

基于多视图与时空一体化异质性感知的交通流量预测

陈波¹, 肖玥², 叶亚琴³, 曹柱³

(1. 武汉中地软件股份有限公司, 湖北 武汉 430073;

2. 汉江师范学院 数学与计算机科学学院, 湖北 十堰 442000; 3. 中国地质大学(武汉) 计算机学院, 湖北 武汉 430078)

摘要: 交通流量预测是智能交通服务的一项重要任务。由于交通网络节点之间存在复杂的时空依赖关系, 并且数据模式呈现时空异质性, 因此该项任务极具挑战性。针对现有交通流量预测模型忽视了数据间的高阶空间信息, 以及时空异质性感知能力不足的问题, 提出一种基于多视图与时空一体化异质性感知的交通流量预测模型(MGSTEAtt)。首先, 从欧氏空间与非欧氏空间两个视角构造基于地理距离和语义相似性的两个掩码矩阵, 并将其输入到空间自注意力模块中同时建模局部和全局空间依赖。其次, 将各节点的时空嵌入作为自注意机制中构造参数矩阵时的输入, 为每个节点构建特定的参数空间, 以实现模型对时空一体化的异质性精细化感知能力。最后, 在杭州和上海地铁数据集上的实验结果表明, MGSTEAtt能有效建模节点间的多种空间关系和时空异质性, 相比基线模型具有更高的预测精度。

关键词: 交通流量预测; 注意力机制; 多视图; 时空一体化异质性感知; 时空大数据

DOI: 10.11907/rjtk.251287

扫描二维码阅读全文: 

中图分类号: TP18

文献标识码: A

文章编号: 1672-7800(2026)009-0081-10

Traffic Flow Prediction Based on Multi-View and Spatiotemporal Integrated Heterogeneity Perception

CHEN Bo¹, XIAO Yue², YE Yaqin³, CAO Zhu³

(1. Wuhan Zhongdy Software Co., LTD, Wuhan 430073, China; 2. School of Mathematics and Computer Science, Hanjiang Normal University, Shiyan 442000, China; 3. School of Computer Science, China University of Geosciences (Wuhan), Wuhan 430078, China)

Abstract: Traffic flow prediction is an important task to realize intelligent transportation services. Given the complex spatiotemporal dependencies between traffic network nodes and the spatiotemporal heterogeneity of data patterns, this task is highly challenging. To address the issues of existing traffic flow prediction models that neglect high-order spatial information between data and have insufficient perception of spatiotemporal heterogeneity, a traffic flow prediction model based on multi-view and integrated spatiotemporal heterogeneity perception (MGSTEAtt) is proposed. First, two masking matrices based on geographic distance and semantic similarity are constructed from the perspectives of Euclidean space and non-Euclidean space, which are input into the spatial self-attention module to simultaneously model local and global spatial dependencies. Secondly, the spatio-temporal embedding of each node is used as the input when constructing the parameter matrix in the self-attention mechanism, and its specific parameter space is constructed for each node to fully consider the spatio-temporal heterogeneity between data. Finally, the experimental results on the Hangzhou and Shanghai metro datasets demonstrate that MGSTEAtt can effectively model the various spatial relationships between nodes and the spatiotemporal heterogeneity, achieving higher prediction accuracy compared to the baseline models.

Key Words: traffic flow prediction; attention mechanism; multi-view; spatiotemporal integrated heterogeneity perception; spatio-temporal big data

收稿日期: 2025-04-21

基金项目: 智能地学信息处理湖北省重点实验室开放研究项目(KLIGIP-2022-B14); 桐兴(湖北)投资咨询有限公司合作项目(2021116490)

作者简介: 陈波(1979-), 男, 硕士, 武汉中地软件股份有限公司高级工程师, 研究方向为城市时空推理; 肖玥(1999-), 女, 硕士, 汉江师范学院数学与计算机科学学院助教, 研究方向为交通流量预测; 叶亚琴(1979-), 女, 博士, CCF专业会员, 中国地质大学(武汉)计算机学院副教授, 研究方向为城市多模态轨迹智能计算; 曹柱(2001-), 男, 中国地质大学(武汉)计算机学院硕士研究生, 研究方向为轨迹生成。本文通讯作者: 叶亚琴、肖玥。

0 引言

随着城市化进程的不断推进,机动车拥有量不断增加的同时,也给道路系统带来了更大压力,造成了如环境污染、交通堵塞等诸多问题。为了有效应对交通拥堵,智能交通系统(Intelligent Transportation System, ITS)在交通管理中得到了广泛应用^[1]。交通流量预测作为ITS中的重要组成部分,能够帮助城市交通管理部门进行交通引导和控制,优化道路资源配置,提高道路通行能力^[2]。

交通预测的目标是在交通网络的结构约束下,基于历史观测数据估计交通网络未来的交通状况。早期的大多数统计模型是从全局角度将交通预测视为一个回归问题,例如历史平均模型(Historical Average, HA)^[3]和自回归积分滑动平均模型(Auto-Regressive Integrated Moving Average Model, ARIMA)^[4]。交通流量预测涉及多种技术,包括时间序列分析,以及传统机器学习模型(如支持向量机、决策树、随机森林等)、神经网络模型(如循环神经网络、卷积神经网络等)、深度学习模型(如Transformer模型、图神经网络等)。交通预测问题通常比其他时间序列预测问题更具挑战性,其涉及到大量高维数据,且数据间具有复杂的动态时空依赖关系。首先,在空间方面,不同位置对预测位置的影响不同,不同位置之间的空间相关性是高度动态的。其次,在时间方面,同一位置不同时间点的观测值呈非线性变化,并且交通数据具有周期性和季节性,如何选择最相关的历史观测值进行预测仍具有挑战性。因此,由于交通数据在空间和时间维度上都具有很强的动态相关性,如何挖掘数据间复杂的时空依赖性和时空异质性是研究的重难点。

尽管基于深度学习的交通流量预测模型已经取得了显著成功,但仍存在一定的局限性:①在交通网络约束下的交通节点间的空间依赖同时存在于欧氏空间和非欧氏空间中,但现有研究在构建时空图时仅考虑了节点在欧氏空间中的空间依赖,而忽略了非欧氏空间中的高阶空间信息,构造出的时空图难以全面表征交通网络的空间特征,也限制了模型提取空间关系的能力;②由于不同时间点、不同位置的节点对目标节点的影响程度不同,交通数据中的时空异质性问题是影响预测精度的关键。满足时空一体化建设要求、实现时空异质性的高精度感知,尚有诸多核心瓶颈亟待攻克。

