

# 基于时空多头图注意力网络的交通流预测

梁秀霞, 夏曼曼\*, 何月阳, 梁涛  
(河北工业大学人工智能与数据科学学院, 天津 300132)

**摘要:** 针对当前路网交通流量预测方法中存在的挖掘复杂的动态时空特性和长距离的空间依赖性能力不足等问题, 基于多头自注意力网络构建一种新型的交通流预测模型. 考虑到交通流在不同周期尺度下呈现出的高度相似性, 以及静态时空特征, 引入日和周这2种周期尺度下的数据张量作为模型输入, 来表达交通流数据的时间相似性, 并通过输入数据的时空位置编码获取其静态时空特征. 考虑到交通流的动态时空特性和长距离的空间依赖性, 主体模型基于多头自注意力机制分别设计时间多头注意力模块和空间多头注意力模块. 时间多头注意力模块利用一个图掩码矩阵获得局部注意力, 并将其融合到一个多头自注意力中, 以提取交通流的动态时间特征. 空间多头注意力模块利用两个图掩码矩阵获得局部注意力和全局注意力, 并将其融合到一个多头自注意力中, 以提取路网节点的动态空间特征和长距离的空间依赖性. 最后, 设计一个门控融合模块自适应地融合交通流数据的时空相关性特征. 在三个真实交通流基准数据集 PEMS04, PEMS07 和 PEMS08 上进行模型的有效性验证, 结果表明, 所建模型在3个数据集上的3个预测精度指标与其他精度最高模型相比, 平均提高了 4.437%, 2.930%, 4.275%.

**关键词:** 智能交通; 多头图注意力网络; 图掩码机制; 特征融合; 时空数据位置嵌入

**基金项目:** 河北省自然科学基金(No.F2021202022)

**中图分类号:** TP273; U491.1

**文献标识码:** A

**文章编号:** 0372-2112(2024)02-0500-10

**电子学报 URL:** <http://www.ejournal.org.cn>

**DOI:** 10.12263/DZXB.20230474

## Traffic Flow Prediction Based on Spatio-Temporal Multi-head Graph Attention Network

LIANG Xiu-xia, XIA Man-man\*, HE Yue-yang, LIANG Tao  
(School of Artificial Intelligence, Hebei University of Technology, Tianjin 300132, China)

**Abstract:** To overcome the drawbacks of current traffic network flow prediction methods, such as the low capability of capturing highly dynamic spatio-temporal correlation and long-term spatial dependence, this paper constructs a novel traffic flow prediction model based on multi-head self-attention network. The model takes the data tensor at daily period and weekly period scales as model inputs to express the temporal similarity of traffic flow data, and obtains its static spatio-temporal characteristics by encoding the spatio-temporal position embedding of the input data. The main model designs temporal multi-head attention module and spatial multi-head attention module respectively based on multi-head self-attention mechanism for considering the dynamic spatio-temporal characteristics of traffic flow and the long-range spatial dependences. The temporal multi-head attention module obtains the local attention using a graph masking matrix and fuses it into a multi-head self-attention to extract the dynamic temporal characteristics of traffic flow. The spatial multi-head attention module obtains the local attention and global attention using two graph masking matrices and fuses them into a multi-head self-attention to extract the dynamic spatial characteristics and long-range spatial dependences of road network nodes. Finally, a gated fusion module is designed to adaptively fuse the spatio-temporal correlation characteristics of traffic flow data. The effectiveness of the proposed model is verified on three real traffic flow benchmark datasets PEMS04, PEMS07 and PEMS08, and the results show that the three prediction accuracy metrics of the proposed model on the three data sets improved by 4.437%, 2.930%, and 4.275% on average compared with the other models with the highest accuracy.

**Key words:** intelligent transportation; multi-head graph attention network; graph masking mechanism; feature fu-

sion; spatio-temporal data position embedding

Foundation Item(s): Natural Science Foundation of Hebei Province (No.F2021202022)

## 1 引言

随着城市化进程的加快,大量人口迅速向城市聚集.在许多城市,私人车辆数量的迅速增加和对公共交通服务日益增长的需求给现有的交通系统带来了巨大的压力.许多城市致力于开发智能交通系统<sup>[1]</sup>,它可以提供高效的交通管理、准确的交通资源配置和优质的交通服务.交通流预测是智能交通系统的重要组成部分,准确、及时地预测交通流已成为大数据时代智能交通领域研究的迫切任务.

早期,交通流预测主要基于参数化方法,例如自回归移动平均(AutoRegressive Integrated Moving Average, ARIMA)模型及以 ARIMA 为基模型衍生的变体模型.该类方法后续预测是由前步预测决定的,因此,在多步预测过程中会不断积累预测误差.

近年来,随着计算性能的提高,许多先进的深度学习模型因其学习能力强、预测性能准确稳定、可提取深度特征等优点被广泛应用于交通流预测. Lv 等人<sup>[2]</sup>建立了一个自编码神经网络提取交通流特征,并取得了优于传统模型的性能.而此类模型基于全连接神经网络,不能有效提取其空间特征,模型的代表能力有限.

考虑到上述局限,部分研究者从空间特征考虑,引入卷积神经网络(Convolution Neural Network, CNN),然而,常规卷积主要面向规则的欧氏空间结构,并不适用于路网拓扑图结构.因此,一些研究者为了将道路网络拓扑结构集成到预测模型中,引入图神经网络(Graph Neural Network, GNN)来捕捉路网交通流数据的时空相关性.自 2005 年图神经网络的概念被首次提出起,图神经网络已经在多个领域的应用中取得了令人满意的结果.这些领域的的数据通常以拓扑图的形式构成.交通网络可以表示为图,其中节点表示交通传感器,边和其权重由传感器之间的连通性和欧几里得距离来确定.所以,图神经网络(Graph Neural Network, GNN)被证明更适合对交通数据的底层图结构进行建模<sup>[4]</sup>,基于 GNN 的方法在交通预测中得到了广泛探索.杨建喜等人<sup>[5]</sup>以多尺度卷积算子及门控循环单元为核心单元,提出一种多周期组件时空神经网络模型,以周、日和近期组成模型的周期组件,通过门控循环单元提取交通数据的时间相关性,通过多尺度卷积核捕获路网节点的空间相关性. Guo 等人<sup>[6]</sup>提出一种基于注意力的时空图神经网络(Attention based Spatial-Temporal Graph Neural Networks, ASTGNN),在时间维度上,设计一个新的自注意力机制

用以捕捉交通流数据的动态时间特征,在空间维度上,设计一个动态图卷积模块,并利用注意力机制捕捉动态空间特征.

