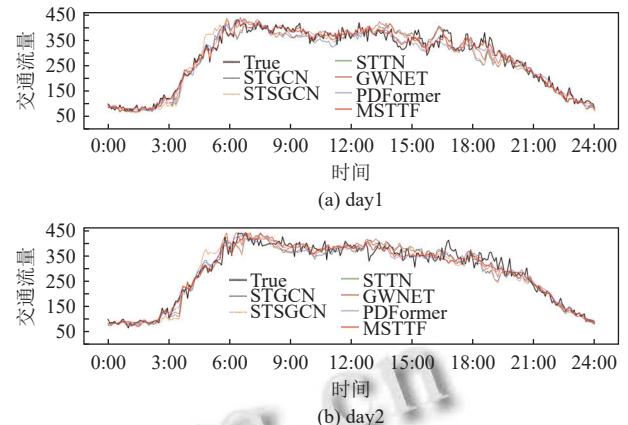


图 7 交通流真实值和各个模型预测值效果展示

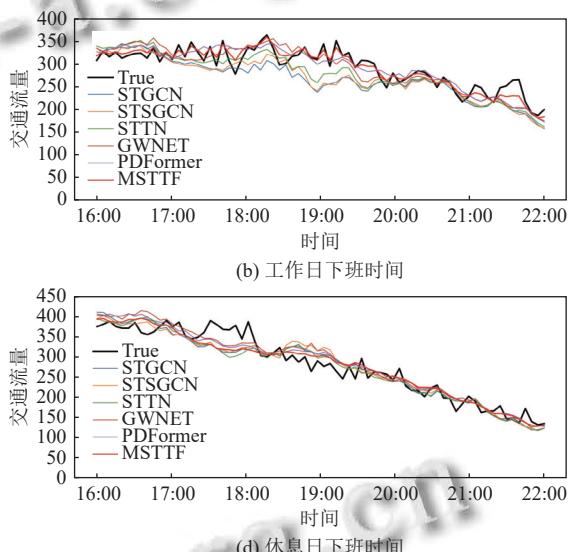


图 8 各模型在工作日和休息日的预测情况

预测时间为 60 min。同时本文还计算了 5 min, 30 min, 60 min 的评价指标,结果如表 4 所示。可以看到, MSTTF 始终优于其余 5 种变体模型,对比 TTF, MSTF, MSTTF, 不难看出, MSTTF 的时间-空间自注意力机制对模型的预测结果影响很大,说明在提取时空特征时,不能简单地考虑单节点间的时空相关性,应该要从路网整体上考虑这种动态长期时空依赖。同时,对比 STF, MTTF,

MSTTF 可以发现, 加入相似空间自注意力后, 模型的预测结果也会有一定的提升, 说明在现实生活中, 存在这种交通状态相似的路网节点。从实验结果来看, 本文设计的时间和空间注意力模块各有所长, 对模型的预

测能力都有各自的贡献。特别需要注意的是, 缺失空间注意力模块会导致精度特别显著下降, 尤其是在长期预测范围内, 这进一步表明了道路网络的空间结构对交通流的长期影响。

表 4 消融实验下的评价指标

模型	评价指标	TF	STF	TTF	MSTF	MTTF	MSTTF
5 min	MAE	17.198	17.375	17.140	17.237	17.196	16.576
	MAPE (%)	11.714	12.042	11.954	12.070	11.767	11.100
	RMSE	27.815	27.758	27.726	27.801	27.788	26.959
30 min	MAE	20.864	19.419	18.783	20.907	20.866	18.267
	MAPE (%)	14.392	13.796	12.669	14.421	14.480	12.235
	RMSE	33.599	31.327	30.694	33.653	33.589	30.076
60 min	MAE	23.719	21.983	21.885	24.941	24.954	19.443
	MAPE (%)	17.642	16.156	14.558	17.658	17.812	13.229
	RMSE	39.630	34.881	34.443	39.569	39.608	31.838
均值	MAE	20.985	19.637	19.259	20.788	21.146	18.121
	MAPE (%)	14.642	14.071	12.991	14.651	14.903	12.264
	RMSE	33.972	31.463	31.023	33.976	33.940	29.847

4 结语

本文提出了一种新的多种时空自注意力融合的交通流预测模型 (MSTTF), 该模型不仅能学习单节点的时间和空间特征, 还通过时间-空间自注意力机制捕获路网中所有节点的长期空间特征, 充分挖掘了交通流的时空特性。与经典的交通流预测模型 STGCN, STGCN, GWNET, STTN 以及 PDFormer 相比, 本文模型能更好拟合交通流的各种变化趋势。对比工作日和休息日的预测结果时, 发现 MSTTF 对休息日这种无规律性的交通流特征学习能力更强。同时在不同时间步的预测结果对比中, 验证了本文所提模型相比于其他模型在长周期预测上有更大优势。在下一步工作中, 我们将继续优化模型的计算开销并研究其他因素对交通流的影响并探索挖掘动态时空相关性其他方法。