针对上述问题,本文提出一种基于多视图与时空一体化异质性感知的交通流量预测模型(Multi-graph and Spatio-temporal Embedding Attention mechanism, MGSTEAtt),模型结合多图和时空嵌入注意力实现了关键技术突破。首先,从欧氏空间和非欧氏空间两个视角构造了基于地理距离和语义相似性的两个掩码矩阵,将其分别输入到地理空间注意模块和语义空间注意模块中,以同时建模节点间

的局部和全局空间依赖;其次,为了能够使模型感知到节点间的时空异质性,将节点的时间、空间嵌入整合到一起成为时空嵌入,并将其作为自注意机制中构造参数矩阵时的输入,为每个节点构建特定的参数空间和时空模式,以充分考虑交通数据中的时空异质性。

本文的主要贡献如下:

(1)提出了一种基于多视图与时空一体化异质性感知的交通流量预测模型,能够融合地理和语义多种关系,建立细粒度时空一体化嵌入学习过程,实现多尺度依赖关系提取和时空异质性感知。

(2)开发了一个节点时空嵌入下的注意力模块,旨在使用自注意力机制从序列特征和上下文交互两个角度建模数据间的时空特征,通过每个节点学习特定的时空模式,进而提升模型对时空异质性的感知能力。

(3)在杭州地铁和上海地铁两个真实世界交通流量数据集上进行了多组基线实验和消融实验,通过对实验结果的分析证明了本文模型的有效性。

1 相关工作

1.1 早期交通流量预测模型

早期的统计模型通常将交通预测视为一个线性问题,并使用固定的理论假设来推断和预测未来的交通状态,例如HA^[3]、ARIMA^[4]、卡尔曼滤波器(Kalman Filter, KF)^[5]等,但其不能有效地处理短期交通流预测的随机性和复杂性。

之后,各种机器学习方法被应用于时间序列预测上。例如,最邻近分类算法(K-Nearest Neighbors, KNN)^[6]和支持向量机(Support Vector Machine, SVM)相较于基于统计理论的交通流量预测方法,在预测性能上有了较大改进。但受限于机器学习自身的表达能力,其难以对大规模交通数据进行高效的建模,在挖掘大数据中深层、复杂、隐含的时空相关性时能力不足。

1.2 基于深度学习的交通流量预测模型

近年来,基于神经网络的深度学习模型由于具有较好的学习和泛化能力而受到学者们青睐。常用的深度学习模型有循环神经网络(Recurrent Neural Network, RNN)^[7]、长短时记忆网络(Long Short-Term Memory, LSTM)^[8]、门控循环单元(Gated Recurrent Unit, GRU)^[9]和卷积神经网络(Convolutional Neural Network, CNN)等。Ma等^[10]首次将LSTM应用于交通领域。Fu等^[11]首次将GRU应用于交通流量预测,实验结果表明,GRU相比ARIMA模型的预测效果更好。Yu等^[12]结合了CNN与LSTM来预测短期和长期交通状态。由于上述方法只适用于欧氏数据的特征提取,而交通网络属于非欧几里得结构,因此图卷积网络(Graph Convolutional Network, GCN)^[13]被引入交通流量预测领域。Yu等^[14]提出时空图卷积网络,在时空卷积块中

集合了图卷积和门控时间卷积。Li等^[15]提出扩散卷积循环神经网络,使用扩散卷积网络来建模交通流预测中信息的扩散过程。Wu等^[16]提出图波网络,组合了GCN和门控时间卷积。Liu等^[17]提出地铁客流量预测模型(Physical-Virtual Collaboration Modeling for Intra-and Inter-Station Metro Ridership Prediction, PVGCN),基于站点和流量序列构造了物理—虚拟图,并使用图卷积门控循环单元捕捉时空关系。Feng等^[18]设计了宏观—微观时空网,并利用图卷积网络与空间注意力分别捕获微观和宏观空间相关性。Chen等^[19]提出一种位置图卷积网络,通过在GCN中添加可学习的矩阵来表示不同节点之间的不同影响水平。Bai等^[20]提出自适应图卷积循环网络(Adaptive Graph Convolutional Recurrent Network, AGCRN),该模型使用GCN提取空间特征,且在进行特征变换时,针对每个节点使用了单独的权重矩阵,以关注节点特有的交通模式。

由于交通网络中不同节点之间的影响程度不同,而GCN无法为不同的相邻节点分配不同权重,从而限制了模型捕获空间信息相关性的能力。因此,有学者将注意力机制引入交通预测领域,进一步提升了模型的预测精度^[21]。Guo等^[22]利用一种基于注意力的时空图卷积网络(Attention Based Spatial-Temporal Graph Convolutional Network, ASTGCN)模型来预测交通流。Zheng等^[23]提出一种基于编解码器的图形多注意网络进行长期流量预测。Bai等^[24]提出时空图序列模型(Spatial-temporal Graph to Sequence Model, STG2Seq),基于交通需求相似度来建模城市的多步乘车需求,并使用基于注意力的输出模块来建模动态时间和通道间的信息。Liu等^[25]提出多步依赖关系网络,通过注意力机制来控制不同时间步长隐藏状态的权重。Yan等^[26]设计了一个空中交通流量预测模型,通过定义多个图来表征机场间的关系,并利用时间感知注意模块来捕获交通流的时间相关性。Xie等^[27]设计了一个时空动态图学习模型(Spatio-Temporal Dynamic Graph Relation Learning, STDGRL),通过可学习参数自适应生成邻接矩阵来捕获动态空间关系,并使用时空嵌入来捕获站点间的不同模式。李正楠等^[28]基于GCN和LSTM研究了考虑时空信息的城市交通流量预测方法。Chen等^[29]基于交通流理论设计了3个交通拥堵指数,以反映节点的运行状态,并设计了一个交通流矩阵,以更好地表示节点之间的关系。Geng等^[30]提出一种时空门控注意力模型,通过门控时间自注意力机制增强对局部和全局时间特征的提取。

虽然基于深度学习的交通流量预测模型取得了阶段性成果,但在高阶时空依赖关系的深度挖掘,以及针对时空一体化要求的时空异质性精准感知两大关键领域依然存在明显短板,亟待取得关键性突破。

2 问题描述

(1) 交通网络。将交通网络表示为一个图 $G = (V, E, A)$, 其中 $V = \{V_1, \dots, V_n\}$ 是节点 ($|V| = N$) 的集合, $E \subseteq V \times V$ 是一组边, A 是网络的邻接矩阵。其中, N 表示图中节点的数量。

(2) 交通流量张量。网络中 N 个节点在 t 时刻的交通流量表示为 $X_t \in \mathbb{R}^{N \times d}$, 其中 d 为交通流量的维数。使用 $X = (X_1, X_2, \dots, X_T) \in \mathbb{R}^{T \times N \times d}$ 表示所有 T 个时间片上所有节点的交通流量张量。

