



计算机工程与应用  
Computer Engineering and Applications  
ISSN 1002-8331, CN 11-2127/TP

## 《计算机工程与应用》网络首发论文

题目: 基于图注意力网络的短时交通流量预测  
作者: 贺佳佳, 黄德启, 王东伟, 张阳婷  
网络首发日期: 2023-09-20  
引用格式: 贺佳佳, 黄德启, 王东伟, 张阳婷. 基于图注意力网络的短时交通流量预测[J/OL]. 计算机工程与应用.  
<https://link.cnki.net/urlid/11.2127.TP.20230920.1347.056>



**网络首发:** 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

**出版确认:** 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

## 基于图注意力网络的短时交通流量预测

贺佳佳, 黄德启, 王东伟, 张阳婷

新疆大学电气工程学院, 乌鲁木齐 830047

**摘要：**交通流量预测是时间序列分析中的一个重要问题，由于道路网络中存在复杂的动态时空依赖性，实现高精度交通流量预测具有挑战性。为了应对这些挑战，提出了一种名为 STBiPGAT 的时空注意力模型，旨在准确捕捉交通流量的时空动态特性。该模型将邻接矩阵、利用节点交通流提取的相关系数矩阵分别与交通流量特征矩阵送入图注意力网络中，以并行方式提取空间局部动态特征与空间隐藏关系，且进行特征融合。随后考虑到节点空间特征向量在时间维度的上下文信息和周期性特性，构造双向 GRU 组件以提取交通流量的前后时间特征。最后引入自注意力机制解决不同时刻输入特征影响的差异，通过全连接层生成预测结果。实验结果在两个真实交通数据集上的评估表明，STBiPGAT 预测误差低于对比模型预测误差，显著提升了预测精度，证实其有效性。

**关键词：**交通流量预测；图注意力网络；注意力机制；时空相关性

**文献标志码：**A      **中图分类号：**TP391      **doi:** 10.3778/j.issn.1002-8331.2306-0130

## Short-time Traffic Flow Prediction Based on Graph Attention Networks

HE Jiajia, HUANG Deqi, WANG Dongwei, ZHANG Yangting

ChinaSchool of Electrical Engineering, Xinjiang University, Urumqi 830047

**Abstract:** Traffic flow prediction is an important problem in time series analysis, and it is challenging to achieve high-accuracy traffic flow prediction due to the complex dynamic spatio-temporal dependencies in road networks. To address these challenges, a spatio-temporal attention model called STBiPGAT is proposed, which aims to accurately capture the spatio-temporal dynamic properties of traffic flows. The model feeds the adjacency matrix, the global correlation matrix extracted using the node traffic flow, and the traffic flow feature matrix into the graph attention network, respectively, to extract the spatial local dynamic features and the spatial hidden relations in a parallel manner, and to perform feature fusion. Subsequently, considering the contextual information and periodicity characteristics of the node spatial feature vectors in the time dimension, a bidirectional GRU component is constructed to extract the before-and-after time features of the traffic flow. Finally, a self-attention mechanism is introduced to resolve the differences in the influence of input features at different moments, and the prediction results are generated through a full connectivity layer. The evaluation of the experimental results on two real traffic datasets shows that the prediction error of STBiPGAT is lower than the prediction error of the comparison model, which significantly improves the prediction accuracy and confirms its effectiveness.

**Key words:** traffic flow prediction; graph attention network; attention mechanism; spatio-temporal correlation

**基金项目：**新疆维吾尔自治区自然科学基金资助项目(2022D01C430)；国家自然科学基金资助项目(51468062)。

**作者简介：**贺佳佳(1998—),男,硕士研究生,研究方向为数据挖掘、交通流量预测,E-mail: 486657983@qq.com; 黄德启(1972—),通信作者,男,博士,副教授,研究方向为智慧交通、交通协同控制; 王东伟(1999—),男,硕士研究生,研究方向为强化学习、交通灯控制; 张阳婷(1999—),女,硕士研究生,研究方向为图像识别、深度学习。

随着城市化进程加速和汽车保有量的不断增加,交通拥堵<sup>[1]</sup>问题日益严重,给人们的出行带来了诸多不便。而准确及时的短时交通流量预测,不仅可以为道路参与者快速提供交通信息,诱导驾驶路径、避免交通拥堵;还可以为交通管控人员提供交通信息,优化道路规划及信号配时方案,合理利用道路,有效避免车流在道路上分配不均,对于缓解交通拥堵有重要意义。目前,对于交通流量预测算法的研究主要集中在统计分析理论、机器学习和深度学习三个方向。

首先,统计分析理论是传统的交通流量预测方法之一。这种方法利用统计模型<sup>[2][3]</sup>和时间序列分析<sup>[4][5]</sup>等工具,基于历史数据和趋势来预测未来交通流量。统计分析理论在一定程度上可以提供可靠的预测结果,但在处理非线性和动态交通场景时存在一些局限性。

其次,机器学习方法已经成为交通流量预测领域的热点研究方向。机器学习技术利用算法和模型从大规模的交通数据中学习和提取特征,以建立预测模型。常见的机器学习方法包括支持向量机(Support Vector Machines,SVM)<sup>[6]</sup>、K近邻(K-Nearest Neighbor,KNN)<sup>[7]</sup>和决策树(Decision Trees)<sup>[8]</sup>等。尽管这些方法具有良好的预测性能,能够应对一定程度上的非线性关系,但机器学习方法对于大量高质量训练数据的需求较高,对输入数据进行特征选择时需要领域知识和专业经验的支持。并且近年来车辆数量呈指数级增长<sup>[9]</sup>,这些传统的机器学习模型无法在当前场景中工作。

最后,深度学习方法近年来在交通流量预测中取得了显著的进展。深度学习技术利用神经网络模型层级结构和表达能力,能够自动学习和提取交通数据中的复杂特征。循环神经网络(Recurrent Neural Networks,RNNs)<sup>[11]</sup>和卷积神经网络(Convolutional Neural Networks,CNNs)<sup>[12][13]</sup>等深度学习模型已经在交通流量预测中得到广泛应用,并且在准确性和预测性能方面取得了显著的改进。

其中一部分学者未考虑交通流量的空间特征,主要分析其在时间序列上的特征。如Zhang等人<sup>[14]</sup>应用门控递归神经网络预测考虑天气条件的城市交通流量,取得了不错的精度。Tian等人<sup>[15]</sup>采用了多尺度时间平滑方法来推断丢失的数据,并利用LSTM来提取交通流量时间特征,进行预测。仅仅考虑时间特征的交通流预测忽略了交通流量与空间之间的关系,而交通流量变化通常受到路段之间的交互作用和影响<sup>[16]</sup>,例如道路拥堵传播效应和交通网络拓扑结构。忽略空间相