然而,基于 GNN 的模型在交通预测方面仍有三个主要限制.首先,路网不同路段的交通流之间存在高度动态的空间相关性<sup>[4]</sup>,而不止是静态的,然而现有的方法主要以静态方式对空间相关性进行建模,这限制了学习动态空间相关性的能力.其次,交通流数据会呈现出明显的周期性,例如,由于工作日和非工作日的周期性,人们的驾驶出行表现出一定的规律性,体现在交通流上,即同一路段每间隔一定周期,交通流状态呈现一定的相似性.第三,由于城市的功能划分,两个距离较远的位置,可能反映出相似的交通模式,这意味着较远位置之间也是具有较强的依赖性的.现有的基于 GNN 的方法通常是在预定义的图结构上进行训练,对路网节点之间相互影响的空间特征挖掘还不够充分,很难捕获长距离节点之间的空间依赖性.

针对上述不足,本文从挖掘交通数据之间高度动态的空间相关性和长距离的空间依赖性的角度出发,构建基于时空多头图注意力神经网络(Spatio-Temporal Multi-head Graph Attention Network, STMGAN)的交通流预测模型,并在 3 个公开数据集验证其预测精度以及预测时长上的多重优势.主要贡献如下:

(1)提出基于时空多头图注意力网络的模型框架,用于交通流预测,充分解决了由交通数据的复杂时空特性引起的问题.

(2)设计一个新的空间多头自注意力模块,利用不同的图掩码机制对交通数据中的节点进行建模,将空间局部邻域信息和空间全局语义信息融合到自注意力模块中,挖掘各个节点之间的动态空间特性和长距离空间依赖性.

(3)设计一种门控融合,以自适应地融合由空间和时间注意力机制提取的信息.

## 2 问题描述与基础模型

由于交通网络数据的拓扑特性,通常将交通网络表示为一张空间图  $G=(V, E, A)$ ,其中  $V$  代表道路节点的集合,  $|V|=N$  为节点的个数,一般情况下,道路节点通常为部署在某个路段上的传感器.  $E$  是连接两个道路节点的边的集合,  $A \in \mathbf{R}^{N \times N}$  是反映图  $G$  内部节点连接情况的邻接矩阵,对于任意的两个节点  $v_i, v_j \in V$  且  $(v_i, v_j) \in E$ ,即节点相连,则邻接矩阵中的元素  $a_{ij}$  为 1,否

则为0.

交通流数据常见的有交通速度、交通流量和交通密度,本文关注交通流量的预测.在时间步长 $t$ 下,交通网络上采集的数据可以形成一个特征矩阵 $\mathbf{X}_{(t)} = (\mathbf{X}_t^1, \mathbf{X}_t^2, \dots, \mathbf{X}_t^N)^T \in \mathbf{R}^{N \times F}$ , $\mathbf{X}_{(t)}$ 表示 $t$ 时刻所有节点的所有特征值.给定交通网络的空间图 $\mathbf{G}$ 和在其上连续 $T_h$ 个历史时间步长收集的数据,交通流预测任务的目标为预测该交通路网在接下来 $T_p$ 个时间步长的交通状态,可以表示为

$$\left[ \mathbf{X}_{(t-T_h+1):t}, \mathbf{G} \right]^f \rightarrow \mathbf{X}_{(t+1):(t+T_p)} \quad (1)$$

其中, $\mathbf{X}_{(t-T_h+1):t}$ 为预测模型所需要的 $T_h$ 个历史时间步数据, $\mathbf{X}_{(t+1):(t+T_p)} \in \mathbf{R}^{N \times F \times T_p}$ 为模型的预测输出, $f$ 为交通流预测模型.

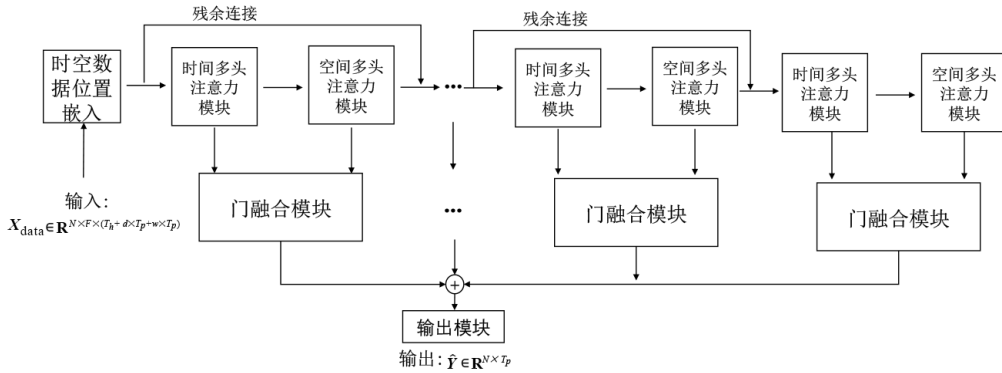


图1 时空多头图注意力网络模型框架

自注意力机制是注意力机制中的一种,多头自注意力<sup>[7]</sup>是实践中采用最广泛的,它能够处理来自不同子空间的信息.本文利用多头自注意力机制设计时间多头注意力模块和空间多头注意力模块,两个模块相互交错连接,以分别挖掘时间和空间相关特性.图1演示了时间多头注意力模块和空间多头注意力模块如何相互协作.

为了制定自注意力操作,本文使用切片表示法,即,对于张量 $\chi \in \mathbf{R}^{N \times F \times T}$ , $t$ 时刻切片是矩阵 $\mathbf{X}_{:,t} \in \mathbf{R}^{N \times F}$ ,节点 $n$ 切片是 $\mathbf{X}_{n,:} \in \mathbf{R}^{F \times T}$ .

### 3.1 数据周期性的处理

交通数据是由人类日常活动产生的,并且是具有周期性的,如图2所示,交通流数据在相邻的天和周中周期性地显示出来,因此在交通流数据预测模型中考虑到数据的周期性是至关重要的,显式地建模周期性是有利于预测的<sup>[8,9]</sup>.为了进一步考虑了隐藏在交通数据中两种类型的周期性模式,即周周期性和日周期性,在预测未来 $T_p$ 时间步长的流量时,除了

## 3 时空多头图注意力网络

时空多头图注意力网络模型架构如图1所示.模型框架由一个时空数据位置嵌入、 $l$ 个时间多头注意力模块、 $l$ 个空间多头注意力模块和 $l$ 个门融合模块组成.交通流原始数据首先经过时空数据位置嵌入,通过图拉普拉斯特征向量对路网结构进行空间位置编码,通过正余弦形式对输入序列的时间进行时间位置编码,由此获得全局静态图结构信息以及时间位置信息.进而通过时间多头注意力模块和空间多头注意力模块,分别捕获时间和空间相关性,最终通过门融合模块,自适应地融合时间和空间特征.为了避免梯度消失的问题,将残差连接从时间多头注意力模块的输入添加到空间多头注意力模块的输出.为了获得最终输出,输出模块将隐藏的特征投影到所需的输出维度.

