

# 基于自适应残差动态融合图注意力网络的交通速度预测

张鲁宁,王景升\*

(中国人民公安大学交通管理学院,北京 100038)

**摘要:**在现代智能交通系统领域里,准确预测交通速度对缓解交通拥堵、提高道路安全以及优化交通管理有着重要意义。为提升现有的交通速度预测模型在中长期预测任务中的性能,本文提出一种自适应残差动态融合图注意力网络的交通速度预测方法,该方法中的双模态图架构通过对自适应邻接矩阵与动态邻接矩阵进行并行处理和动态融合,可以捕捉路网静态拓扑和动态时空关联特征;采用门控时间卷积实现特征筛选,并利用多头注意力机制增强时空特征表达能力,设计动态特征融合单元,通过残差连接保留静态拓扑信息,结合跨层多尺度特征融合避免特征退化。实验结果显示,与 Graph WaveNet 相比,本模型在 METR-LA 和 PEMS-BAY 数据集 60 min 预测任务中均方根误差分别减少 22.5% 和 22.6%。该模型能够实时适应交通状态的变化,为交通管理部门提供精准的速度预测,辅助拥堵疏导、动态路径规划和突发事件响应,具备较高的实际应用价值。

**关键词:**城市交通;交通速度预测;双模态图融合;动态图卷积网络;残差机制;注意力机制

**中图分类号:**TP391;U491 **文献标志码:**A

**引用格式:**张鲁宁,王景升. 基于自适应残差动态融合图注意力网络的交通速度预测[J]. 山东大学学报(理学版),2026,61(5):90-101,113.

## Traffic speed prediction study based on adaptive residual dynamic fusion graph attention network

ZHANG Luning, WANG Jingsheng\*

(School of Traffic Management, People's Public Security University of China, Beijing 100038, China)

**Abstract:** In the field of modern intelligent transportation systems, accurate prediction of traffic speed is of great significance to alleviate traffic congestion, improve road safety, and optimize traffic management. To improve the performance of existing traffic speed prediction models in medium and long-term prediction tasks, this paper proposes an adaptive residual dynamic fusion graph attention network for traffic speed prediction, in which the bimodal graph architecture can capture static topology and dynamic spatio-temporal correlation features of the road network through parallel processing and dynamic fusion of adaptive and dynamic adjacency matrices. Applying gated temporal convolution to realize the feature screening, and using multi-head attention mechanism to enhance the spatio-temporal feature expression ability, designing dynamic feature fusion unit, retaining static topological information through residual connection, and combining cross-layer multi-scale feature fusion to avoid feature degradation. The experimental results show that the root mean square error of this model is reduced by 22.5% and 22.6% compared with Graph WaveNet in the 60 min prediction task for the METR-LA and PEMS-BAY datasets, respectively. The model can adapt to the changes of the traffic state in real time, provide accurate speed prediction for the traffic management department, and assist in the congestion diversion, dynamic path planning, and emergency response. The model has high practical application value.

**Key words:** city traffic; traffic speed prediction; bimodal map fusion; dynamic graph convolutional network; residual connection; attention mechanism

## 0 引言

机动车保有量的持续增长在提升出行便利性的同时,也加剧了交通拥堵与事故风险。面对有限的道路

收稿日期:2024-10-28;网络出版时间:2025-09-03

基金项目:国家重点研发计划资助项目(2023YFB4302701)

第一作者:张鲁宁(2001—),男,硕士研究生,研究方向为智能交通、交通预测。E-mail:201921310034@stu.ppsuc.edu.cn

\*通信作者:王景升(1970—),男,硕士,副教授,研究方向为智能交通、信号控制。E-mail:wjs1970@vip.163.com

资源与不断增长的交通需求,发展智能交通系统成为关键。其中,交通预测作为智能交通系统的核心技术,对于优化出行效率、提升道路安全具有至关重要的作用。早期研究主要基于统计学模型,如运用自回归整合移动平均(auto regressive integrated moving average, ARIMA)<sup>[1]</sup>模型分析高速公路交通时间序列数据,其建模过程透明且可解释性强,在短时预测场景表现良好,然而受限于线性假设,难以捕捉交通数据中的非线性模式。随着技术进步,传统机器学习方法如 $k$ 近邻<sup>[2]</sup>、支持向量机(support vector machine, SVM)<sup>[3]</sup>、决策树<sup>[4]</sup>等广泛应用于交通预测领域,但这类方法在识别动态和复杂的交通模式方面仍存在局限。

为突破传统方法的局限,深度学习模型被引入交通预测领域。研究者采用循环神经网络(recurrent neural network, RNN)<sup>[5]</sup>构建时序依赖模型,有效捕捉序列数据的短期动态特征,然而在处理长时交通序列时,RNN面临严重的梯度爆炸问题。长短期记忆网络<sup>[6]</sup>通过门控机制选择性记忆关键信息,提升模型长期预测能力,但其建模能力仅局限于时间维度,不适用于具有复杂空间依赖性的路网分析。为此,卷积神经网络(convolutional neural network, CNN)<sup>[7]</sup>被引入交通预测领域中。然而,CNN的欧几里得卷积核机制面临根本性局限——交通网络的非规则拓扑结构使得传统“邻域”概念失效。Cui等<sup>[8]</sup>引入图卷积网络直接编码路网拓扑,显著提升空间依赖性建模的准确性,但仍存在难以捕获多尺度层级特征和无法响应动态变化图结构的问题。Li等<sup>[9]</sup>开发多尺度图卷积架构,通过分层邻域聚合策略实现跨尺度特征提取。Wu等<sup>[10]</sup>借助数据驱动的方法来学习并更新图结构,解决图结构不贴近实际路网情况的问题。针对动态图处理这一具体挑战,学者提出的图注意力网络<sup>[11]</sup>以及动态图卷积网络<sup>[12]</sup>提高了预测的准确度。图注意力网络通过计算节点间注意力系数实现邻居特征加权聚合,多头注意力机制的引入显著增强模型对多模式特征的捕获能力与鲁棒性。王博文等<sup>[13]</sup>提出融合时空模块的图注意力模型,因依赖固定图结构,多步预测中误差随步长放大。Zhu等<sup>[14]</sup>提出的改进图注意力模型提高预测准确度,创新性构建动态时空注意力机制,实验证明其在多步预测中具有卓越性能。

然而,随着预测时间步长不断增加,现有模型的准确性和稳定性会明显下降。Zhang等<sup>[15]</sup>提出的深度时空残差网络虽擅长捕捉短期时空依赖,却因依赖链断裂导致误差呈指数累积。Wu等<sup>[16]</sup>提出的混合深度学习模型受限于注意力机制的历史模式偏差及RNN长时捕捉缺陷。Cini等<sup>[17]</sup>虽提升了预测精度,但算法计算复杂度过高且对数据噪声敏感。Ji等<sup>[18]</sup>提出的基于时空自监督学习(spatio temporal self supervised learning, ST-SSL)的交通预测框架,通过自适应数据增强任务建模时空异质性,但随着预测时间步数延长,模型对历史时空依赖的捕捉能力下降,导致误差逐步上升。Wang等<sup>[19]</sup>提出基于时空关联的多变量预测模型,利用历史注意力机制增强模型表现,但交通变量的非线性耦合关系未被充分捕捉,长期预测中变量间的协同变化偏离实际。Li等<sup>[20]</sup>提出OpenCity融合Transformer与图神经网络,其固定1小时的时间窗分割导致细粒度波动特征捕获缺失,削弱了长期趋势预测精度。孙大盟等<sup>[21]</sup>改进的图卷积网络通过动态聚合与残差设计增强空间建模能力,却因静态邻接矩阵限制致使超过60 min预测的周期特征误差增加7.5%,这凸显了动态空间关联缺失的累积效应。Zhang等<sup>[22]</sup>提出多尺度动态图卷积网络,其静态与自适应图的线性耦合机制忽视空间依赖强度的非线性演化,导致30 min以上预测的空间结构失真。Cai等<sup>[23]</sup>提出的双图Transformer虽借时间聚合捕捉长期依赖,但为缓解计算负担被迫舍弃关键短期波动,难以平衡预测精度与效率。综上所述,以往模型存在动态空间关联建模缺陷、梯度消失、特征混淆、中长期预测中误差累积、计算效率低等问题。

基于以上研究,本文提出双模态图融合的自适应残差动态融合图注意力网络(adaptive residual dynamic graph attention networks, ARDGAT)用于交通速度预测。实证表明该模型在短/中长期预测任务中的性能显著优于基线模型。在模型应用层面,通过中长期速度预测生成拥堵预警窗口支撑动态限速与匝道控制策略,实时输出路段级速度波动趋势赋能车载导航高精度动态路径规划,优化路网主动管理响应。本文的创新点主要体现在以下3个方面:

(1)为优化动态空间建模并提升计算效率,设计双模态图并行架构,提出自适应邻接矩阵与动态邻接矩阵的并行架构。前者捕获路网固有连接特性,后者利用扩散图卷积和图生成器实时建模交通流动态变化,通过可学习参数实现动态权重融合,在维持计算效率的同时显著增强长时交通流动态捕捉能力。

(2)针对传统时序模型混淆突发波动与周期特征的缺陷,本文设计门控时空卷积与多头注意力协同架构,时间维度采用膨胀因果卷积捕获多尺度周期规律,空间维度利用多头注意力机制并行提取差异化空间模式,通过时空协同机制有效分离稳态规律与瞬时扰动,显著增强模型对时空特征的捕捉能力。

(3) 针对深层网络特征退化与静态拓扑丢失问题,构建残差跨层多尺度特征融合架构,通过动态特征融合单元把静态拓扑特征以残差形式保留,避免因梯度消失导致路网物理连接信息被稀释,采用跨层跳跃连接整合浅层微观波动与深层宏观趋势,构建多粒度时空依赖的联合表征。

# 1 图结构建模

## 1.1 双模态图结构

相比于传统单动态邻接矩阵方法,本研究提出双模态图结构并行架构,构建自适应邻接矩阵  $A_{apt}$  与动态邻接矩阵  $A_{dyn}$  实现空间依赖的多角度建模。其中,自适应邻接矩阵  $A_{apt}$  依靠可学习的节点嵌入捕捉路网固有的静态拓扑结构,能够稳定地反映出长期交通模式;而动态邻接矩阵  $A_{dyn}$  则运用扩散图卷积和 Gumbel Softmax 技术,对交通速度变化中所蕴含的动态空间关联特征进行实时建模。这 2 种邻接矩阵并非进行简单串联,而是分别经过空间卷积模块处理后通过自适应权重开展动态融合,让模型可以根据实际交通状况自动调整静态和动态特征的比重。

## 1.2 自适应邻接矩阵

将路网拓扑结构表示为  $G=(V,E)$ ,节点集合为  $V$ ,边集合为  $E$ ,初始邻接矩阵为  $A$ 。图 1(a) 是 PEMS-BAY 和 METR-LA 数据集基于道路连接与地理位置的初始邻接矩阵热力图,它仅能反映路网的静态拓扑连接情况,既没办法适应实际交通数据的动态变化,又可能会遗漏潜在的节点关系;图 1(b) 为动态邻接矩阵生成示意图。

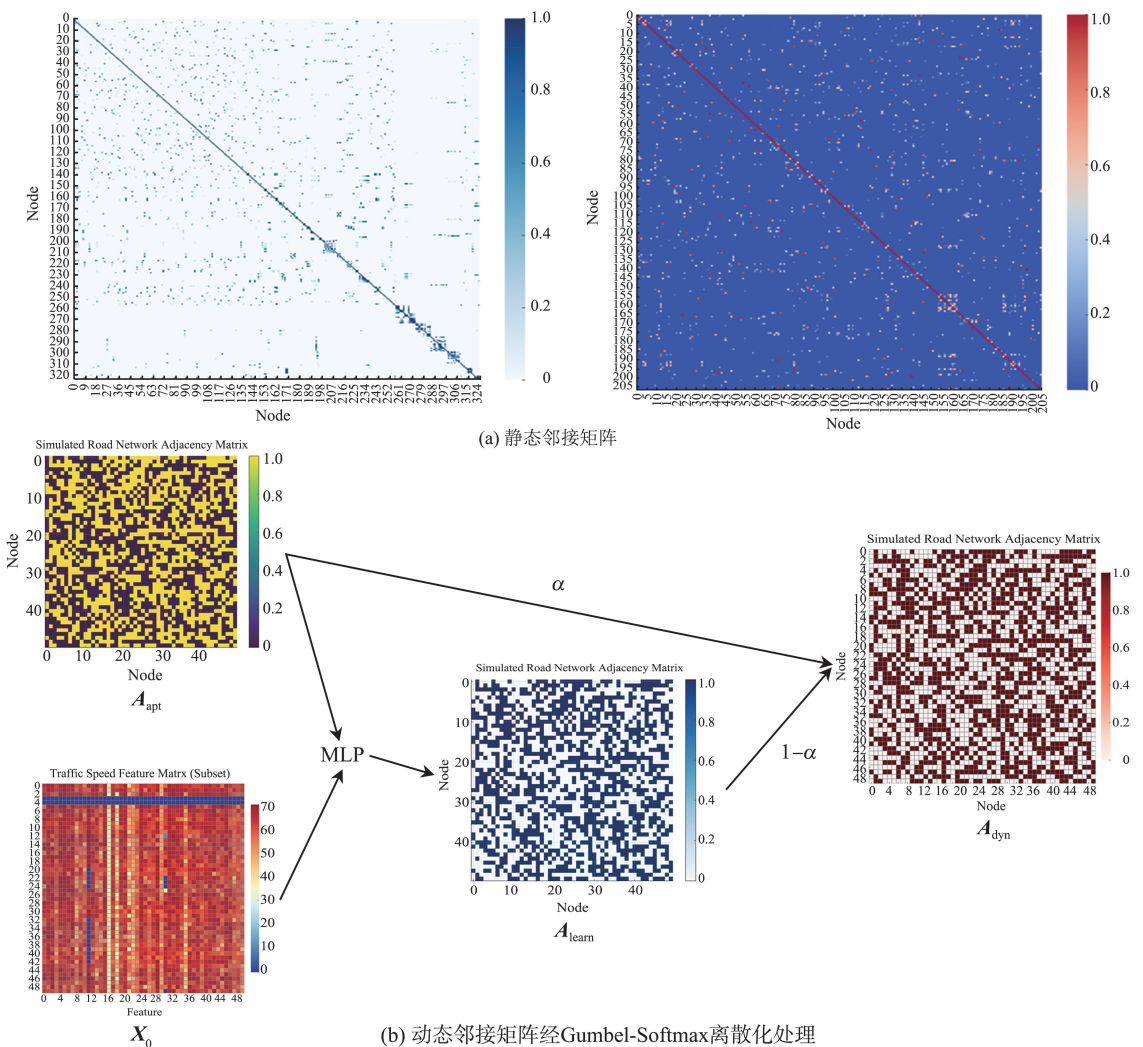


图 1 邻接矩阵热力图  
Fig.1 Heat map of the adjacency matrix

自适应邻接矩阵可通过动态学习节点间的隐含关系,捕捉交通流随时间的动态传播模式。本文定义自适应邻接矩阵为  $\mathbf{A}_{\text{apt}} \in \mathbf{R}^{N \times N}$ ,  $N$  为节点数,它能依据训练结果自动调整权重,用来替代传统预定义的静态邻接矩阵,其表示为

$$\mathbf{A}_{\text{apt}} = \text{Softmax}(\text{Relu}(\mathbf{E}_1 \mathbf{E}_2)), \quad (1)$$

$$\mathbf{E}_2 = \mathbf{E}_1^T, \quad (2)$$

其中,  $\mathbf{E}_1 \in \mathbf{R}^{N \times C}$  和  $\mathbf{E}_2 \in \mathbf{R}^{C \times N}$  是每个节点初始随机且能够学习的参数向量,  $C$  代表嵌入维度,  $\mathbf{E}_1$  是源节点嵌入矩阵,它的每行表示路网里一个节点的嵌入,  $\mathbf{E}_2$  为目标节点嵌入矩阵,即  $\mathbf{E}_1$  的转置矩阵。

如图 2 所示,将  $\mathbf{E}_1$  和  $\mathbf{E}_2$  相乘,得到源节点和目标节点之间的空间依赖权值。该权值能够表征 2 个节点之间的潜在关联和相似性,可准确反映出路网节点间的空间依赖关系。使用 Relu 激活函数消除弱连接以保证权重非负,采用 Softmax 函数对自适应邻接矩阵进行归一化处理,生成概率分布形式的邻接权重。

