

中文引用格式:周新民,徐天,李达,等. 基于自适应时空多头图注意力网络的交通流量预测模型[J]. 中国安全科学学报, 2025, 35(11):149-156.

英文引用格式:ZHOU Xinmin, XU Tian, LI Da, et al. Traffic flow prediction model based on adaptive spatio-temporal multi-head graph attention network[J]. China Safety Science Journal, 2025, 35(11):149-156.

# 基于自适应时空多头图注意力网络的 交通流量预测模型<sup>\*</sup>

周新民<sup>1,2</sup>教授, 徐天<sup>3</sup>, 李达<sup>3</sup>, 王隆鑫<sup>4</sup>, 胡江华<sup>3</sup>, 王伟<sup>4</sup>

(1 湖南工商大学 人工智能与先进计算学院, 湖南 长沙 410205; 2 湘江实验室, 湖南 长沙 410205;  
3 湖南工商大学 前沿交叉学院, 湖南 长沙 410205; 4 湖南工商大学 计算机学院, 湖南 长沙 410205)

中图分类号:X951

文献标志码:A

DOI: 10.16265/j.cnki.issn1003-3033.2025.11.0469

基金项目:国家社会科学基金资助(21BGL231);湘江实验室重大项目(23XJ01001)。

**【摘要】** 为改善城市交通安全并解决交通拥堵问题,提出自适应时空多头图注意力网络的交通流预测模型,该模型通过三阶段架构实现精细化交通流预测,首先,采用时空编码模块对交通流数据的时空信息编码,构建交通流图;其次,对交通流图的流量属性和图结构层面进行自适应增强;然后,利用双分支多头空间自注意力模块挖掘全局与局部空间的内在关联,解耦复杂空间依赖关系,结合时间异质性与时间多头注意力机制,多角度捕捉时间序列中的复杂动态关系,同时,引入多层感知器(MLP)挖掘现有流量与未来流量的深层联系;最后,以真实世界数据集为例验证模型的有效性。结果表明:该模型在交通流预测中的关键指标平均绝对误差(MAE)和平均绝对百分比误差(MAPE)结果优于现有方法,分别最高降低5.12%和19.11%,同时,模型的预测值与真实值高度吻合,尤其在突发车流波动场景下,预测结果能较好地反映实际交通流量的变化。通过优化时空特征提取能力,模型能够有效降低预测误差,从而提升交通流量预测的性能,这证明了模型的有效性。

**【关键词】** 时空多头图注意力网络; 自适应; 多头空间自注意力; 异质性; 时空数据

## Traffic flow prediction model based on adaptive spatio-temporal multi-head graph attention network

ZHOU Xinmin<sup>1,2</sup>, XU Tian<sup>3</sup>, LI Da<sup>3</sup>, WANG Longxin<sup>4</sup>, HU Jianghua<sup>3</sup>, WANG Wei<sup>4</sup>

(1 School of Artificial Intelligence and Advanced Computing, Hunan University of Technology and Business, Changsha Hunan 410205, China; 2 Xiangjiang Laboratory, Changsha Hunan 410205, China;  
3 School of Frontier Crossover Studies, Hunan University of Technology and Business, Changsha Hunan 410205, China; 4 School of Computer Science, Hunan University of Technology and Business, Changsha Hunan 410205, China)

**Abstract:** An adaptive spatio-temporal multi-head graph attention network was proposed for traffic-flow prediction to improve urban traffic safety and alleviate congestion. A three-stage architecture was designed to achieve refined prediction. Firstly, a spatio-temporal information from traffic flow data was encoded by a spatio-temporal encoding and a traffic-flow graph was constructed. Subsequently, the traffic-flow graph is

adaptively enhanced at both the flow attribute and graph structure levels. Then, a two-branch multi-head spatial self-attention mechanism was employed to mine intrinsic associations at global and local spatial scales, thereby decoupling complex spatial dependencies. By temporal heterogeneity and a multi-head temporal attention mechanism, the complex dynamic relationships in time-series data were capture from multiple perspectives. A multilayer perceptron (MLP) was incorporated to explore deep connections between present and future traffic flows. The model's effectiveness was validated using real-world datasets. The results demonstrate that the proposed model outperforms existing methods in key metrics for traffic flow prediction, with reductions in Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) of up to 5.12% and 19.11%, respectively. Furthermore, a high degree of agreement was observed between the predicted and true values, particularly in scenarios with sudden traffic fluctuations. The predictions reflect the actual changes effectively. By optimizing spatio-temporal feature extraction capabilities, prediction errors are effectively reduced and the performance of traffic flow prediction is improved, which demonstrates the effectiveness of the model.

**Keywords:** spatio temporal multi head graph attention network; adaptive; multi-head spatial self-attention; heterogeneity; spatio-temporal data

## 0 引言

随着城市扩张与居民数量增长,道路交通承载能力面临考验,这不仅增加了交通事故风险,更造成区域性交通拥堵问题<sup>[1]</sup>。作为城市交通信息的关键生成主体,车辆生成的大量时空数据为交通流量分析提供了有效支持,准确预测区域的进出车辆流量,有助于优化出行路径选择,缓解高峰期车辆拥堵压力,更能降低交通事故发生概率,进而提高城市交通安全水平<sup>[2]</sup>。如何利用复杂的时空数据实现精确的交通流量预测,是一个至关重要的问题。