(3) 问题定义。形式上, 给定一个交通系统上观察到的交通流量张量 X , 本文的目标是从之前 T 步的流量观测值中学习一个映射函数 f , 以预测未来 T' 步的交通流量。数学形式的定义如下:

$$[X_{(t-T+1)}, \dots, X_t; G] \xrightarrow{f} [Y_{(t+1)}, \dots, Y_{(t+T')}] \quad (1)$$

3 模型构建

本文提出的 MGSTAtt 模型框架如图 1 所示。其基于编码器—解码器架构, 主要包括时空嵌入模块、时间自注意力模块、空间自注意力模块、趋势构建模块和时间交互注意模块 5 部分。在编码器中, 首先采用时间自注意层来捕捉时间特征, 然后使用由地理空间自注意和语义空间自注意模块组成的空间自注意层, 同时捕捉局部和全局空间依赖。特别地, 使用节点时间嵌入和空间嵌入整合成的时空嵌入来引导注意力进行时空异质性感知。在解码器中, 原始输入首先通过一个趋势构建模块提取趋势信息, 以增强模型对流量局部变化的感知, 然后通过时空嵌入引导下的时间自注意层和空间自注意层提取时空特征。此外, 为了捕捉历史交通状态和预测状态之间的注意力分布关系, 在解码器中还构建了时间交互注意模块。最终经由预测模块中的线性层得到预测输出值。

3.1 编码器

3.1.1 时空嵌入

时空嵌入模块的计算流程如图 2 所示。

(1) 时间嵌入。首先, 将序列中固有的时间属性记为 T_t^m , 并通过独热编码整形为 d 维向量。其次, 为了考虑相对于输入时间的位置信息, 采用三角函数进行位置嵌入。计算公式如下:

$$PE_{pos}(i) = \begin{cases} \sin(pos/10\,000^{i/d}), & \text{if } i \text{ is even} \\ \cos(pos/10\,000^{i/d}), & \text{if } i \text{ is odd} \end{cases} \quad (2)$$

将时间属性和时间位置两个向量进行拼接, 得到时间嵌入 $TE \in \mathbb{R}^d$ 。

(2) 空间嵌入。对于邻接矩阵 A , 首先计算归一化的拉普拉斯矩阵 $\Delta = I - D^{-1/2}AD^{-1/2}$, 其中 D 是度矩阵, I 是单位

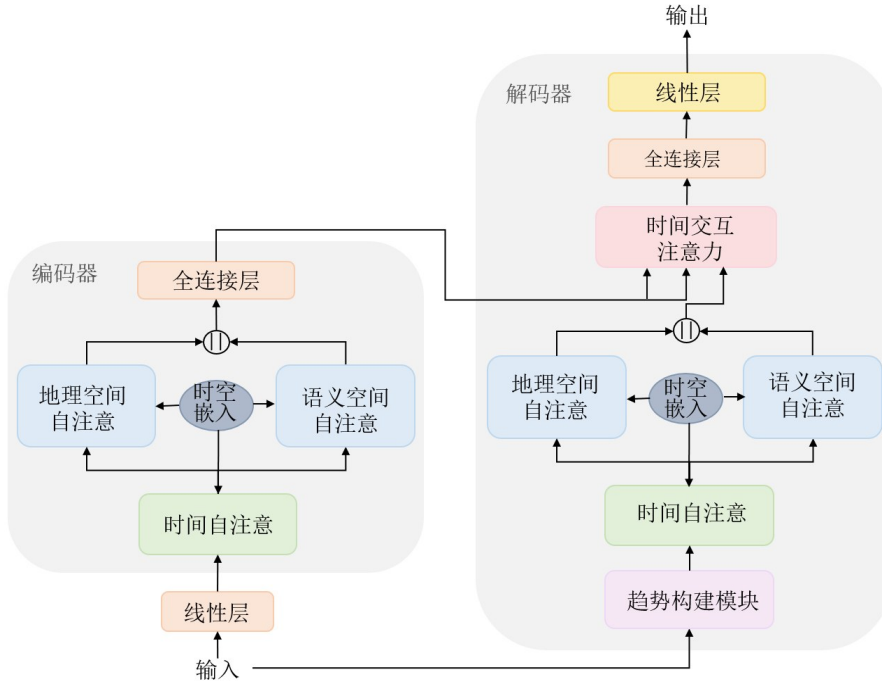


Fig. 1 Overall structure of the MGSTEAtt model

图1 MGSTEAtt模型总体结构

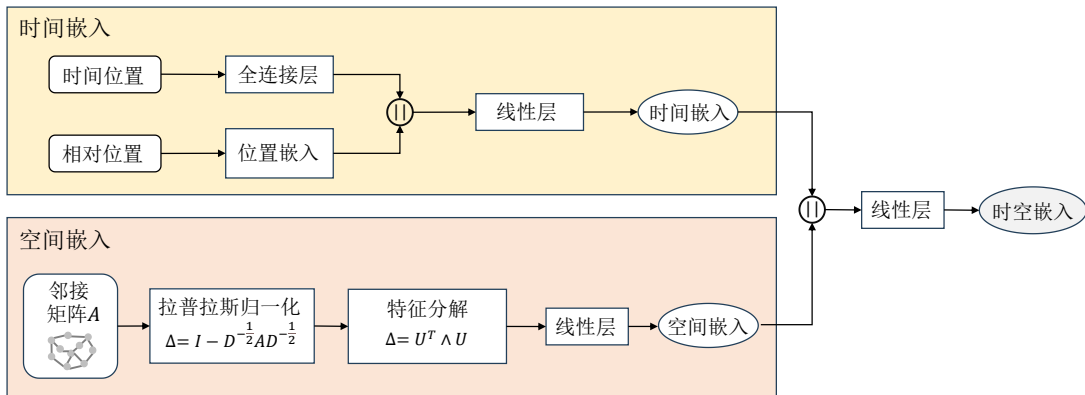


Fig. 2 Spatio-temporal embedding diagram

图2 时空嵌入示意图

矩阵。然后执行特征分解 $\Delta = U^T \Lambda U$, 得到特征值矩阵 Λ 和特征向量矩阵 U 。在计算特征映射后, 通过可学习的线性变换生成节点 V_i 的最终空间嵌入 $SE \in \mathbb{R}^d$ 。

(3) 时空嵌入。将时间嵌入 $TE \in \mathbb{R}^d$ 和空间嵌入 $SE \in \mathbb{R}^d$ 进行拼接, 再经过一个线性层后得到节点 i 在 t 时刻的时空嵌入 STE_i^t 。计算公式如下:

$$STE = \text{Linear}(TE, SE) \quad (3)$$

3.1.2 时间自注意力

在时间建模方面, 基于注意力的方法通过关注每个时间位置以有效地学习长期依赖关系, 并且可以实现并行化序列处理。

自注意力(Self-attention)是注意力机制的一种变体, 相比传统注意力机制, 自注意力主要利用特征本身的信息来计算注意力系数, 因此可以减少对外部信息的依赖, 更好地捕获数据内在的相关性。多头注意力机制可提高注