考虑过去 $T_h$ 时间步长中的历史记录外,本文还引入了另外两个数据源,即周周期性数据源、日周期性数据源.

图2展示了2018年1月1日至7日路段11的交通流流量图,2018年1月8日至14日路段11的交通流流量图.

#### (1) 周周期性数据源

引入过去 $w$ 周内同一天的交通记录的 $T_p$ 切片作为周周期性数据源 $\chi_w$ ,如图3的 $\chi_w$ 部分所示.

$$\chi_w = (\mathbf{X}_{t_0-7*w*q+1}, \dots, \mathbf{X}_{t_0-7*w*q+T_p}, \mathbf{X}_{t_0-7*(w-1)*q+1}, \dots, \mathbf{X}_{t_0-7*(w-1)*q+T_p}, \dots, \mathbf{X}_{t_0-7*q+1}, \dots, \mathbf{X}_{t_0-7*q+T_p}) \in \mathbf{R}^{N \times F \times w * T_p} \quad (2)$$

其中, $t_0$ 代表预测切片的时间, $q$ 代表一天的时间(24小时).

#### (2) 日周期性数据源

引入过去连续 $d$ 天中每一天的交通记录的 $T_p$ 切片作为日周期性数据源 $\chi_d$ ,如图3的 $\chi_d$ 部分所示.

$$\chi_d = (\mathbf{X}_{t_0-d*q+1}, \dots, \mathbf{X}_{t_0-d*q+T_p}, \mathbf{X}_{t_0-(d-1)*q+1}, \dots, \mathbf{X}_{t_0-(d-1)*q+T_p}, \dots, \mathbf{X}_{t_0-q+1}, \dots, \mathbf{X}_{t_0-q+T_p}) \in \mathbf{R}^{N \times F \times d * T_p} \quad (3)$$

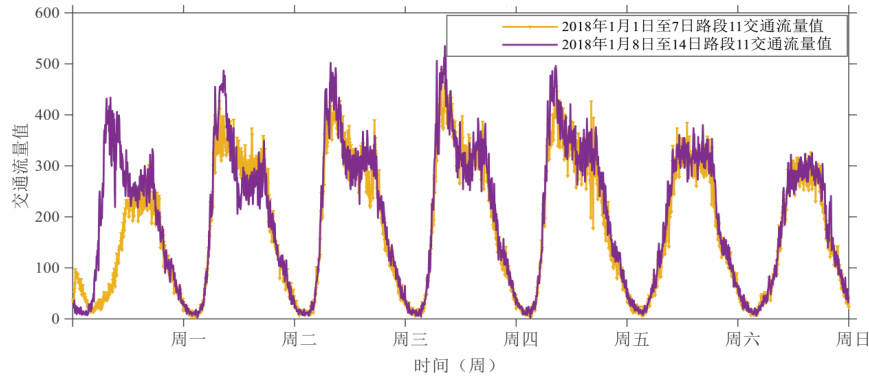


图2 路段11 PEMS04交通流量示例图

其中,  $t_0$  代表预测切片的时间,  $q$  代表一天的时间(24小时).

在获得周周期张量  $\chi_w$  和日周期张量  $\chi_d$ , 本文将它们与过去  $T_h$  时间步长的张量  $\chi$  拼接起来, 即得  $\chi_{\text{data}} = (\chi \parallel \chi_d \parallel \chi_w) \in \mathbf{R}^{N \times F \times (T_h + d^* T_p + w^* T_p)}$  作为模型的输入.

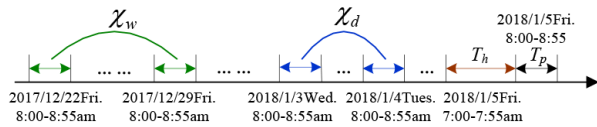


图3 构建周期序列输入的示例图

### 3.2 时空数据位置嵌入

#### (1) 时间数据位置嵌入

在本文模型中, 动态完全由多头自注意机制建模. 注意力通过加权和函数建立了输入和目标之间的依赖关系, 这将导致序列的位置发生变化时, 注意力机制的计算结果不会产生任何变化. 然而, 次序信息在时间序列建模任务中起着重要作用, 因为邻近的序列往往更相关, 例如, 下午 5:00 的交通流量比下午 4:00 的交通流更能预测 5:30 的流量. 因此, 对模型引入次序信息可能会有效地提高模型预测的精度. 为此, 本文为  $\chi_{\text{data}}^{(0)}$  中的每个元素表示引入了位置编码机制, 在序列输入到时间多头注意力模块之前, 对序列进行位置编码, 得到序列的位置嵌入. 为了简单起见, 本文为位置  $t$  处的输入元素选择固定位置嵌入, 如式(4)(5)所示, 并且每个向量维度  $d: 1 \leq d \leq d_{\text{model}}$ .

$$E_{\text{tpe}}(t, 2d) = \sin(t/10\,000^{2d/d_{\text{model}}}) \quad (4)$$

$$E_{\text{tpe}}(t, 2d+1) = \cos(t/10\,000^{2d/d_{\text{model}}}) \quad (5)$$

其中,  $t$  是输入中每个元素的时间信息. 当输入包括周周期张量和日周期张量时, 引入时间位置嵌入有助于模型更好地识别  $\chi, \chi_w, \chi_d$  之间的相对位置关系.

#### (2) 空间数据位置嵌入

除了动态变化的交通条件外, 每个空间节点还与一些静态特征相关联, 这些静态特征主要由空间特征

决定. 为了表示道路网络的结构, 本文使用图拉普拉斯特征向量, 它可以更好地描述图中节点之间的距离, 并保留图的全局结构信息<sup>[10]</sup>.

首先, 获得归一化拉普拉斯矩阵, 之后进行特征值分解, 获得特征值  $\lambda$  和特征向量矩阵  $U$ , 具体如式(6)所示, 最后, 选取前  $k$  个最小非 0 特征值对应的特征向量作为空间位置嵌入, 其中, 经实验验证  $k=8$  最佳.

$$\Delta = I - D^{-1/2} A D^{-1/2} = U^T \Lambda U \quad (6)$$

其中,  $A$  是邻接矩阵,  $D$  是度矩阵,  $I$  是单位矩阵.