交通流量预测是通过分析历史和当前的交通数据,预测未来某一时段内区域的交通流量,如车流量的流入和流出。这类预测任务通常归类于时间序列分析范畴,自回归积分移动平均法(Autoregressive Integrated Moving Average Model, ARIMA)<sup>[3]</sup>和卡尔曼滤波<sup>[4]</sup>等是比较典型的方法。随着机器学习技术的发展,支持向量机<sup>[5]</sup>开始用于交通流量预测场景中,LIN Xianfu等<sup>[6]</sup>研究发现,结合ARIMA和自回归条件异方差模型的优点,可以更好地处理复杂的交通流量数据,提高预测的准确性和可靠性。递归神经网络(Recurrent Neural Network, RNN)<sup>[7]</sup>也广泛用于从大量序列信息中捕获时间相关性。

近年来,各种深度学习技术的进步显著推动了交通流量预测研究的发展,如ZHANG Rongji等<sup>[8]</sup>提出一种基于遗传算法优化的时间卷积神经网络模型,通过优化超参数提高预测精度;ASHISH等<sup>[9]</sup>将

注意力机制用于预测领域,该模型取得了显著的性能提升。YE Xiaofei等<sup>[10]</sup>利用时间卷积捕捉长期依赖关系和注意力机制来提高出租车需求预测的准确性。吴莹莹等<sup>[11]</sup>结合卷积神经网络和门控循环单元,通过引入注意力机制增强模型对关键特征的捕捉能力,提高预测精度。YU Bing等<sup>[12]</sup>提出一种基于图卷积网络的深度学习框架(Spatial-Temporal Graph Convolutional Network, ST-GCN),通过结合图卷积和因果卷积,能同时处理交通数据的空间和时间特性。BAI Lei等<sup>[13]</sup>结合图卷积和RNN,并引入自适应机制来捕捉交通数据的时空依赖。FANG Zheng等<sup>[14]</sup>基于图神经网络的模型,如时空图常微分方程网络,通过道路邻接关系和时间序列相似性构建邻接矩阵,试图同时处理时间和空间依赖性。

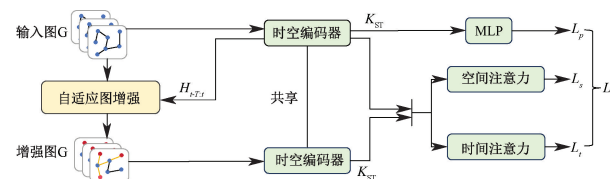
综上,国内外学者对交通流量预测的研究已取得一些成果,但是,基于机器学习的方法是将路网中的交通要素建模为离散个体,没有考虑道路之间的空间关联性。目前,对于交通流的预测大多基于单一区域的预测,缺乏对不同区域内交通分布不均的空间异质性建模,且交通系统中不同位置之间的时空依赖性动态相关的,因为这些依赖性会随着出行模式和突发事件的影响而发生变化。因此,笔者拟提出基于自适应时空多头图注意力网络的交通流预测模型,对获取的交通流量数据作图增强处理,再分别处理时间、空间数据,在4个数据集上验证模型的有效性,以期提高交通安全管理水平。

## 1 交通流量图构建

交通流预测的准确度受交通数据的质量、气象条件、交通事故等外部因素引发的动态不确定性,以及路网中各节点间时空关联性的影响。为此,将城市划分为  $N = I \times J$  个互不重叠的地理网格,记作  $s_n (1 \leq n \leq N)$ 。交通流量图记为  $G = (V, E, A, X_{t-T:t})$ , 其中,  $V = \{s_1, s_2, \dots, s_N\}$  表示空间区域的集合,  $E$  是连接  $V$  中 2 个区域边的集合。图  $G$  的邻接矩阵  $A$  为  $A \in \mathbb{R}^{N \times N}$ , 表示节点间连接情况, 用张量  $X_{t-T:t} \in \mathbb{R}^{T \times N \times 2} = (X_{t-T}, X_{t-T+1}, \dots, X_t)$  表示前  $T$  个时间步中区域的交通流入和流出数据。

## 2 交通流量预测模型构建

为有效分析并预测交通流时空特征,设计一种交通流量预测模型,其框架如图 1 所示,其中,时空编码器用于捕捉和编码交通流数据中的时空数据,图增强模块可自适应增强交通流的流量属性和图结构特征。多层感知器(Multilayer Perceptron, MLP)将高层次时空特征转换为交通流量的预测结果。该模型的核心由双分支多头空间自注意力模块捕捉不同空间尺度的依赖关系,跨尺度的空间建模。多头时序对比注意力模块学习交通流的动态时序模式,使模型可精准地捕捉动态变化的交通流量特征,更贴合真实的交通流量预测场景。



注:  $L$  为损失,  $H$  为交通信息增强中的矩阵。

图 1 交通流预测模型框架

Fig. 1 Traffic flow prediction model framework

### 2.1 时空编码模块

为更精确地捕捉交通流量特征,在时空自监督学习(Spatio-Temporal Self-Supervised Learning, ST-SSL)的基础上<sup>[15]</sup>,提出一种新的编码模型,该模型包括时间卷积(Temporal Convolution, TC)和空间卷积(Spatial Convolution, SC)。

在 TC 部分,引入一维因果卷积,在时间轴上应用门控机制。TC 将交通流作为输入,对每个区域生成包含时间信息的输出:

$$(H_{t-T_0}, H_{t-T_0+1}, \dots, H_t) = \text{TC}(X_{t-T}, X_{t-T+1}, \dots, X_t) \quad (1)$$

式中:  $H_t$  为时间步  $t$  时的嵌入矩阵,  $H_t \in \mathbb{R}^{N \times D}$ ;  $D$  为嵌入维度;  $T_0$  为经卷积操作后输出嵌入序列的长度。为捕捉不同区域的空间关联,设计基于图消息传递机制的 SC 编码器:

$$E_t = \text{SC}(H_t, A) \quad (2)$$

式中  $A$  为图  $G$  的区域邻接矩阵。SC 编码器整合地理信息得到所有区域的嵌入  $(E_{t-T_0}, E_{t-T_0+1}, \dots, E_t)$ 。