意力的泛化性。多头注意力是利用多个查询向量 $Q = [q_1, q_2, \dots, q_n]$ 并行地从输入信息 $(K, V) = [(k_1, v_1), (k_2, v_2), \dots, (k_n, v_n)]$ 中选取多组信息, 从而让不同的注意力头学习到不同空间的语义信息。

由于上述映射关系对所有节点都使用共享参数矩阵进行计算, 这种共享模式虽然可以聚集空间节点信息, 降低过拟合风险, 但也会忽视时空一体化异质性感知。基于此, 本文提出使用节点嵌入信息来引导注意力机制的学习。具体来说, 通过为每个注意力头构建一个多层感知机, 并将时空嵌入作为输入, 使用由其生成的权重矩阵来代替共享参数, 进而将查询、键和值的学习与节点各自的时空条件联系起来, 以获取特定节点的模式。计算公式如下:

$$W_q = \text{Relu}(W_q^1 x + b_q^1) W_q^2 + b_q^2 \quad (4)$$

$$W_k = \text{Relu}(W_k^1 x + b_k^1) W_k^2 + b_k^2 \quad (5)$$

$$W_v = \text{Relu}(W_v^1 x + b_v^1) W_v^2 + b_v^2 \quad (6)$$

其中, x 为时空嵌入(Spatio-temporal Embedding, STE), W_q^1, W_k^1, W_v^1 为权重矩阵, b_q^1, b_k^1, b_v^1 为偏置, $W_k^2, W_v^2, W_q^2, b_k^2, b_v^2, b_q^2$ 为可学习的参数。

获取了融合时空嵌入的查询、键和值矩阵后,进行时间自注意力计算,为了保证模型的稳定性,还采用了残差^[31]和层归一化方法。计算公式如下:

$$H^l = \text{LayerNorm}(H^{l-1} + H^l) \quad (7)$$

3.1.3 空间自注意力

在自注意力计算中,每个节点与所有节点进行交互,这相当于将空间图视为全连通图。然而,只有少数节点对之间的交互是必要的,如欧氏空间中距离较近的节点对和非欧氏空间中功能相似的节点对。基于此,本文引入两个图掩码矩阵 M_{geo} 和 M_{sem} ,以同时捕获交通数据中的局部和全局空间依赖关系。

从近距离的角度出发,距离较近的节点之间相关性通常更强。本文通过高斯核函数来定义基于距离的邻接矩阵^[32]:

$$A_{ij}^{Dis} = \begin{cases} \exp\left(-\frac{Dis(i,j)^2}{\sigma^2}\right), & \exp\left(-\frac{Dis(i,j)^2}{\sigma^2}\right) \geq \epsilon \\ 0, & otherwise \end{cases} \quad (8)$$

其中, $Dis(i,j)$ 是节点 i 与节点 j 之间的距离, σ 是距离的标准差。 ϵ 是权重系数的阈值,根据经验取值为 0.1。之后,选择节点权重系数大于阈值的节点作为其地理邻居,将当前节点与其地理邻居之间的权值分别设置为 1 和 0,以构造地理掩码矩阵 M_{geo} 。通过该方式,可以屏蔽彼此相距较远的节点对之间的注意力。

从长远来看,引入历史交通数据来建模序列间的相似性。相似性矩阵定义如下:

$$A_{ij}^{sem} = \begin{cases} \exp\left(-\frac{\|x_{his}^i - x_{his}^j\|^2}{\min(\|x_{his}^i\|^2, \|x_{his}^j\|^2)}\right), & w_{ij} \geq \epsilon \\ 0, & otherwise. \end{cases} \quad (9)$$

其中, x_{his}^i 是第 i 个节点的历史流量序列, ϵ 是权重系数的阈值,根据经验取值为 0.3。之后,选择每个节点相似度最高的 K 个节点作为其语义邻居,将当前节点与其语义邻居之间的权值分别设置为 1 和 0,以构造语义掩码矩阵 M_{sem} 。通过该方式,可以找到由于具有相似的城市功能而表现出相似交通模式的遥远节点对之间的高阶依赖关系。

在提取空间特征时,空间自注意力模块也使用时空嵌入引导的方式来计算权重矩阵,并加入了两个掩码矩阵,分别构造了地理空间自注意力和语义空间自注意力。通过地理和语义的自注意力计算,空间自注意力模块可以同时整合短期地理邻域和远程语义邻域信息。

3.2 解码器

3.2.1 趋势构建模块

流量的趋势是指一段时间内,流量的持续变化方向和趋势性,考虑流量的趋势信息可以加强模型对网络中流量变化的感知能力。趋势信息可以通过平滑滤波器进行提取,计算公式如下:

$$Trend_t = \frac{1}{p} \sum_{i=1}^p X_{t+i} \quad (10)$$

3.2.2 时间交互注意力模块

为了获得历史交通需求,并预测交通状态间的注意力分布关系,即让解码器能沿着时间维度自适应地关注编码特征,设计了时间交互注意力模块。时间交互注意力模块将编码器的输出作为键和值,并将时空嵌入和历史序列作为查询。

4 实验与结果分析

4.1 数据集

为了验证 MGSTAtt 模型的有效性,本文在杭州地铁和上海地铁两个数据集上进行了实验。数据集详情如下:

(1) 杭州地铁数据集。时间范围为 2019 年 1 月 1 日—2019 年 1 月 25 日,包括 80 个节点,每个节点测量值被聚合为 15 min 的窗口。

(2) 上海地铁数据集。时间范围为 2016 年 7 月 1 日—2016 年 9 月 30 日,包括 288 个节点,每个节点测量值被聚合为 15 min 的窗口。

4.2 模型评价指标

本文采用平均绝对误差(Mean Absolute Error, MAE)、平均绝对百分比误差(Mean Absolute Percentage Error, MAPE)和均方根误差(Root Mean Squared Error, RMSE) 3 个指标来评估模型性能。各指标具体定义如下:

$$MAE = \frac{1}{T} \sum_{t=1}^T |\tilde{y}_t - y_t| \quad (11)$$

$$MAPE = \frac{1}{T} \sum_{t=1}^T \left| \frac{\tilde{y}_t - y_t}{y_t} \right| \quad (12)$$

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (\tilde{y}_t - y_t)^2} \quad (13)$$

其中, $\tilde{y}_t$ 表示 t 时刻的预测值, y_t 表示 t 时刻的地面真实值。上述 3 个指标都可以反映预测值与真实值之间的偏离程度,该值越小,说明预测结果越准确,模型性能越好。

