

基于增强时空 Transformer 的交通流预测方法

叶宝林^{1,2,4†}, 苗永超^{1,2,3}, 王 征^{1,2,3}

1. 嘉兴大学 全省多模态感知与智能系统重点实验室, 浙江 嘉兴 314001;
2. 嘉兴大学 嘉兴市智慧交通重点实验室, 浙江 嘉兴 314001;
3. 浙江理工大学 信息科学与工程学院, 杭州 310018;
4. 嘉兴数字城市实验室有限公司, 浙江 嘉兴 314001)

摘要: 精准预测路网交通流是保障智能交通系统高效运行的基础. 针对现有方法存在难以有效建模交通流数据中复杂非线性时空动态依赖关系的问题, 提出一种基于增强时空 Transformer (ESTformer) 的交通流预测方法. 该方法设计多尺度时间 Transformer 和增强空间 Transformer 分别建模交通流序列数据之间的时间依赖关系和路网不同节点之间的空间依赖关系. 多尺度时间 Transformer 构建短期门控卷积网络捕获交通流数据中的短期时间依赖关系, 并引入时间多头自注意力机制捕获长期动态时间依赖关系. 增强空间 Transformer 通过对偶变换增强键向量的特征表达能力, 利用时变掩码矩阵动态更新键向量, 提高模型同时捕获节点特征和边特征的能力. 在 4 个真实交通流数据集上的测试结果表明, 与基线方法相比, 所提基于 ESTformer 的交通流预测方法具有更优越的预测性能. 相比于 13 种基线方法在不同数据集上表现最佳者, 所提方法在 12 个时间步上的平均绝对误差 (MAE) 和均方根误差 (RMSE) 分别改进了 1.14%-3.88%, 0.36%-1.78%.

关键词: 智能交通; 交通流预测; 时空依赖性; 时空 Transformer; 卷积神经网络; 对偶变换

中图分类号: TP181 文献标志码: A

DOI: 10.13195/j.kzyjc.2025.1161

引用格式: 叶宝林, 苗永超, 王征. 基于增强时空 Transformer 的交通流预测方法 [J]. 控制与决策.

Traffic flow prediction method based on enhanced spatio-temporal Transformer

YE Bao-Lin^{1,2,4†}, MIAO Yong-chao^{1,2,3}, WANG Zheng^{1,2,3}

1. Provincial Key Laboratory of Multimodal Perceiving and Intelligent Systems, Jiaxing University, Jiaxing 314001, China;
2. Jiaxing Key Laboratory of Smart Transportations, Jiaxing University, Jiaxing 314001, China;
3. School of Information Science and Engineering, Zhejiang Sci-Tech University, Hangzhou 310018, China;
4. Zhejiang Jiaxing Digital City laboratory Co.Ltd., Jiaxing 314001, China)

Abstract: Accurate prediction of road network traffic flow is the foundation for ensuring the efficient operation of intelligent transportation systems. Aiming at the problem that existing methods are difficult to effectively model the complex nonlinear spatio-temporal dynamic dependency relationships in traffic flow data, a traffic flow prediction method based on enhanced spatio-temporal Transformer (ESTformer) is proposed. This method designs a multi-scale temporal Transformer and an enhanced spatial Transformer to respectively capture the temporal dependency relationships among traffic flow sequence data and the spatial dependency relationships among different nodes. The multi-scale time Transformer builds a short-term gated convolutional network to capture short-term time-dependent relationships in traffic flow data, and introduces a time multi-head self-attention mechanism to capture long-term dynamic time-dependent relationships. The enhanced space Transformer enhances the feature expression ability of the key vector through dual transformation and dynamically updates the key vector using time-varying mask matrix, thereby improving the model's ability to simultaneously capture node features and edge features. The test results on four real traffic flow datasets show that, compared with the baseline method, the proposed traffic flow prediction method based on ESTformer has superior prediction performance. Compared with the 13 baseline methods that performed best on different datasets, the mean absolute error (MAE) and root mean square error (RMSE) of the proposed method

收稿日期: 2025-11-08; 录用日期: 2026-02-13.

基金项目: 国家自然科学基金项目 (61603154); 浙江省自然科学基金项目 (LTGS23F030002).

†通信作者. E-mail: yebaolin@zjxu.edu.cn.

improved by 1.14%-3.88% and 0.36%-1.78%, respectively, at 12 time steps.

Keywords: intelligent transportation; traffic flow prediction; spatio-temporal dependence; spatio-temporal Transformer; convolutional neural networks; dual transformation

0 引言

及时精准地预测路网交通流量信息不仅是设计有效交通诱导策略的基础,也是开发先进交通信号控制系统提升路网通行效率的关键^[1-3].由于路网中的交通流数据通常具有复杂时空关联关系^[4],利用历史交通流量数据预测未来交通流信息的核心是构建能够精准挖掘路网中交通流量特征信息的交通流预测模型.历史平均、自回归移动平均模型、向量自回归模型等^[5-6]早期基于统计学的交通流预测方法假设交通流量数据之间的关系呈线性变化,无法建模路网交通流数据中包含的非线性特征信息.为捕获交通流数据中的非线性特征信息,研究人员利用基于支持向量机、K近邻等^[7]机器学习方法对交通流序列数据进行建模.然而,基于机器学习的方法需手动筛选特征,且预测精度依赖于预设的参数.于是循环神经网络(RNN)^[8]及其变体LSTM^[9]和GRU^[10]被引入交通流预测领域,用于建模交通流时序数据之间的非线性时间关联关系.然而,基于RNN的交通流预测方法通过链式结构传递隐藏信息,在处理长序列交通量数据时存在梯度消失或梯度爆炸的问题.另外,由于当前路段的交通状态会同时影响附近多个路段的未来交通状态,对同时具有时间和空间关联关系的交通流数据,RNN仅建模交通流数据中的时间相关性而忽略了空间相关性导致模型预测精度下降.随着深度学习技术的不断发展,深度卷积神经网络^[11]被应用于提取路网交通流数据中的空间特征信息,并通过融合时间序列模型去捕获时间特征信息.然而,基于CNN的交通流预测模型只能处理欧几里得空间数据,难以对非欧几里得空间中的路网交通流数据进行建模.

由于图神经网络(GNN)^[4]提供了一种结构可拓展且表达能力强的框架,能够自适应地学习交通流数据中复杂的非线性时空特征信息,构建基于GNN及其变体的交通流预测模型是交通预测领域一个热点研究问题^[12].例如,为精准捕获路网交通流数据中的时空特征信息,Yu等^[13]提出一种基于时空图卷积网络(STGCN)的交通流预测方法.Li等^[14]提出基于扩散卷积循环神经网络(DCRNN)与Bai等^[15]提出基于自适应图卷积循环网络(ACGRN)的预测方法均采用图神经网络GNN捕获路网中的空间特征信息,采用循环神经网络RNN提取交通流序列数

据中的时间特征信息.Li等^[16]提出基于动态图卷积循环网络的交通流预测方法,构建反映节点相似性的动态邻接矩阵以增强空间结构自适应表达能力,并融合RNN和GNN捕捉路网交通流数据中的时空特征信息.尽管RNN和GRU等方法在捕获交通流数据中的时间特征信息方面表现不俗,但也存在梯度不稳定等局限性.因此,Wu等^[17]基于图波网(GWN)设计了一种基于节点嵌入学习的自适应邻接矩阵去捕获交通节点之间隐藏的空间依赖性,并通过融合扩张因果卷积网络捕获节点间的时间依赖性.然而,随着膨胀率增加,该方法会丢失短期信息.Guo等^[18]提出的注意力机制的时空图卷积网络(ASTGCN)和Li等^[19]提出的时空融合图神经网络(STFGNN)均是利用路网中节点邻接关系或时间序列相似性等先验信息构建邻接矩阵.然而,这类方法构建的邻接矩阵主要是刻画节点之间的静态空间关联关系,未能捕获节点之间的动态空间关联关系,这在一定程度上限制了基于GNN的交通流预测模型在不同交通场景中的适应性与泛化能力.

多头注意力机制作为Transformer架构的核心模块,因其能够有效建模交通流数据中的动态依赖关系,基于GNN和Transformer融合模型的交通流预测方法得到了广泛关注和应用.Wang等^[20]提出了一种基于时空图神经网络(STGNN)的交通流预测方法.该方法先设计一个带有位置注意力机制的GNN建模路网节点之间的空间关联关系,再构建GRU和Transformer分别建模路网中的短期和长期时间关联关系.Zheng等^[21]提出的图多注意力网络(GMAN)利用多头注意力机制动态建模节点之间的空间相关性,并利用多头自注意力刻画非线性时间特征.然而,该方法对短期时间模式的捕捉相对不足,导致预测精度受限.Luo^[22]等提出基于动态图卷积网络(AFDGCN)的交通流预测方法,将动态图学习与GRU相结合同步建模路网交通流数据中的时空关联关系,并引入时空注意力机制捕捉长距离时空特征信息.Yan^[23]等提出名为“流量转换器”的交通预测算法(Tformer),通过多头注意机制和堆叠层使转换器可以学习序列数据中的动态和分层特征,使用全局编码器和全局-局部解码器来提取和融合时空特征信息.然而,上述方法中空间多头注意力的键向量通常基于节点特征生成,忽略了边属性(如静