参考文献

- Zaki JF, Ali-Eldin A, Hussein SE, et al. Traffic congestion prediction based on hidden Markov models and contrast measure. *Ain Shams Engineering Journal*, 2020, 11(3): 535–551. [doi: [10.1016/j.asej.2019.10.006](https://doi.org/10.1016/j.asej.2019.10.006)]
- Ye YN, Chen L, Xue F. Passenger flow prediction in bus transportation system using ARIMA models with big data. *Proceedings of the 2019 International Conference on Cyber-enabled Distributed Computing and Knowledge Discovery (CyberC)*. Guilin: IEEE, 2019. 436–443. [doi: [10.1109/CyberC.2019.00081](https://doi.org/10.1109/CyberC.2019.00081)]
- Razali NAM, Shamsaimon N, Ishak KK, et al. Gap, techniques and evaluation: Traffic flow prediction using machine learning and deep learning. *Journal of Big Data*, 2021, 8(1): 152. [doi: [10.1186/s40537-021-00542-7](https://doi.org/10.1186/s40537-021-00542-7)]
- Hou Y, Deng ZY, Cui HK. Short-term traffic flow prediction with weather conditions: Based on deep learning algorithms and data fusion. *Complexity*, 2021, 2021: 6662959. [doi: [10.1155/2021/6662959](https://doi.org/10.1155/2021/6662959)]
- Yang LJ, Yang Q, Li YH, et al. K-nearest neighbor model based short-term traffic flow prediction method. *Proceedings of the 18th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES)*. Wuhan: IEEE, 2019. 27–30. [doi: [10.1109/DCABES48411.2019.00014](https://doi.org/10.1109/DCABES48411.2019.00014)]
- Tang JJ, Chen XQ, Hu Z, et al. Traffic flow prediction based on combination of support vector machine and data denoising schemes. *Physica A: Statistical Mechanics and its Applications*, 2019, 534: 120642. [doi: [10.1016/j.physa.2019.03.007](https://doi.org/10.1016/j.physa.2019.03.007)]
- Zhang WB, Yu YH, Qi Y, et al. Short-term traffic flow prediction based on spatio-temporal analysis and CNN deep learning. *Transportmetrica A: Transport Science*, 2019, 15(2): 1688–1711. [doi: [10.1080/23249935.2019.1637966](https://doi.org/10.1080/23249935.2019.1637966)]
- Chen XY, Xie XS, Teng D. Short-term traffic flow prediction based on ConvLSTM model. *Proceedings of the 5th IEEE Information Technology and Mechatronics Engineering Conference (ITOEC)*. Chongqing: IEEE, 2020. 846–850. [doi: [10.1109/ITOEC49072.2020.9141783](https://doi.org/10.1109/ITOEC49072.2020.9141783)]
- Zhao LN, Wen XY, Wang YP, et al. A novel hybrid model of ARIMA-MCC and CKDE-GARCH for urban short-term