4.3 实验设置

实验在 Ubuntu 20.04 环境下进行,使用单个 NVIDIA RTX 3080 GPU, CUDA 12.0, Python 环境版本为 3.6.13。针对模型训练,本文选择 MAE 作为损失函数。将数据集中用于训练、验证和测试的数据比例划分为 7:1:2,并在分区数据集上滑动一个大小为 $T + T'$ 的窗口,生成时间序列数据。在两个数据集中使用相同的超参数,其中注意力的头

数为4,训练轮数设置为100。采用Adam方法进行优化,初始学习率为0.001,权重衰减为0.000 2。

4.4 基线模型对比实验

实验设置了9种基线模型进行比较,以实现对未来4个步长流量状况的预测。表1、表2分别为杭州地铁数据集和上海地铁数据集上基线实验的结果。可以看出,预测难度随着步长的增加而增加。其中,Auto Encoder(AE)、Feedforward Neural Network(FNN)、Recurrent Neural Networks(RNN)和Sequence to Sequence(Seq2Seq)等模型的预测精度并不理想。

在其他基线模型中,ASTGCN利用注意力机制探寻节点间的动态时空关系,可以有效提升预测精度。然而,

ASTGCN、STG2Seq只在一种图上进行图卷积,无法充分提取节点间的多种空间依赖关系。AGCRN关注到了不同节点的特性,并考虑了节点之间的复杂关系,取得了较为理想的预测结果。然而,其图由模型自适应生成,可解释性较差。PVGCN整合多图以促进时空表征学习,却忽略了时空异质性问题。STDGRL学习了站点间的动态空间关系,取得了较好的预测效果,但由参数生成的自适应生成图的可解释性较差。

与上述基线模型相比,本文模型全面考虑了节点间潜在的空间关系,并系统解决了交通数据中的时空异质性问题。通过结合这些优势,本文模型实现了最佳的预测性能。

Table 1 Prediction performance on Hangzhou metro dataset

表1 杭州地铁数据集上的预测结果

模型	15min			30min			45min			60min		
	RMSE	MAE	MAPE/%	RMSE	MAE	MAPE/%	RMSE	MAE	MAPE/%	RMSE	MAE	MAPE/%
AE	87.28	40.58	40.08	87.28	37.64	38.09	65.73	36.56	40.03	70.36	42.23	37.04
FNN	63.59	34.18	32.37	87.56	44.36	44.46	115.20	56.79	60.84	138.66	67.81	86.01
RNN	77.99	38.37	43.83	78.88	38.21	42.82	79.60	38.56	43.14	80.61	39.12	43.61
Seq2Seq	72.52	40.30	59.12	61.16	34.00	43.80	59.31	32.97	38.78	59.99	33.02	37.99
ASTGCN	62.39	35.84	45.06	67.14	39.06	44.45	63.52	34.70	40.33	104.46	49.62	55.05
STG2Seq	39.52	23.80	27.09	40.72	24.72	29.51	43.36	25.98	33.59	46.05	26.50	37.93
AGCRN	45.30	25.10	26.95	44.96	25.28	26.43	46.50	26.22	26.67	49.60	27.47	25.32
PVGCN	37.76	22.68	23.70	39.34	23.33	23.81	40.95	24.22	24.45	42.61	24.93	25.49
STDGRL	39.54	23.26	20.91	40.43	23.77	21.41	42.87	24.89	22.30	45.17	25.83	25.70
MGSTAtt	36.62	22.45	23.00	38.10	22.86	23.70	39.46	23.50	24.21	40.30	24.12	25.00

Table 2 Prediction performance on Shanghai metro dataset

表2 上海地铁数据集上不同模型的预测性能

模型	15min			30min			45min			60min		
	RMSE	MAE	MAPE/%	RMSE	MAE	MAPE/%	RMSE	MAE	MAPE/%	RMSE	MAE	MAPE/%
AE	141.1	53.86	51.05	117.27	49.28	46.83	102.41	46.90	45.66	100.99	46.29	43.50
FNN	77.18	35.96	34.87	108.52	46.31	46.63	146.33	58.25	62.02	176.56	68.71	76.80
RNN	168.6	65.12	87.25	173.08	67.74	85.57	175.90	68.15	83.95	177.42	68.94	84.67
Seq2Seq	149.8	62.96	79.49	94.28	44.25	62.28	90.24	41.61	46.83	88.51	41.01	43.78
ASTGCN	83.14	43.69	70.00	90.67	49.34	82.00	81.29	43.35	74.39	86.76	47.22	81.18
STG2Seq	47.19	24.98	23.26	50.58	26.17	26.79	52.68	26.75	28.49	56.81	28.22	34.30
AGCRN	54.80	26.16	28.62	54.94	26.71	30.19	57.77	27.70	29.48	60.58	28.71	30.06
PVGCN	45.97	23.29	26.83	47.83	24.16	27.23	52.02	25.33	27.92	55.27	26.29	28.69
STDGRL	46.86	23.72	21.43	49.29	24.37	21.66	52.90	25.42	22.48	57.39	26.58	23.41
MGSTAtt	45.20	22.78	26.10	46.32	23.35	25.89	48.13	24.21	27.23	50.21	24.86	27.79

4.5 消融实验

为了验证两种掩码矩阵的有效性,本文设计了掩码矩阵的不同组合实验。图3、图4显示了模型在杭州地铁数

据集和上海地铁数据集上不同邻接矩阵下的预测性能。由图可知,只使用基于距离关系构造的掩码矩阵或只使用流量序列相似性关系构造的邻接矩阵预测效果并不理想,

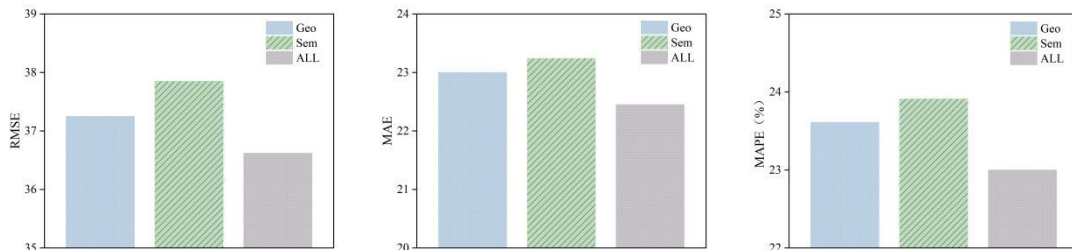


Fig. 3 Prediction performance of different adjacency matrix combinations on Hangzhou metro dataset