最后, 通过简单地将上述嵌入向量求和, 得到时空数据位置嵌入的输出  $\chi_{\text{emb}}$ :

$$\chi_{\text{emb}} = \chi_{\text{data}} + E_{\text{spe}} + E_{\text{tpe}} \quad (7)$$

### 3.3 时间多头注意力模块

在不同的时间切片中, 交通状况之间存在相关性(例如, 周期性的、趋势性的), 并且相关性在不同的情况下有所不同. 因此, 本文使用了一个时间多头注意力模块来挖掘不同节点的动态时间模式, 如图 4 所示. 形式上, 对于节点  $n$ , 在时间维度上应用多头自注意力机制, 首先被映射到三个特征空间, 获得查询向量  $Q_n^{(k)}$ , 键向量  $K_n^{(k)}$  和价值向量  $V_n^{(k)}$ , 接着获得节点  $n$  的所有时间序列之间的时间相关性  $T_n^{(k)}$ , 以及第  $k$  个注意力分数  $\beta_n^{(k)}$ , 最终将不同的注意力机制拼接, 获得节点  $n$  的所有时间序列之间的时间相关性, 具体过程如下所示.

$$T_n^{(k)} = \frac{\langle (Q_n^{(k)}), (K_n^{(k)})^T \rangle}{\sqrt{d'}} \quad (8)$$

$$\beta_n^{(k)} = \text{softmax}(T_n^{(k)}) \quad (9)$$

$$\text{TSA}(Q_n^{(k)}, K_n^{(k)}, V_n^{(k)}) = \sum_{k=1}^K \beta_n^{(k)} \times V_n^{(k)} \quad (10)$$

其中,  $\langle \cdot, \cdot \rangle$  代表内积操作,  $Q_n^{(k)} = X_{:,n}; W_Q^{(k)}, K_n^{(k)} = X_{:,n}; W_K^{(k)}, V_n^{(k)} = X_{:,n}; W_V^{(k)}, W_Q^{(k)}, W_K^{(k)}, W_V^{(k)} \in \mathbf{R}^{d \times d'}$  是可学习参数.

对于式(10)的多头时间自注意力, 节点  $n$  的所有时间片刻之间会进行交互, 然而, 越近的时间对下一时刻的影响越大. 因此, 引入了一个掩码矩阵  $M_{\text{adjt}}$ , 以捕获近距离的时间相关性.

$M_{\text{adjt}}$  是由二进制组成,当只有两个时间片段的距离小于阈值 $\lambda$ 时,值设置为1,否则为0.在此基础上,进一步设计一个时间多头注意力,即时间局部自注意力,可以定义为

$$\text{AdjTSA}(Q_n^{(k)}, K_n^{(k)}, V_n^{(k)}) = \prod_{k=1}^K \left\{ \sum (\text{softmax}(T_n^{(k)} \odot M_{\text{adjt}}) \times V_n^{(k)}) \right\} \quad (11)$$

其中, $\odot$ 为Hadamard乘积, $\prod$ 为串联操作.

为降低模型的计算复杂度,本文将异构注意力融合到一个多头自注意力块中,具体而言,注意力头部包括两种类型,即TSAH和AdjTSAH,分别对应上述两种类型的注意力机制.这些注意力的结果被拼接并投影以获得输出,形式上如式(12)所示:

$$Z_T = \oplus (\text{TSA}, \text{AdjTSA}) W^{(T)} \quad (12)$$

其中, $\oplus$ 代表拼接操作, $W^{(T)}$ 为可学习参数.

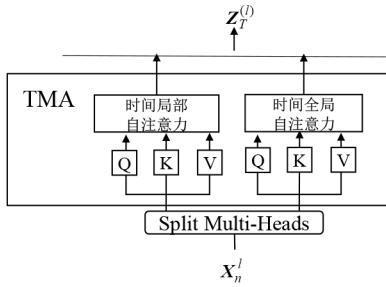


图4 时间多头注意力模块图

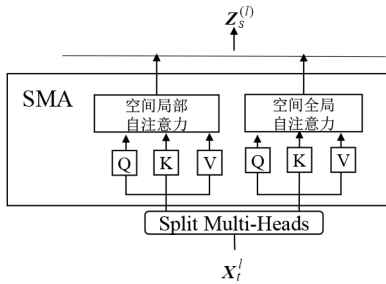


图5 空间多头注意力模块

### 3.4 空间多头注意力模块

一条道路的交通状况会受到其他道路不同程度的影响,且这种影响随着时间的推移是高度动态相关的.为了对这些特性进行建模,本文设计了一个空间多头注意力模块,如图5所示.若使用简单的自注意力机制,每个节点会与所有节点交互,然而,只有少数节点之间的交互是必要的,主要为邻近节点和非邻近但处于相似功能区的节点.非邻近节点间的相似性,如果采用欧氏距离来度量,由于非邻近的缘故,其欧氏距离会很大,则会被认为非邻近节点间的作用关系较弱.因此,引入了两个图掩码矩阵 $M_{\text{adj}}$ 和 $M_{\text{glos}}$ ,使其同时捕获交通数据中邻近节点空间相关性和全局节点空间相关性.

对于邻近节点空间相关性,本文构造一个二进制

邻近掩码矩阵 $M_{\text{adj}}$ ,当两个节点之间的距离(即图中的跳数)小于阈值 $\beta$ 时,值为1,否则为0.由此可以消除彼此距离较远的节点的注意力.对于全局节点空间相关性,本文构造一个二进制全局掩码矩阵 $M_{\text{glos}}$ ,使用权重动态时间扭曲(Weighted Dynamic Time Warping, WDTW)算法<sup>[11]</sup>计算节点之间历史交通流的相似性,为每个节点选择相似度最高的 $K$ 个节点作为全局邻居,将全局邻居和当前节点的二进制掩码矩阵值设置为1,否则为0,由此可以找到非邻近但因处于相似功能区而表现出的相似交通流模式的节点.

在以上两个图掩码矩阵的基础上,本文进一步设计了两个空间多头注意力,即空间局部自注意力和空间全局自注意力(AdjSSA和GloSSA),定义如下:

$$S_t^{(k)} = \frac{\langle (Q_t^{(k)}, (K_t^{(k)}))^T \rangle}{\sqrt{d'}} \quad (13)$$

$$\text{AdjSSA}(Q_t^{(k)}, K_t^{(k)}, V_t^{(k)}) = \prod_{k=1}^K \left\{ \sum (\text{softmax}(S_t^{(k)} \odot M_{\text{adj}}) \cdot V_t^{(k)}) \right\} \quad (14)$$

$$\text{GloSSA}(Q_t^{(k)}, K_t^{(k)}, V_t^{(k)}) = \prod_{k=1}^K \left\{ \sum (\text{softmax}(S_t^{(k)} \odot M_{\text{glos}}) \cdot V_t^{(k)}) \right\} \quad (15)$$

其中, $Q_t^{(k)} = X_{t,:} W_Q^{(k)}$ ,  $K_t^{(k)} = X_{t,:} W_K^{(k)}$ ,  $V_t^{(k)} = X_{t,:} W_V^{(k)}$ ,  $W_Q^{(k)}, W_K^{(k)}, W_V^{(k)} \in \mathbf{R}^{d \times d'}$ 是可学习参数.