态特征、动态权重) 对空间交互的影响. Hu^[24] 等提出基于动态时空图变换网络 (DST-GTN) 的交通流预测方法, 通过在低通和全通滤波器中采用自适应权重来细化全局和局部时空特征表示. 另外, 研究人员也提出了基于偶动态时空图卷积网络^[25]、时空同步图卷积神经网络^[26]、时空自适应 Transformer 的方法进行交通流预测^[27]. 尽管上述基于 GNN 和 Transformer 的交通流预测方法取得了一些有价值的研究结果, 但仍存在一定的局限性. 首先, 基于 GNN 的交通流预测模型存在过度平滑问题, 无法通过堆叠多层网络提取交通数据中隐藏的时空特征, 导致基于 GNN 的模型在捕获长期时空依赖性方面存在一定的局限性. 其次, 基于 GNN 的交通流预测方法主要使用单一神经网络捕获交通流量数据中短期时间相关性或者长期时间相关性, 无法同时捕获短期和长期时间相关性. 另外, 基于时间 Transformer 的交通流预测方法虽然擅长捕获长期时间依赖关系, 但对短期突变的时间特征不敏感. 基于空间 Transformer 的交通流预测方法中, 传统键向量仅编码自身节点特征, 忽略了边的动态特征与静态特征. 为解决上述问题, 本文提出了一种基于增强时空 Transformer (Enhance spatio-temporal transformer, ESTformer) 模型的交通流预测方法, 使用 ESTformer 建模交通流量数据中复杂的时空关联关系. 本文的主要贡献包括:

(1) 提出一种新的基于增强时空 Transformer (ESTformer) 模型的交通流预测方法, 该方法可以同步建模路网交通流数据中的局部和全局时空关联关系. ESTformer 模型通过跳跃连接, 堆叠多层时空 Transformer 使所提方法可以更有效地捕获深层次隐藏时空特征而不会产生过平滑问题.

(2) 提出了一种新的多尺度时间 Transformer (multi-scale temporal transformer, MSTformer) 模型, 可以分层提取交通流数据中的短期和长期时间关联关系. MSTformer 模型中的短期门控卷积 (short-term gated convolutional, STGC) 在使用多尺度卷积核捕获短期交通流量变化规律的同时, 利用门控机制保留了重要历史交通信息. 另外, 在 MSTformer 中引入多头时间注意力机制, 提高了其挖掘长期动态时间关联关系的能力.

(3) 提出了一种新的增强空间 Transformer (enhanced spatial transformer, ESformer) 模型, 可以捕获路网中不同节点之间动态变化的全局空间关联关系. 与传统多头自注意力机制不同, 所提增强空间 Transformer 模型使用对偶变换增强键向量的特征表

达能力, 同时构建时变掩码矩阵动态更新键向量, 使得键向量能够同时编码节点特征与边特征.

1 基本定义

1.1 交通拓扑图

在交通流预测任务中, 可以将实际路网表示成带有权重的有向拓扑图 $G = (V, E, A)$, 其中 V 表示路网中所有顶点 (也即传感器节点) 组成的集合, $|V| = N$ 表示传感器数量, E 表示边的集合, $|E| = M$ 表示有向路段的数量, $A \in \mathbb{R}^{N \times N}$ 表示反映顶点关联关系的邻接矩阵. 交通流预测本质上是根据 N 个节点在 T 个历史时间步的交通流量数据 $\tilde{\mathbf{X}}_T = \{X_1, X_2, \dots, X_t, \dots, X_T\} \in \mathbb{R}^{N \times C \times T}$ (C 表示交通特征) 和图 G 学习一个函数 $f(\cdot)$ 预测未来 P 个时间步的流量 $\hat{\mathbf{X}}_P = \{X_{T+1}, X_{T+2}, \dots, X_{T+P}\} \in \mathbb{R}^{N \times C \times P}$. 其中, $X_t = (x_t^1, x_t^2, \dots, x_t^N) \in \mathbb{R}^{N \times C}$, $x_t^i \in \mathbb{R}^C$ 表示第 i 个节点在第 t 个时间步的流量特征. 上述过程可由式 (1) 描述:

$$\hat{\mathbf{X}}_P = f(\tilde{\mathbf{X}}_T; G). \quad (1)$$

式中: $f(\cdot)$ 代表可以学习的非线性变换函数.

1.2 对偶变换

交通拓扑图通过对偶变换可转换为超图^[25], 对偶变换的原理是将带有权重的有向交通拓扑图 $G = (V, E, A)$ 的节点和边分别映射为超图 $G_h = (V_h, E_h, \mathbf{H})$ 的边和节点. 其中, $|V_h| = M$, $|E_h| = N$, $\mathbf{H} \in \mathbb{R}^{M \times N}$ 是超图 G_h 的关联矩阵, 可同时描述交通拓扑图和超图的结构信息, 具体定义如下:

$$\mathbf{H}_{ij} = \begin{cases} 1, & \text{如果 } v_i \in e_j \\ 0, & \text{其它.} \end{cases} \quad (2)$$

式中: $v_i \in V_h, e_j \in E_h$. 可将 \mathbf{H} 分解为 $\mathbf{H} = \mathbf{H}_{\text{sn}} + \mathbf{H}_{\text{dn}}$, 矩阵 $\mathbf{H}_{\text{sn}} \in \mathbb{R}^{M \times N}$ 和 $\mathbf{H}_{\text{dn}} \in \mathbb{R}^{M \times N}$ 中的非零元素分别代表交通拓扑图中有向边的源节点和目标节点.

在交通拓扑图中边的特征通常由不同节点之间的距离或者是否相连来定义. 然而, 实际上边的特征信息不仅包含当前时刻平均速度、路段排队长度、实时流量等路网动态交通特征信息, 也包含路网拓扑结构、车道数等路网静态特征信息. 因此, 本文利用交通拓扑图中边和节点的特征重新定义超图的超节点特征 $\tilde{\mathbf{X}}_h$, 如式 (3) 所示:

$$\tilde{\mathbf{X}}_h = \{[(\omega_1 \odot \mathbf{H}_{\text{sn}})\tilde{\mathbf{X}}_T \oplus (\omega_2 \odot \mathbf{H}_{\text{dn}})\tilde{\mathbf{X}}_T]; \mathbf{A}_{\text{dis}}\}. \quad (3)$$

式中: $\tilde{\mathbf{X}}_T \in \mathbb{R}^{N \times C \times T}$ 表示交通图节点特征, \odot 表示哈达玛积, \oplus 表示求和操作, $\{;\}$ 表示矩阵拼接操作.

交通拓扑图中边和节点之间的相关性可以通过可学习的权重矩阵 ω_1 和 ω_2 捕获. $(\omega_1 \odot \mathbf{H}_{\text{in}})\tilde{\mathbf{X}}_T$ 和 $(\omega_2 \odot \mathbf{H}_{\text{in}})\tilde{\mathbf{X}}_T$ 表示有向边的动态特征, $\mathbf{A}_{\text{dis}} \in \mathbb{R}^{N \times N}$ 表示根据路网中不同节点之间距离构建的邻接矩阵, 代表有向边的静态特征. 通过式 (4) 也可以将超图的超节点特征变换回交通拓扑图.

$$\bar{\chi} = (\omega_3 \odot \mathbf{H})^\top \tilde{\mathbf{X}}'_h. \quad (4)$$

式中: \top 代表矩阵转置操作, ω_3 是可学习的权重矩阵, \odot 表示哈达玛积, $\tilde{\mathbf{X}}'_h$ 为更新后的超图的超节点特征, $\bar{\chi}$ 为更新后交通拓扑图的节点特征.

2 增强时空 Transformer

如图 1 所示, 本文所提增强时空 Transformer 模

型框架由一个输入层, L 个堆叠的时空 Transformer 层和两个预测层组成. 其中, 时空 Transformer 层由多尺度时间 Transformer 和增强空间 Transformer 组成. 输入层为卷积神经网络, 主要是将输入的交通流时间序列数据转换到高维空间, 提高网络的预测能力. 此外, 考虑到交通流数据呈现出明显的周期性变化规律, 在原始输入数据的时间维度上引入日周期嵌入和周周期嵌入, 分别提取日周期与周周期特征, 提高 ESTformer 捕获周期性特征的能力. 多尺度时间 Transformer 用于提取短期和长期时间关联关系, 增强空间 Transformer 用于提取全局动态空间关联关系. 预测层用来输出模型预测未来 P 个时间步的交通流量.