时空编码模块每个结构单元由 TC→SC→TC 的序列构成,得到时间维度  $T'$  的嵌入矩阵序列  $(K_{t-T'}, K_{t-T'+1}, \dots, K_t)$ , 最终生成  $K_{ST} \in \mathbb{R}^{N \times D}$ ,  $K_{ST}$  的每一行  $k_n \in \mathbb{R}^D$  表示区域  $s_n$  的最终嵌入。

### 2.2 自适应图增强模块

考虑传统的预测方法在处理时空关系时,忽略了时间与空间的异质性关系。为此,提出一种基于异质性关系的图结构优化方法,如图 2 所示,将研究区域划分出 5 个子区域,其中有 2 个区域(学校)的建筑群相似度高,区域的连接关系由异质性程度高低决定,实线表示区域异质性低,虚线表示异质性高。因此,优化交通流图的交通信息增强与图拓扑结构。



图 2 交通流数据的异质性连接

Fig. 2 Heterogeneous connectivity of traffic flow data

#### 2.2.1 交通信息增强

对于区域  $s_n$ , 构建嵌入序列  $(h_{t-T,n}, h_{t-T+1,n}, \dots, h_{t,n})$ , 该序列涵盖从  $(H_{t-T}, H_{t-T+1}, \dots, H_t)$  中提取的  $T$  个连续时间步长数据。再构建一个综合的嵌入表示:

$$y_n = \sum_{\tau=t-T}^t q_{\tau,n} \cdot h_{\tau,n} \quad (3)$$

$$q_{\tau,n} = h_{\tau,n}^T \cdot w_0 \quad (4)$$

式中:  $q_{\tau,n}$  为聚合权重, 量化特定时间步的交通模式与区域整体交通规律的相关性, 根据时间步长范围  $(t-T, t)$  的索引  $\tau$  确定;  $h_{\tau,n}$  为区域  $s_n$  在时间步长  $\tau$  的嵌入向量;  $w_0 \in \mathbb{R}^D$  为一个通过学习调整的参数。

在该模块中,使用从伯努利分布  $\rho_{\tau,n} \sim \text{Bern}(1 - q_{\tau,n})$  派生的屏蔽概率  $\rho_{\tau,n}$  对区域  $s_n$  在时间步长  $\tau$  的数据去噪,排除不相关的交通量。 $\rho_{\tau,n}$  的高值表示在时间  $\tau$  时的交通量  $x_{\tau,n}$  与区域  $s_n$  的交通模式的关联性较弱,因此,该时刻的数据更有可能被过滤掉。经过数据增强处理后,交通流数据的表示形式为  $\tilde{X}_{t-T:t}$ 。

### 2.2.2 图拓扑结构优化

经过交通信息增强后,优化区域交通流的拓扑结构,剔除相关性较低的区域间连接。使用一种度量 2 个区域异质性水平的方法:

$$p_{m,n} = \frac{\mathbf{y}_m^T \mathbf{y}_n}{\|\mathbf{y}_m\| \|\mathbf{y}_n\|} \quad (5)$$

式中  $p_{m,n}$  为区域异质性度量,较低的意味着区域  $s_m$  与  $s_n$  的异质性程度高。对于任意 2 个空间相邻区域  $s_m$  和  $s_n$ ,若异质性程度高,则隐藏连接边  $E$ ,掩蔽概率  $\rho_{m,n}$  由伯努利分布  $\rho_{m,n} \sim \text{Bern}(1 - p_{m,n})$  计算得出。若异质性程度低,则在  $s_m$  和  $s_n$  之间建立新的边。原始图  $G$  经过图增强模块优化后得到增强的  $\tilde{G} = (V, \tilde{E}, \tilde{A}, \tilde{X}_{t-T+1:t})$ , 其中,  $\tilde{X}_{t-T:t}$  为去偏交通量输入,  $\tilde{E}, \tilde{A}$  为去噪后的图结构。

## 2.3 双分支多头空间自注意力

在增强后的交通流图基础上,为使区域嵌入能够识别交通数据中邻近节点和全局节点的空间相关性,使用双分支多头空间自注意力模块,如图 3 所示,该模块使用多头自注意力结构,在处理位置嵌入时能并行关注多个特征子区域。

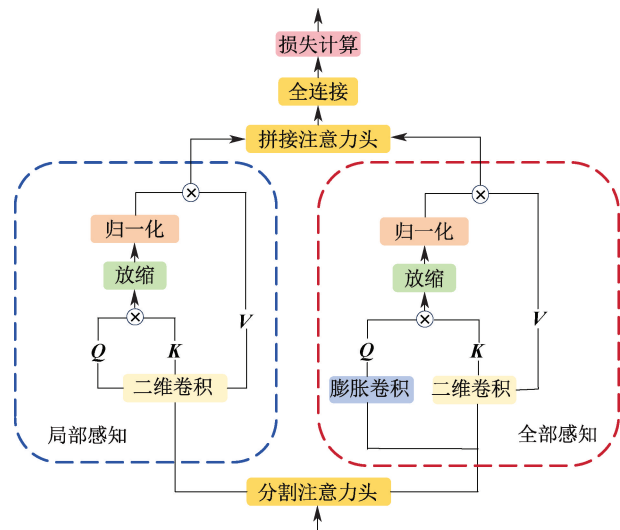


图 3 双分支多头空间自注意力模块

Fig. 3 Two-branch multi-head spatial self-attention module

### 2.3.1 局部-全局分支设计

双分支多头空间自注意力模块的输入为原始  $\mathbf{K}_{ST} \in \mathbb{R}^{N \times D}$  或增强  $\tilde{\mathbf{K}}_{ST} \in \mathbb{R}^{N \times D}$ 。局部分支使用二维卷积分别获取查询矩阵  $\mathbf{Q}$ 、键矩阵  $\mathbf{K}$  和值矩阵  $\mathbf{V}$  的值,再通过注意力机制捕获局部空间相关性。