关性可能导致预测结果的不准确性,特别是在具有复杂路网结构的城市或区域。

因此,部分研究者开始研究如何对交通流的空间特征进行有效建模,将时间特征和空间特征融合进行交通流预测。Wu等人<sup>[17]</sup>提出结合CNN和LSTM的新型深度架构来预测未来交通流量,利用一维CNN捕获空间交通流特征,利用两个LSTM进行短期挖掘交通流量的可变性和周期性。Ma等人<sup>[18]</sup>通过二维时空矩阵将时空交通动态转换为描述交通流时空关系的图像,并将CNN应用于图像。由于CNN主要适用于处理结构化的数据,而交通道路呈拓扑结构,交通流量具有非结构化数据,故以上方法利用CNN强行提取交通流量的空间特征会损失一部分有效信息。针对于非结构化信息特征提取这个问题,产生了图神经网络一系列模型,其核心思想是将临近节点的信息聚合到中心节点,从而提取图结构中的空间特征,这正适用于拓扑结构的交通数据。如Yu等人<sup>[19]</sup>为融合时空两方面的特征,结合图卷积和门控时间卷积构造了时空卷积块,可以提取有用的空间特征和基本时间特征。Guo等人<sup>[20]</sup>提出ASTGCN模型,采用堆叠模型结构,将多个STGCN和注意力机制组合在一起,形成一个多层结构,以进一步提高模型的精度和鲁棒性。Fang等人<sup>[21]</sup>提出了一种名为时空图常微分方程网络(STGODE)的方法,基于语义邻接矩阵和静态空间邻接矩阵的组合将常微分方程(ODE)纳入GCN,它利用基于张量的常微分方程(ODE)来捕捉时空动态,其中语义邻接矩阵也使用DTW计算。

上述利用GCN进行交通流量预测任务的模型确实能达到很高精度,然而仅依靠静态的邻接矩阵无法提取交通路网动态性。为了解决提取交通流量中的动态空间特性问题,本文提出一个新颖模型:基于并行提取时空动态特征的图注意力网络(STBiGAT)。该模型一方面采用可更加准确捕捉节点和边之间动态关系的图注意力网络<sup>[22]</sup>(Graph Attention Network,GAT),引入多头注意力增加模型鲁棒性;另一方面利用相关系数矩阵与交通流量矩阵同步进行深度挖掘交通流的空间特征。将同步进行的两模块特征融合起来送入双向门控循环单元(Bidirectional Gated Recurrent Unit,BiGRU)提取时间特征,经时间自注意力层送入全连接网络中,获得预测的交通流量值。

## 1 相关工作

### 1.1 问题描述和定义

本文将道路网络表示为  $G=(V,E,A)$ ，其中  $V$  表示道路网络内的一组  $N$  个节点， $E$  表示节点之间连通性的一组边， $A \in R^{N \times N}$  表示一个加权邻接矩阵，这里使用顶点之间的距离来分配初始权重。历史交通信息在交通网络上可以表示为一个序列  $X=(X^1, X^2, \dots, X^t)$ ，其中每个  $X^{(t)} \in R^{N \times 1}$  代表  $N$  个节点在时间  $t$  的交通流量信息。因此，交通网络的短时交通流预测问题可以表示公式 (1)：

$$Y_i^{(t)} = f(G, X_i^{(t)}) \quad (1)$$

式中， $Y_i^{(t)}$  表示第  $i$  节点经过模型预测  $f(\bullet)$  得到的第  $t$  个时间段交通流数据； $G$  为分布节点拓扑结构的表示。

### 1.2 节点关系分析

在城市区域中，各路段交通流量受到其他路段的历史时刻交通流量影响，并且相邻的路段之间存在一定空间相关性。这意味着当某一条路段的交通状态发生变化时，其相邻路段的交通流量也会相应地发生变化。为了衡量各节点交通流之间的相关性，本文使用皮尔逊相关系数 (Person Correlation Coefficient, PCC) [23] 进行计算。皮尔逊相关系数是一种统计量，用于衡量两个变量之间线性关系的强度，其取值范围为-1 到 1。具体而言，对于变量  $x$  和变量  $y$ ，可以使用下述公式 (式 (2)) 计算它们之间的相关系数：

$$P_{x,y} = \frac{cov(x,y)}{\sigma_x \sigma_y} \quad (2)$$

式中， $cov(x,y)$  表示变量  $x$  和变量  $y$  的协方差， $\sigma_x$  和  $\sigma_y$  分别表示两个变量的标准差。

同样，为了研究交通流量在不同节点的相关性，在这里选取公共数据集中 9 个有连接的节点，通过皮尔逊相关计算方法研究各节点在空间上的关系，并将结果以热力图的形式进行可视化展示，如图 1 所示。

热力图展示了相关系数的分布情况，用颜色来表示不同程度相关性。较高的相关系数对应着较深的颜色，

表示节点之间的交通流量具有较强相关性。相反，较低的相关系数对应着较浅的颜色，表示节点之间的交通流量相关性较弱。通过热力图，可以识别出在空间上相邻节点之间可能存在较高的交通流量相关性，而非相邻节点之间可能存在较低的相关性。因此，利用皮尔逊相关系数计算方法构建节点相关矩阵，有助于帮助模型理解城市交通网络中节点之间的隐藏空间关系。

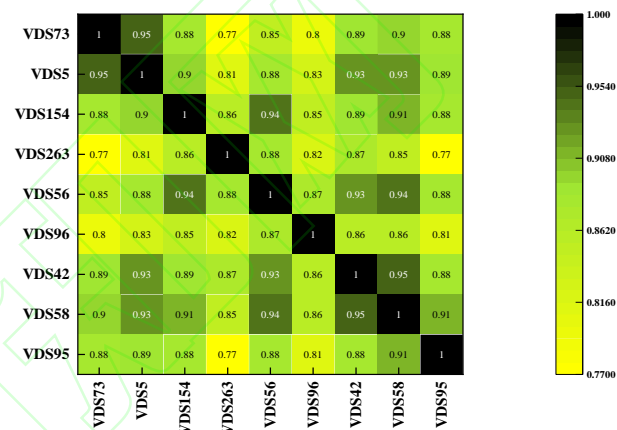


图 1 节点关系热力图 (修改图片为清晰的.emf 矢量图)

Fig.1 Heat map of node relationships

## 2 交通流预测模型构建

### 2.1 STBiPGAT 总体框架

针对交通流时空动态信息挖掘不够深入的特点，本文提出了结合多头注意力机制的 STBiPGAT 模型，主要由空间特征提取模块、时间特征提取模块、自注意力机制层和全连接层组成，总体架构如图 2 所示。该模型采用了多头注意力机制的图注意力网络 (GAT)，能够在不依赖先验知识情况下对动态空间依赖性进行建模，提取路网空间特征。此外，模型还引入了特征融合机制，综合利用了 GAT 子网络的输出和相关系数矩阵，通过 Hamada 乘积操作 (Hamada Product)，将两种特征进行融合，以增强模型对节点关系和节点特征建模能力。利用双向门控循环单元 BiGRU 进行时间建模，以捕捉时序数据中的依赖关系。最终，该模型通过自注意力机制对特征进行加权，利用全连接网络输出预测结果。