图3 不同掩码矩阵组合在杭州地铁数据集上的预测性能

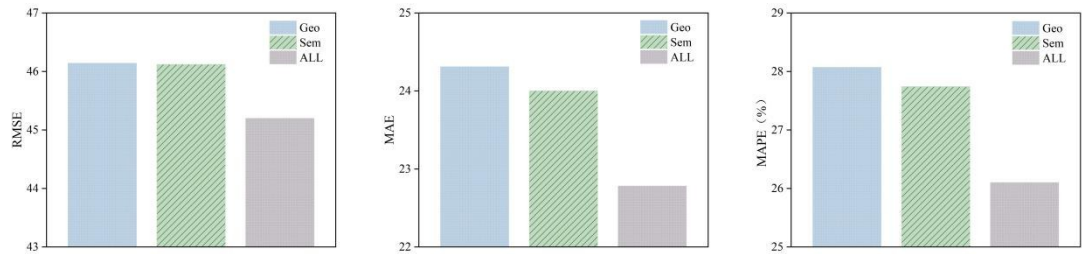


Fig. 4 Prediction performance of different adjacency matrix combinations on Shanghai metro dataset

图 4 不同掩码矩阵组合在上海地铁数据集上的预测性能

这可能是由于其只能分别反映网络中的欧氏空间关系或非欧氏空间关系。当在模型中同时使用距离掩码矩阵和相似性掩码矩阵时,模型的预测结果有较为显著的提升。

表 3、表 4 进一步展示了模型在不同邻接矩阵下多个步长的预测结果。

Table 3 Multi-step prediction performance comparison of different adjacency matrix combinations on Hangzhou metro dataset

表 3 不同掩码矩阵组合在杭州地铁数据集上的多步预测性能比较

掩码组合	15min			30min			45min			60min		
	RMSE	MAE	MAPE/%	RMSE	MAE	MAPE/%	RMSE	MAE	MAPE/%	RMSE	MAE	MAPE/%
M_{geo}	37.25	23.00	23.61	38.20	23.44	23.76	40.03	24.15	24.52	42.84	25.36	25.84
M_{sem}	37.85	23.24	23.91	39.05	23.75	24.13	40.87	24.31	24.59	42.77	25.12	26.08
$M_{geo} + M_{sem}$	36.62	22.45	23.00	38.10	22.86	23.70	39.46	23.50	24.21	40.30	24.12	25.00

Table 4 Multi-step prediction performance comparison of different adjacency matrix combinations on Shanghai metro dataset

表 4 不同掩码矩阵组合在上海地铁数据集上的多步预测性能比较

掩码组合	15min			30min			45min			60min		
	RMSE	MAE	MAPE/%	RMSE	MAE	MAPE/%	RMSE	MAE	MAPE/%	RMSE	MAE	MAPE/%
M_{geo}	46.14	24.31	28.07	49.91	25.93	28.50	56.10	27.84	29.46	63.12	29.85	30.86
M_{sem}	46.12	24.00	27.74	47.35	24.20	27.89	48.85	25.22	28.47	52.31	26.14	29.47
$M_{geo} + M_{sem}$	45.20	22.78	26.10	46.32	23.35	25.89	48.13	24.21	27.23	50.21	24.86	27.79

为了研究时空异质对预测效果的影响,本文设计了 3 个变体:①no-STE:不将时空嵌入作为输入以生成注意力机制的参数矩阵;②no-TE:时空嵌入中去除时间嵌入,只保留空间嵌入;③no-SE:时空嵌入中去除空间嵌入,只

保留时间嵌入。由图 5、图 6 可知,不使用时空嵌入时预测结果最差,从而验证了考虑空间异质性和时间异质性的必要性。表 5、表 6 进一步展示了模型在不同邻接矩阵下多个步长的预测结果。

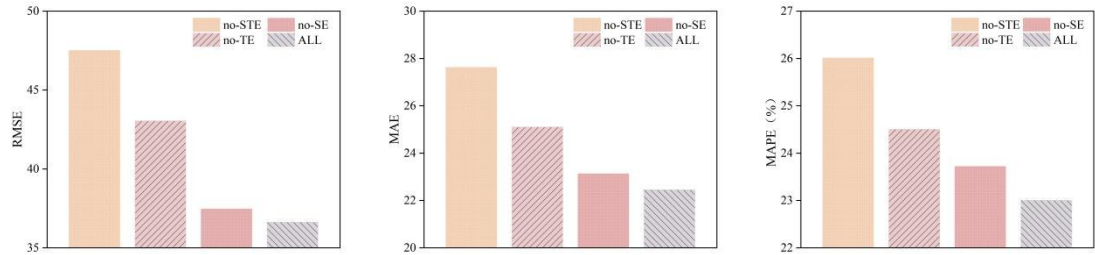


Fig. 5 Prediction performance of different spatio-temporal embedding combinations on Hangzhou metro dataset

图 5 不同时空嵌入组合在杭州地铁数据集上的预测性能

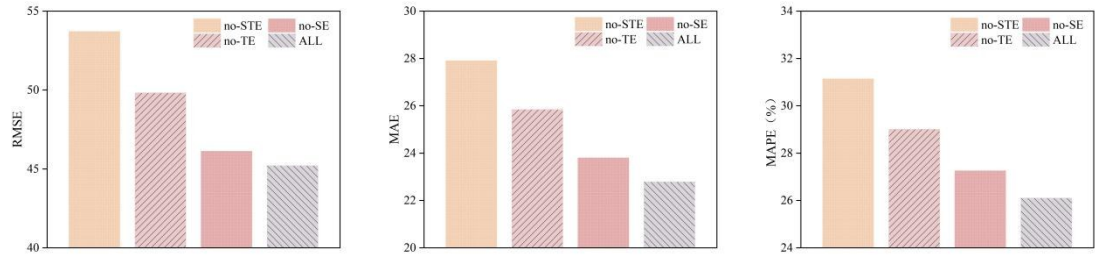


Fig. 6 Prediction performance of different spatio-temporal embedding combinations on Shanghai metro dataset