为降低模型的计算复杂度,本文将异构注意力融合到一个多头自注意力块中,具体而言,注意力头部包括两种类型,即AdjSSAH和GloSSAH,分别对应上述两种类型的注意力机制.这些注意力的结果被拼接并投影以获得输出,形式上如下所示:

$$Z_S = \oplus (\text{AdjSSA}, \text{GloSSA}) W^{(S)} \quad (16)$$

### 3.5 时空融合

道路在特定时间步长的交通状况与其先前的值和其他道路的交通状况相关.若直接将交通流数据通过空间特征提取模块和时间特征提取模块进行预测,则不能有效捕获交通流的时空交融性.为此,本文设计了一个时空融合模块来融合时空特征,主要采用门控融合来自适应地融合空间和时间特征,如图6所示,使模型可以更有效地学习数据中的时空交融性.

在第 $l$ 块时间多头注意力模块和空间多头注意力模块的输出分别为 $Z_T^{(l)}, Z_S^{(l)}$ ,其融合如下:

$$Z^{(l+1)} = m \odot Z_S^{(l)} + (1 - m) \odot Z_T^{(l)} \quad (17)$$

$$m = \sigma(Z^{(l)} W_{m,1} + (Z^{(l)})^T W_{m,2} + b_m) \quad (18)$$

其中, $W_{m,1} \in \mathbf{R}^{d \times d}$ ,  $W_{m,2} \in \mathbf{R}^{d \times d}$ ,  $b_m \in \mathbf{R}^d$ 是可学习的参数, $\sigma(\cdot)$ 表示sigmoid激活函数.

与门控循环神经网络(Gate Recurrent Unit, GRU)类似,门控融合机制可以控制神经网络中信息的传递.如图6所示,在时空融合的模块中,本层输入 $Z^{(l)}$ 为上一

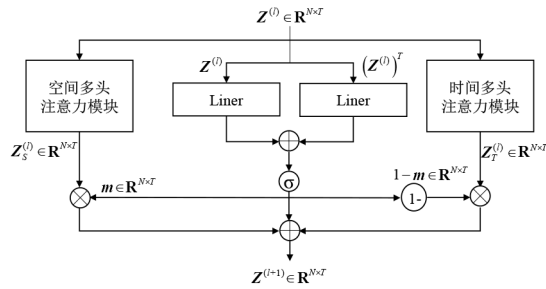


图6 时空融合模块

层的输出,首先  $\mathbf{Z}^{(l)}$  分别进入时间多头注意力模块和空间多头注意力模块获得时间特征和空间特征,同时  $\mathbf{Z}^{(l)}$  也进入两个Liner层.需要注意的是,进行空间和空间特征提取时,由于考虑的维度不同,需要将时间多头注意力模块和空间多头注意力模块对应的线性层的输入进行转置.从时间多头注意力模块和空间多头注意力模块输出的结果分别和权重矩阵进行点乘,然后将两者相加进行最后的输出,该步骤的计算式如式(17)所示,其中权重矩阵  $\mathbf{m}$  是通过式(18)获得.  $\mathbf{Z}^{(l)}$  在进入Liner层之前由于时间提取特征和空间提取特征获得的维度不同,需要先进行转置,接着通过激活函数进行非线性激活,最后进行按位相加得到门控融合权重  $\mathbf{m}$ .

### 3.6 输出模块

对每一层门控融合模块获得的输出进行相加,得到了一个最终的隐藏表示  $\mathbf{Z}^{(l+1)}$ .最后,将这个最终的隐藏表示通过一个两层的全连接前馈神经网络模块进行预测,表达式为

$$\hat{\mathbf{Y}} = \text{ReLU}(\mathbf{Z}^{(l+1)} \mathbf{W}_0) \mathbf{W}_1 \quad (19)$$

其中,  $\mathbf{Z}^{(l+1)}$  为所有层的特征隐藏表示之和,  $\mathbf{W}_0$  和  $\mathbf{W}_1$  为可训练的参数矩阵.

## 4 实验设计与结果分析

### 4.1 基准数据集

实验采用交通流预测领域的3个基准数据集,分别为PEMS04, PEMS07, PEMS08,3个数据集分别来自加利福尼亚州的3个地区,由加州运输绩效测量系统(PeMS)收集.实验选择3个数据集中的交通流量特征构建预测任务,采用每隔5 min汇总过的数据,每天共有288个时间点数据.采用Z-Score方法对交通流数据统一进行标准化.数据集具体描述如表1所示.

### 4.2 基线模型

模型STMGAN将与以下代表性基线模型进行比较:

(1)SVR<sup>[12]</sup>,即支持向量机.它利用支持向量机进行回归预测.

(2)LSTM<sup>[13]</sup>,即长短期记忆网络.它是一种特殊的

表1 基准数据集描述

数据集	节点数	边数	时间范围	时间点数
PEMS04	307	340	2018-01-01 ~2018-02-08	16 992
PEMS07	883	866	2017-05-01 ~2017-08-31	28 224
PEMS08	170	295	2016-07-01 ~2017-08-31	17 856

RNN模型.

(3)DCRNN<sup>[14]</sup>,即扩散卷积递归神经网络.它将交通网络建模为有向图,并使用双向随机游走膨胀卷积来学习空间表示,再将其输入GRU以捕获时间变化.

(4)STGCN<sup>[15]</sup>,即时空图卷积神经网络.它使用两个时空卷积模块建模交通流预测.

(5)Graph WaveNet<sup>[16]</sup>,即图小波网络模型.它将图卷积与膨胀卷积相结合来捕获时空相关性.

(6)ASTGCN<sup>[17]</sup>,即基于注意力时空图卷积网络.它由三个独立的组件组成,分别对交通流数据的当前时刻、日周期、周周期进行建模,再对三个组件的输出进行加权求和,最终获得预测值.

(7)ASTGNN<sup>[6]</sup>,即基于注意力的时空图神经网络.它设计了一个新的自注意机制,能有效捕捉交通数据的时空动态特性.

### 4.3 实验设置

数据集以基线模型相同的比例进行划分,即按6:2:2的比例将PEMS04, PEMS07, PEMS08这3个数据集拆分成训练集、验证集和测试集.

使用历史1 h(即  $T_h=12$ )的数据预测未来1 h(即  $T_p=12$ )的交通流流量.实验环境为NVIDIA GeForce RTX 3090 GPU和80 GB CPU, PyTorch 1.10.1和Python 3.9.7.隐藏维度  $d$  是在{16, 32, 64, 128, 256}寻找最佳预测结果,设置时空层数  $l$  为6,设置全局邻居的节点个数  $K$  为5,设置时间多头注意力模块中的掩码阈值  $\lambda$  为8,设置时间多头注意力模块中的掩码阈值  $\beta$  为5,设置时间多头自注意力的多头数为8,其中TSAH为6, AdjTSAH为2,设置空间多头自注意力的多头为8,其中AdjSSAH为4, GloSSAH为4,各项参数设置如表2.使用Adam优化器训练本文的模型,学习率为0.001,批量大小为16,为减少训练误差,每次实验重复10次,

表2 各项参数设置

符号	数值	符号	数值
$l$	6	$K$	5
TSAH	6	AdjSSAH	4
AdjTSAH	2	GloSSAH	4
$\lambda$	8	$\beta$	5

每次重复 300 个迭代.