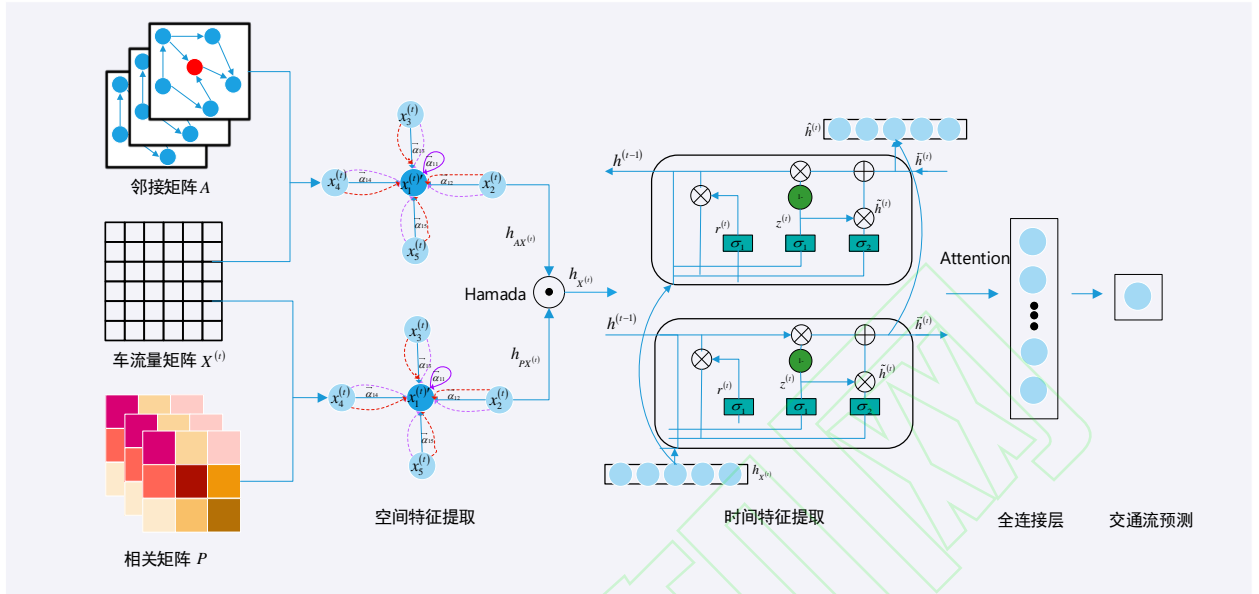


图2 总体架构图

Fig.2 Overall architecture diagram

## 2.2 空间特征提取

### 2.2.1 空间依赖建模

交通网络通常用图结构来描述, 本文选择一种对图结构数据进行操作的新型神经网络——图注意力层 (Graph Attention Layer)。该网络不像 GCN 只能提取路网静态特恒, 具有结构依赖性, 带有图注意力层的网络可以直接应用于归纳学习问题, GAT 的实现与之相同。其输入是  $t$  时刻  $N$  个节点的所有特征  $X^{(t)} = \{x_1^{(t)}, x_2^{(t)} \dots x_N^{(t)}, x_i^{(t)} \in \mathbb{R}^{d^{(0)}}\}$ , 其中  $N$  表示节点数量,  $d^{(0)}$  表示节点特征数量。在这里从等式 (3) 到等式 (5) 制定 GAT 工作原理:

$$\begin{aligned} e_{ij}^{(t)} &= \alpha(Wx_i^{(t)}, Wx_j^{(t)}) \\ &= \text{LeakyReLU}(a^T [Wx_i^{(t)} \parallel Wx_j^{(t)}]) \end{aligned} \quad (3)$$

式中参数矩阵  $W \in \mathbb{R}^{d^{(l+1)} \times d^{(l)}}$  对每个节点的特征进行线性变换,  $e_{ij}^{(t)}$  代表了  $t$  时刻节点  $v_j$  对节点  $v_i$  的重要程度,  $a \in \mathbb{R}^{2d^{(l+1)}}$  是一个单层的前馈神经网络, 在该网络的输出层上加入了激活函数  $\text{LeakyReLU}$ 。为了更好地为每个节点分配注意力权重, 并使得不同节点之间权重可比较, 将所有邻居计算出的相关度进行统一的归一化处理, 表达式如 (4):

$$\alpha_{ij}^{(t)} = \text{softmax}_j(e_{ij}^{(t)}) = \frac{\exp(e_{ij}^{(t)})}{\sum_{v_k \in \tilde{N}(v_i)} \exp(e_{ik}^{(t)})} \quad (4)$$

在计算完节点注意力权重之后, 可采用注意力机制思想来加权求和节点的特征向量, 从而生成每个节点的新特征向量  $x_i^{(t)'}$ , 如式 (5):

$$x_i^{(t)'} = \sigma \left( \sum_{v_j \in \tilde{N}(v_i)} \alpha_{ij}^{(t)} Wx_j^{(t)} \right) \quad (5)$$

为了进一步稳定学习过程, 设计了多头注意力。K 个独立的注意力分别有各自的参数, 且其输出可以合成两种方式拼接 (公式 (6)) 和平均 (公式 (7))。在这项工作中, 对隐藏层使用连接, 对最后一层使用平均操作, 最终  $t$  时刻  $i$  节点的新特征  $x_i^{(t)'}$ 。聚合所有节点的新特征得到  $X^{(t)'} = \{\bar{x}_1^{(t)'}, \bar{x}_2^{(t)'}, \dots, \bar{x}_i^{(t)'}, \bar{x}_i^{(t)'} \in \mathbb{R}^{d^{(l+1)}}\}$ 。

$$x_i^{(t)'} = \parallel_{k=1}^K \sigma \left( \sum_{v_j \in \tilde{N}(v_i)} \alpha_{ij}^{(t)(k)} W^{(k)} x_j^{(t)} \right) \quad (6)$$

式中  $\parallel$  表示拼接工作,  $\alpha_{ij}^{(t)(k)}$  是第  $k$  组注意力机制计算出的权重系数,  $W^{(k)}$  是对应的学习参数。以上过程如图 3 所示:

$$x_i^{(t)'} = \sigma \left( \frac{1}{K} \sum_{k=1}^K \sum_{v_j \in \tilde{N}(v_i)} \alpha_{ij}^{(t)(k)} W^{(k)} x_j^{(t)} \right) \quad (7)$$