图 6 不同时空嵌入组合在上海地铁数据集上的预测性能

为了验证模型主要组成模块的有效性,本文设计了两个变体:①no-SSA:去除空间自注意模块;②no-TSA:去除

时间自注意模块。实验结果如图 7、图 8 所示。若在预测过程中不使用时间自注意模块或空间自注意模块,预测结

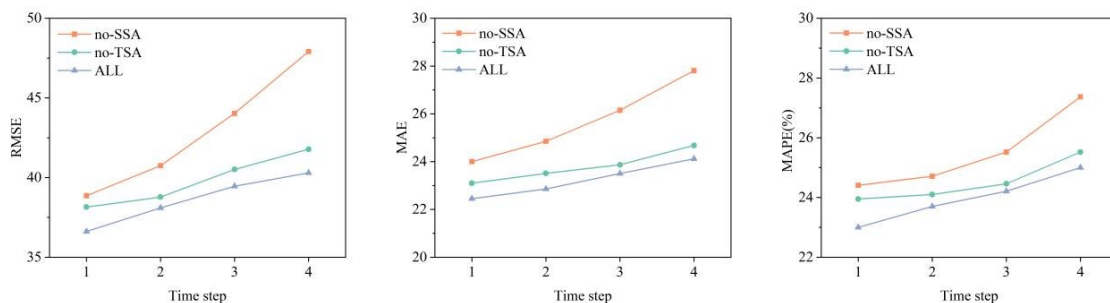
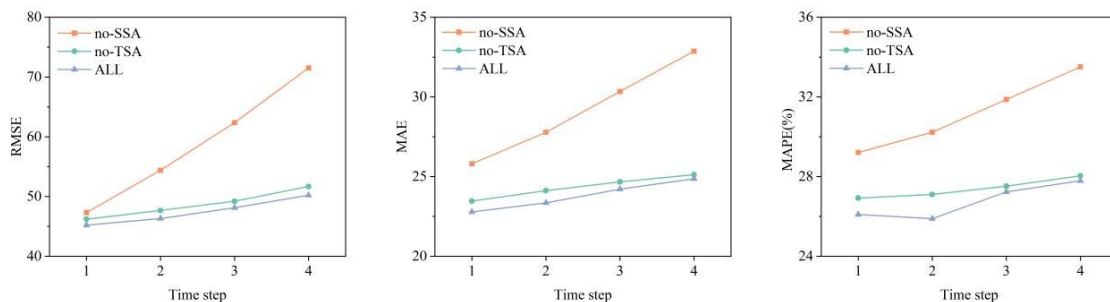
Table 5 Multi-step prediction performance comparison of different spatio-temporal embedding combinations on Hangzhou metro dataset**表 5 不同时空嵌入组合在杭州地铁数据集上的多步预测性能比较**

时空嵌入组合	15min			30min			45min			60min		
	RMSE	MAE	MAPE/%	RMSE	MAE	MAPE/%	RMSE	MAE	MAPE/%	RMSE	MAE	MAPE/%
no - STE	47.51	27.62	26.01	47.50	27.89	26.12	48.21	28.86	27.34	54.13	31.81	31.32
no - TE	43.04	25.10	24.50	43.41	25.35	25.45	44.65	25.74	25.23	45.23	26.32	27.05
no - SE	37.46	23.13	23.72	38.34	23.78	24.10	40.03	24.20	24.71	41.72	24.86	25.89
ours	36.62	22.45	23.00	38.10	22.86	23.70	39.46	23.50	24.21	40.30	24.12	25.00

Table 6 Multi-step prediction performance comparison of different spatio-temporal embedding combinations on Shanghai metro dataset**表 6 不同时空嵌入组合在上海地铁数据集上的多步预测性能比较**

时空嵌入组合	15min			30min			45min			60min		
	RMSE	MAE	MAPE/%	RMSE	MAE	MAPE/%	RMSE	MAE	MAPE/%	RMSE	MAE	MAPE/%
no - STE	53.70	27.90	31.14	57.21	28.95	31.16	62.10	30.87	32.64	66.51	32.35	35.76
no - TE	49.82	25.84	29.00	53.11	26.21	28.81	57.31	26.72	29.74	60.10	28.35	30.85
no - SE	46.12	23.80	27.25	48.23	24.23	27.51	49.68	25.10	28.00	51.13	25.83	28.91
ours	45.20	22.78	26.10	46.32	23.35	25.89	48.13	24.21	27.23	50.21	24.86	27.79

果并不理想,表明在预测过程中考虑数据间的时间依赖及空间依赖是至关重要的。

**Fig. 7 Prediction performance of different attention module combinations on Hangzhou metro dataset****图 7 不同注意力模块在杭州地铁数据集上的预测性能****Fig. 8 Prediction performance of different attention module combinations on Shanghai metro dataset****图 8 不同注意力模块在上海地铁数据集上的预测性能**

4.6 超参数分析

本文在杭州地铁数据集上对超参数进行实验,分别评估了注意力机制中权重矩阵隐藏维度大小以及多头注意力头数 n 对模型预测的影响。实验中改变了研究中的某一个超参数,并固定了其他超参数。参数调整总体上对预测结果影响并不大,即模型对参数不敏感,具有一定的稳定性,数值上仍具有最优值。如图9所示,当注意力机制的隐藏维数为16时,模型的预测误差最小,这可能是由于当隐藏维数过大时模型容易出现过拟合问题,进而影响预测性能。图10显示了注意力头数对预测性能的影响,该参数设置没有线性特征。当注意力头数为2时,模型的预测性相对较好,但随着注意力头数增加,特征融合的复杂度提升,模型性能反而下降。

5 结语

本文针对城市交通流量预测任务的非欧氏时空依赖挖掘欠缺,以及时空异质性感知能力不足的问题,提出一种基于多视图和时空一体化异质性感知的交通流量预测模型MGSTEAtt。该模型有机融合多图数据,并运用时空嵌入注意力攻克了交通流量预测的关键技术难题。具体而言,模型通过构建基于地理和语义两个视角的掩码矩阵,同时捕获局部和全局的空间依赖,以提升模型对节点间潜在空间关系的感知能力。同时将节点的时空嵌入整合到时空注意力机制中,为每个节点学习其特定的时空模式,以一体化考虑数据中的时空异质性。在两个真实世界