在全局分支中,引入膨胀卷积,模型可在较少层数内学习全局信息。将膨胀卷积生成的  $\mathbf{Q}$  与二维卷积获取的  $\mathbf{K}$  相乘,得到全局感知注意力权重,最后与  $\mathbf{V}$  相乘得到全局感知结果。过程为:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d}}\right) \mathbf{V} \quad (6)$$

$$h_i = \text{Attention}(\mathbf{Q}\mathbf{W}_i^Q, \mathbf{K}\mathbf{W}_i^K, \mathbf{V}\mathbf{W}_i^V) \quad (7)$$

$$\mathbf{M}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \sum_{i=1}^n h_i \mathbf{W}^o \quad (8)$$

$$\mathbf{M}_g, \mathbf{M}_l = \text{Split}(\mathbf{M}(\mathbf{Q}, \mathbf{K}, \mathbf{V})) \quad (9)$$

$$\mathbf{M}_g(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \bigoplus (g_1, g_2, \dots, g_i) \mathbf{W}_g^o \quad (10)$$

$$\mathbf{M}_l(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \bigoplus (l_1, l_2, \dots, l_j) \mathbf{W}_l^o \quad (11)$$

$$\mathbf{M}_a(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \mathbf{M}_g(\mathbf{Q}, \mathbf{K}, \mathbf{V}) \oplus \mathbf{M}_l(\mathbf{Q}, \mathbf{K}, \mathbf{V}) \quad (12)$$

式中:  $d$  为特征维度;  $\mathbf{W}^o$  为总投影矩阵;  $\mathbf{W}_i^Q, \mathbf{W}_i^K, \mathbf{W}_i^V$  为  $\mathbf{Q}, \mathbf{K}, \mathbf{V}$  的投影矩阵;  $\mathbf{M}(\mathbf{Q}, \mathbf{K}, \mathbf{V})$  为多头自注意力公式;  $i$  为全局注意力头数;  $j$  为局部注意力头数;  $n = (i+j)$  为总头数;  $\mathbf{W}_g^o, \mathbf{W}_l^o$  为全局与局部注意力的投影矩阵;  $\bigoplus$  为特征拼接,相加得到全局与局部融合  $\mathbf{M} = (m_1, m_2, \dots, m_n) \in \mathbb{R}^{N \times D}$ 。

### 2.3.2 对比学习异构增强

为捕获不同区域流量之间相互依赖关系,并保持异构区域的能力,在注意力机制后引入一个对比学习组件,同一批次中,原始输入与其对应增广数据构成正样本对比组,同时,将该原始输入与其他样本组合为负样本对比组。计算损失函数:

$$L_s = \sum_{n=1}^N -\ln \frac{\exp(\cos(m_n, \tilde{m}_n)/\tau)}{\sum_{n'=1}^N \exp(\cos(m_n, \tilde{m}_{n'})/\tau)} \quad (13)$$

式中:  $L_s$  为损失函数;  $\cos(\cdot)$  为计算余弦相似度;  $\tau$  为温度参数。

## 2.4 多头时序对比注意力模块

为处理交通流量在不同时间段的相互影响,构建多头时序对比注意力模块,如图 4 所示。将时间异质性注入到时间感知模块中。通过编码器融合区域表征  $\mathbf{k}_{t,n}$  和  $\tilde{\mathbf{k}}_{t,n}$ :

$$\mathbf{X}_{t,n} = \gamma_1 \odot \mathbf{k}_{t,n} + \gamma_2 \odot \tilde{\mathbf{k}}_{t,n} \quad (14)$$



式中:  $\odot$  为按元素相乘;  $\gamma_1, \gamma_2$  为学习参数。

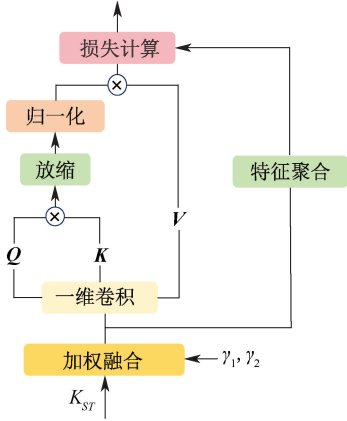


图 4 多头时序对比注意力模块

Fig. 4 Multi-head temporal contrastive attention module

获得  $n$  的查询  $Q_n^T$ 、键  $K_n^T$  和值  $V_n^T$  矩阵:

$$Q_n^T = X_{t,n} W_Q^T, K_n^T = X_{t,n} W_K^T, V_n^T = X_{t,n} W_V^T \quad (15)$$