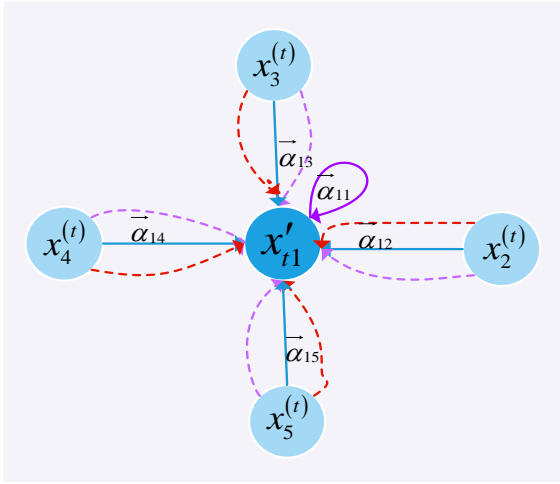


图3 多头注意力机制

Fig.3 Multi-headed attention mechanism

### 2.2.2 节点关系特征提取

为了更好地挖掘交通道路之间的空间特征，仅依靠邻接矩阵可能不足够。因此，使用已有的时间序列数据，并利用皮尔逊相关系数计算函数来提取不同节点特征之间的关联度。通过这种方式构建全局相关系数矩阵  $P$ ，以捕捉交通流量数据中的潜在空间特征。假设任意两节点的交通流时间序列分别为  $x = \{x^1, x^2 \dots x^T\}$  和  $y = \{y^1, y^2 \dots y^T\}$ ，则节点之间的相关系数通过式 (8) 求得。

$$P_{x,y} = \frac{\sum_{i=1}^T (x^i - \bar{x})(y^i - \bar{y})}{\sqrt{\sum_{i=1}^T (x^i - \bar{x})^2} \sqrt{\sum_{i=1}^T (y^i - \bar{y})^2}} \quad (8)$$

式中， $x^i$ 、 $y^i$  分别表示交通流时间序列  $x$ 、 $y$  在  $i$  时刻的值， $P_{x,y}$  表示不同节点之间的相关性程度， $T$  表示时间序列长度， $\bar{x}$ 、 $\bar{y}$  分别为两序列样本的平均值。

### 2.2.3 特征融合

为了提高交通流量预测准确性，本模型采用了两个方面的建模方法，以进一步挖掘节点之间的空间关系。首先，通过利用邻接矩阵  $A$  和车流量矩阵  $X^{(t)}$  来进行卷积操作，通过图注意力网络 GAT 来获取道路的局部动态特征  $h_{AX}^{(t)}$ ，这一方面的建模方法可以帮助模型理解道路之间的关联性和空间依赖性；另一方面，模型利用全局相关系数矩阵  $P$  和车流量矩阵  $X^{(t)}$  进行非线性运算，从而获取节点之间的关联程度特征  $h_{PX}^{(t)}$ 。这一建模方法通过计算节点之间的相关性，能够捕捉节点之间的关联程度。将两个方面的建模方法并行进

行计算，并将它们提取的特征进行融合，整个过程如图 4 所示。最终得到的特征矩阵  $h_{X^{(t)}}$  表示通过式 (9) 定义。

$$h_{X^{(t)}} = \sigma(h_{AX^{(t)}} \odot h_{PX^{(t)}}) \quad (9)$$

式中  $\sigma(\bullet)$  为  $\text{relu}(\bullet)$  激活函数， $\odot$  为哈达玛乘积，表示矩阵对应元素相乘。

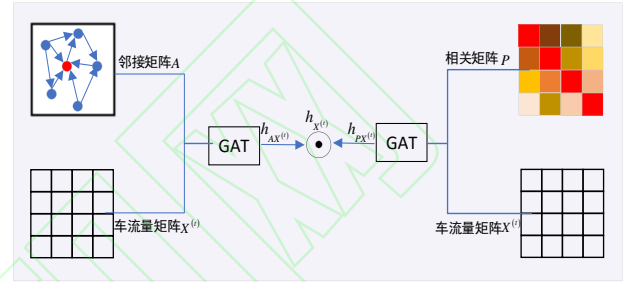


图4 空间特征融合组件

Fig.4 Spatial feature fusion component

## 2.3 时间特征提取

### 2.3.1 双向门控循环单元 BiGRU

交通流的形成是由人们出行需求驱动的。在特定的时间和空间，人们出行活动呈现出一定规律性和趋势，例如上下班通勤、商业活动、学校放学等。这些出行需求导致了交通流在不同时间点上的聚集和流动，形成了时间相关性。GRU 通过引入更新门和重置门，通过门控机制来决定隐藏状态的更新和重置程度。这使得 GRU 能够灵活地决定是否将过去的信息传递给当前时刻，以及在多大程度上保留过去的信息，适合提取时间特征。

对于交通流时间序列而言，中心检测点交通流不仅与历史时刻交通流相关，还会受到未来时刻交通流一定影响。而单向 GRU 并不能获得当前交通流前后时间特征，故本文采用双向长短期记忆网络 BiGRU。BiGRU 结构包括两个独立 GRU 隐藏层，分别处理前向和后向时间序列数据，同时考虑前后时段的数据动态特征。将获取具有空间特征的特征矩阵  $h_{X^{(t)}}$  代入到前向 GRU 公式中，可得到前向输出特征  $\vec{h}^{(t)}$ ，具体过程如下所示：

$$\vec{z}^{(t)} = \sigma(\vec{W}_{xz} \cdot h_{X^{(t)}} + \vec{W}_{zh} \cdot \vec{h}^{(t-1)} + \vec{b}_z) \quad (10)$$

$$\vec{r}^{(t)} = \sigma(\vec{W}_{xr} \cdot h_{X^{(t)}} + \vec{W}_{rh} \cdot \vec{h}^{(t-1)} + \vec{b}_r) \quad (11)$$

$$\vec{h}^{(t)'} = \tanh(\vec{W} \cdot h_{X^{(t)}} + \vec{W} \cdot (\vec{r}^{(t)} \odot \vec{h}^{(t-1)}) + \vec{b}) \quad (12)$$

$$\vec{h}^{(t)} = (1 - \vec{z}^{(t)}) \odot \vec{h}^{(t-1)} + \vec{z}^{(t)} \odot \vec{h}^{(t)'} \quad (13)$$

式中  $\bar{z}^{(t)}$  表示更新门,  $\bar{r}^{(t)}$  表示复位门。  $\bar{h}^{(t-1)}$  和  $\bar{h}^{(t)}$  分别是记忆单元在时刻  $(t-1)$  的输出和隐藏层在时刻  $t$  的输出;  $\bar{w}$  和  $\bar{b}$  分别表示相应的权重系数和偏置项;  $\odot$  表示哈达玛乘积;  $\sigma(\bullet)$  是激活函数  $\tanh(\bullet)$  是双曲正切函数。