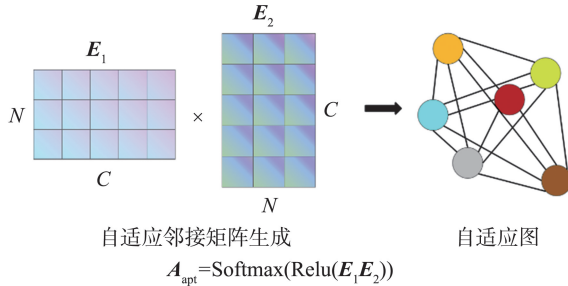


图 2 自适应邻接矩阵生成示意图

Fig.2 Schematic of adaptive neighborhood matrix generation

### 1.3 动态邻接矩阵

在现实交通场景中,路段间的关联性随着交通流的时空状态而改变。为了探索更深层次的空间相关性,本文使用图生成器构建一个具有时空特征的邻接矩阵,来模拟路网中非相邻路段产生的临时性强关联情况。传统的动态邻接矩阵通过节点特征相似度直接生成,仅能反映出交通流的单步关联状况,而本方法提出的动态邻接矩阵可捕捉交通流级联传播效应同时抑制非必要的弱关联噪声,具体构建原理如图 1(b) 所示。首先将特征矩阵  $\mathbf{X}_0 \in \mathbf{R}^{C \times N \times T}$  ( $T$  为时间步) 和  $\mathbf{A}_{\text{apt}}$  输入图生成器,进行  $K$  阶扩散图卷积处理,通过扩散步长  $k$  显式捕捉路网中远距离节点间的潜在关联,解决复杂路网的间接空间依赖难以建模的问题,表示为

$$\text{GCN}(\mathbf{X}_0, \mathbf{A}_{\text{apt}}) = \sum_{k=0}^K (\mathbf{A}_{\text{apt}}^k \cdot \mathbf{X}_0) \mathbf{W}, \quad (3)$$

其中,  $\mathbf{W}$  是图卷积层的权重矩阵,  $k$  为扩散步长,  $K$  是扩散步长最大值,  $\text{GCN}(\mathbf{X}, \mathbf{A}_{\text{apt}})$  表示扩散图卷积层,  $\mathbf{A}_{\text{apt}}^k$  是邻接矩阵  $\mathbf{A}_{\text{apt}}$  的  $k$  次方,通过建模扩散  $k$  步后节点间的连接关系,表征交通流传播的多步空间依赖性。

随后使用多层感知机 (multilayer perceptron, MLP) 将扩散图卷积提取的特征映射为可解释的节点关联概率,捕捉高峰时段车流突变与平峰期稳态车流的差异化关联,实现数据驱动的图结构学习,并生成离散连接概率矩阵  $\mathbf{A}' \in \mathbf{R}^{N \times N}$ :

$$\mathbf{A}' = \text{Softmax}(\text{MLP}(\text{GCN}(\mathbf{X}_0, \mathbf{A}_{\text{apt}}))). \quad (4)$$

Softmax 函数用于将最终的输出转换成一个概率分布,对每个节点的特征归一化,双 tanh 函数激活将特征压缩至  $[-1, 1]$ ,防止梯度爆炸,其具体结构为

$$\mathbf{H}_1 = \tanh(\text{BN}(\mathbf{W}_1 \cdot \text{GCN}(\mathbf{X}_0, \mathbf{A}_{\text{apt}}) + \mathbf{b}_1)), \quad (5)$$

$$\mathbf{H}_2 = \tanh(\text{BN}(\mathbf{W}_2 \cdot \mathbf{H}_1 + \mathbf{b}_2)), \quad (6)$$

$$\mathbf{A}' = \text{Softmax}(\mathbf{W}_3 \cdot \mathbf{H}_2 + \mathbf{b}_3), \quad (7)$$

其中, BN 表示批量归一化层,  $\mathbf{H}_1 \in \mathbf{R}^{N \times 2N}$  为 MLP 第一层的输出,  $\mathbf{H}_2 \in \mathbf{R}^{N \times N}$  为 MLP 第二层的输出,  $\mathbf{W}_1 \in \mathbf{R}^{N \times 2N}$ ,  $\mathbf{W}_2 \in \mathbf{R}^{2N \times N}$ ,  $\mathbf{W}_3 \in \mathbf{R}^{N \times N}$  分别为第一、二、三层的权重矩阵,  $\mathbf{b}_1 \in \mathbf{R}^{2N}$ ,  $\mathbf{b}_2 \in \mathbf{R}^N$ ,  $\mathbf{b}_3 \in \mathbf{R}^N$  分别为第一、二、三层的偏置向量。

由于  $\mathbf{A}' \in \mathbf{R}^{N \times N}$  为离散矩阵, Gumbel Softmax 将 MLP 输出的离散连接概率矩阵  $\mathbf{A}'$  离散化为 0—1 分布

的邻接矩阵  $\mathbf{A}_{\text{learn}}$ 。该操作通过抑制特定时段下的无效连接,避免全连接结构引入的噪声干扰,从而提升模型在传感器缺失场景中的稳定性。具体公式如下:

$$\mathbf{A}_{\text{learn}} = \text{GumbelSoftmax}(\mathbf{A}') = \text{Softmax} \frac{(\log(\mathbf{A}') - \log(-\log(g)))}{\tau}, \quad (8)$$

其中,  $g$  是从  $\text{Gumbel}(0, 1)$  分布中采样的随机噪声,  $\tau$  是温度参数,控制输出分布的离散程度,通过控制参数实现直通估计。 $\mathbf{A}_{\text{learn}} \in \mathbf{R}^{N \times N}$  为图生成器生成的动态邻接矩阵,可以模拟路网节点之间生成的动态关联。

引入可学习自适应参数因子  $\alpha$ ,通过反向传播动态调节不同时间段下静态拓扑与实时交通特征的融合比例,平衡静态拓扑与动态变化的贡献。将  $\mathbf{A}_{\text{learn}} \in \mathbf{R}^{N \times N}$  与  $\mathbf{A}_{\text{apt}}$  加权相加处理得到  $\mathbf{A}_{\text{dyn}} \in \mathbf{R}^{N \times N}$ :

$$\mathbf{A}_{\text{dyn}} = \alpha \mathbf{A}_{\text{apt}} + (1 - \alpha) \mathbf{A}_{\text{learn}}, \quad (9)$$

通过式(9)得到的  $\mathbf{A}_{\text{dyn}}$  本质上是一个时变权重矩阵。

## 2 自适应残差动态融合图注意力网络

### 2.1 模型框架

自适应残差动态融合图注意力网络模型架构如图3所示,包含矩阵生成层、时空特征堆叠层和输出层。首先,输入特征矩阵  $\mathbf{X}_0$  经扩散图卷积层与图生成器处理,输出自适应邻接矩阵  $\mathbf{A}_{\text{apt}}$  和动态邻接矩阵  $\mathbf{A}_{\text{dyn}}$ ,同时通过全连接层对  $\mathbf{X}_0$  进行维度变换,生成特征矩阵  $\mathbf{X}'_0$ ;然后,时空特征堆叠层采用层级渐进处理机制接收  $\mathbf{X}'_0$  和邻接矩阵,第  $k$  层融合第  $k-1$  层的输出特征与高层反馈信息,每一层中分别经由时间卷积在时间维度上进行信息处理,空间维度建模(并行使用图卷积网络和图注意力网络分别处理  $\mathbf{A}_{\text{apt}}$  和  $\mathbf{A}_{\text{dyn}}$ ,提取静态拓扑与动态关联特征)和动态特征融合(协同优化两类空间特征,实现路网静态结构保护与动态时空关联建模的平衡),该分层架构通过残差跳跃连接传递多尺度特征,使底层低阶特征向顶层高阶特征渐进抽象;最终,输出层整合各层特征,经全连接网络生成预测结果矩阵  $\mathbf{Z}$ 。

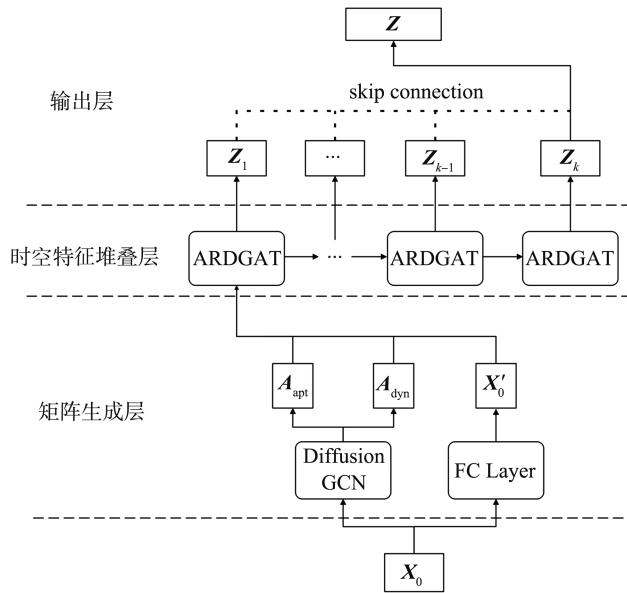


图3 模型架构图

Fig.3 Model architecture diagram

### 2.2 矩阵生成层

输入特征矩阵  $\mathbf{X}_0$  通过一个全连接层进行特征维度变换,计算公式如下:

$$\mathbf{X}'_0 = \text{Relu}(\mathbf{W}\mathbf{X}_0 + b), \quad (10)$$

其中,  $\mathbf{X}'_0$  为输出特征矩阵,  $\mathbf{W}$  和  $b$  为可学习的参数。

为模拟信息在图中的多步扩散过程来增强节点特征,采用扩散图卷积层获取路网中不相邻节点之间的空间依赖关系:

$$\text{DGCN}(\mathbf{X}, \mathbf{A}_{\text{dyn}}) = \sum_{k=0}^K (\mathbf{A}_{\text{apt}}^k \cdot \mathbf{X}'_0) \mathbf{W}_0, \quad (11)$$