式中  $W_Q^T, W_K^T, W_V^T \in \mathbb{R}^{D \times D}$  为学习参数。在时间维度上对节点  $n$  使用注意力操作:

$$A_n^T = \frac{(Q_n^T)(K_n^T)^T}{\sqrt{d'}} \quad (16)$$

将  $A_n^T$  与值矩阵  $V_n^T$  相乘,得到输出:

$$t_n = \text{softmax}(A_n^T V_n^T) \quad (17)$$

将  $t_n$  进行拼接后得到  $T_n = \{t_1, t_2, \dots, t_n\}$ 。线性变换  $T_n$  得到最终输出  $O_{t,n} = T_n W_p$ , 其中  $W_p$  为一个可学习的线性投影矩阵。为使模块理解输入特征的整体特性,保留重要特征,聚合处理融合特征,定义  $r_t$  为:

$$r_t = \sigma\left(\frac{1}{N} \sum_{n=1}^N X_{t,n}\right) \quad (18)$$

式中  $\sigma(*)$  为 sigmoid 函数。为提升模型对不同时间步长下城市交通流的区分能力,用对比学习将相同时间步的区域嵌入视为正样本对,不同时间步的嵌入视为负样本对。定义损失函数:

$$L_t = -\left(\sum_{n=1}^N \ln f(O_{t,n}, r_t) + \sum_{n=1}^N \ln(1 - f(O_{t',n}, r_t))\right) \quad (19)$$

式中:  $t$  和  $t'$  分别为不同时间段;  $f(*)$  为已定义的函数,定义为:  $f(O_{t,n}, r_t) = \sigma(O_{t,n}^T W r_t)$ ;  $W \in \mathbb{R}^{N \times N}$  为可训练的矩阵。

## 2.5 MLP

将区域  $s_n$  的嵌入  $k_n \in K_{ST}$  输入到 MLP 中,得到未来时间  $t+1$  的交通流预测结果:

$$\hat{l}_{t+1,n} = \text{MLP}(k_n) \quad (20)$$

式中  $\hat{l}_{t+1,n}$  为 MLP 模型的预测结果。通过最小化损失函数优化模型:

$$L_p = \sum_{n=1}^N \lambda |l_{t+1,n}^i - \hat{l}_{t+1,n}^i| + (1 - \lambda) |l_{t+1,n}^o - \hat{l}_{t+1,n}^o| \quad (21)$$

式中  $l_{t+1,n}^i, l_{t+1,n}^o$  分别为流入和流出的实际初始值。 $\lambda$  用于调整不同交通流的权重。整合式 (13) 和式 (19) 中的损失,获得最终损失:

$$L = L_s + L_t + L_p \quad (22)$$

通过反向传播算法训练模型,直到  $L$  收敛。

## 3 预测模型试验与分析

### 3.1 数据集信息及试验设置

试验的 4 个数据集见表 1, 其中, NYCBike1<sup>[17]</sup>, NYCBike2<sup>[16]</sup> 数据集记录了骑行时长、起始结束点和起始结束时间。NYCTaxi<sup>[16]</sup>, BJTaxi<sup>[17]</sup> 数据集记录了乘客上下车时间、地点和行程距离。

表 1 试验数据集信息

Table 1 Information of experimental datasets

| 数据集      | 时间间隔/h | 区域面积  | 车辆数量              | 时间范围                  |
|----------|--------|-------|-------------------|-----------------------|
| NYCBike1 | 1      | 16×8  | $6.8 \times 10^3$ | 2014-04-01—2014-09-30 |
| NYCBike2 | 0.5    | 10×20 | $2.6 \times 10^6$ | 2016-07-01—2016-08-29 |
| NYCTaxi  | 0.5    | 10×20 | $22 \times 10^6$  | 2015-01-01—2015-03-   |
| BJTaxi   | 0.5    | 32×32 | $34 \times 10^3$  | 2015-03-01—2015-06-30 |

选取预测时间前 2 h 和 3 d 的交通流量数据预测下一个时间段的交通流量。按照 7 : 1 : 2 的比例划分训练集、验证集和测试集。

在试验评估中,使用 2 个常见的指标:平均绝对误差损失 (Mean Absolute Error, MAE) 和平均绝对百分比误差 (Mean Absolute Percentage Error, MAPE), 分别评估模型预测精度。

### 3.2 试验结果分析

为更好地分析模型的预测精度,与 8 个基线模型进行对比,包含 2 个经典预测模型,即 ARIMA 和支持向量回归 (Support Vector Regression, SVR); 3 个时空融合预测模型,即 ST-ResNet, ST-GCN, 图多

注意力网络 (Graph Multi-Attention Network, GMAN);3 个基于异质性的时空预测模型,即自适应图卷积循环网 (Adaptive Graph Convolutional Recurrent Network, AGCRN),时空同步图卷积网络 (Spatial-Temporal Synchronous Graph Convolutional Network, STSGCN),ST-SSL。

试验中,将特征维度  $D$  设定为 64,学习率设置为 0.001,batch-size 设置为 32,epoch 设置为 100。部分模型采用权威期刊中的试验结果<sup>[15]</sup>。各数据集上不同模型的 MAE 和 MAPE 值见表 2。

由表 2 可知:经典时序模型 ARIMA 和 VAR 在

处理时空耦合性特征的交通数据时,预测精度最低。与 3 个时空融合预测模型、3 个基于异质性的时空预测方法相比,自适应时空多头图注意力网络的各项结果均为最优。在 NYCBike1 预测结果中,自适应时空多头图注意力网络输入流量的 MAPE 为 23.14%,较次优模型 ST-SSL 降低 0.92%,在 NYCBike2 数据集中输入流量的 MAPE 从 22.57% 降至 22.28%。在车辆最多的 NYCTaxi 和面积最大的 BJTaxi 数据集都展现出领先效果,证明自适应时空多头图注意力网络在交通流预测中的优势。