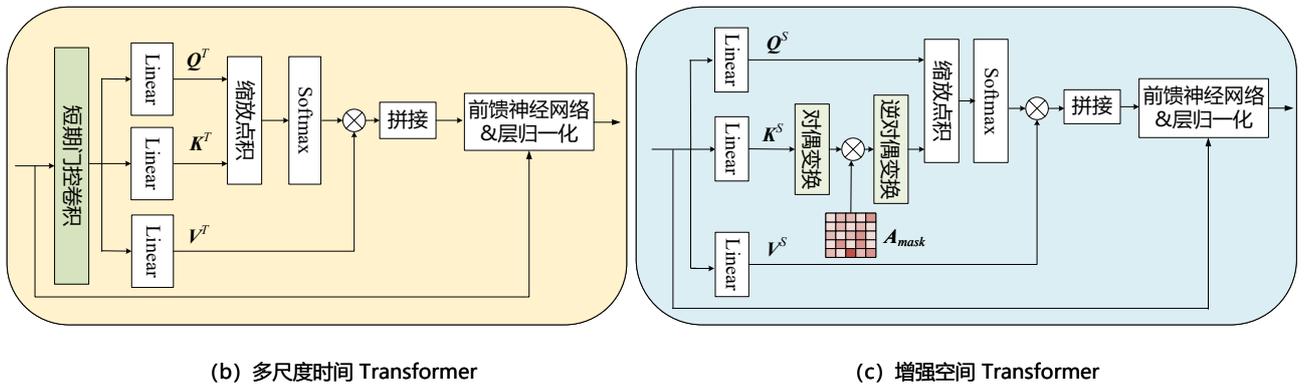
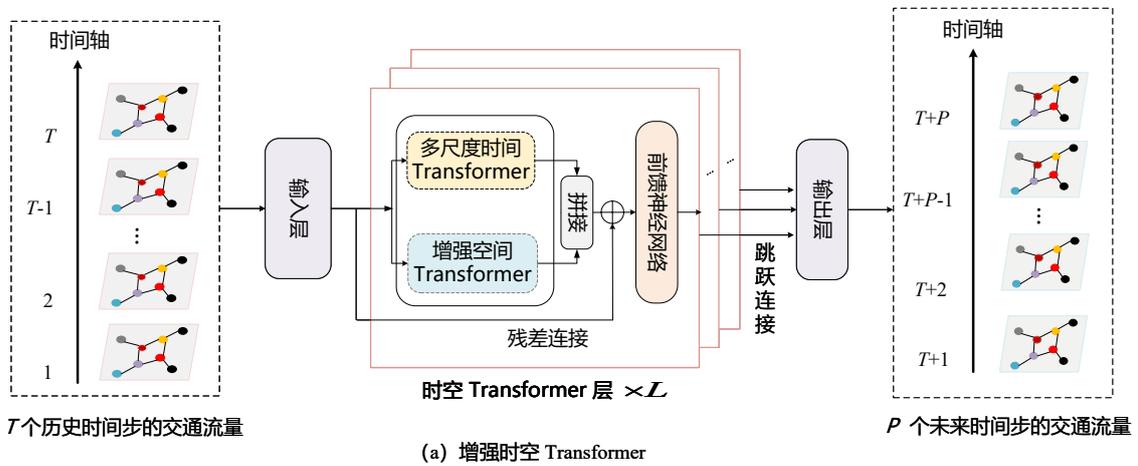


图1 增强时空 Transformer 模型总框架

2.1 多尺度时间 Transformer

路网中各交叉口未来的交通流量不仅受短期历史交通流的影响, 也与长期历史交通流存在明显的关联关系. 例如, 某区域早高峰 7:30-8:00 时段出现的交通拥堵, 不仅受其相邻区域 7:15-7:30 时段的交通流影响, 可能也受其他区域 6:30-7:30 时段的交通流影响. 考虑到传统 Transformer 中的时间多头自注意力机制虽然可以有效捕获交通流数据中的全局时间相关性时, 但因忽略了局部时间领域的上下文信

息导致难以有效捕获其局部时间相关性. 因此, 为同时捕获路网交通流数据中的短期和长期时间关联关系, 本文设计了一个多尺度时间 Transformer (MSTformer) 模型分层提取交通数据之间的短期和长期时间关联关系. 如图 1(b) 所示, 首先构建一个短期门控卷积网络捕获路网交通流数据的短期时间关联关系. 然后, 利用时间 Transformer 多头注意力机制捕获交通流数据的长期时间关联关系.

短期门控卷积网络使用具有更小感受野的二维

卷积捕获输入交通流时间序列数据中若干相邻时间步之间的关联关系, 并不触及较远历史时间步, 使模型聚焦于捕获短期时间相关性. 使用门控机制进一步筛选和强化短期关键特征, 并通过层归一化操作确保模型训练的稳定性. 然后, 使用残差连接学习交通流数据中包含的复杂非线性关系, 并保留输入数据中的原始特征. 给定输入时间序列数据 $\tilde{\mathbf{X}}$, 通过短期门控卷积网络后的输出如下式 (5)-(6) 所示:

$$\mathbf{Y}_0 = \varphi(\mathbf{C}_1(\tilde{\mathbf{X}}) + \mathbf{b}_1) \odot \theta(\mathbf{C}_2(\tilde{\mathbf{X}}) + \mathbf{b}_2), \quad (5)$$

$$\hat{\mathbf{Y}}_0 = \text{LayerNorm}(\mathbf{Y}_0 + \tilde{\mathbf{X}}). \quad (6)$$

式中: $\tilde{\mathbf{X}}$ 为输入数据, $\hat{\mathbf{Y}}_0$ 为短期门控卷积网络的输出; φ 为 sigmoid 激活函数; θ 为双曲正切 tanh 激活函数, C_1, C_2 表示二维卷积操作; b_1, b_2 表示可学习的偏置矩阵, \odot 表示哈达玛积; LayerNorm 表示层归一化操作.

在使用短期门控卷积网络提取路网交通流数据的短期时间关联关系后, 利用时间 Transformer 多头注意力机制进一步挖掘交通流数据的长期时间关联关系. 时间 Transformer 的多头自注意力机制通过计算查询向量 (Q) 和键向量 (K) 的相似度, 可以捕获交通流序列数据中任意两个时间步的关联关系. 首先, 将短期门控卷积网络的输出 $\hat{\mathbf{Y}}_0$ 映射到三个高维子空间, 得到一个三元组 $(\mathbf{Q}_i^T, \mathbf{K}_i^T, \mathbf{V}_i^T)$, 三元组中各变量如式 (7) 所示:

$$\begin{aligned} \mathbf{Q}_i^T &= \hat{\mathbf{Y}}_0 \mathbf{W}_q^T \\ \mathbf{K}_i^T &= \hat{\mathbf{Y}}_0 \mathbf{W}_k^T \\ \mathbf{V}_i^T &= \hat{\mathbf{Y}}_0 \mathbf{W}_v^T \end{aligned} \quad (7)$$

式中: $\mathbf{Q}_i^T, \mathbf{K}_i^T, \mathbf{V}_i^T$ 分别表示时间 Transformer 第 i 个头的查询向量, 键向量和值向量, $\mathbf{W}_q^T, \mathbf{W}_k^T, \mathbf{W}_v^T$ 为可学习的权重矩阵. 然后, 利用式 (8)-(9) 可计算出多尺度时间 Transformer 第 i 个头的输出 $head_i^T$:

$$h_i^T = \text{Softmax}\left(\frac{\mathbf{Q}_i^T \mathbf{K}_i^{T\top}}{\sqrt{d_k}}\right), \quad (8)$$

$$head_i^T = h_i^T \mathbf{V}_i^T. \quad (9)$$

式中: h_i^T 为第 i 个头的自注意力权重, Softmax 为激活函数, $\sqrt{d_k}$ 为比例因子. 如式 (10) 所示, 对各注意力头的输出 $head_i^T$ 进行加权求和可得第 i 层多尺度时间 Transformer 的输出 $\Theta_{\text{MSTformer}}^i$:

$$\Theta_{\text{MSTformer}}^i = [\{head_1^T; \dots; head_i^T; \dots\}] \mathbf{W}_i^T. \quad (10)$$

式中: \mathbf{W}_i^T 为可学习的时间权重矩阵; $\{;\}$ 表示拼接操作.

2.2 增强空间 Transformer

受限于路网空间结构和车辆行驶速度, 交通流

在路网中不同交通节点之间传递时存在一定的滞后性, 且不同时段滞后的时间长短并不相同, 对相邻节点的影响也不同. 例如, 上游交通节点在某个时间段发生交通拥堵, 经过若干时间步后可能会传递到下游交通节点. 但上游交通节点在高峰时间段发生交通拥堵, 对下游交通节点的影响, 相比于其在平峰时段发生交通拥堵对下游交通节点的影响并不相同. 也即, 不同时间段不同交通节点间的空间关联关系对交通流的影响并不相同. 因此, 为有效建模路网中不同交通节点之间动态变化的空间关联关系, 本文提出一个增强空间 Transformer (ESformer) 模型在多个子空间中同时捕获节点特征与边特征. 如图 1(c) 所示, 多头注意力机制将输入数据 $\tilde{\mathbf{X}}$ 映射到三个高维子空间, 得到一个三元组 $(\mathbf{Q}_i^S, \mathbf{K}_i^S, \mathbf{V}_i^S)$. 三元组中各变量如式 (11) 所示:

$$\begin{aligned} \mathbf{Q}_i^S &= \tilde{\mathbf{X}} \mathbf{W}_q^S \\ \mathbf{K}_i^S &= \tilde{\mathbf{X}} \mathbf{W}_k^S \\ \mathbf{V}_i^S &= \tilde{\mathbf{X}} \mathbf{W}_v^S \end{aligned} \quad (11)$$

式中: $\mathbf{Q}_i^S, \mathbf{K}_i^S, \mathbf{V}_i^S$ 分别代表空间 Transformer 的第 i 个头的查询向量, 键向量和值向量, $\mathbf{W}_q^S, \mathbf{W}_k^S, \mathbf{W}_v^S$ 是可学习的权重矩阵. 标准空间 Transformer 模型中, 键向量仅基于节点特征生成, 忽略了边 (如节点间距、流量方向) 的动态与静态特征. 受双动态时空图卷积网络^[25]启发, 将对偶变换用于键向量. 如式 (12) 所示, 利用 1.2 节定义的超图关联矩阵对键向量 \mathbf{K}_i^S 进行加权与拼接可增强其特征表达能力.

$$\tilde{\mathbf{K}}_i^S = \{[(\omega_4 \odot \mathbf{H}_{\text{sn}}) \mathbf{K}_i^S \oplus (\omega_5 \odot \mathbf{H}_{\text{dn}}) \mathbf{K}_i^S]; \mathbf{A}_{\text{dis}}\}. \quad (12)$$

其中, ω_4 和 ω_5 表示可学习权重矩阵, \oplus 表示求和操作, $\{;\}$ 表示矩阵拼接操作, $\mathbf{A}_{\text{dis}} \in \mathbb{R}^{N \times N}$ 表示根据路网中不同节点之间的距离构建的邻接矩阵, 代表有向边的静态特征.