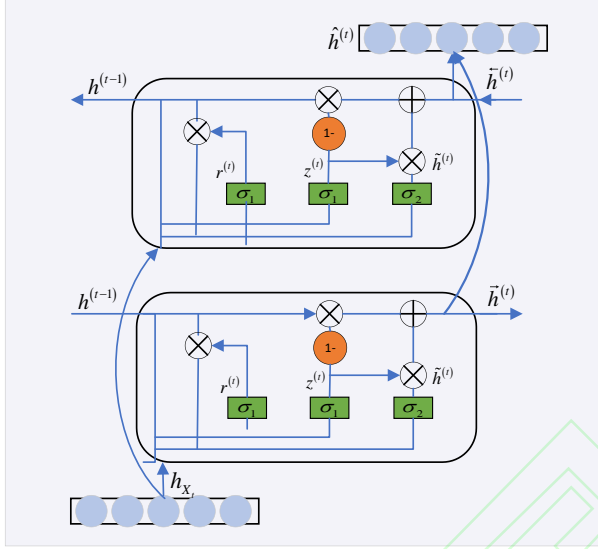


图5 BiGRU 结构图

Fig.5 BiGRU structure diagram

同理,通过下一层 GRU 网络的反向训练也可得到后向训练过程中的输出  $\bar{h}^{(t)}$ 。当两个方向的 GRU 都被训练后,双向隐藏状态的输出通过公式 (14) 被链接为新特征  $\hat{h}^{(t)}$ , 其中  $\oplus$  为拼接操作。

$$\hat{h}^{(t)} = \bar{h}^{(t)} \oplus h^{(t)} \quad (14)$$

### 2.3.2 自注意力机制

自注意力机制 (Self-Attention Mechanism) 是一种深度学习中的技术,用于处理具有序列结构的数据。在自注意力机制中,每个序列元素都会与序列中所有其他元素进行交互,以计算对其他元素的重要性得分,从而使得整个序列中元素都能够参与到计算中。本文将自注意力机制引入到 BiGRU 网络中,通过自注意力机制计算不同时刻交通流输入特征的重要性,继而对输入特征给予不同的权重,突出重要特征,从而提高模型性能,其计算公式如下:

$$\varepsilon^{(t)} = \tanh(W_w \hat{h}^{(t)} + b_w) \quad (15)$$

$$\alpha^{(t)} = \frac{\exp(u_w^T \varepsilon^{(t)})}{\sum_i \exp(u_w^T \varepsilon^{(i)})} \quad (16)$$

$$h^{(t)} = \sum_i \alpha^{(i)} \hat{h}^{(i)} \quad (17)$$

式中,  $W_w$ 、 $b_w$ 、 $u_w$  是训练模型所设的参数,  $u_w$  对应于自注意力机制中的 Query;  $\varepsilon^{(t)}$  表示  $t$  时刻

BiGRU 网络隐藏状态  $\hat{h}^{(t)}$  所得到的注意力评分值,其对应于自注意力机制中的 Key;  $\alpha^{(t)}$  表示每个时刻隐藏状态  $\hat{h}^{(t)}$  对应的注意力权重系数,隐藏状态  $\hat{h}^{(t)}$  对应于自注意力机制中的 Value;  $h^{(t)}$  表示最终加权求和得到的特征向量。

经过自注意力机制层后,将得到的时空特征向量  $h_t$  输入到全连接网络中,计算得到  $t$  时刻的交通流输出  $y_t$ , 计算过程如式 (15):

$$y^{(t)} = h^{(t)} W_{hy} + b_y \quad (18)$$

## 3 实验结果及分析

### 3.1 实验数据集

本文在真实世界的两个数据集 PeMSD4 和 PeMSD8 上进行了实验,以验证和评估 STBiPGAT 模型性能。PeMSD4 数据集包含了美国加州旧金山湾区的路网节点交通状态数据,而 PeMSD8 数据集则包含了美国加州圣贝纳迪诺县的路网节点交通状态数据。这两个数据集的数据均以每 5 分钟的时间间隔进行汇总,并提供了多个时间段的观测数据。详细的数据集信息如表 1 所示。

表 1 所用 PeMS 数据集信息

数据集	节点数	数据长度	时间跨度
PeMSD4	307	16992	2018/1/1-2018/2/28
PeMSD8	170	17856	2016/7/1-2016/8/31

当数据集中存在缺失点时,本文采用线性插值法进行恢复缺失数据。此外,数据通过 Z-Score 方法标准化。最后将数据集划分为训练集和测试集两个部分,比例为 7:3。

### 3.2 评价指标

本文将平均绝对误差 (MAE) 和均方根误差 (RMSE) 作为实验性能指标,其数学表达式分别如下:

$$MAE = \frac{1}{k} \sum_{i=1}^k |\hat{y}_i - y_i| \quad (19)$$

$$RMSE = \sqrt{\frac{1}{k} \sum_{i=1}^k (\hat{y}_i - y_i)^2} \quad (20)$$

式中,  $\hat{y}_i$  表示预测值,  $y_i$  表示真实值,  $k$  表示样本数。

### 3.3 实验设置

本文基于 Pytorch 框架实现 STBiPGAT 模型,使用 0.5 个小时的历史数据来预测后面 5 分钟的交通流量。所有实验均在 Windows 服务器 (CPU: AMD Ryzen7 5800H, GPU: NVIDIA GeForce RTX 3060) 上进行训练和测试。通过实验,参数设置如下:使用 Adam 优化

器训练模型,其中 epoch 数为 400,初学习率为 0.001, batchsize 设为 64。将 BiLSTM 输入层神经元个数设为 64,隐藏层神经元个数设为 64,隐藏层个数设为 4,

为了防止过拟合,影响模型泛化能力,设置 Dropout=0.3。

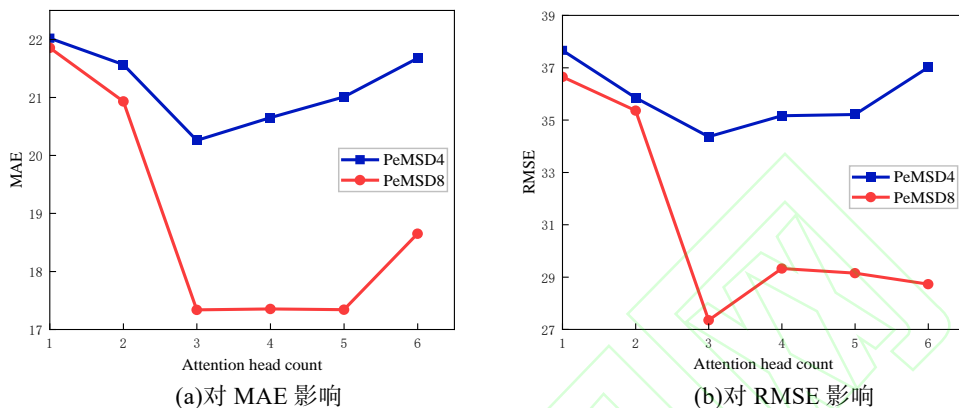


图6 注意力头数的影响

Fig.6 Effect of attentional head count

由于图注意力网络 (GAT) 的多头注意力头数对模型性能影响很大,增加注意力头数可以提高模型表征能力和泛化能力,但每增加一个注意力头,模型的计算量就会相应增加,过多的注意力头数可能导致计算资源消耗过大,使得模型训练和推理效率下降。因此,选择适量的注意力头数非常重要。本文在给定隐藏层维度为 12 的基础上,进行 1-6 个头数实验的对比,其结果如图 6 所示。经图所知,尽管在两个数据集上的表现头数为 4 不全部是最优,但是综合性能和计算量来看,头数为 4 的模型能够在保持较低计算复杂度的同时获得令人满意的模型性能。