其中,  $\text{DGCN}(\mathbf{X}, \mathbf{A}_{\text{dyn}})$  表示扩散图卷积层,  $\mathbf{A}_{k_{\text{apt}}} \cdot \mathbf{X}'_0$  计算特征矩阵  $\mathbf{X}_0$  在  $k$  步扩散后的累积影响, 通过权重矩阵  $\mathbf{W}_0$  进行转换产生新的节点特征表示, 并通过求和得到最终的节点表示。

### 2.3 时空特征堆叠层

#### 2.3.1 门控时间卷积模块

门控时间卷积模块核心设计目标是从复杂交通速度序列提取多尺度时间依赖, 并动态筛选关键特征, 包含 1D 扩张因果卷积(图 4)以及门控机制 2 个部分。与递归神经网络相比, 扩张卷积具有以下优势: 可通过增加层的深度显著扩展感受野, 不会引入 RNN 固有的梯度爆炸问题, 能用较少层数捕捉长距离依赖关系并减轻梯度爆炸风险, 在不引入未来信息情况下, 只在当前步骤插入零进行预测以保持因果性。在实际操作中这种卷积通过在输入序列上滑动并应用不同步长的滤波器来执行, 对于一个一维输入序列  $x \in \mathbf{R}^D$  和一个滤波器  $f \in \mathbf{R}^k$ , 扩张因果卷积的操作表示为

$$x * f(t) = \sum_{s=0}^{p-1} f(s) \cdot x(t-d * s), \quad (12)$$

其中,  $d$  是扩张因子, 用来控制滤波器跳过的间隔距离, 即控制时间窗口的感知范围, 从而学习短期和中长期时间粒度的交通模式,  $f(t)$  代表在时间步处的输出,  $f(s)$  代表滤波器的权重,  $p$  为卷积核的大小, 即滤波器中的权重数,  $s$  为索引变量。

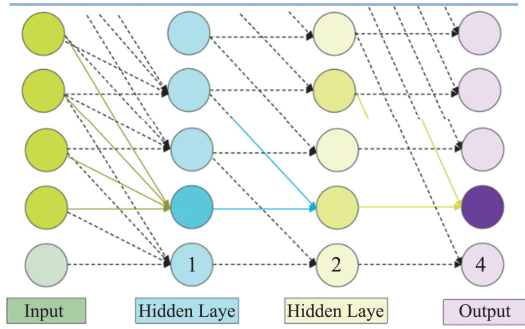


图 4 1D 扩张因果卷积示意图  
Fig.4 Schematic of 1D dilation causal convolution

输入序列经过时间卷积模块(temporal convolutional network, TCN)提取时间序列特征后, 分别通过激活函数 tanh 和 sigmoid 后相乘实现门控机制。sigmoid 函数的输出相当于门控信号, 决定 tanh 输出传递到网络的下一层的信息, 输入特征矩阵为  $\mathbf{X}$ , 门控时间卷积表示为

$$\mathbf{H}_t = f(\mathbf{W}_1 \mathbf{X} + \mathbf{c}) \odot \sigma(\mathbf{W}_2 \mathbf{X} + \mathbf{d}), \quad (13)$$

其中,  $\mathbf{W}_1$  和  $\mathbf{W}_2$  为 TCN 卷积层的权重矩阵,  $\mathbf{c}$  和  $\mathbf{d}$  是添加到 TCN 卷积层输出上的偏置项,  $f$  为用于输出的激活函数, 负责引入非线性特性,  $\sigma$  为 sigmoid 激活函数,  $\odot$  表示元素间的乘积,  $\mathbf{H}_t$  为输出的特征矩阵。

#### 2.3.2 空间卷积模块

单一空间建模不能同时兼顾路网物理约束和交通流动态传播, 使得预测结果和实际情况存在偏离。为了充分捕捉路网静态拓扑和动态时空之间的关联, 本模块采用并行双分支架构, 即分别运用图卷积网络和图注意力网络对门控时间卷积模块所得到的自适应邻接矩阵加以处理。这样既确保长期交通模式和路网拓扑结构不会被忽略, 又能够捕捉到物理上不连通但交通流强相关的节点关联性。动态邻接矩阵经过 Gumbel-Softmax 离散化之后输入到 GAT 中, 通过谱域中图卷积网络学习自适应邻接矩阵特征, 该特征矩阵  $\mathbf{H}'$  可以通过一阶切比雪夫多项式展开来进行近似:

$$\mathbf{H}' = (\mathbf{I}_N + \mathbf{D}^{1/2} \mathbf{A} \mathbf{D}^{-1/2}) \mathbf{X}_0 \mathbf{W}, \quad (14)$$

其中,  $\mathbf{A}$  是邻接矩阵,  $\mathbf{D}$  是度矩阵,  $\mathbf{I}_N$  是单位矩阵,  $\mathbf{W}$  是图卷积核,  $\mathbf{X}_0$  是经过 TCN 输出特征矩阵。

将公式中归一化邻接矩阵  $\mathbf{D}^{1/2} \mathbf{A} \mathbf{D}^{-1/2}$  替换为自适应邻接矩阵  $\mathbf{A}_{\text{apt}}$ , 可得到 GCN 在更高层次上处理后的输出特征矩阵  $\mathbf{H}_a$  为

$$\mathbf{H}_a = (\mathbf{I}_N + \mathbf{A}_{\text{apt}}) \mathbf{H}_t \mathbf{W}. \quad (15)$$

在动态图结构卷积上, 使用多头注意力网络 GAT 对提取动态图结构的特征, 生成时变空间权重, GAT 结构如图 5 所示。

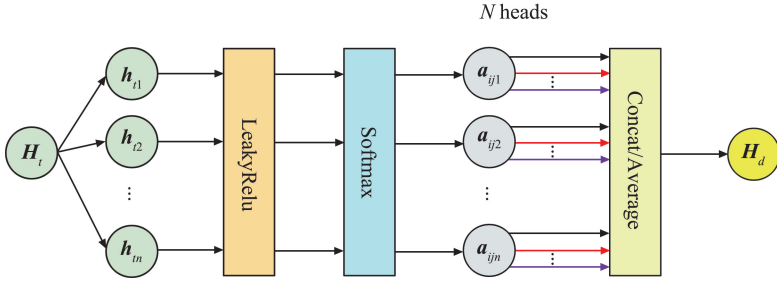


图 5 GAT 模型结构图

Fig.5 Structure of GAT model

计算节点  $j$  在节点  $i$  上的注意力值  $e_{ij}$ , 并将  $e_{ij}$  归一化处理得到注意力得分  $a_{ij}$ :

$$e_{ij} = \text{LeakyRelu}(\mathbf{a}^T [\mathbf{W} \mathbf{h}_{t,i} \parallel \mathbf{W} \mathbf{h}_{t,j}]), \tag{16}$$

$$a_{ij} = \frac{\exp(e_{ij})}{\sum_{j \in N_i} \exp(e_{ij})}, \tag{17}$$

其中,  $\mathbf{a}^T$  为矩阵乘法函数的转置,  $\mathbf{h}_{t,i}$  是节点  $i$  在  $t$  上的特征表示向量,  $\mathbf{h}_{t,j}$  是节点  $j$  在  $t$  上的特征表示向量, LeakyRelu 为激活函数,  $\mathbf{W}$  为将输入特征转化为输出特征的权重矩阵。

在多头图注意力网络中, 通过计算  $n$  组注意力系数并采用加权平均方法提取空间特征, 得到输出  $\mathbf{H}_d$ :

$$\mathbf{H}_d = \sigma \left( \frac{1}{K} \sum_{m=1}^M \sum_{j \in N_i} a_{ij}^m \mathbf{W} \mathbf{h}_{t,j}^m \right), \tag{18}$$