为使模型在捕获路网交通节点之间的空间相关性时重点关注中心节点附近的节点, 以及虽然距离中心节点较远但表现出类似功能的节点, 设计一个参数化的时变掩码矩阵来动态更新键向量的超节点特征. 如式 (13) 所示, 设计一个全局节点嵌入向量 ξ 生成自适应空间掩码矩阵 ξ_{mask} .

$$\xi_{\text{mask}} = \text{Softmax}(\xi \odot \xi^T). \quad (13)$$

式中: Softmax 为激活函数. 为使具有相似流量模式的超节点之间信息可以相互传播, 将时变掩码矩阵用于超节点之间相似性度量, 具体如式 (14) 所示:

$$\bar{\mathbf{K}}_i^S = \tilde{\mathbf{K}}_i^S \odot \xi_{\text{mask}}. \quad (14)$$

式中: $\bar{\mathbf{K}}_i^S$ 为同时包含边特征与节点特征的键向量, 它融合了边的动态特征与静态特征, 用于提升注意力

权重分配的准确性. 为保证增强后的键向量输入维度与查询向量一致, 利用式 (15) 的反对偶变换, 将增强后的超节点特征映射回原始节点特征空间.

$$\hat{\mathbf{K}}_i^S = (\boldsymbol{\omega}_6 \odot \mathbf{H}) \mathbf{K}_i^S. \quad (15)$$

式中: $\boldsymbol{\omega}_6$ 表示可学习权重矩阵, \mathbf{H} 表示交通路网图对应的超图关联矩阵. 然后, 利用式 (16)-(17) 可计算出空间 Transformer 第 i 个头的输出 $head_i^S$:

$$h_i^S = \text{Softmax}\left(\frac{\mathbf{Q}_i^S \hat{\mathbf{K}}_i^S}{\sqrt{d_k}}\right), \quad (16)$$

$$head_i^S = h_i^S \mathbf{V}_i^S. \quad (17)$$

式中: h_i^S 为注意力权重, $\sqrt{d_k}$ 为预设的比例因子. 如式 (18) 所示, 对各个注意力头的输出 $head_i^S$ 进行加权求和可得第 i 层增强空间 Transformer 的输出 $\Theta_{\text{ESformer}}^i$:

$$\Theta_{\text{ESformer}}^i = [\{head_1^S; \dots; head_i^S; \dots\}] \mathbf{W}_i^S. \quad (18)$$

式中: \mathbf{W}_i^S 是可学习的空间权重矩阵, $\{\}; \}$ 表示矩阵拼接操作. 最后, 将多尺度时间 Transformer 的输出与增强空间 Transformer 的输出通过式 (19) 进行拼接, 得到第 i 层时空 Transformer 的输出 $\bar{\mathbf{Y}}_i$.

$$\bar{\mathbf{Y}}_i = \{\Theta_{\text{MSTformer}}^i; \Theta_{\text{ESformer}}^i\} \mathbf{W}_i. \quad (19)$$

式中: \mathbf{W}_i 表示可学习的权重矩阵.

2.3 输入层和输出层

输入层为卷积神经网络, 主要是用来将输入流量数据 $\tilde{\mathbf{X}}_T$ 转换到更高维度空间, 如式 (20) 所示.

$$\hat{\mathbf{X}} = C_3(\tilde{\mathbf{X}}_T). \quad (20)$$

式中: C_3 为卷积操作. 此外, 考虑到交通流数据呈现出明显的周期性变化规律, 在原始输入数据的时间维度上引入日周期嵌入和周周期嵌入, 分别提取日周期与周周期特征, 提高 ESTformer 捕获周期性特征的能力.

输出层为 2 个卷积网络层, 如式 (21)-(22) 所示, 将每一层时空 Transformer 的输出 $\bar{\mathbf{Y}}_i$ 执行跳跃连接并融合得到深层时空特征 $\hat{\mathbf{Y}}_{\text{hid}}$. 然后, 对融合后的时空特征进行卷积操作获得模型最终输出:

$$\hat{\mathbf{Y}}_{\text{hid}} = \sum_{i=1}^L FC_i(\bar{\mathbf{Y}}_i), \quad (21)$$

$$\tilde{\mathbf{Y}} = C_5(C_4(\hat{\mathbf{Y}}_{\text{hid}})). \quad (22)$$

式中: L 表示时空 Transformer 的总层数, FC_i 表示第 i 层的全连接操作, C_4 和 C_5 均表示卷积操作, $\tilde{\mathbf{Y}}$ 表示模型输出.

2.4 损失函数

本文选取对异常值具有更好鲁棒性的 Huber 函

数作为损失函数, 其定义如下式 (23) 所示. 式中: \mathbf{Y} 是交通流量的真实值, $\tilde{\mathbf{Y}}$ 是模型对未来交通流量的预测值. ϖ 是一个超参数, 用于控制 Huber 损失函数在异常值处的公式选择.

$$L(\mathbf{Y}, \tilde{\mathbf{Y}}) = \begin{cases} \frac{1}{2}(\mathbf{Y} - \tilde{\mathbf{Y}})^2, & |\mathbf{Y} - \tilde{\mathbf{Y}}| \leq \varpi \\ \varpi|\mathbf{Y} - \tilde{\mathbf{Y}}| - \frac{1}{2}\varpi^2, & |\mathbf{Y} - \tilde{\mathbf{Y}}| > \varpi \end{cases} \quad (23)$$

本文所提基于增强时空 Transformer (ESTformer) 的交通流预测算法, 其训练流程如算法 1 所示.

算法 1 基于 ESTformer 的交通流预测算法

输入: 历史交通流量数据 $\tilde{\mathbf{X}}_T$, 交通拓扑图 G , 训练轮数为 N_0 , ESTformer 网络层数为 L .

输出: 学习的 ESTformer 网络.

1. 初始化 ESTformer 网络训练参数 θ ;
2. 对偶变换生成超图关联矩阵 \mathbf{H} ;
3. 构造训练集 $\mathbf{D}_{\text{train}}$ 和验证集 $\mathbf{D}_{\text{valid}}$;
4. for epoch=0; epoch < N_0 ; epoch++ do
5. 随机选取一个输入样本 $\tilde{\mathbf{X}}_T$;
6. 经过日周期嵌入、周周期嵌入与卷积操作获得高维输入特征 $\hat{\mathbf{X}}$;
7. for $l=0$; $l < L$; $l++$ do
8. 利用第 l 层多尺度时间 Transformer 获得时间特征 $\Theta_{\text{MSTformer}}^l$;
9. 利用第 l 层增强空间 Transformer 获得空间特征 $\Theta_{\text{ESformer}}^l$;
10. 利用公式(19)获得经过时空 Transformer 融合后的时空特征 $\bar{\mathbf{Y}}_l$;
11. end for
12. 利用公式(22)获得预测值 $\tilde{\mathbf{Y}}$;
13. 利用预测值 $\tilde{\mathbf{Y}}$ 和真实值 \mathbf{Y} 通过公式(23)计算损失 $L(\mathbf{Y}, \tilde{\mathbf{Y}})$;
14. 利用梯度下降法更新模型可学习参数 θ ;
15. end for

3 实验与结果分析

为验证本文所提基于增强时空 Transformer (ESTformer) 的交通流预测方法的有效性, 选取 13 种基线方法作为对照组. 另外, 为与基线模型的测试条件保持一致, 选取美国加利福尼亚交通运输部公开的高速公路数据集 PEMS03、PEMS04、PEMS07 和 PEMS08 作为测试数据进行测试实验. 最后, 设计一系列的消融实验, 对所提增强时空 ESTformer 模型中不同模块的有效性进行了分析和评估.

3.1 实验数据

本次实验使用的 4 个测试数据集 PEMS03、

PEMS04, PEMS07, PEMS08 由 39000 多个传感器采集得到, 每个传感器 30 秒采集一次数据. 如表 1 所示, 上述 4 个测试数据集已按 5 分钟为间隔对原始采样数据进行了整合. 为确保测试实验的准确性, 本文将训练集、验证集、测试集的比例设置为 6:2:2. 实验测试过程中, 以 5 分钟为时间步长, 使用过去 1 小时 (12 个时间步) 的交通流量数据来预测未来 1 小时内的交通流量.