### 3.4 Baselines

为了验证模型有效性,将本文提出的 STBiPGAT 模型与以下经典和最先进的学习模型进行了比较:(1) HA 模型是基于指数平滑方法的一种扩展,用于对具有季节性变化的时间序列数据进行预测和分析;(2) BPNN 模型可以通过学习历史交通数据的模式和规律,对未来交通流量的变化趋势进行预测;(3) GCN 模型能够利用路网和交通流量数据构建交通图,并通过应用图卷积和池化操作从交通图中提取关键信息,以用于进行交通流预测;(4) GAT 利用自适应注意力机制可以根据节点之间的关系自动调整权重,从而更准确地捕捉交通图中的重要特征;(5) GRU 是利用历史交通流量数据获取特征的一种循环神经网络;(6) CNN-TCN 组合模型主要通过 CNN 层进行局部特征提取,并将其连接到 TCN 层进行时序建模。(7) GCN-BiLSTM 是一个组合预测模型,GCN 用于捕捉交通网络中节点和边之间的关系,而 BiLSTM 则用于

处理交通流量时间序列中的时序性信息。(8) MSTGCN 是一种基于多尺度时空图卷积网络的学习算法,可以用于交通流预测和建模。(9) ASTGCN 使用卷积神经网络和注意力层提取时空特性。(10) STGODE 模型通过基于张量的常微分方程捕获时空动态,构建更深层次的网络并同步利用时空特征进行交通流预测。

### 3.5 实验结果对比与分析

在实验中,本文在 PeMSD4 和 PeMSD8 数据集上进行性能验证。从表二可以看出,STBiPGAT 模型在两个数据集上都取得了相对最好的性能。而仅考虑时间特征的 HA、BPNN 和 LSTM 模型由于忽略了路网空间特征而表现不佳,同样,仅考虑空间特征的 GCN 和 GAT 等模型也有局限性。因此,时间和空间特征结合对于准确交通流预测是必要的。CNN-TCN 网络虽然提取了时空特征,但其将路网拓扑结构强行转换为规则的欧式数据,损失了部分空间特征。GCN-BiLSTM 网络仅提取了空间静态特征,且无法区分相连节点的重要程度,导致预测效果不佳。本文提出的 STBiPGAT 模型通过将图注意力整合到 BiGRU 架构中,同时融入了相关系数矩阵,可有效地学习路段之间复杂空间动态关系,并捕获交通流时间模式,提升了预测精度。具体节点实验结果如图 7 (a)、(b) 所示。

表2 不同模型的性能指标对比

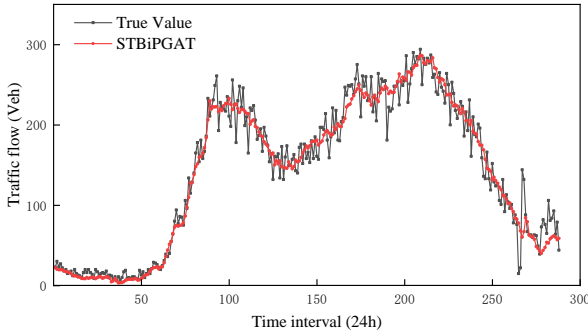
Table 2 Comparison of performance metrics for different models

模型	PeMSD4		PeMSD8	
	MAE	RMSE	MAE	RMSE
HA	38.24	56.68	31.92	47.42
BPNN	36.31	51.31	31.25	44.06

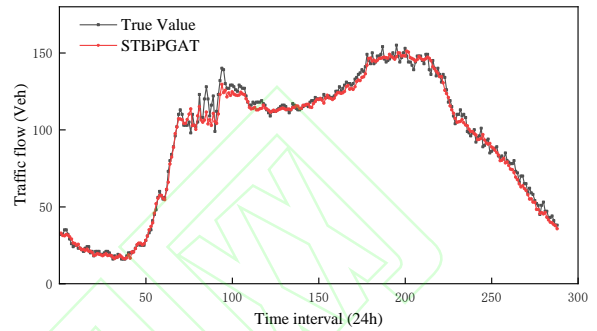


GCN	31.98	47.58	29.17	42.86
GAT	26.21	41.42	28.29	42.17
GRU	28.07	40.35	25.23	35.35
CNN-TCN	25.12	39.17	20.40	29.89
GCN-BiLSTM	24.73	38.05	20.22	30.63

MSTGCN	24.99	38.04	19.55	29.58
ASTGCN	21.98	34.57	18.03	27.74
STGODE	21.20	<b>32.96</b>	16.59	25.89
<b>STBiPGAT</b>	<b>19.69</b>	<b>33.02</b>	<b>15.78</b>	<b>25.05</b>



(a) PeMSD4 数据集 104 号节点



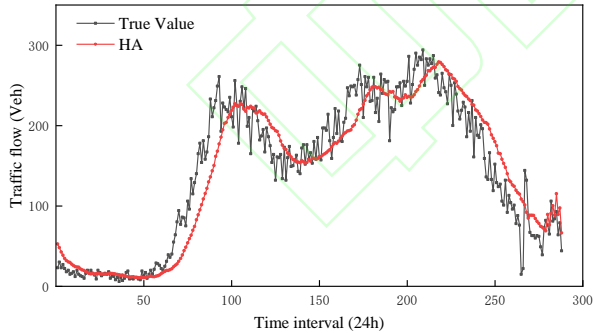
(b) PeMSD8 数据集 109 号节点

图 7 STBiPGAT 预测结果

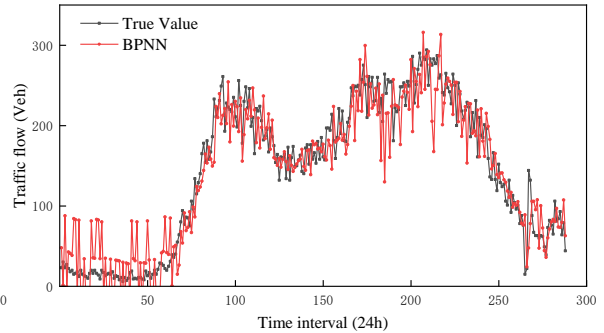
Fig.7 STBiPGAT prediction results

其与近年流行的 MSTGCN 模型相比,在 PeMSD4 数据集上使 MAE 值和 RMSE 值分别降低了 21.21%和 13.20%;它还优于另一个同时考虑时间和空间特征的 ASTGCN 模型,分别使 MAE 值和 RMSE 值降低了 10.42%和 4.49%;与 STGODE 相比,所提出模型在两个数据集的评价指标值仅有 PeMSD4 数据集的 RMSE

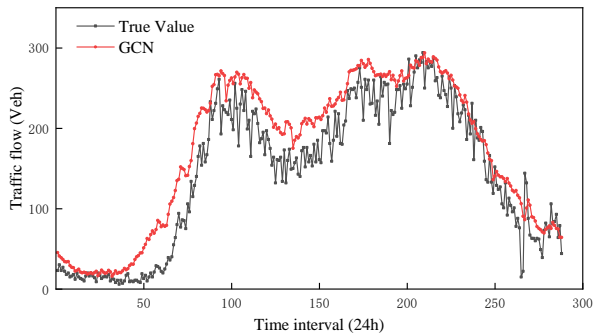
值较高,其余评价指标值均更低。图 8 展示使用上述模型对 PeMSD4 数据集第 104 节点的预测结果。本文的 STBiPGAT 方法考虑了时空动态特征,能有效捕捉节点演化趋势,从而描述节点在时空上的行为。因此,本文模型相对于其他模型预测更精确,是一种有前途的交通预测方法。