表 2 各数据集上不同模型的 MAE 和 MAPE 值

Table 2 MAE and MAPE values of different models on various datasets

| 数据集      | 指标   | 类型 | ARIMA | SVR   | ST-ResNet | STGCN | GMAN  | AGCRN | STSGCN | ST-SSL       | 本文模型         |
|----------|------|----|-------|-------|-----------|-------|-------|-------|--------|--------------|--------------|
| NYCBike1 | MAE  | 流入 | 10.66 | 7.27  | 5.53      | 5.33  | 6.77  | 5.17  | 5.81   | <u>4.90</u>  | <b>4.85</b>  |
|          |      | 流出 | 11.33 | 7.98  | 5.74      | 5.59  | 7.17  | 5.47  | 6.10   | <u>5.21</u>  | <b>5.14</b>  |
|          | MAPE | 流入 | 33.05 | 25.39 | 25.46     | 26.92 | 31.72 | 25.59 | 26.51  | 24.06        | <b>23.14</b> |
|          |      | 流出 | 35.03 | 27.42 | 26.36     | 27.69 | 34.74 | 26.63 | 27.56  | <u>24.60</u> | <b>24.33</b> |
| NYCBike2 | MAE  | 流入 | 8.91  | 12.82 | 5.63      | 5.21  | 5.24  | 5.18  | 5.25   | <u>5.06</u>  | <b>4.96</b>  |
|          |      | 流出 | 8.70  | 11.48 | 5.26      | 4.92  | 4.97  | 4.79  | 4.94   | <u>4.73</u>  | <b>4.62</b>  |
|          | MAPE | 流入 | 28.86 | 46.52 | 32.17     | 27.73 | 27.38 | 27.14 | 29.26  | <u>22.57</u> | <b>22.28</b> |
|          |      | 流出 | 28.22 | 41.91 | 30.48     | 26.83 | 26.75 | 26.17 | 28.02  | <u>21.55</u> | <b>21.17</b> |
| NYCTaxi  | MAE  | 流入 | 20.86 | 52.16 | 13.48     | 13.12 | 15.09 | 12.13 | 13.69  | <u>12.22</u> | <b>12.03</b> |
|          |      | 流出 | 16.80 | 41.71 | 10.78     | 10.35 | 12.06 | 9.87  | 10.75  | <u>10.00</u> | <b>9.84</b>  |
|          | MAPE | 流入 | 21.49 | 65.10 | 24.83     | 21.01 | 22.73 | 18.78 | 22.91  | <u>17.10</u> | <b>16.43</b> |
|          |      | 流出 | 21.23 | 64.06 | 24.42     | 20.78 | 21.97 | 18.41 | 22.37  | <u>17.11</u> | <b>16.69</b> |
| BJTaxi   | MAE  | 流入 | 21.48 | 52.77 | 12.12     | 12.34 | 13.13 | 12.30 | 12.72  | <u>11.63</u> | <b>11.50</b> |
|          |      | 流出 | 21.60 | 52.74 | 12.16     | 12.41 | 13.20 | 12.38 | 12.79  | <u>11.60</u> | <b>11.49</b> |
|          | MAPE | 流入 | 23.12 | 65.51 | 15.50     | 16.66 | 18.67 | 15.61 | 17.22  | <u>15.35</u> | <b>15.24</b> |
|          |      | 流出 | 20.67 | 65.51 | 15.57     | 16.76 | 18.84 | 15.75 | 17.35  | <u>15.42</u> | <b>15.33</b> |

注:加粗数值表示最高得分,下划线数值表示次优得分。

3.3 消融试验对比分析

为验证自适应时空多头图注意力网络中各子模块的有效性,设计 4 种变体消融模型,分别是①SA-R (Structure Augmentation-Rand) 变体,用随机边去除和添加增强替换异质引导的图结构增强;②TA-R (Traffic-level Augmentation-Rand) 变体,用随机流量掩码增强替换异质引导的信息增强;③SA-FNN (Spatial Attention-Feed-Forward Neural Network) 变体,用前馈神经网络模块替换双分支多头空间自注意力模块;④TA-FC (Temporal Attention Fully Connected Layer) 变体,用全连接层替换多头时序对比注意力模块。

从图 5 可以看出,修改不同的模块都会使模型预测结果变差。双分支多头空间自注意力和多头时序对比注意机制对预测准确性的提升有显著作用,

表明多头注意力机制能够有效整合交通流数据中的时空特征。此外,文中模型在各项指标上始终优于 SA-R 和 TA-R,表明自适应异质性引导的流量层面和图结构层面增强的有效性。

3.4 预测结果可视化分析

为进一步研究自适应时空多头图注意力网络的预测效果,图 6 为 4 个数据集中各随机抽取一个节点样本在 300 个时间步的交通输入输出真实值与预测值。图中预测曲线与真实数据拟合,表明有可靠的交通流预测能力,在 BJTaxi 与 NYCTaxi 数据集中稳定识别常规与突变流量波动特征,验证了模型在复杂交通流中的泛化性能。但由于随机噪声的影响,其在局部细节处的预测精度仍不够准确。