表1 测试数据集

| 数据集 | 节点 | 时间步 | 时间范围 |
|--------|-----|-------|-----------------------|
| PEMS03 | 358 | 26208 | 2018/09/01-2018/11/30 |
| PEMS04 | 307 | 16992 | 2018/01/01-2018/02/28 |
| PEMS07 | 883 | 28224 | 2017/05/01-2017/08/31 |
| PEMS08 | 170 | 17856 | 2016/07/01-2016/08/31 |

3.2 实验设置及评价指标

本实验在处理器为 13th Gen Intel (R) Core (TM) i7-13700K, GPU 为 NVIDIA GeForce RTX 3090, 安装 Ubuntu23.02 系统的台式机上, 训练过程中模型的参数设置如表 2 所示. 采用式 (24)-(25) 定义的平均绝对误差 (mean absolute error, MAE) 和均方根误差 (root mean square error, RMSE) 作为评价指标评估模型的预测性能.

$$\text{RMSE} = \sqrt{\frac{1}{N * T} \sum_{n=1}^N \sum_{t=1}^T (\mathbf{Y}_t^n - \tilde{\mathbf{Y}}_t^n)^2}, \quad (24)$$

$$\text{MAE} = \frac{1}{N * T} \sum_{n=1}^N \sum_{t=1}^T |\mathbf{Y}_t^n - \tilde{\mathbf{Y}}_t^n|. \quad (25)$$

3.3 与基线方法的对比实验

为评估所提模型 ESTformer 的预测性能, 选取 HA^[5]、LSTM^[9]、GRU^[10]、STGCN^[13]、DCRNN^[14]、

表2 实验参数

| 数据集 | PEMS03 | PEMS04 | PEMS07 | PEMS08 |
|----------------|--------|--------|--------|--------|
| 学习率 | 0.01 | 0.01 | 0.01 | 0.01 |
| 批处理大小 | 8 | 8 | 8 | 8 |
| 优化器 | Adam | Adam | Adam | Adam |
| 模型嵌入维度 | 64 | 64 | 64 | 64 |
| 多头注意力头数 | 4 | 4 | 4 | 4 |
| Transformer的层数 | 6 | 6 | 6 | 6 |

AGCRN^[15]、GWN^[17]、ASTGCN^[18]、AFDGCN^[22]、Tformer^[23]、DDSTGCN^[25]、STSGCN^[26] 和 BiSTAT^[27] 等 13 种方法作为基线方法进行对比实验.

表 3-表 4 展示了不同预测模型在四个测试数据集上进行 5 分钟 (1 个时间步长)、15 分钟 (3 个时间步长)、30 分钟 (6 个时间步长)、45 分钟 (9 个时间步长)、60 分钟 (12 个时间步长) 以及 60 分钟内的平均预测性能. 值得注意的是, 实际测试时, 每次均对 12 个不同预测步长进行测试. 虽然表 3-表 4 仅展示了 5 个不同预测步长下的预测结果, 但表中最后 2 列的平均值为 12 个不同预测步长下预测结果的平均值. 其中粗体表示最佳预测结果, 下划线的表示次优结果. 不难发现, 在 4 个测试数据集上, 所提方法 ESTformer 的平均预测性能整体上优于其他所有 13 种基线方法. 也即, 不论是短期预测 (5 分钟, 15 分钟, 30 分钟), 还是长期预测 (45 分钟, 60 分钟), 所提方法 ESTformer 均具有更好的预测性能.

进一步分析, 不难发现在测试数据集 PEMS03 上, 虽然以 5 分钟作为预测步长时, 所提方法 ESTformer 的预测性能稍低于基线方法 DDSTGCN, 但相较于所有基线方法中表现最好的 DDSTGCN, 所提 ESTformer 方法在 12 个不同预测步长上的平均预测指标 MAE 和 RMSE 分别改进了 1.14% 和

表3 本文方法与基线方法在数据集 PEMS03 和 PEMS04 上的预测性能比较

| 数据集 | 模型 | 5分钟 | | 15分钟 | | 30分钟 | | 45分钟 | | 60分钟 | | 平均值 | |
|-------------|---------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|-------|
| | | MAE | RMSE | MAE | RMSE |
| PEMS03 | HA | 19.98 | 30.64 | 23.57 | 36.48 | 29.13 | 44.98 | 34.60 | 53.21 | 40.10 | 61.15 | 30.03 | 47.17 |
| | LSTM | 13.75 | 21.52 | 16.68 | 26.11 | 19.96 | 30.79 | 23.49 | 35.43 | 27.31 | 40.31 | 20.59 | 31.41 |
| | GRU | 13.68 | 21.26 | 16.56 | 25.68 | 19.94 | 30.39 | 23.59 | 35.11 | 27.50 | 40.11 | 20.62 | 31.66 |
| | DCRNN | 13.01 | 21.62 | 15.45 | 25.75 | 17.98 | 29.80 | 20.46 | 33.36 | 23.12 | 37.18 | 18.31 | 30.07 |
| | STGCN | 17.55 | 30.04 | 18.45 | 31.26 | 19.54 | 32.58 | 20.43 | 33.78 | 21.70 | 35.29 | 19.63 | 32.72 |
| | GWN | 13.36 | 22.77 | 16.06 | 27.37 | 19.25 | 32.29 | 22.57 | 36.96 | 26.12 | 41.89 | 19.81 | 32.88 |
| | ASTGCN | 13.28 | 22.10 | 15.24 | 25.25 | 16.71 | 27.64 | 18.31 | 30.13 | 20.25 | 33.04 | 16.96 | 28.15 |
| | AGCRN | 14.31 | 25.10 | 14.61 | 25.84 | 15.65 | 27.51 | 16.33 | 28.68 | 17.42 | 30.23 | 15.70 | 27.63 |
| | STSGCN | 14.52 | 24.69 | 16.40 | 27.69 | 18.05 | 30.31 | 19.61 | 32.63 | 21.27 | 35.13 | 18.20 | 30.46 |
| | BiSTAT | 13.12 | 22.27 | 14.18 | 24.22 | 15.43 | 26.50 | 16.47 | 28.32 | 17.51 | 30.04 | 15.49 | 26.55 |
| | Tformer | 13.48 | 21.64 | 14.90 | 24.07 | 16.25 | 26.10 | 16.87 | 27.05 | 17.90 | 28.36 | 16.07 | 25.87 |
| | DDSTGCN | 12.33 | 19.66 | | | | | | | | 16.86 | | |
| | AFDGCN | 13.02 | 22.53 | 14.12 | 24.64 | 15.16 | 26.74 | 15.97 | 28.23 | | 29.23 | 15.14 | 26.66 |
| Ours | | | 13.47 | 22.63 | 14.69 | 24.79 | 15.78 | 26.89 | 16.68 | 27.51 | 14.75 | 25.01 | |

表3(续)

| 数据集 | 模型 | 5分钟 | | 15分钟 | | 30分钟 | | 45分钟 | | 60分钟 | | 平均值 | |
|--------|-------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | | MAE | RMSE |
| PEMS04 | HA | 26.18 | 38.78 | 30.70 | 45.01 | 37.47 | 54.36 | 44.20 | 63.62 | 50.84 | 72.66 | 38.56 | 55.83 |
| | LSTM | 18.78 | 28.36 | 22.06 | 32.83 | 26.12 | 38.33 | 30.34 | 43.92 | 35.12 | 28.34 | 26.90 | 39.86 |
| | GRU | 18.56 | 28.17 | 21.71 | 32.54 | 25.79 | 38.00 | 30.15 | 43.54 | 34.89 | 39.52 | 26.63 | 39.52 |
| | DCRNN | 17.76 | 28.20 | 20.64 | 32.24 | 23.90 | 36.82 | 27.29 | 41.53 | 30.84 | 46.37 | 24.44 | 37.53 |
| | STGCN | 21.59 | 32.97 | 23.44 | 35.31 | 25.45 | 37.84 | 27.70 | 40.72 | 30.45 | 44.13 | 25.91 | 38.44 |
| | GWN | 18.08 | 28.92 | 20.91 | 33.05 | 24.52 | 38.29 | 28.36 | 43.49 | 32.55 | 49.15 | 25.24 | 18.21 |
| | ASTGCN | 17.81 | 28.46 | 19.57 | 31.04 | 21.20 | 33.45 | 23.08 | 36.10 | 25.57 | 39.46 | 21.62 | 34.13 |
| | AGCRN | 18.57 | 30.19 | 18.89 | 30.93 | 19.66 | 32.30 | 20.24 | 33.44 | 21.35 | 34.96 | 19.76 | 32.45 |
| | STSGCN | 18.12 | 29.11 | 19.80 | 31.58 | 21.30 | 33.83 | 22.86 | 36.15 | 24.47 | 38.46 | 21.51 | 34.13 |
| | BiSTAT | 17.62 | 28.17 | 18.34 | 29.46 | 19.21 | 30.81 | 20.00 | 31.96 | 20.98 | 33.30 | 19.32 | 30.90 |
| | Tformer | 17.82 | 27.19 | 18.96 | | 19.82 | | 20.72 | | 21.68 | | 19.92 | |
| | DDSTGCN | | 27.30 | 18.40 | 29.37 | 19.64 | 31.16 | 20.69 | 32.53 | 21.69 | 33.86 | 19.66 | 31.11 |
| | AFDGCN | 17.44 | 28.12 | | 29.72 | | 31.04 | | 32.17 | | 33.29 | | 31.11 |
| | Ours | 16.97 | | 17.84 | 28.92 | 18.71 | 30.37 | 19.49 | 31.54 | 20.12 | 32.45 | 18.74 | 30.38 |