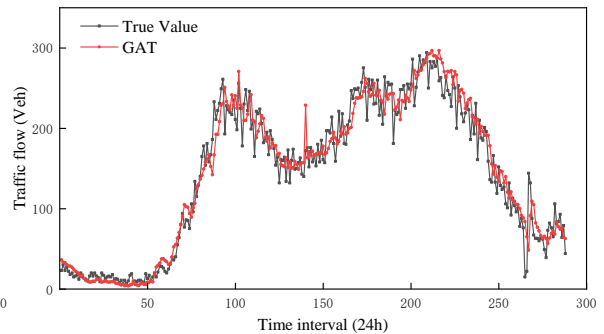
(a) HA 模型



(b) BPNN 模型



(c) GCN 模型



(d) GAT 模型

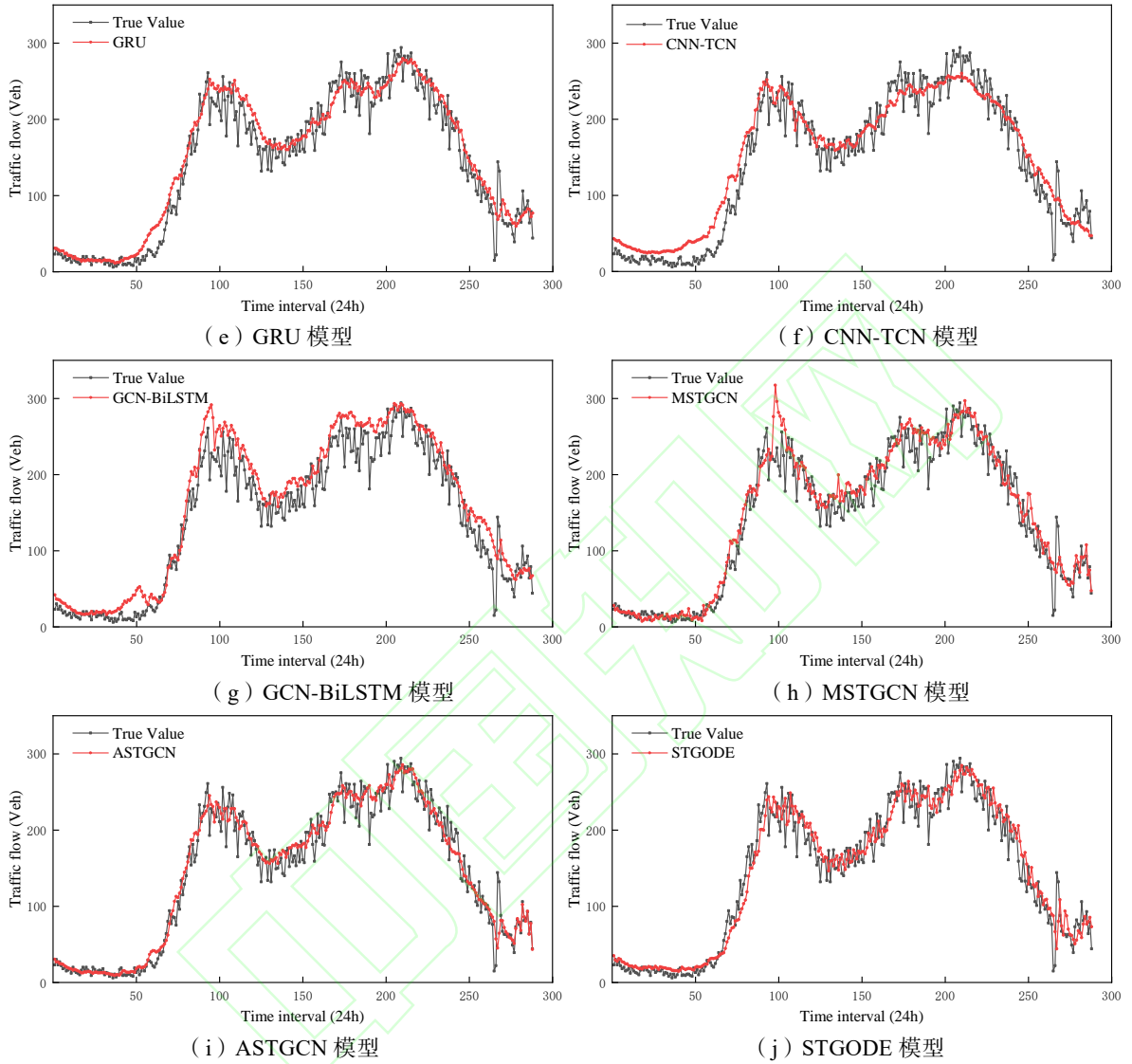


图 8 PeMSD4 数据集 104 号节点使用不同模型的预测结果

Fig.8 The prediction results of different models using the PeMSD4 dataset

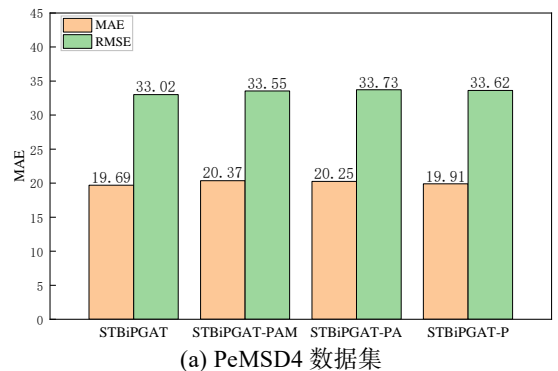
### 3.6 消融分析

为进一步探究本文模型的优越性,下面进行消融对比实验。基于 STBiPGAT 模型设计了 3 个变体,详情如下。

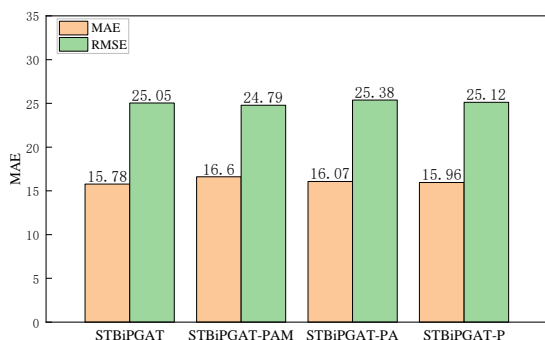
(1) STBiPGAT-PAM: 移除相关系数矩阵特征融合模块、时间自注意力机制模块、空间多头注意力模块。