#### 4.4 评价指标

选取平均绝对误差 (Mean Absolute Error, MAE)、均方根误差 (Rooted Mean Square Error, RMSE) 与平均绝对百分比误差 (Mean Absolute Percentage Error, MAPE) 作为实验验证指标. 具体计算公式如下.

(1) 平均绝对误差:

$$E_{MAE} = \frac{1}{S} \sum_{s=1}^S |\hat{y}_s - y_s| \quad (20)$$

(2) 均方根误差:

$$E_{RMSE} = \sqrt{\frac{1}{S} \sum_{s=1}^S (\hat{y}_s - y_s)^2} \quad (21)$$

(3) 平均绝对百分比误差:

$$E_{MAPE} = \frac{100\%}{S} \sum_{s=1}^S \left| \frac{\hat{y}_s - y_s}{y_s} \right| \quad (22)$$

其中,  $S$  为预测值的数量,  $\hat{y}_s$  为第  $s$  个预测值,  $y_s$  为其对应的真实观测值.

#### 4.5 实验结果与分析

##### 4.5.1 模型预测精度

表 3 展示了下一个小时 (12 步预测) 的预测结果, 最好的结果用粗体字表示, 其次加下划线表示.

表 3 STMGAN 模型与相关基线模型的预测评价指标

Model	PeMS04			PeMS07			PeMS08		
	MAE	RMSE	MAPE/%	MAE	RMSE	MAPE/%	MAE	RMSE	MAPE/%
SVR	28.656	44.592	19.152	32.976	50.148	15.432	23.256	36.156	14.712
LSTM	26.808	40.74	22.332	29.712	45.324	14.136	22.188	33.588	18.744
DCRNN	22.74	36.576	14.748	23.64	36.516	12.276	18.18	28.176	11.232
STGCN	21.756	34.764	13.872	22.896	35.436	11.988	17.844	27.12	11.208
Graph WaveNet	19.356	31.716	13.308	21.216	34.116	9.072	15.072	23.856	9.516
ASTGCN	22.421	34.252	15.871	21.221	34.100	9.051	15.072	24.800	9.512
ASTGNN	<u>18.600</u>	<u>31.032</u>	<u>12.636</u>	<u>20.616</u>	<u>34.02</u>	<u>8.856</u>	<u>14.976</u>	<u>24.708</u>	<u>9.492</u>
STMGAN	<b>18.420</b>	<b>30.252</b>	<b>12.192</b>	<b>19.824</b>	<b>32.652</b>	<b>8.412</b>	<b>13.656</b>	<b>23.592</b>	<b>9.084</b>

从表 3 中可以看出:

(1) 本文的 STMGAN 在所有数据集中的所有指标都显著优于所有基线方法下的指标. 与第二好的方法相比, STMGAN 的 MAE, MAPE, RMSE 平均降低了 4.437%, 2.930%, 4.275%.

(2) SVR 和 LSTM 是预测性能最差的, DCRNN, STGCN, Graph WaveNet, ASTGNN 这些基于图的模型比 SVR 和 LSTM 这些只考虑了时间特征的模型好, 表明了道路网络信息对交通预测的重要性, 即空间特征对交通流预测很重要.

(3) DCRNN 是一种基于 RNN 的典型时空预测方法, 预测精度比基于 GNN 模型的低, 表明捕捉长时间相关性的能力对交通流预测的重要性.

(4) 基于注意力的模型中 ASTGNN 是最好的基线, 预测性能比基于 GCN 基线 STGCN 和 ASTGCN 以及基于 TCN 最好的基线 Graph WaveNet 好, 这是由于当加深网络时 GCN 会出现过平滑, 无法捕获长时间相关性, 且基于 GCN 和 TCN 的模型对同阶邻域上不同的邻居分配的权重是完全相同的, 限制了模型对空间信息相关性的捕捉能力. ASTGCN 由三个独立的组件组成, 分别对交通流数据的当前时刻、日周期、周周期进行建模, 利用 1D CNN 来捕获时间特征, 利用 GCN 来捕获空间特征, 然而, 受卷积核大小的限制, 1D CNN 很难捕捉到长

期的时间相关性, 且由于是对每个周期数据分别进行建模, 无法挖掘每个周期模式数据之间的相关性, 而本文的模型 STMGAN 对周期性数据进行显式的建模, 能充分挖掘每个周期模式之间的相关性, 因此预测性能低于本文的模型 STMGAN. ASTGNN 是结合 GCN 和自注意力模块来聚合邻居信息的, 无法捕捉非邻近节点的相似性. 而本文的模型 STMGAN 利用掩码机制, 获取了全局空间相似性, 通过多头注意力捕捉数据的时空动态相关性, 并取得了更大的精确度.

##### 4.5.2 预测时长分析

在 PEMS08 和 PEMS04 数据集上以 10 min 为时间间隔增长, 使预测时长增长至 1 h, 对比 STMGAN 与各个基线模型随时间增长的性能变化情况, 实验结果如图 7 和图 8 所示.

通常, 随着预测时间步长增大, 预测任务将变得更加困难, 然而, 从图 7 和图 8 可以看出, 本文的 STMGAN 的交通流预测性能下降最小, 验证了所建模型不同预测时长下的一致优越性.

DCRNN 是一种典型的基于 RNN 的时空图数据预测方法. 由于 RNN 捕捉长时间相关性的能力有限, 其长期预测精度远低于本文模型 STMGAN.