表4 本文方法与基线方法在测试数据集 PEMS07 和 PEMS08 上的预测性能比较

| 数据集 | 模型 | 5分钟 | | 15分钟 | | 30分钟 | | 45分钟 | | 60分钟 | | 平均值 | |
|--------|-------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | | MAE | RMSE |
| PEMS07 | HA | 29.95 | 43.87 | 35.67 | 52.06 | 44.06 | 64.01 | 52.32 | 75.69 | 60.44 | 87.04 | 45.35 | 65.78 |
| | LSTM | 19.02 | 32.46 | 23.54 | 35.26 | 28.55 | 41.94 | 33.40 | 48.34 | 38.87 | 55.44 | 29.23 | 43.57 |
| | GRU | 19.03 | 28.94 | 23.62 | 35.39 | 28.38 | 42.04 | 33.37 | 48.48 | 39.07 | 55.64 | 29.25 | 43.70 |
| | DCRNN | 17.94 | 27.95 | 21.97 | 33.94 | 25.68 | 39.16 | 29.38 | 43.95 | 33.29 | 49.09 | 26.14 | 39.56 |
| | STGCN | 31.64 | 50.66 | 32.52 | 51.71 | 33.26 | 52.74 | 34.38 | 53.96 | 36.01 | 55.63 | 33.62 | 53.03 |
| | GWN | 18.07 | 28.82 | 21.71 | 34.77 | 25.89 | 41.04 | 30.11 | 47.08 | 34.82 | 53.72 | 26.58 | 41.84 |
| | ASTGCN | 19.35 | 30.07 | 23.94 | 36.88 | 28.77 | 43.89 | 33.76 | 50.56 | 39.37 | 58.07 | 29.59 | 45.52 |
| | AGCRN | 19.10 | 30.93 | 19.89 | 32.68 | 21.29 | 34.91 | 22.09 | 36.48 | 23.51 | 38.45 | 21.26 | 34.99 |
| | STSGCN | 19.08 | 31.22 | 22.02 | 35.78 | 24.72 | 40.66 | 27.13 | 44.63 | 29.82 | 48.81 | 24.92 | 40.84 |
| | BiSTAT | 17.75 | 28.43 | 19.19 | 31.29 | 20.62 | 33.80 | 21.76 | 35.63 | 22.93 | 37.29 | 20.64 | 33.69 |
| | Tformer | 17.73 | | 19.40 | | 20.67 | | 21.86 | | 23.06 | | 20.75 | |
| | DDSTGCN | | 27.96 | 19.72 | 31.83 | 21.57 | 34.82 | 23.04 | 37.00 | 24.49 | 39.02 | 21.55 | 34.64 |
| | AFDGCN | 17.65 | 28.29 | | 31.47 | | 34.04 | | 35.86 | | 37.41 | | 33.97 |
| | Ours | 16.83 | 27.07 | 18.28 | 30.25 | 19.61 | 32.82 | 20.59 | 34.59 | 21.58 | 36.08 | 19.58 | 32.73 |
| PEMS08 | HA | 21.16 | 31.33 | 25.15 | 37.00 | 31.09 | 45.32 | 37.02 | 53.46 | 42.92 | 61.42 | 32.06 | 47.52 |
| | LSTM | 14.42 | 21.49 | 17.39 | 25.94 | 20.81 | 30.89 | 24.56 | 35.59 | 28.69 | 40.74 | 21.53 | 32.06 |
| | GRU | 14.66 | 21.53 | 17.41 | 25.78 | 20.94 | 30.68 | 24.51 | 35.33 | 28.47 | 40.35 | 21.55 | 31.28 |
| | DCRNN | 13.58 | 20.77 | 15.95 | 24.64 | 18.32 | 28.44 | 20.66 | 31.81 | 23.04 | 35.07 | 18.60 | 28.65 |
| | STGCN | 19.22 | 27.92 | 20.53 | 29.92 | 22.21 | 32.24 | 24.05 | 34.62 | 26.40 | 37.36 | 22.61 | 32.65 |
| | GWN | 13.24 | 21.06 | 15.58 | 25.27 | 18.28 | 30.01 | 21.07 | 34.38 | 24.18 | 38.85 | 18.74 | 30.46 |
| | ASTGCN | 15.35 | 23.28 | 17.12 | 26.19 | 19.11 | 29.18 | 21.03 | 31.91 | 23.47 | 35.22 | 19.42 | 29.72 |
| | AGCRN | 14.46 | 22.40 | 14.92 | 23.42 | 15.92 | 25.21 | 16.72 | 26.64 | 17.86 | 28.34 | 16.02 | 25.31 |
| | STSGCN | 14.84 | 22.80 | 16.18 | 25.02 | 17.38 | 27.15 | 18.45 | 28.81 | 19.54 | 30.34 | 17.45 | 27.15 |
| | BiSTAT | 13.55 | 21.26 | | 22.70 | | 24.21 | | 25.51 | 16.65 | 26.98 | 14.98 | 24.31 |
| | Tformer | 13.99 | 20.87 | 15.01 | 22.70 | 15.64 | | 16.40 | | 17.34 | | 15.78 | |
| | DDSTGCN | | | 14.26 | | 15.41 | 24.37 | 16.45 | 26.02 | 17.38 | 27.30 | 15.47 | 24.35 |
| | AFDGCN | 13.33 | 20.81 | 14.19 | 22.65 | 14.96 | 24.38 | 15.65 | 25.58 | | 26.80 | | 24.36 |
| | Ours | 12.81 | 19.91 | 13.93 | 21.99 | 14.73 | 23.65 | 15.61 | 24.99 | 16.33 | 26.09 | 14.83 | 23.70 |

0.36%。在测试数据集 PEMS04 上,相较于所有基线方法中表现最好的 AFDGCN 和 Tformer,所提方法 ESTformer 在 12 个不同预测步长上的平均预测指标 MAE 和 RMSE 分别改进了 1.57% 和 0.52%。在数据集 PEMS07 上,相较于所有基线方法中表现最

好的 AFDGCN 和 Tformer,所提方法 ESTformer 在 12 个不同预测步长上的 MAE 和 RMSE 分别改进了 3.88% 和 1.65%。在测试数据集 PEMS08 上,相较于所有基线方法中表现最好的 AFDGCN 和 Tformer,所提方法 ESTformer 在 12 个不同预测步长上的

MAE 和 RMSE 分别改进了 1.26% 和 1.78%。这表明, 相比于基线方法, 所提基于 ESTformer 的交通流预测方法可以学习到路网中不同节点之间更多隐藏的时空相关性, 具有更强的交通流预测能力。

另外, 实验结果表明, 传统方法 HA 的预测性能远比不上基于深度学习的交通流预测方法。基于 LSTM 或 GRU 的方法因仅考虑交通流量数据中的时间特征忽略了空间依赖性, 导致整体预测效果不佳。DCRNN 和 AGCRN 是基于循环神经网络 RNN 的交通流预测方法, 但因 RNN 在处理长时间序列数据时容易出现梯度消失或爆炸问题, 导致当预测时间步长较大时, 难以有效捕捉交通流量数据中的长期时间关联关系。STGCN 和 GWN 是基于卷积神经网络 CNN 的交通流预测方法, 分别使用 CNN 和 TCN 对时间相关性进行建模, 但由于使用较小感受野的卷积核更擅长于捕获短期时间相关性导致它们很难做出准确的长期预测。与上述方法相比, 本文所提基于增强时空 Transformer 的交通流预测方法具有更优的预测性能, 其主要原因是所提方法同时融合了短期门控卷积神经网络捕获短期时间关联关系和时间 Transformer 模型捕获长期时间关联关系的优势。

由于考虑了周期性因素, 相比于其他基线方法, 加入了注意力机制的 DDSTGCN 在数据集 PEMS03 上具有较好预测效果。Tformer 则因使用全局编码器和全局-局部解码器提取全局和局部空间特征, 在 PEMS04、PEMS07 和 PEMS08 等数据集上表现优于其他基线方法。AFDGCN 通过引入自注意力机制增强模型捕获相邻节点之间特征, 在 PEMS07 和 PEMS08 数据集上具有优于其他基线方法的预测性能。不同于基于自注意力机制的基线方法 AFDGCN 和 Tformer 相比, 本文所提基于 ESTformer 的预测方法对空间 Transformer 的键向量进行偶变换, 使键向量具有更强的特征表达能力, 能够同时编码节点属性与边属性, 提升了注意力权重分配的准确性。例如, 若两节点距离较远 (静态特征弱), 但流量变化高度相似 (动态特征强), 模型仍会为其分配高注意力权重。若两节点距离近且流量变化高度相似, 则权重进一步强化, 形成局部-全局特征的平衡。

3.4 消融实验

为了更好地评估所构建的 ESTformer 模型中不同模块的有效性, 将以下四种变体模型在数据集 PEMS04 和 PEMS08 上进行测试并对实验结果进行分析讨论。(1) w/o Dt(Dual transformer): 该变体模型

在空间多头自注意力机制中移除了对键向量的对偶变换操作;(2) w/o Mask: 该变体模型在空间多头自注意力机制中移除了自学习掩码矩阵;(3) w/o ESformer: 该变体模型在空间多头自注意力移除了对偶变换和自学习掩码矩阵;(4)w/o MSTformer: 该变体模型使用普通的时间 Transformer 代替多尺度时间 Transformer。各变体模型在 PEMS04 和 PEMS08 数据集上进行测试时的实验结果如表 5 所示。

表5 不同方法 60 分钟 (12 个预测步) 预测结果的平均值

| 模型 | PEMS04 | | PEMS08 | |
|---------------|--------------|--------------|--------------|--------------|
| | MAE | RMSE | MAE | RMSE |
| w/o Dt | 19.10 | 30.48 | 15.04 | 24.01 |
| w/o Mask | 19.36 | 31.27 | 15.21 | 24.84 |
| w/o ESformer | 20.95 | 33.62 | 14.99 | 24.05 |
| w/o MSTformer | 18.80 | 30.46 | 15.05 | 23.99 |
| ESTformer | 18.74 | 30.38 | 14.83 | 23.70 |