(2) STBiPGAT-PA: 移除相关系数矩阵特征融合模块、时间自注意力机制模块。

(3) STBiPGAT-P: 移除相关系数矩阵特征融合模块。



(a) PeMSD4 数据集



(b) PeMSD8 数据集

图9 消融实验

Fig.9 Ablation experiment

由图9(a)中可以看出,STBiPGAT模型在PeMSD4数据集上实现了最佳性能,相比于STBiPGAT-P、STBiPGAT-PA、STBiPGAT-PAM变体模型,其MAE值分别降低了1.11%、2.84%和3.45%,RMSE值分别降低了1.82%、2.15%和1.61%,在PeMSD8数据集上的实验结果也呈现类似的趋势,这些结果直接验证了所引入模块的有效性。由此可见适当多头数的注意力机制能够从多个角度捕捉不同相邻道路关联度,可以有效增强模型鲁棒性;自注意力机制能够捕获节点不同时刻的重要性,自适应地调整权重;而相关系数矩阵特征融合模块能帮助模型深度挖掘节点间的隐藏关系,进一步增强模型表达能力。综上,本文提出的STBiPGAT模型中各模块都有效果,能够更好地帮助模型完成交通流预测任务。

## 4 总结

本文提出了基于深度学习框架STBiPGAT的交通流量预测模型,集成了GAT和BiGRU,共同学习路网的时空依赖性。模型输入是时间序列图,每个时间步的图由节点和边组成。在空间依赖性提取块中,多头注意力机制的GAT建模节点与邻居之间的关系,同时融合全局相关系数矩阵,深度挖掘空间依赖关系,捕捉节点局部结构。在时间特征提取块中,BiLSTM编码每个时间步的节点状态,获取全面时间特征,捕捉节点动态演化趋势。STBiPGAT将空间依赖性提取块和时间特征提取块输出融合,并通过全连接层输出预测结果。实验证明,STBiPGAT在动态空间网络预测任务上具有更佳准确率和稳定性。未来交通流预测模型还将考虑更多因素,例如天气、节假日等。这些因素对于交通流状况产生重大影响,而且往往具有突发性或周期性。因此,将这些因素融入到模型中,可以提高模型鲁棒性和准确性,以帮助预测交通流量变化,并提供更加个性化、智能化的出行建议和交通管

理方案。

## 参考文献:

- [1] 杨旭东. 智能交通:城市拥堵治理的新路径[J]. 智慧中国, 2020, No.51(5): 64-66.  
YANG X D. Intelligent Transportation: A New Path for Urban Congestion Management[J]. Wisdom China, 2020, No.51(5): 64-66.
- [2] SMITH B L, DEMETSKY M J. Traffic flow forecasting: comparison of modeling approaches[J]. Journal of Transportation Engineering, 1997, 123(4): 261-266.
- [3] LI L, LIN W H, LIU H. Type-2 fuzzy logic approach for short-term traffic forecasting[C]//IEEE Proceedings-Intelligent Transport Systems. IET Digital Library, 2006, 153(1): 33-40.
- [4] AHMED M S, COOK A R. Analysis of freeway traffic time-series data by using Box-Jenkins techniques[J]. Transportation Research Record, 1979, 772: 1-9.
- [5] STATHOPOULOS A, KARLAFTIS G M. A multivariate state space approach for urban traffic flow modeling and prediction[J]. Transportation Research Part C, 2003, 11(2): 121-135.
- [6] YAN H, QI Y, YU D J. Short-term traffic flow prediction based on a hybrid optimization algorithm[J]. Applied Mathematical Modelling, 2022, 102: 385-404.
- [7] LIN G, LIN A, GU D. Using support vector regression and K-nearest neighbors for short-term traffic flow prediction based on maximal information coefficient[J]. Information Sciences, 2022, 608: 517-531.
- [8] 沈夏炯, 张俊涛, 韩道军. 基于梯度提升回归树的短时交通流预测模型[J]. 计算机科学, 2018, 45(6): 222-227.  
SHEN X Q, ZHANG J T, HAN D J. Short-term traffic flow prediction model based on gradient boosting regression tree[J]. Computer Science, 2018, 45(6): 222-227.
- [9] KASHYAP A A, RAVIRAJ S, DEVARAKONDA A, et al. Traffic flow prediction models-A review of deep learning techniques[J]. Cogent Engineering, 2022, 9(1): 2010510.
- [10] 王庆, 田可可, 朱昌锋, 等. 融合多因素的短时交通流预测研究[J]. 计算机工程与应用, 2022, 58(21).  
WANG Q, TIAN K K, ZHU C F, et al. Short term traffic flow prediction based on Multi-Factors[J]. Journal of Computer Engineering & Applications, 2022, 58(21).
- [11] WANG X X, XU L H, CHEN K X. Data-driven short-term forecasting for urban road network traffic based on data processing and LSTM-RNN[J]. Arab. J. Sci. Eng. 2019, 44: 3043-3060.
- [12] MEHDI M Z, KAMMOUN H M, BENAYED N G, et al. Entropy-based traffic flow labeling for CNN-based traffic congestion prediction from meta-parameters[J]. IEEE Access, 2022, 10: 16123-16133.
- [13] ISLAM Z, ABDEL-ATY M, MAHMOUD N. Using CNN-LSTM to predict signal phasing and timing aided by High-Resolution detector data[J]. Transportation research part C: emerging technologies, 2022, 141: 103742.
- [14] ZHANG D, KABUKA M R. Combining weather condition data to predict traffic flow: a GRU-based deep learning approach[J]. IET Intelligent Transport Systems, 2018, 12(7): 578-585.
- [15] TIAN Y, ZHANG K, LI J, et al. LSTM-based traffic flow prediction with missing data[J]. Neurocomputing, 2018,

- 318: 297-305.
- [16] 杨柳.深度学习在交通运输中应用综述[J].信息技术与信息化,2022(2):192-195.  
YANG L. A review of deep learning applications in transportation[J]. Information Technology and Informatization, 2022(2):192-195.
- [17] WU Y, TAN H. Short-term traffic flow forecasting with spatial-temporal correlation in a hybrid deep learning framework[J]. arXiv preprint arXiv:1612. 01022, 2016.
- [18] MA X, DAI Z, HE Z, et al. Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction[J]. Sensors, 2017, 17(4): 818.
- [19] YU B, YIN H, ZHU Z X. Spatio-temporal graph convolutional networks: a deep learning frame work for traffic forecasting[C]//Proceedings of the 27th International Joint Conference on Artificial Intelligence. Stock- holm, Sweden, 2018, 3634-3640.
- [20] 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. Palo Alto, USA: AAAI Press, 2019, 33(1): 922-929.
- [21] FANG Z, LONG Q, SONG G, et al. Spatial-temporal graph ode networks for traffic flow forecasting[C]//Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining. 2021: 364-373.
- [22] VELICKOVIC P, CUCURULL G, CASANOVA A, et al. Graph attention networks[J]. arXiv preprint arXiv, 2017, abs/1710.10903.
- [23] OKWONU F Z, ASAJU B L, ARUNAYE F I. Breakdown analysis of pearson correlation coefficient and robust correlation methods[C]//IOP Conference Series: Materials Science and Engineering. IOP Publishing, 2020, 917(1): 012065.