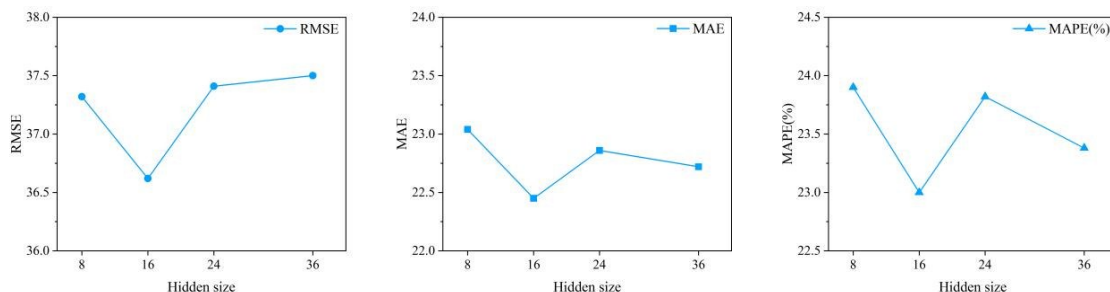


Fig. 9 Effect of hidden size on model performance in attention mechanisms

图9 注意力机制中隐藏维度大小对模型性能的影响

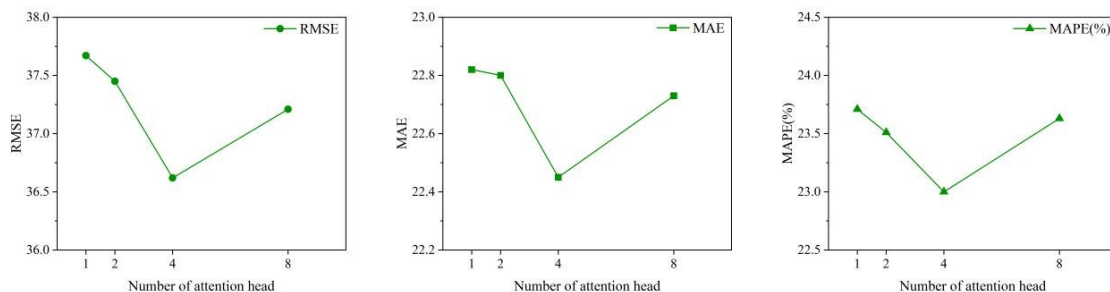


Fig. 10 Effect of number of attentional heads on model performance

图10 注意力头数对模型性能的影响

的交通数据集上进行实验验证,结果表明,本文模型相较于其他主流算法,在预测准确度上有所提升。

未来将从以下两方面完善相关工作:①在考虑节点的时空属性时采用时空嵌入方法,在处理实时性较强的数据时效果不够理想,未来可以探寻更加高效的信息处理方式;②自注意力机制虽然可以提升模型的长期预测效果,但也会增加模型的复杂度。因此,如何对自注意力计算进行简化以降低模型复杂度值得进一步研究。

参考文献:

- [1] GUEVARA L, AUAT C F. The role of 5G technologies: challenges in smart cities and intelligent transportation systems [J]. Sustainability, 2020, 12(16): 6469.
- [2] CHIU Y C, ZHENG H, VILLALOBOS J, et al. Modeling no-notice mass evacuation using a dynamic traffic flow optimization model [J]. Iie Transactions, 2007, 39(1): 83-94.
- [3] LIU J, GUAN W. A summary of traffic flow forecasting methods [J]. Journal of Highway and Transportation Research and Development, 2004, 21(3): 82-85.
- [4] BOX G E P, JENKINS G M, REINSEL G C, et al. Time series analysis: forecasting and control [M]. Hoboken: John Wiley & Sons, 2015.
- [5] BISHOP G, WELCH G. An introduction to the Kalman filter [EB/OL]. https://www.cs.cornell.edu/courses/cs4758/2013sp/materials/kalman_notes.pdf.
- [6] COVER T, HART P. Nearest neighbor pattern classification [J]. IEEE Transactions on Information Theory, 1967, 13(1): 21-27.
- [7] ZAREMBA W, SUTSKEVER I, VINYALS O. Recurrent neural network regularization [DB/OL]. <https://arxiv.org/pdf/1409.2329v1>.
- [8] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural Computation, 1997, 9(8): 1735-1780.
- [9] CHUNG J, GULCEHRE C, CHO K H, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling [DB/OL]. <https://arxiv.org/pdf/1412.3555v1>.
- [10] MA X, TAO Z, WANG Y, 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.
- [11] FU R, ZHANG Z, LI L. Using LSTM and GRU neural network methods for traffic flow prediction [C]// 2016 31st Youth Academic Annual Conference of Chinese Association of Automation, 2016: 324-328.
- [12] YU H, WU Z, WANG S, et al. Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks [J]. Sensors, 2017, 17(7): 1501.
- [13] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks [DB/OL]. <https://arxiv.org/pdf/1609.02907>.
- [14] YU B, YIN H, ZHU Z. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting [C]// Proceedings of the 27th International Joint Conference on Artificial Intelligence, 2018: 3634-3640.
- [15] LI Y, YU R, SHAHABI C, et al. Diffusion convolutional recurrent neural network: data-driven traffic forecasting [C]// Vancouver: International Conference on Learning Representations, 2018.
- [16] WU Z, PAN S, LONG G, et al. Graph WaveNet for deep spatial-temporal graph modeling [DB/OL]. <https://arxiv.org/pdf/1906.00121>.
- [17] LIU L, CHEN J, WU H, et al. Physical-virtual collaboration modeling for intra-and inter-station metro ridership prediction [J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 23(4): 3377-3391.
- [18] FENG S, WEI S, ZHANG J, et al. A macro-micro spatio-temporal neural network for traffic prediction [J]. Transportation Research Part C: Emerging Technologies, 2023, 156: 104331.
- [19] CHEN Z, LU Z, CHEN Q, et al. Spatial-temporal short-term traffic

- flow prediction model based on dynamical-learning graph convolution mechanism[J]. Information Sciences, 2022, 611: 522–539.
- [20] 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.
- [21] 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, 2017: 6000–6010.
- [22] GUO S, LIN Y, FENG N, et al. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting[C]// Proceedings of the AAAI Conference on Artificial Intelligence, 2019: 922–929.
- [23] ZHENG C, FAN X, WANG C, et al. GMAN: a graph multi-attention network for traffic prediction[C]// Proceedings of the AAAI Conference on Artificial Intelligence, 2020: 1234–1241.
- [24] BAI L, YAO L, KANHERE S S, et al. STG2Seq: spatial-temporal graph to sequence model for multi-step passenger demand forecasting [C]// Proceedings of the 28th International Joint Conference on Artificial Intelligence, 2019: 1981–1987.
- [25] LIU D, WANG J, SHANG S, et al. MSDR: Multi-step dependency relation networks for spatial temporal forecasting [C]// Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2022: 1042–1050.
- [26] YAN Z, YANG H, GUO D, et al. Improving airport arrival flow prediction considering heterogeneous and dynamic network dependencies[J]. Information Fusion, 2023, 100: 101924.
- [27] XIE P, MA M, LI T, et al. Spatio-temporal dynamic graph relation learning for urban metro flow prediction [J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(10): 9973–9984.
- [28] LI Z N, ZHAO Z H. Urban traffic flow prediction considering spatiotemporal information based on GCN and LSTM[J]. Journal of Jilin University (Information Science Edition), 2025, 43(1): 187–184.
李正楠, 赵智辉. 结合 GCN 和 LSTM 考虑时空信息的城市交通流量预测[J]. 吉林大学学报(信息科学版), 2025, 43(1): 187–184.
- [29] CHEN J, ZHENG L, HU Y, et al. Traffic flow matrix-based graph neural network with attention mechanism for traffic flow prediction[J]. Information Fusion, 2024, 104: 102146.
- [30] GENG Z, XU J, WU R, et al. STGAFormer: spatial-temporal gated attention transformer based graph neural network for traffic flow forecasting [J]. Information Fusion, 2024, 105: 102228.
- [31] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]// Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770–778.
- [32] SHUMAN D I, NARANG S K, FROSSARD P, et al. The emerging field of signal processing on graphs: extending high-dimensional data analysis to networks and other irregular domains[J]. IEEE Signal Processing Magazine, 2013, 30(3): 83–98.

(责任编辑:黄健)