STGCN 和 Graph WaveNet 是两种典型的基于 CNN 的方法. 在时间维度上, 它们使用 1D CNN 或 TCN 来捕

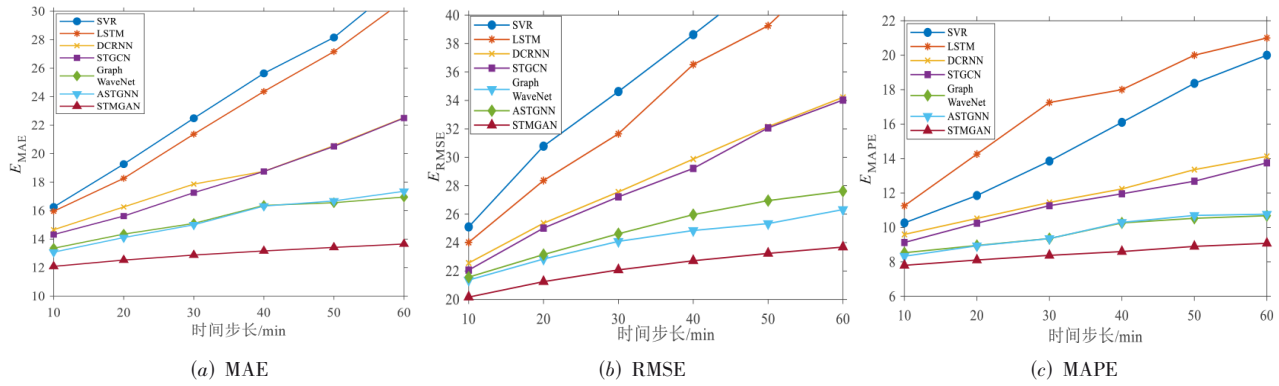


图 7 基线模型在 PEMS08 数据集上不同时间步长下的预测精度

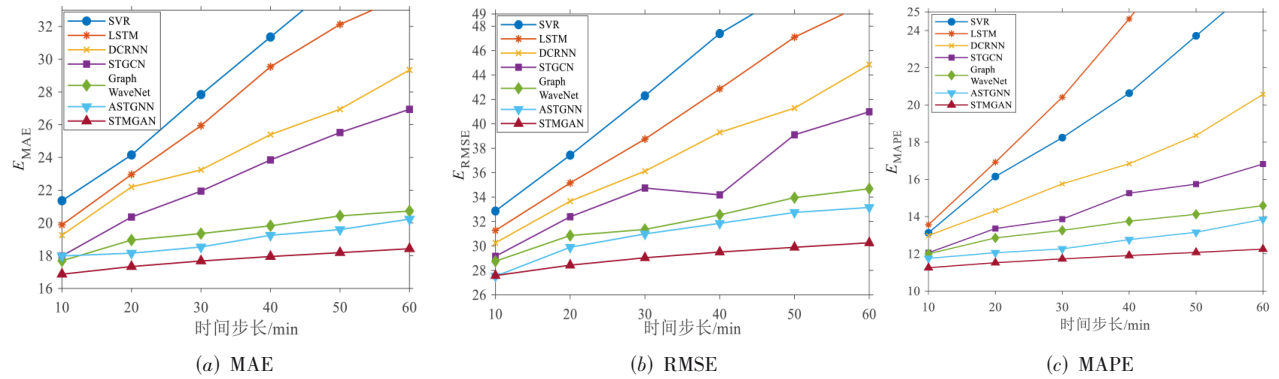


图 8 基线模型在 PEMS04 数据集上不同时间步长下的预测精度

捉时间相关性. 受限于卷积核的大小, 1D CNN 很难处理长期的时间信息, 而自注意允许序列在聚焦全局这产生了一个有效的全局感受野, 从而使本文模型 STMGAN 能够做出准确的长期预测.

ASTGNN 是基于注意力机制的方法, 会产生一个全局感受野. 整个城市的交通流共享该模型的注意力, 会使该模型专注于交通量大的主要城市道路, 或者注意力分布分散. 然而, 交通量低的节点应该关注模式相似的节点, 而不是流量大的节点. 此外, 距离太远的节点对当前节点的影响很小, 如果模型分散地关注所有节点, 则其性能将减弱. 而在 STMGAN 中引入  $M_{adj_s}$  和  $M_{glos}$ , 会使注意力集中在周围节点和远处相似功能区的节点, 因此, STMGAN 不仅考虑了邻近节点空间相关性, 而且寻找全局相似功能区域来捕获长距离空间依赖性, 提高了预测精度.

### 4.5.3 消融实验

对 PEMS08 数据进行了消融实验, 以验证本文模型关键组件对预测结果的有效性. 将不含不同组件的 STMGAN 命名如下:

(1)w/o 周期: 不带周期性嵌入的 STMGAN. 仅使用交通流原始数据, 不再对周期性进行建模.

(2)w/o 掩码: 不带掩码机制的 STMGAN. 删除注意力机制中的三个掩码矩阵  $M_{adj_t}$ ,  $M_{adj_s}$ ,  $M_{glos}$ , 即无法捕捉长期的时空依赖性.

(3)w/GCN: 使用图卷积网络 (GCN) 取代空间多头注意力模块, 无法捕捉动态和长期的空间依赖性.

(4)w/o 融合: 不带时空融合的 STMGAN. 对输出的时空数据直接相加操作.

以数据集 PEMS08 为例, 将每个实验重复 10 次, 每次重复 300 个迭代, 图 9 展示了各个消融实验的评价指标结果的平均值. 模型 w/o 周期与 STMGAN 模型相比, MAE, RMSE, MAPE 分别升高 1.86%, 0.500%, 2.130%, 表明挖掘数据的周期性对预测结果的有效性. 模型 w/o 掩码是预测性能下降最多的, 与 STMGAN 模型相比, MAE, RMSE, MAPE 分别升高 16.469%, 13.305%, 17.790%, 表明使用掩码矩阵识别邻近节点和相似性节点的重要性, 能使模型性能得到大的提升. 模型 w/GCN 与 STMGAN 模型相比, MAE, RMSE, MAPE 分别升高 6.253%, 3.510%, 4.580%, 表明多头自注意力网络在捕获动态和长范围空间依赖性方面优于 GCN. 模型 w/o 融合去除了自适应门控融合, 节点的时空信息进行无差别聚合传播, 与 STMGAN 模型相比, MAE, RMSE,

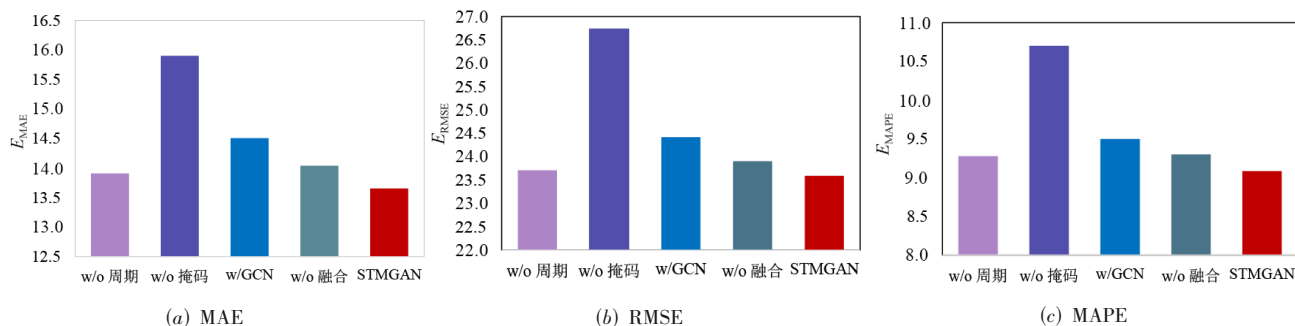


图9 模型STMGAN各组件对准确性的影响

MAPE 分别升高 2.812%, 1.322%, 2.378%, 表明门控机制自适应聚合信息的有效性。

## 5 结论

针对交通流复杂的动态时空特性,本文提出基于时空多头图注意力神经网络模型,并在3个真实基准数据集上进行验证,得到主要结论如下:

(1) 提出时空多头图神经网络模型框架,并在三个真实交通流基准数据集 PEMS04 和 PEMS07 和 PEMS08 上进行模型的有效性验证,结果表明,所建模型框架在3个数据集上的3个预测评价指标与其他精度最高模型相比,平均降低了 4.437%, 2.930%, 4.275%, 揭示了 STMGAN 模型在短期交通流预测领域的潜力。

(2) 相比仅对周期数据进行独立建模的 ASTGCN 预测结果,STMGAN 在3个数据集上的 MAE, RMSE, MAPE 平均降低 13.046%, 7.592%, 14.163%, 说明 STMGAN 通过显式地构建周周期性数据源和日周期性数据源,以及对输入数据进行时空位置编码嵌入,能充分挖掘交通流数据的时间相似性和静态时空特征。

(3) 相比 ASTGNN, STGCN, Graph WaveNet 的预测结果,STMGAN 表现出更优异的精度,说明 STMGAN 通过构建空间多头注意力模块,利用两个掩码矩阵分别获取局部邻接节点的交通流特征信息和长距离节点的交通流相关性信息,同时为降低模型的计算复杂度,将异构注意力融合到一个多头自注意力块中,能解决路网交通流的动态空间信息和长距离空间依赖性不足等问题,实现对交通流数据动态空间信息的有效挖掘。

(4) 相比基于 RNN 的典型时空图数据预测方法 DCRNN 的预测结果,STMGAN 在3个数据集上的 MAE, RMSE, MAPE 平均降低 25.277%, 17.589%, 30.182%, 说明 STMGAN 通过设计时间多头注意力模块,能实现对交通流数据动态时间信息的有效挖掘。

(5) 针对时间信息和空间信息的异质性融合,设计

了时空融合模块,通过与消融实验做对比,说明 STMGAN 实现了对时空信息的动态自适应融合。

未来工作中,我们将把 STMGAN 应用于其他时空预测任务,如风电预测<sup>[18]</sup>。此外,我们将探索交通预测中的预训练技术,以解决数据不足的问题。

## 参考文献

- [1] 李长乐, 王硕, 岳文伟, 等. 面向空地一体化交通的虚拟车道: 发展阶段与关键技术[J]. 电子学报, 2022, 50(5): 1255-1265.
- LI C L, WANG S, YUE W W, et al. Virtual lanes for air-ground integrated transportation systems: Evolution and key techniques[J]. Acta Electronica Sinica, 2022, 50(5): 1255-1265. (in Chinese)
- [2] LV Y S, DUAN Y J, KANG W W, et al. Traffic flow prediction with big data: A deep learning approach[J]. IEEE Transactions on Intelligent Transportation Systems, 2015, 16(2): 865-873.
- [3] GORI M, MONFARDINI G, SCARSELLI F. A new model for learning in graph domains[C]//Proceedings of 2005 IEEE International Joint Conference on Neural Networks. Piscataway: IEEE, 2005: 729-734.
- [4] JIANG W W, LUO J Y, HE M, et al. Graph neural network for traffic forecasting: The research progress[J]. ISPRS International Journal of Geo-Information, 2023, 12(3): 100.
- [5] 杨建喜, 郁超顺, 李韧, 等. 基于多周期组件时空神经网络的路网通行速度预测[J]. 交通运输系统工程与信息, 2021, 21(3): 112-119, 139.
- YANG J X, YU C S, LI R, et al. Traffic network speed prediction via multi-periodic-component spatial-temporal neural network[J]. Journal of Transportation Systems Engineering and Information Technology, 2021, 21(3): 112-119, 139. (in Chinese)
- [6] GUO S N, LIN Y F, WAN H Y, et al. Learning dynamics

- and heterogeneity of spatial-temporal graph data for traffic forecasting[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 34(11): 5415-5428.
- [7] 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. New York: ACM, 2017: 6000-6010.
- [8] LIN Z Q, FENG J, LU Z Y, et al. DeepSTN+: Context-aware spatial-temporal neural network for crowd flow prediction in metropolis[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 1020-1027.
- [9] ZONOOZI A, KIM J-J, LI X L, et al. Periodic-CRN: A convolutional recurrent model for crowd density prediction with recurring periodic patterns[C]//Proceedings of the 27th International Joint Conference on Artificial Intelligence. Stockholm: AAAI Press, 2018: 3732-3738.
- [10] HOANG V T, JEON H J, YOU E S, et al. Graph representation learning and its applications: A survey[J]. Sensors, 2023, 23(8): 4168.
- [11] JEONG Y S, JEONG M K, OMITAOMU O A. Weighted dynamic time warping for time series classification[J]. Pattern Recognition, 2011, 44(9): 2231-2240.
- [12] DRUCKER H, BURGESS C J C, KAUFMAN L, et al. Support vector regression machines[C]//Proceedings of the 9th International Conference on Neural Information Processing Systems. New York: ACM, 1996: 155-161.
- [13] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [14] LI Y G, YU R, SHAHABI C, et al. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting[EB/OL]. (2017)[2023]. <http://arxiv.org/abs/1707.01926.pdf>.
- [15] 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. New York: ACM, 2018: 3634-3640.
- [16] WU Z H, PAN S R, LONG G D, et al. Graph WaveNet for deep spatial-temporal graph modeling[EB/OL]. (2019) [2023]. <https://arxiv.org/abs/1906.00121>.
- [17] GUO S N, LIN Y F, FENG N, et al. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 922-929.
- [18] JIANG J W, HAN C K, WANG J Y. BUAA\_BIGSCity: Spatial-temporal graph neural network for wind power

forecasting in Baidu KDD CUP 2022[EB/OL]. (2022) [2023]. <https://arxiv.org/abs/2302.11159>.

#### 作者简介



**梁秀霞** 女, 1972年6月生于天津市。现为河北工业大学人工智能与数据科学学院教授、硕士生导师。获河北省教育厅奖一项, 在国内外发表学术论文50余篇。主要研究方向为高等过程控制、智能控制、检测技术与装置。

E-mail: lxx68@163.com



**夏曼曼** 女, 1998年11月出生于河南省周口市。现为河北工业大学人工智能与数据科学学院硕士研究生。主要研究方向为智慧交通。

E-mail: 1826807320@qq.com