其中,  $a_{ij}^m$  为节点  $i$  在节点  $j$  上第  $m$  组注意力值,  $\mathbf{h}_{t,j}^m$  为在第  $m$  个注意力头中节点  $j$  在  $t$  上的特征表示向量,  $M$  为注意力头数,  $\sigma$  为 sigmoid 激活函数。

### 2.3.3 动态特征融合单元

动态特征融合单元(图 6)是模型的核心模块, 旨在通过改进传统门控循环单元(gated recurrent unit, GRU)自适应地融合静态特征与动态特征, 并结合残差连接保留关键信息, 避免深层网络忘记初始图结构。融合图卷积网络和图注意力网络输出矩阵, 生成综合时空表征  $\mathbf{H}_g$ 。为区分多源输入特征矩阵区别, 用  $\mathbf{H}_d$  和  $\mathbf{H}_a$  分别表示动态和静态特征输入。

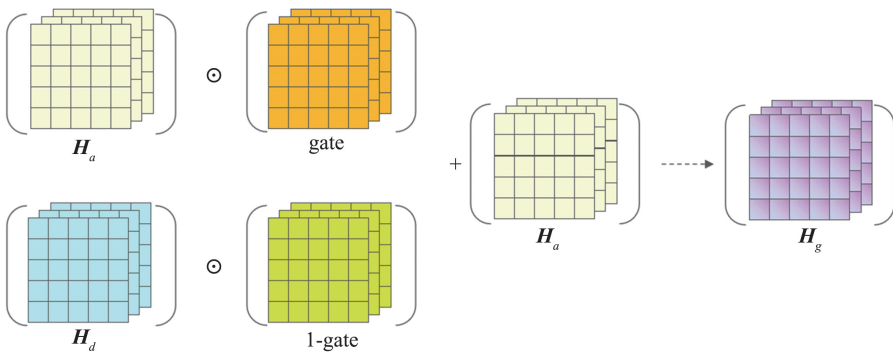


图 6 动态特征融合单元示意图

Fig.6 Schematic diagram of dynamic feature fusion unit

将静态特征  $\mathbf{H}_a$  和动态特征  $\mathbf{H}_d$  沿特征维度拼接, 生成联合特征  $[\mathbf{H}_a \parallel \mathbf{H}_d]$ , 通过权重矩阵  $\mathbf{W}_z$  和偏置项  $\mathbf{b}$  进行线性变换, 并应用 Sigmoid 激活函数  $\sigma$  生成门控值 gate:

$$\text{gate} = \sigma([\mathbf{H}_a \parallel \mathbf{H}_d] \mathbf{W}_z + \mathbf{b}). \tag{19}$$

利用门控值对静态特征  $\mathbf{H}_a$  和动态特征  $\mathbf{H}_d$  进行加权求和, 将原始静态特征  $\mathbf{H}_a$  作为残差项直接叠加到融合结果中, 避免深层网络训练中的梯度消失,  $\odot$  为元素级乘法, 公式如下:

$$\mathbf{H}_g = \text{gate} \odot \mathbf{H}_a + (1 - \text{gate}) \odot \mathbf{H}_d + \mathbf{H}_a. \tag{20}$$

损失函数  $L$  对静态特征  $\mathbf{H}_a$  的梯度为

$$\frac{\partial L}{\partial \mathbf{H}_a} = \frac{\partial L}{\partial \mathbf{H}_g} \odot (\text{gate} + \mathbf{H}_a \odot \frac{\partial \text{gate}}{\partial \mathbf{H}_a} + 1), \tag{21}$$

$$\frac{\partial \text{gate}}{\partial \mathbf{H}_a} = \text{gate} \odot (1 - \text{gate}) \odot \mathbf{W}_z[:, :, \dim(\mathbf{H}_a)], \quad (22)$$

其中  $\mathbf{W}_z[:, :, \dim(\mathbf{H}_a)]$  为权重矩阵中对应静态特征的分块。依据梯度传播公式(21),残差项  $\mathbf{H}_a$  将恒等映射引入梯度流,确保静态拓扑信息在深层网络中持续传递,从而有效缓解梯度消失问题,从根本上缓解梯度消失问题。

## 2.4 输出层

将低阶和高阶的时空特征堆叠层  $\{Z_1, Z_2, \dots, Z_k\}$  通过跨层跳跃连接整合多尺度特征,作为输出层  $Z$ 。最终,筛选后的特征经过2层全连接网络完成预测,第一层进行特征维度变换,第二层输出未来交通速度的预测结果,可得

$$Z = \text{Concat}(Z_1, Z_2, \dots, Z_k), \quad (23)$$

其中  $\text{Concat}(\cdot)$  表示沿特征维度的跳跃连接操作。

# 3 实例验证

## 3.1 数据集及实验环境参数介绍

选用数据集 METR-LA 和 PEMS-BAY 开展训练与验证,基本信息如表1所示。METR-LA 和 PEMS-BAY 是交通预测领域被广泛使用的基准数据集,覆盖洛杉矶高速公路以及湾区主要干道的交通数据,主要由表达交通网络结构和空间关系的邻接矩阵和记录路段速度变化情况的特征矩阵组成,包含工作日和周末等不同的交通运行模式,能保证数据的时空连续性以及场景的多样性。将2个数据集按照7:1:2的比例划分为训练集、验证集和测试集。

表1 数据集  
Table 1 Parameters of the datasets

数据集	节点个数	数据粒度	记录时间
METR-LA	207	5 min	2012-03-01—06-30
PEMS-BAY	325	5 min	2017-01-01—05-31

本实验在 Microsoft Windows 11 操作系统上运行,使用 NVIDIA GeForce RTX 4090 GPU 和 Python 3.9 的 Pytorch 深度学习框架进行模型训练与测试。交通预测模型基于 60 min 的历史数据窗口,利用 12 个历史观测数据来预测未来 15 min、30 min 和 60 min 的交通情况。通过在预设区间内进行网格搜索验证,确定输入批量大小为 32,空间图神经网络中的隐藏单元数量为 32,注意力头数设置为 4。训练过程中采用 Adam 优化算法,  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ , 初始学习率  $\gamma = 0.001$ , 正则化超参数  $\lambda = 0.0001$ , 最大训练次数为 100, 当验证集损失连续 10 次未下降时触发早停机制,自动调整学习率与权重衰减。

## 3.2 性能评价指标

采用预测值与真实值之间的均方根误差 RMSE、平均绝对百分比误差 MAPE 和平均绝对误差 MAE 评估交通预测模型性能, RMSE、MAPE 和 MAE 公式如下:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \hat{X}_i)^2}, \quad (24)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{X_i - \hat{X}_i}{X_i} \right| \times 100\%, \quad (25)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |X_i - \hat{X}_i|, \quad (26)$$

其中,  $X_i$  为交通速度的实际值,  $\hat{X}_i$  为交通速度的预测值。

## 3.3 模型参数设定实验

### 3.3.1 注意力头数实验

在 ARDGAT 模型中,注意力头数会影响预测精度。为选择最佳注意力头数分别在 METR-LA 和 PEMS-BAY 数据集上进行预测时长为 60 min 的实验,分别设置注意力头数为 1、2、3、4、5、6、7、8。实验结果如图7所示,随着注意力头数的改变,模型评价指标 MAPE 和 RMSE 呈现先减少后增加的趋势。究其原因,当注意力头数较少时无法充分捕捉数据特征,此时增加头数提高了模型的表示能力,能更好地学习数据中的复杂模