(1) w/o Dt: 移除对偶变换操作后, 仅使用掩码矩阵更新键向量, 导致在 2 个数据集上的误差评价指标 MAE 和 RMSE 均有所上升。在数据集 PEMS04 和 PEMS08 上, MAE 分别增加了 1.88% 和 1.39%, RMSE 分别增加了 0.33% 和 1.31%。

(2) w/o Mask: 移除自学习掩码矩阵后, 模型难以及时动态更新空间 Transformer 中键向量的超节点特征, 导致在数据集 PEMS04 和 PEMS08 上, MAE 分别增加了 3.31% 和 2.56%, RMSE 分别增加了 2.92% 和 4.81%。

(3)w/o ESformer: 仅使用普通空间多头自注意力提取空间依赖性时, 在数据集 PEMS04 和 PEMS08 上, MAE 分别增加了 5.82% 和 1.06%, RMSE 分别增加了 10.7% 和 1.48%。这表明相比普通空间 Transformer, 增强空间 Transformer 通过对键向量进行对偶变换和动态更新, 增强了键向量的特征表达能力, 有效提高了 ESTformer 模型的预测精度。

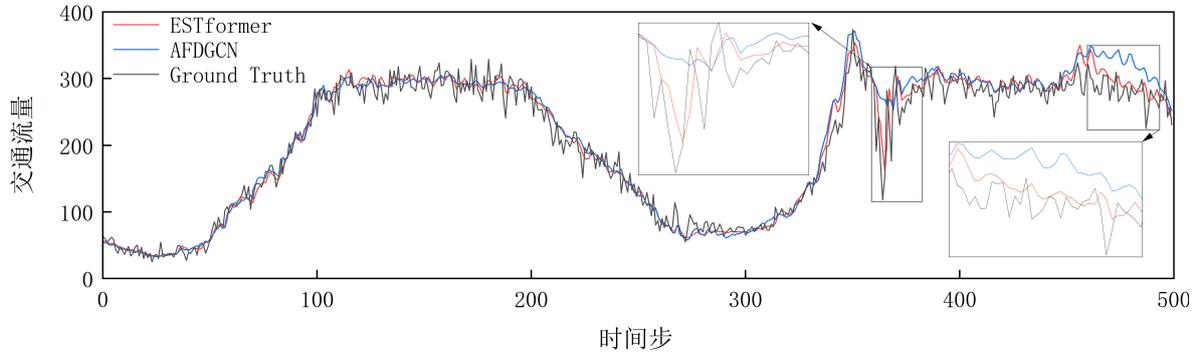
(4) w/o MSTformer: 仅使用普通时间 Transformer 提取时间相关性时, 该变体模型相较于 ESTformer, 在数据集 PEMS04 和 PEMS08 上, MAE 分别上升了 0.31% 和 1.46%, RMSE 分别上升了 0.26% 和 1.22%。这说明在时间 Transformer 前加入短期门控卷积 (STGC) 模块, 增强了 ESTformer 模型捕获交通流数据中短期时间关联关系的能力。

3.5 预测结果可视化

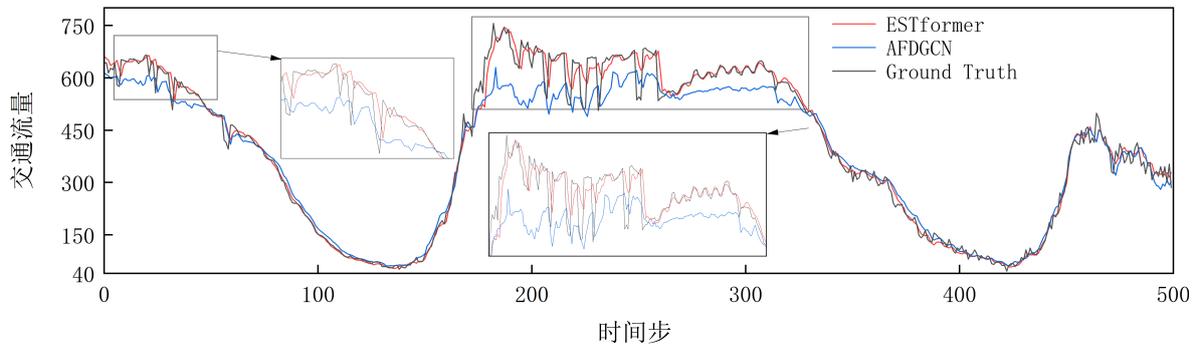
为进一步展示所提基于 ESTformer 的交通流预测方法的预测性能, 在数据集 PEMS03 和 PEMS07 上进行测试和流量可视化分析。如图 2 所示, 选择综

合预测效果最好的基线方法 AFDGCN 和所提基于 ESTformer 的预测方法进行流量预测,并展示了两种方法在 500 个时间步长的预测结果和真实交通流量数据的对比情况.从图 2(a)中可以看出,在 PEMS03 上,ESTformer 在流量高峰期的预测性能整体优于基线方法 AFDGCN.从图 2(b)可以看出,在 PEMS07

数据集上,ESTformer 的预测效果明显优于基线方法 AFDGCN,且在预测高峰期流量的预测值明显更接近真实流量.这展示了所提基于 ESTformer 的交通流预测方法在进行长时预测时的优势,并进一步凸显了所提 ESTformer 模型从交通流量数据中学习隐藏时空特征的能力.



(a) PEMS03 上第 205 个采样节点预测值与真实值曲线



(b) PEMS07 上第 334 个采样节点预测值与真实值曲线

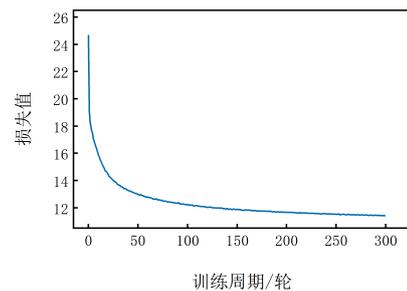
图2 ESTformer 和 AFDGCN 在不同数据集上的交通流量预测结果

3.6 收敛性与计算效率分析

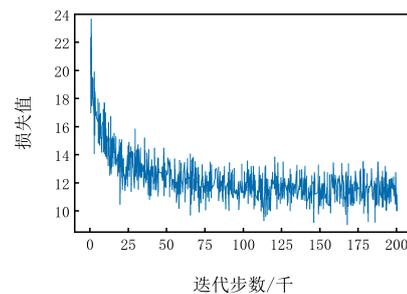
为了评估所提 ESTformer 模型在进行训练时模型的收敛能力,在 PEMS08 数据集上进行损失函数可视化分析.如图 3 所示,在 300 个周期的训练中,选择使用 Huber 损失函数计算每个训练步长的损失值.从图 3(a)中可以看 ESTformer 模型在前 50 个训练周期损失值快速下降,随后则缓慢下降.在 200 个训练周期后,损失值基本趋于平稳且维持在 12 左右.每个训练周期,模型会经历多次迭代.模型训练过程中损失值随迭代步数的变化趋势如图 3(b) 示.

可以直观的看到每次参数更新后损失值的变化.训练过程中前 2.5 万次迭代,损失快速下降,后续损失值在^[10,14]区间内波动.结合图 3(a) 训练周期平均损失的持续下降,说明局部波动未阻碍整体收敛,表明模型在迭代过程中持续学习,最终收敛到稳定状态,具备良好的收敛趋势.

为与基线方法 STSGCN 和 AFDGCN 进行公平比较,在 PEMS08 上进行实验时,使用相同的软硬件



(a) 损失函数值随训练周期变化趋势



(b) 损失函数值随迭代步数变化趋势

图3 ESTformer 在数据集 PEMS08 上的损失函数值随训练周期和迭代步数变化曲线

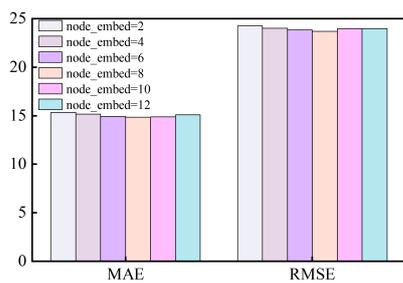
环境来进行模型效率分析. 表6展示了所提基于ESTformer的预测方法与两种基线方法每一轮的训练时间和推理时间以及内存开销等指标. 在PEMS08上, ESTformer每一轮的训练时间和所需内存略高于基线模型STSGCN和AFDGCN. 但ESTformer模型每一轮的推理时间却低于基线模型.

表6 在PEMS08上计算效率比较

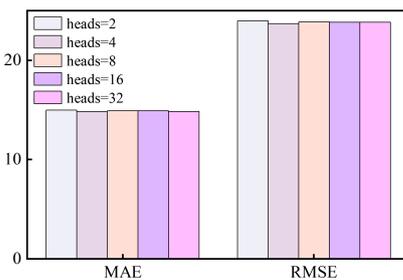
| 模型 | 训练时间 (s/轮) | 推理时间 (s) | 内存开销 (MB) |
|-----------|--------------|-------------|-------------|
| STSGCN | 45.76 | 8.12 | 2620 |
| AFDGCN | 50.29 | 7.32 | 3540 |
| ESTformer | 51.03 | 5.02 | 3640 |

3.7 超参数实验

(1) ESTformer模型中2个重要参数分别是时变掩码的节点嵌入大小和多头自注意力机制的头数. 时变掩码的节点嵌入大小不仅影响掩码质量, 而且决定了空间多头自注意力模块参数的多样性. 多头自注意力机制通过并行处理多个独立的注意力头, 能够同时关注交通流数据序列数据的不同部分, 从而捕获更丰富的时空特征. 图4(a)展示了预测时间区间选为60分钟时, 在数据集PEMS08上, 不同节点嵌入维度大小对ESTformer模型平均预测结果的影响. 可以看出, 在数据集PEMS08上模型的最佳节点嵌入维度为8. 适当增加节点嵌入大小可以使掩码矩阵中包含更多的参数信息, 提高模型捕获时空特征的能力. 但当节点嵌入数过大时, 参数数量过多, 模型容易出现过拟合, 导致模型预测精度有所下降.