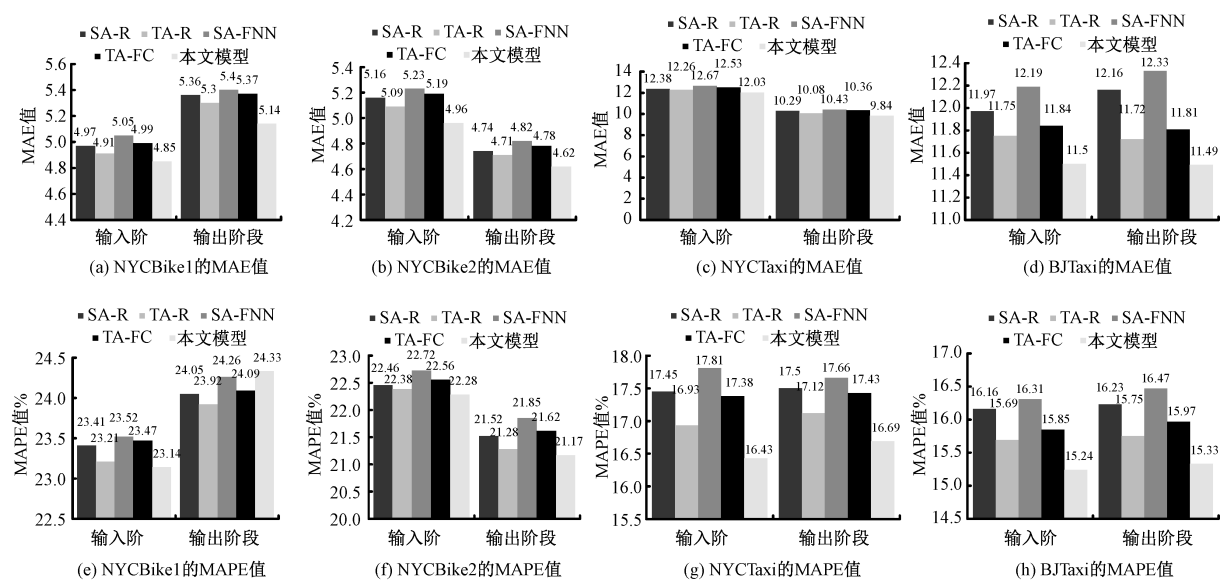


图 5 自适应时空多头图注意力网络的消融试验

Fig. 5 Ablation study on the adaptive spatio-temporal multi-head graph attention network

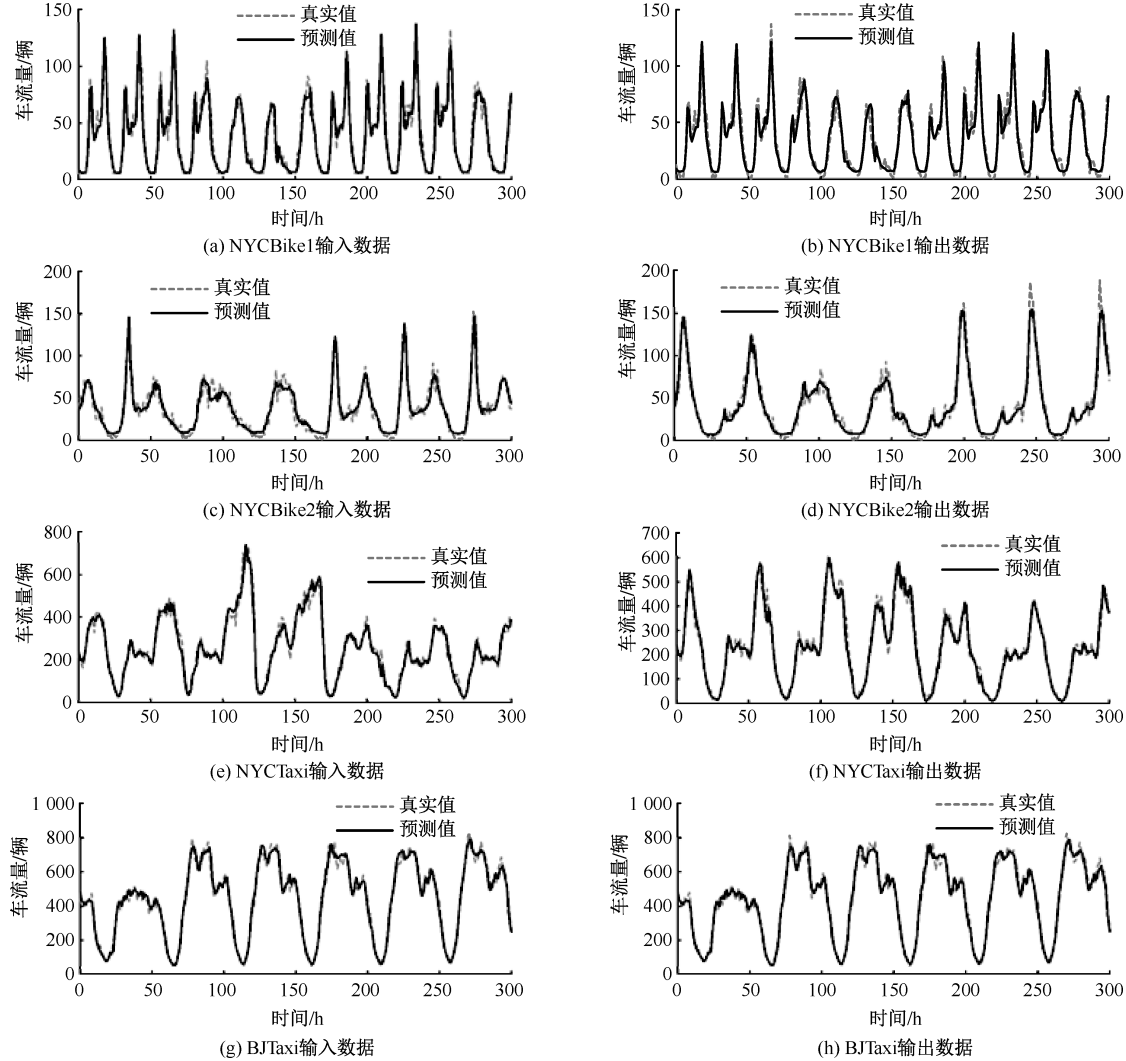


图 6 各数据集预测值与其真实值的对比曲线

Fig. 6 Comparative curves of predicted versus actual values across multiple datasets