式,但注意力头数过多又会导致过拟合风险提高,出现注意力稀疏,本文将注意力头数设定为4。

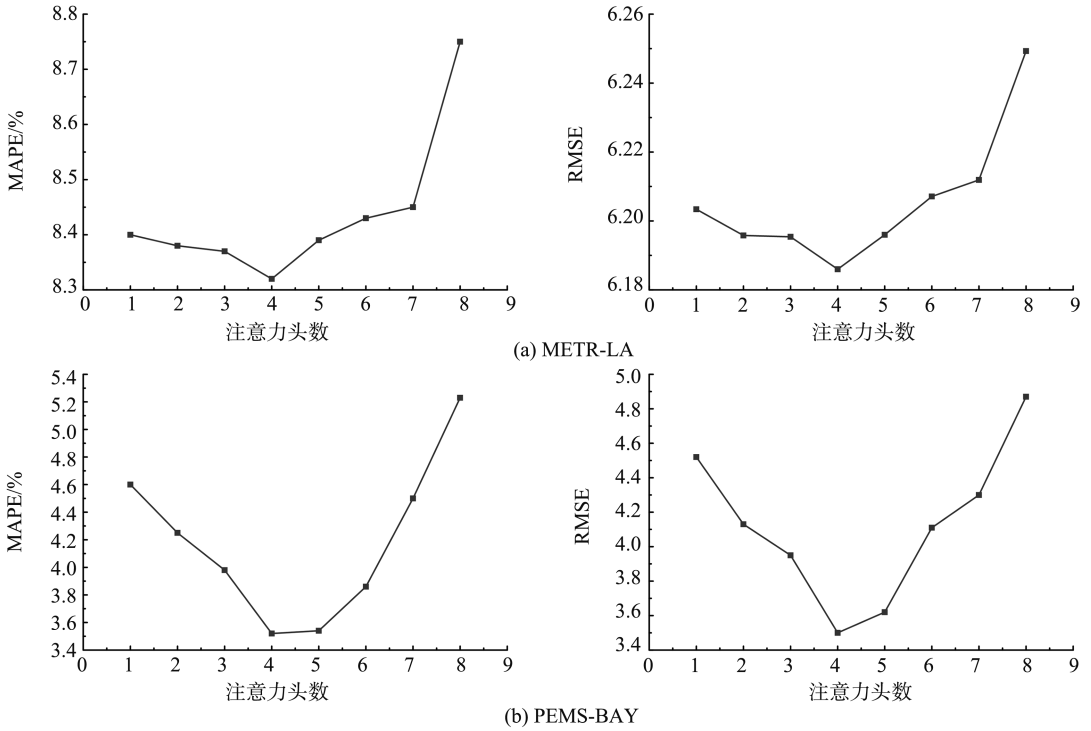


图7 不同注意力头数下的模型评价结果

Fig.7 Evaluation results of models with different attention head settings

### 3.3.2 模型隐藏单元实验

确定模型隐藏单元数量对于降低实验误差非常重要,本研究借助二分法分别在 METR-LA 和 PEMS-BAY 数据集上开展实验,探究不同隐藏单元数对预测性能的影响。在预测时长为 60 min 的情况下分别设置隐藏单元个数为 8、16、32、64、128、256 进行实验,实验结果如图 8 所示,当隐藏单元设置为 32 时模型评价指标 MAPE 和 RMSE 达到最低。

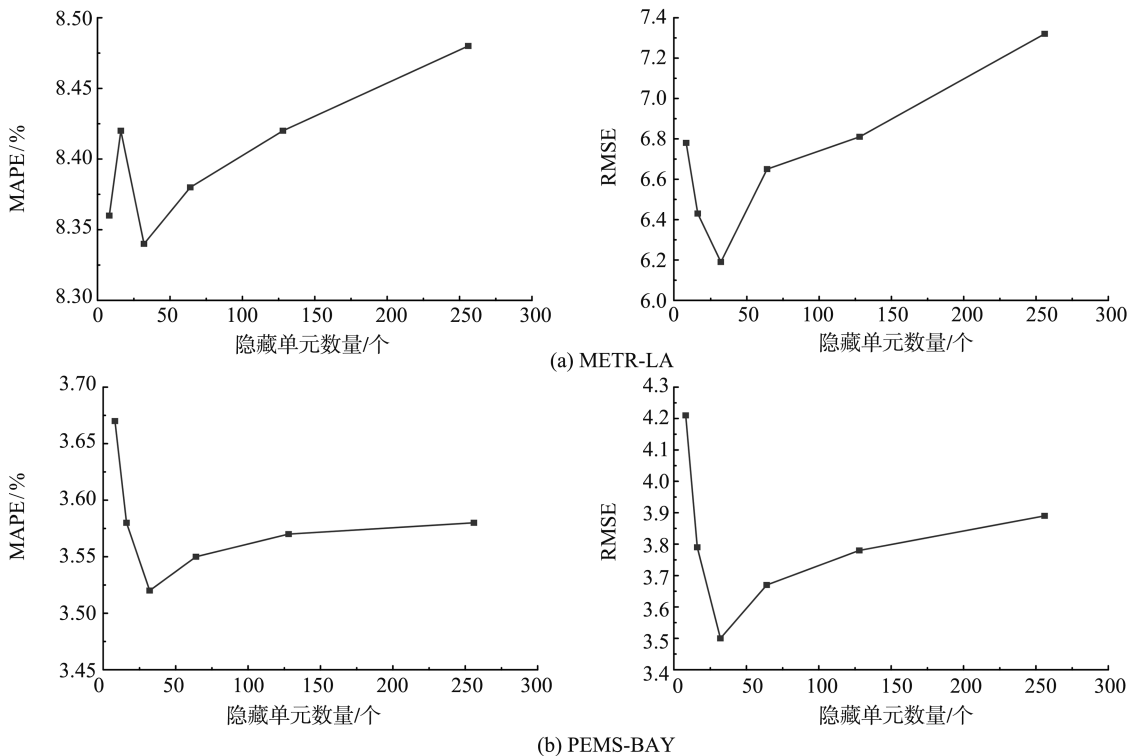


图8 不同隐藏单元数量下的模型评价结果

Fig.8 Evaluation results of the model with different settings of the number of hidden units

### 3.4 消融实验

针对 ARDGCN 模型和 ARDGAT 模型进行消融实验,ARDGCN 为 ARDGAT 保留动态图结构和残差连接并去除 GAT 模块后的模型。与 Graph WaveNet 相比,ARDGCN 通过图神经网络并行处理静态图和动态图,而 Graph WaveNet 仅采用静态图结构建模空间依赖,实验结果如表 2 和图 9 所示。

表 2 ARDGCN 和 ARDGAT 模型消融实验预测性能比较表

Table 2 Comparison of predictive performance of ARDGCN and ARDGAT model ablation experiments

数据集	模型	15 min		30 min		60 min	
		RMSE	MAPE/%	RMSE	MAPE/%	RMSE	MAPE/%
METR-LA	ARDGCN	4.67	6.25	5.98	8.37	8.02	11.21
	ARDGAT	4.44	5.99	5.16	6.94	6.19	8.32
PEMS-BAY	ARDGCN	2.42	2.53	3.94	3.88	5.15	5.03
	ARDGAT	2.17	2.19	2.80	2.77	3.50	3.52

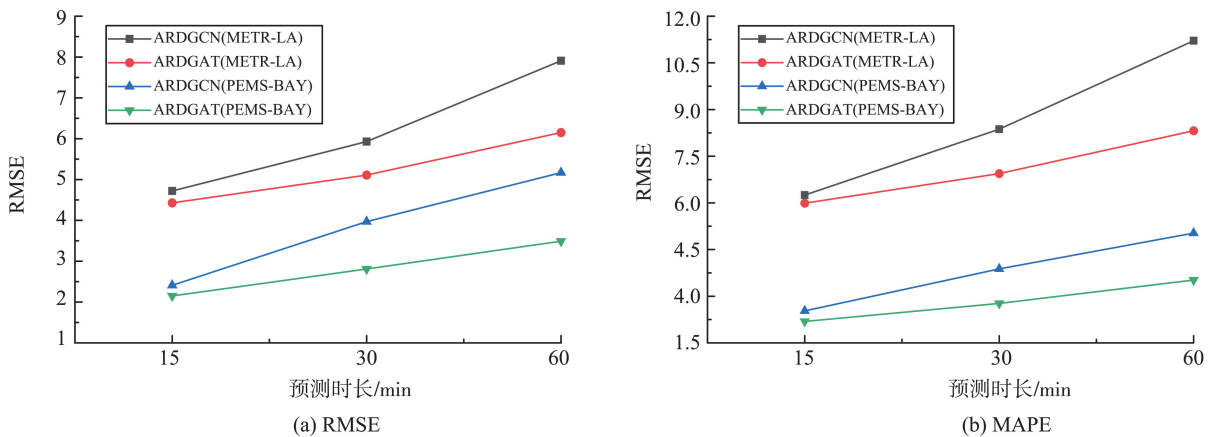


图 9 ARDGCN 和 ARDGAT 模型消融实验预测性能比较图

Fig.9 Comparison of the prediction performance of ARDGCN and ARDGAT ablation experiments

在 2 个数据集 (METR-LA 和 PEMS-BAY) 上的所有预测时段 (15 min、30 min、60 min), ARDGAT 模型的 RMSE 和 MAPE 都优于 ARDGCN 模型。在预测时长 15 min 时, 2 个模型预测误差差别不大, 当预测时长为 30 min 和 60 min 时, ARDGCN 模型误差显著上升, ARDGAT 模型误差上升幅度不大, 证明其在处理长时间依赖性时更具鲁棒性。这是因为相比于 ARDGCN 模型, ARDGAT 模型采用多头图注意力网络捕捉动态图空间依赖关系, 注意力机制为不同节点和边赋予自适应的权重, 使模型能够有效区分不同节点间的相关性, 从而提高预测的准确度。