(a) 不同节点嵌入维度的预测结果



(b) 不同注意力头数的预测结果

图4 在数据集PEMS08上ESTformer使用不同嵌入节点维度和不同注意力头数时的预测结果

(2) 图4(b)展示了不同注意力头数对ESTformer模型预测精度的影响. 可以看出, 由于不同的注意力头可以独立地学习到不同的注意力权重分配方式, 通过适当地增加“头”数有助于ESTformer模型更好地捕获交通流数据中的时空特征. 当时间和空间多头注意力“头”数均为2时, 模型的预测性能较差. 在“头”数为4时, 所提ESTformer模型的整体预测性能较好. 但当继续增加头数时, 模型的预测精度并没有得到明显提升, 且随着注意力“头”数的增加, 模型的计算复杂度也会相应增加.

4 结论

为有效捕获城市路网交通流序列数据中的复杂时空关联关系, 本文提出了一种基于增强时空Transformer的交通流预测方法, 用于预测路网中未来的交通流量状况. 在时间建模方面, 所提方法采用融合短期门控卷积和时间多头自注意力的多尺度时间Transformer增强模型分层提取交通流数据之间的短期和长期时间关联关系的能力. 在空间建模方面, 所提方法采用增强空间Transformer捕获路网中不同节点之间动态变化的全局空间关联关系. 为验证所提基于ESTformer的交通流预测方法有效性, 在4个真实高速公路交通流量数据集上进行了实验测试. 测试结果表明, 所提基于ESTformer的交通流预测方法预测性能优于最佳基线方法, 且在长时预测方面具有更明显的优势. 在未来的研究工作中, 拟建模天气、交通事故等因素对交通流量预测的影响, 并进一步优化所提模型的网络结构, 以便设计更精准的交通流预测方法, 提升路网交通状态感知能力.

参考文献 (References)

- [1] Ye B L, Wu P, Li L X, et al. Deep reinforcement learning based on spatiotemporal information for network-wide traffic signal coordination control[J]. *IEEE Open Journal of Intelligent Transportation Systems*, 2025, 6: 1424-1438.
- [2] Liu Y, Kulcsár B A, Wu J M, et al. Emerging AI-driven smart and sustainable mobility[J]. *Transportation Research — Part E: Logistics and Transportation Review*, 2025, 198: 104126.
- [3] Ye B L, Wu P, Li L X, et al. Multi-intersection traffic signal control based on dynamic spatiotemporal memory enhanced learning[J]. *Control Engineering Practice*, 2025, 165: 106606.
- [4] Ye B L, Zhang M J, Li L X, et al. A survey of traffic flow prediction methods based on long short-term memory networks[J]. *IEEE Intelligent Transportation Systems Magazine*, 2024, 16(5): 87-112.
- [5] Pan B, Demiryurek U, Shahabi C. Utilizing real-world transportation data for accurate traffic prediction[C]. *IEEE the 12th International Conference on Data Mining*.

- Brussels, 2012: 595-604.
- [6] 谷振宇, 陈聪, 郑家佳, 等. 基于时空图卷积循环神经网络的交通流预测[J]. *控制与决策*, 2022, 37(3): 645-653.
(Gu Z Y, Chen C, Zheng J J, et al. Traffic flow prediction based on STG-CRNN[J]. *Control and Decision*, 2022, 37(3): 645-653.)
- [7] 祁朵, 毛政元. 基于自适应时序剖分与KNN的短时交通流量预测[J]. *地球信息科学学报*, 2022, 24(2): 339-351.
(Qi D, Mao Z Y. Short-term traffic flow prediction based on adaptive time slice and KNN[J]. *Journal of Geo-Information Science*, 2022, 24(2): 339-351.)
- [8] 刘建伟, 宋志妍. 循环神经网络研究综述[J]. *控制与决策*, 2022, 37(11): 2753-2768.
(Liu J W, Song Z Y. Overview of recurrent neural networks[J]. *Control and Decision*, 2022, 37(11): 2753-2768.)
- [9] Ma X L, Tao Z M, Wang Y H, et al. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data[J]. *Transportation Research — Part C: Emerging Technologies*, 2015, 54: 187-197.
- [10] Dey R, Salem F M. Gate-variants of gated recurrent unit (GRU) neural networks[C]. *IEEE the 60th International Midwest Symposium on Circuits and Systems*. Boston, 2017: 1597-1600.
- [11] 刘金培, 罗瑞, 陈华友, 等. 基于多尺度3D-CNN-CBAM的空气质量指数时空预测[J]. *控制与决策*, 2025, 40(2): 404-412.
(Liu J P, Luo R, Chen H Y, et al. Spatio-temporal forecasting of air quality index based on multi-scale 3D-CNN-CBAM[J]. *Control and Decision*, 2025, 40(2): 404-412.)
- [12] 叶宝林, 戴本岙, 苗永超, 等. 基于自适应时空同步建模的交通流预测[J]. *计量学报*, 2025, 46(6): 802-812.
(Ye B L, Dai B A, Miao Y C, et al. Adaptive spatio-temporal synchronous modeling based traffic flow prediction[J]. *Acta Metrologica Sinica*, 2025, 46(6): 802-812.)
- [13] Yu B, Yin H T, Zhu Z X. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting[C]. *Proceedings of the 27th International Joint Conference on Artificial Intelligence*. Stockholm, 2018: 3634-3640.
- [14] Li Y, Yu R, Shahabi C, et al. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting[C]. *6th International Conference on Learning Representations*. Vancouver, Canada, 2018: 2079-2094.
- [15] Bai L, Yao L, Li C, et al. Adaptive graph convolutional recurrent network for traffic forecasting[J]. *Advances in neural information processing systems*, 2020, 33: 17804-17815.
- [16] Li F X, Feng J, Yan H, et al. Dynamic graph convolutional recurrent network for traffic prediction: Benchmark and solution[J]. *ACM Transactions on Knowledge Discovery from Data*, 2023, 17(1): 1-21.
- [17] Wu Z H, Pan S R, Long G D, et al. Graph WaveNet for deep spatial-temporal graph modeling[C]. *Proceedings of the 28th International Joint Conference on Artificial Intelligence*. Macao, 2019: 1907-1913.
- [18] Guo S N, Lin Y F, Feng N, et al. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting[C]. *Proceedings of the AAAI Conference on Artificial Intelligence*. Hawaii, 2019: 922-929.
- [19] Li M Z, Zhu Z X. Spatial-temporal fusion graph neural networks for traffic flow forecasting[c]. *Proceedings of the AAAI Conference on Artificial Intelligence*. Online, 2021: 4189-4196.
- [20] Wang X Y, Ma Y, Wang Y Q, et al. Traffic flow prediction via spatial temporal graph neural network[C]. *Proceedings of the Web Conference*. Taipei, 2020: 1082-1092.
- [21] Zheng C P, Fan X L, Wang C, et al. GMAN: A graph multi-attention network for traffic prediction[C]. *Proceedings of the AAAI Conference on Artificial Intelligence*. New York, 2020: 1234-1241.
- [22] Luo X L, Zhu C J, Zhang D T, et al. Dynamic graph convolutional network with attention fusion for traffic flow prediction[C]. *ECAI 2023*. Kraków, 2023: 1633-1640.
- [23] Yan H Y, Ma X L, Pu Z Y. Learning dynamic and hierarchical traffic spatiotemporal features with transformer[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 23(11): 22386-22399.
- [24] Hu T, Li S, Wang R, et al. DSTGTN: A Dynamic Spatio-Temporal Graph Transformer Network for Subway Flow Prediction[C]. *2025 5th International Conference on Artificial Intelligence and Industrial Technology Applications*. Xi'an China, 2025: 1914-1917.
- [25] Sun Y F, Jiang X H, Hu Y L, et al. Dual dynamic spatial-temporal graph convolution network for traffic prediction[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 23(12): 23680-23693.
- [26] Song C, Lin Y F, Guo S N, et al. Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting[C]. *Proceedings of the AAAI Conference on Artificial Intelligence*. New York, 2020: 914-921.
- [27] Chen C L, Liu Y B, Chen L, et al. Bidirectional spatial-temporal adaptive transformer for urban traffic flow forecasting[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2023, 34(10): 6913-6925.

作者简介

叶宝林 (1984–), 男, 副教授, 博士, 主要研究方向为深度学习、强化学习、多模态数据融合、复杂系统建模、优化与控制相关理论和应用, E-mail: yebaolin@zjxu.edu.cn;

苗永超 (2000–), 男, 硕士, 主要研究方向为交通流预测、智慧交通, E-mail: miaoyc0308@163.com;

王征 (2002–), 男, 硕士生, 主要研究方向为交通流预测、智慧交通, E-mail: wangzq202511@163.com.