4 结 论

1) 针对城市交通流量预测问题,设计一种自适应时空多头注意力网络的交通流预测模型。试验结果表明:该模型对城市交通流量预测有着较高的准确性,且精度优于各基线模型。

2) 消融试验进一步证实,各个模块能有效提升预测准确性,尤其是双分支多头空间自注意机制和多头时序对比注意机制起重要作用。在可视化结果中该模型不仅能准确预测正常交通流结果,还在稀

疏车流和密集车流场景下表现出良好的适应性。

3) 目前,极端场景(如大规模事故,天气变化)公开的数据样本稀缺,且与常规环境下的数据差异巨大,难以进行交通流预测,所以自适应时空多头注意力网络未考虑在极端场景下的预测能力,在极端场景使用时存在局限性。未来将多源动态数据(如天气、兴趣点数据,突发事件)融入模型,提升模型对复杂交通场景的泛化能力;同时,探索图神经网络与复杂路网信息的结合,进一步拓展其在智能交通系统中的实际应用场景。

参 考 文 献

[1] 周新民,金江涛,鲍娜娜,等. 基于注意力卷积长短时记忆模型的城市出租车流量预测[J]. 中国安全科学学报, 2024, 34(7): 153-162.  
ZHOU Xinmin, JIN Jiangtao, BAO Nana, et al. Urban taxi traffic flow prediction based on attentive ConvLSTM-ResNet model[J]. China Safety Science Journal, 2024, 34(7): 153-162.

[2] 胡立伟,贺雨,侯智,等. 山区高速公路交通事故风险多维度耦合研究[J]. 中国安全科学学报, 2024, 34(5): 17-27.  
HU Liwei, HE Yu, HOU Zhi, et al. Multi-dimensional coupling study on traffic accident risk of highway in mountainous areas[J]. China Safety Science Journal, 2024, 34(5): 17-27.

[3] ALGHAMDI T, ELGAZZAR K, BAYOUMI M, et al. Forecasting traffic congestion using ARIMA modeling[C]. 2019 15<sup>th</sup> International Wireless Communications & Mobile Computing Conference (IWCMC). IEEE, 2019: 1 227-1 232.

[4] GUO Jianhua, HUANG Wei, WILLIAMS B M. Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification[J]. Transportation Research Part C: Emerging Technologies, 2014, 43: 50-64.

[5] WU C H, HO J M, LEE D T. Travel-time prediction with support vector regression[J]. IEEE Transactions on Intelligent Transportation Systems, 2004, 5(4): 276-281.

[6] LIN Xianfu, HUANG Yuzhang. Short-term high-speed traffic flow prediction based on ARIMA-GARCH-M model[J]. Wireless Personal Communications, 2021, 117(4): 3 421-3 430.

[7] CONNOR J T, MARTIN R D, ATLAS L E. Recurrent neural networks and robust time series prediction[J]. IEEE Transactions on Neural Networks, 1994, 5(2): 240-254.

[8] ZHANG Rongji, SUN Feng, SONG Ziwen, et al. Short-term traffic flow forecasting model based on GA-TCN[J]. Journal of Advanced Transportation, 2021, 2021(1): DOI:10.1155/2021/1338607.

[9] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. arXiv, 2017: DOI: 10.48550/arXiv.1706.03762.

[10] YE Xiaofei, HAO Yu, YE Qiming, et al. Demand forecasting of online car - hailing by exhaustively capturing the temporal dependency with TCN and attention approaches[J]. IET Intelligent Transport Systems, 2024, 18(12): 2 565-2 575.

[11] 吴莹莹,赵丽宁,袁志鑫,等. 基于注意力机制的 CNN-GRU 船舶交通流预测模型[J]. 大连海事大学学报, 2023, 49(1): 75-84.  
WU Yingying, ZHAO Lining, YUAN Zhixin, et al. CNN-GRU ship traffic flow prediction model based on attention mechanism[J]. Journal of Dalian Maritime University, 2023, 49(1): 75-84.

[12] YU Bing, YIN Haoteng, ZHU Zhanxing. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting[J]. arXiv, 2017: DOI:10.24963/ijcai.2018/505.

[13] BAI Lei, YAO Lina, LI Can, et al. Adaptive graph convolutional recurrent network for traffic forecasting[J]. Advances in Neural Information Processing Systems, 2020, 33: 17 804-17 815.

[14] FANG Zheng, LONG Qingqing, SONG Guojie, et al. Spatial-temporal graph ode networks for traffic flow forecasting[C]. Proceedings of the 27<sup>th</sup> ACM SIGKDD Conference on Knowledge Discovery & Data Mining, 2021: 364-373.

[15] JI Jiahao, WANG Jingyuan, HUANG Chao, et al. Spatio-temporal self-supervised learning for traffic flow prediction[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2023: 4 356-4 364.

[16] YAO Huaxiu, TANG Xianfeng, WEI Hua, et al. Revisiting spatial-temporal similarity: a deep learning framework for traffic prediction[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019: 5 668-5 675.

[17] ZHANG Junbo, ZHENG Yu, QI Dekang. Deep spatio-temporal residual networks for citywide crowd flows prediction[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2017: 1 655-1 661.



作者简介: 周新民 (1977—),男,湖南新邵人,博士,教授,博士生导师,主要从事新型智慧城市、商务智能与大数据、人工智能大模型等方面的研究。E-mail:zhouxinmin2699@163.com。