### 3.5 实验结果

本研究将所提出的模型与另外 6 种预测模型进行比较分析, 所有对比基线模型均采用原文推荐的最优参数配置或引用原文已发表数据, 并保持与 ARDGAT 相同的训练、验证、测试集划分, 使用 60 min 历史窗口预测未来 15—60 min 状态, 采用 RMSE、MAPE、MAE 3 种指标评估模型性能, 用于对比的 6 种模型如下所示。

- (1) ARIMA: 采用传统时间序列分析方法, 通过拟合模型到已知数据预测未来交通速度。
- (2) DCRNN: 融合扩散卷积网络和循环神经网络特点, 在空间维度利用图卷积模拟节点间的空间连接与交互, 在时间维度通过 GNN 及其变体捕捉时间序列动态变化。
- (3) STGCN: 利用图卷积技术分析空间结构并应用二维卷积处理时间序列, 以多层次方式综合时空数据。
- (4) T-GCN: 通过结合图卷积网络和时间序列分析的方式, 在空间维度上借助图卷积来捕捉节点间的空间依赖性, 在时间维度上运用门控循环单元揭示节点的时间相关特性。
- (5) Graph WaveNet: 结合图卷积网络与波形网络, 在空间维度使用图卷积网络捕捉交通网络中各节点间的空间关联性, 在时间维度采用深层因果卷积网络结构捕捉时间序列数据的长期依赖关系。
- (6) STAWnet: 利用时空注意力机制捕捉时空信息, 在空间维度通过空间注意力机制提取不同节点间的

空间关联性,在时间维度利用时间注意力机制动态调整模型对时间序列数据的关注程度以适应不同时间点重要性变化。

表 3 呈现出本研究提出模型与经典模型<sup>[24]</sup>在预测性能方面的对比结果。图 10(a)和图 10(b)则分别展示了预测时长为 60 min 时,ARDGAT 模型在 METR-LA 数据集和 PEMS-BAY 数据集上真实值与预测值的对比情况。

表 3 模型在两个数据集上预测性能比较表  
Table 3 Comparison of the predictive performance of various models on the two datasets

数据集	模型	15 min			30 min			60 min		
		RMSE	MAPE/%	MAE	RMSE	MAPE/%	MAE	RMSE	MAPE/%	MAE
METR-LA	ARIMA	8.21	9.60	3.99	10.45	12.70	5.15	13.20	17.40	6.90
	DCRNN	5.38	7.30	2.77	6.45	8.80	3.15	7.60	10.50	3.60
	STGCN	5.74	7.62	2.88	7.24	9.57	3.47	9.40	12.70	4.59
	T-GCN	5.26	7.81	3.03	6.12	9.45	3.52	7.31	11.80	4.30
	Graph WaveNet	5.15	6.90	2.69	6.22	8.37	3.07	7.37	10.01	3.53
	STAWnet	5.22	6.98	2.70	6.14	8.22	3.04	7.16	9.82	3.44
	ARDGAT	4.44	5.99	2.58	5.16	6.94	2.92	6.19	8.32	3.31
PEMS-BAY	ARIMA	3.30	3.50	1.62	4.76	5.40	2.33	6.50	8.30	3.38
	DCRNN	2.95	2.90	1.38	3.97	3.90	1.74	4.74	4.90	2.07
	STGCN	2.96	2.90	1.36	4.27	4.17	1.81	5.69	5.79	2.49
	T-GCN	2.83	3.14	1.50	3.40	3.76	1.73	4.35	4.95	2.18
	Graph WaveNet	2.74	2.73	1.30	3.70	3.67	1.63	4.52	4.63	1.95
	STAWnet	2.78	2.76	1.31	3.70	3.67	1.61	4.36	4.47	1.89
	ARDGAT	2.17	2.19	1.27	2.80	2.77	1.54	3.50	3.52	1.75

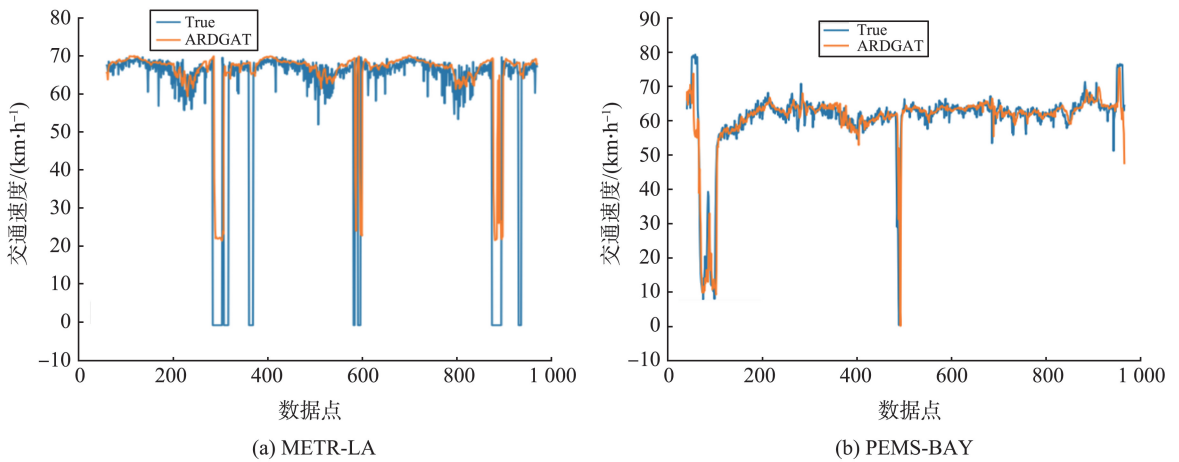


图 10 预测时长 60 min 时模型在 METR-LA 数据集上真实值与预测值对比图

Fig.10 Plot of true versus predicted values of the model on the METR-LA dataset for a prediction duration of 60 min

结果显示在中期和长期预测指标上,本研究模型显著优于对比组里的其它 4 种模型。ARIMA 主要依靠时间序列数据的历史值来对未来值进行预测,由于仅凭借时间序列的历史值和线性关系,无法捕捉交通数据的全部复杂性,因此 ARIMA 在此研究中预测性能较差;DCRNN 借助循环网络结构捕捉时间序列的长期依赖性,并且考虑到了交通网络中的空间特征;T-GCN 模型不但继承了 GCN 在捕捉空间依赖性方面的优势,还通过 GRU 有效地对时间序列动态变化进行建模,但受限于 GRU 的递归结构,难以建模超长时间依赖,导致预测步长增大时误差累积加剧;STGCN 通过 TCN 更有效地对长期时间依赖性进行建模,但采用静态图的结构建模,致使在捕捉交通网络中动态变化的空间依赖性方面存在限制;Graph WaveNet 引入了自适应邻接矩阵,让模型在训练过程中能够动态学习和更新节点之间的关系,但预设的自适应邻接矩阵难以完全反映交通网络中随时间变化的动态空间相关性。本文提出的模型为动态图和自适应图结构,并且引入空间和时间注意力机制给不同空间和时间段赋予权重,这是 Graph WaveNet 和 STGCN 没有考虑到的,使得 ARDGAT 在交通速度的中长期预测任务上表现更加准确。

## 4 结论

本文提出用于交通速度预测的新框架 ARDGAT,其核心创新在于构建并行双模态图架构:一方面通过图注意力网络处理动态邻接矩阵,结合扩散图卷积捕获远距离空间依赖;另一方面设计门控时间卷积与时间注意力机制,显式建模多尺度周期性规律。该架构创新性地引入残差连接防止静态拓扑信息被深层网络稀释,并通过跨层跳跃连接解决循环神经网络梯度问题,实验分析验证了其理论有效性与应用可行性。

虽然 ARDGAT 表现出优越的预测性能,但仍存在模型计算复杂度较高,尚未通过国内交通数据验证本土适应性,对时空数据分布敏感,传感器故障导致的连续数据缺失会引发预测偏移等局限。针对上述问题开展的后续研究对 ARDGAT 应用于真实交通场景预测具有重要意义。