- traffic flow prediction. IET Intelligent Transport Systems, 2022, 16(2): 206–217. [doi: [10.1049/itr2.12138](https://doi.org/10.1049/itr2.12138)]
- 10 Li ZH, Chen CL, Min Y, *et al.* Dynamic hidden Markov model for metropolitan traffic flow prediction. Proceedings of the 92nd IEEE Vehicular Technology Conference (VTC2020-Fall). Victoria: IEEE, 2020. 1–5. [doi: [10.1109/VTC2020-Fall49728.2020.9348555](https://doi.org/10.1109/VTC2020-Fall49728.2020.9348555)]
- 11 Raskar C, Nema S. Metaheuristic enabled modified hidden Markov model for traffic flow prediction. Computer Networks, 2022, 206: 108780. [doi: [10.1016/j.comnet.2022.108780](https://doi.org/10.1016/j.comnet.2022.108780)]
- 12 Sun ZY, Hu YJ, Li W, *et al.* Prediction model for short-term traffic flow based on a K-means-gated recurrent unit combination. IET Intelligent Transport Systems, 2022, 16(5): 675–690. [doi: [10.1049/itr2.12165](https://doi.org/10.1049/itr2.12165)]
- 13 Xing YM, Ban XJ, Liu X, *et al.* Large-scale traffic congestion prediction based on the symmetric extreme learning machine cluster fast learning method. Symmetry, 2019, 11(6): 730. [doi: [10.3390/sym11060730](https://doi.org/10.3390/sym11060730)]
- 14 丁男, 高壮林, 许力, 等. 基于数据优先级和交通流密度的异构车联网数据链路层链路调度算法. 计算机学报, 2020, 43(3): 526–536. [doi: [10.11897/SP.J.1016.2020.00526](https://doi.org/10.11897/SP.J.1016.2020.00526)]
- 15 卢海鹏, 韩莹, 张凯, 等. 基于 VMD-BiLSTM-BLS 模型的短时交通流预测. 计算机系统应用, 2022, 31(5): 238–245. [doi: [10.15888/j.cnki.csa.008469](https://doi.org/10.15888/j.cnki.csa.008469)]
- 16 张维, 袁绍欣, 陶建军, 等. 基于多元因素的 Bi-LSTM 高速公路交通流预测. 计算机系统应用, 2021, 30(6): 184–190. [doi: [10.15888/j.cnki.csa.007969](https://doi.org/10.15888/j.cnki.csa.007969)]
- 17 Li ZJ, Li CH, Cui X, *et al.* Short-term traffic flow prediction based on recurrent neural network. Proceedings of the 2021 International Conference on Computer Communication and Artificial Intelligence (CCAI). Guangzhou: IEEE, 2021. 81–85. [doi: [10.1109/CCAI50917.2021.9447466](https://doi.org/10.1109/CCAI50917.2021.9447466)]
- 18 Li MZ, Zhu ZX. Spatial-temporal fusion graph neural networks for traffic flow forecasting. Proceedings of the 35th AAAI Conference on Artificial Intelligence. AAAI, 2021. 4189–4196. [doi: [10.1609/aaai.v35i5.16542](https://doi.org/10.1609/aaai.v35i5.16542)]
- 19 Chen C, Xu YB, Zhao JX, *et al.* Combining random forest and graph wavenet for spatial-temporal data prediction. Intelligent and Converged Networks, 2022, 3(4): 364–377. [doi: [10.23919/ICN.2022.0024](https://doi.org/10.23919/ICN.2022.0024)]
- 20 王雨松, 吴向东, 尤晨欣, 等. 基于 DWT-GCN 的短时交通流预测. 计算机系统应用, 2022, 31(9): 306–312. [doi: [10.15888/j.cnki.csa.008682](https://doi.org/10.15888/j.cnki.csa.008682)]
- 21 Song C, Lin YF, Guo SN, *et al.* Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting. Proceedings of the 34th AAAI Conference on Artificial Intelligence. New York: AAAI, 2020. 914–921. [doi: [10.1609/aaai.v34i01.5438](https://doi.org/10.1609/aaai.v34i01.5438)]
- 22 Guo K, Hu YL, Sun YF, *et al.* Hierarchical graph convolution network for traffic forecasting. Proceedings of the 35th AAAI Conference on Artificial Intelligence. AAAI, 2021. 151–159. [doi: [10.1609/aaai.v35i1.16088](https://doi.org/10.1609/aaai.v35i1.16088)]
- 23 Vijayalakshmi B, Ramar K, Jhanjhi NZ, *et al.* An attention-based deep learning model for traffic flow prediction using spatiotemporal features towards sustainable smart city. International Journal of Communication Systems, 2021, 34(3): e4609. [doi: [10.1002/dac.4609](https://doi.org/10.1002/dac.4609)]
- 24 Reza S, Ferreira MC, Machado JJM, *et al.* A multi-head attention-based Transformer model for traffic flow forecasting with a comparative analysis to recurrent neural networks. Expert Systems with Applications, 2022, 202: 117275. [doi: [10.1016/j.eswa.2022.117275](https://doi.org/10.1016/j.eswa.2022.117275)]
- 25 Yan HY, Ma XL, Pu ZY. Learning dynamic and hierarchical traffic spatiotemporal features with Transformer. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(11): 22386–22399. [doi: [10.1109/TITS.2021.3102983](https://doi.org/10.1109/TITS.2021.3102983)]
- 26 Wen YJ, Xu P, Li ZH, *et al.* RPConformer: A novel transformer-based deep neural networks for traffic flow prediction. Expert Systems with Applications, 2023, 218: 119587. [doi: [10.1016/j.eswa.2023.119587](https://doi.org/10.1016/j.eswa.2023.119587)]
- 27 Xu MX, Dai WR, Liu CM, *et al.* Spatial-temporal Transformer networks for traffic flow forecasting. arXiv: 2001.02908, 2020.
- 28 Jiang JW, Han CK, Zhao WX, *et al.* PDFormer: Propagation delay-aware dynamic long-range Transformer for traffic flow prediction. Proceedings of the 37th AAAI Conference on Artificial Intelligence. Washington: AAAI, 2023. 4365–4373.
- 29 Saha S, Haque A, Sidebottom G. Deep sequence modeling for anomalous ISP traffic prediction. Proceedings of the 2022 IEEE International Conference on Communications. Seoul: IEEE, 2022. 5439–5444. [doi: [10.1109/ICC45855.2022.9838262](https://doi.org/10.1109/ICC45855.2022.9838262)]

(校对责编: 牛欣悦)