### 参考文献:

- [1] WEI Honglei, SONG Qingbiao, DAN Chonghong, et al. Performance evaluation of ARIMA, autoformer, and symmetric LSTNFCL models for traffic accident emergency prediction[J]. *Symmetry*, 2025, 17(5):639-639.
- [2] ZHOU Xiao, GUO Ling, LI Rui, et al. Intelligent teaching recommendation model for practical discussion course of higher education based on naive bayes machine learning and improved  $k$ -NN data mining algorithm[J]. *Information*, 2025, 16(6):512.
- [3] KHALLADI A S, OUESSAI A, KECHE M. Road traffic classification from nighttime videos using the multihead self-attention vision transformer model and the SVM[J]. *Automatic Control and Computer Sciences*, 2024, 58(5):544-554.
- [4] 薛红军,陈广交,李鑫民,等. 基于决策树理论的交通流参数短时预测[J]. *交通信息与安全*, 2016, 34(3):64-71.  
XUE Hongjun, CHEN Guangjiao, LI Xinmin, et al. Short-term prediction of traffic flow parameters based on decision tree theory[J]. *Journal of Traffic Information and Safety*, 2016, 34(3):64-71.
- [5] SANHITA D, KUMAR A M, ARKA D. A recurrent neural network model for predicting two-leader car-following behavior [J]. *Transportation Letters*, 2024, 16(5):461-475.
- [6] XU W, BLANCAFLOR E, ABISADO M. Performance and improvement of deep learning algorithms based on LSTM in traffic flow prediction[J]. *Discover Applied Sciences*, 2025, 7(4):278.
- [7] PRAKASH N, SHARAN B, ANIL K, et al. Time series prediction for traffic flow forecasting using CNN[J]. *Asian Journal of Research in Computer Science*, 2025, 18(5):442-449.
- [8] CUI Z Y, KRISTIAN H, RUIMIN K, et al. Traffic graph convolutional recurrent neural network: a deep learning framework for network-scale traffic learning and forecasting[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2019, 21(11):4883-4894.
- [9] LI G H, MULLER M, THABET K A, et al. Can GCNs go as deep as CNNs? [C] // 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2020:9266-9275.
- [10] WU Zonghan, PAN Shirui, LONG Guodong, et al. Graph wavenet for deep spatial temporal graph modeling [C] // Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI). San Francisco, CA: Morgan Kaufmann, 2019:1907-1913.
- [11] VELIKOVI P, CUCURULL G, CASANOVA A, et al. Graph attention networks[EB/OL]. (2018-02-04)[2025-04-28]. <https://doi.org/10.48550/arXiv.1710.10903>.
- [12] MANESSI F, ROZZA A, MANZO M. Dynamic graph convolutional networks[J]. *Pattern Recognition*, 2020, 97:107000.
- [13] 王博文,王景升. 基于图注意力网络的交通流量时空预测模型[J]. *公路交通科技*, 2022, 39(6):153-160.  
WANG Bowen, WANG Jingsheng. A spatio-temporal prediction model of traffic flow based on graph attention network[J]. *Highway and Transportation Science and Technology*, 2022, 39(6):153-160.
- [14] ZHU Mingting, ZHU Xianqiang, ZHU Cheng. STGATP: a spatio-temporal graph attention network for long-term traffic prediction[J]. *Lecture Notes in Computer Science*, 2021, 12893:255-266.
- [15] ZHANG Junbo, ZHENG Yu, QI Dekang. Deep spatio-temporal residual networks for citywide crowd flows prediction[EB/OL]. (2017-10-01)[2025-04-29]. <https://doi.org/10.48550/arXiv.1610.00081>.
- [16] WU Yuankai, TAN Huachun, QIN Lingqiao, et al. A hybrid deep learning-based traffic flow prediction method and its understanding[J]. *Transportation Research Part C: Emerging Technologies*, 2018, 95:32-47.

- Recognition, 2025, 159:111115.
- [23] XIE Qin, ZHANG Qinghua, XIA Shuyin, et al. GBG++: a fast and stable granular ball generation method for classification [J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2024, 8(2):2022-2036.
- [24] 肖振国,陈林书,孙少杰,等. 基于代数粒的聚类方法[J]. 计算机工程与科学,2024,46(1):150-158.  
XIAO Zhenguo, CHEN Linshu, SUN Shaojie, et al. A clustering method based on algebraic granularity [J]. Computer Engineering & Science, 2024, 46(1):150-158.
- [25] 徐伟华,黄旭东,蔡可. 基于粒计算的多源信息融合方法综述[J]. 数据采集与处理,2023,38(2):245-261.  
XU Weihua, HUANG Xudong, CAI Ke. Review of multi-source information fusion methods based on granular computing [J]. Journal of Data Acquisition and Processing, 2023, 38(2):245-261.
- [26] CHEN Qiangqiang, HE Linjie, DIAO Yanan, et al. A novel neighborhood granular meanshift clustering algorithm [J]. Mathematics, 2023, 11(1):207.
- [27] JIANG Hailiang, CHEN Yumin, KONG Liru, et al. An LVQ clustering algorithm based on neighborhood granules [J]. Journal of Intelligent & Fuzzy Systems, 2022, 43(5):6109-6122.
- [28] CHEN Yumin, ZHU Shunzhi, LI Wei, et al. Fuzzy granular convolutional classifiers [J]. Fuzzy Sets and Systems, 2022, 426:145-162.
- [29] LI Wei, WEI Zhongnan, CHEN Yumin, et al. Fuzzy granular hyperplane classifiers [J]. IEEE Access, 2020, 8:112066-112077.
- [30] 郑晨颖,陈颖悦,侯贤宇,等. 一种邻域粒的模糊 C 均值聚类算法[J]. 山东大学学报(理学版),2024,59(5):35-44.  
ZHENG Chenying, CHEN Yingyue, HOU Xianyu, et al. A neighbourhood granular fuzzy C-means clustering algorithm [J]. Journal of Shandong University (Natural Science), 2024, 59(5):35-44.
- [31] 吴海,牛娇娇,铁文彦,等. 基于粒概念网络的概念格构造方法[J]. 山东大学学报(理学版),2025,60(12):21-31.  
WU Hai, NIU Jiaojiao, TIE Wenyan, et al. Concept lattice construction method based on granular concept network [J]. Journal of Shandong University (Natural Science), 2025, 60(12):21-31.
- [32] HE Linjie, CHEN Yumin, ZHONG Caiming, et al. Granular elastic network regression with stochastic gradient descent [J]. Mathematics, 2022, 10(15):2628.
- [33] JIANG Hailiang, CHEN Yumin, JIANG Hongbo, et al. A granular sigmoid extreme learning machine and its application in a weather forecast [J]. Applied Soft Computing, 2023, 147:110799.

(编辑:陈丽萍)

(上接第 101 页)

- [17] CINI N, AYDIN Z. A deep ensemble approach for long-term traffic flow prediction [J]. Arabian Journal for Science and Engineering, 2024, 49(9):12377-12392.
- [18] JI Jiahao, WANG Jingyuan, HUANG Chao, et al. Spatio-temporal self-supervised learning for traffic flow prediction [EB/OL]. (2024-03-6) [2025-05-01]. <https://doi.org/10.48550/arXiv.2212.04475>.
- [19] WANG Zhaohuan, XU Yi, HAN Liangzhe, et al. Multivariate long-term traffic forecasting with graph convolutional network and historical attention mechanism [C] // Knowledge Science, Engineering and Management Lecture Notes in Computer Science. Cham: Springer, 2023, 14120:112-123.
- [20] LI Zhonghang, XIA Long, SHI Lei, et al. OpenCity: open spatio-temporal foundation models for traffic prediction [EB/OL]. (2024-08-16) [2025-05-01]. <https://arxiv.org/abs/2408.10269>.
- [21] 孙大盟,欧阳安杰,何立明. 基于改进 STGCN 深度学习框架的交通速度预测 [J]. 计算机技术与发展,2024,34(11):133-139.  
SUN Dameng, OUYANG Anjie, HE Liming. Traffic speed prediction based on improved STGCN deep learning framework [J]. Computer Technology and Development, 2024, 34(11):133-139.
- [22] ZHANG Dongping, LAN Hao, WANG Mingting, et al. Graph convolutional networks with multi-scale dynamics for traffic speed forecasting [J]. Applied Soft Computing, 2025:174.
- [23] CAI Jiongbiao, WANG H C, HU Kun. LCDFormer: long-term correlations dual-graph transformer for traffic forecasting [J]. Expert Systems With Applications, 2024, 249:123721.
- [24] TIAN Chenyu, CHAN Waikin. Spatial-temporal attention wavenet: a deep learning framework for traffic prediction considering spatial-temporal dependencies [J]. IET Intelligent Transport Systems, 2021, 15(4):549-561.

(编辑:唐秋)