

注意力融合的特征增强动态图卷积交通流预测模型

戴卿鹏, 杨 昊, 文 武

(成都信息工程大学 计算机学院, 四川 成都 610225)

摘要: 实时、准确的交通流量预测对城市管控与治理等方面具有重要的现实应用价值, 而现有方法在动态时空依赖建模和特征提取方面存在不足。为进一步提高预测精度, 提出了注意力融合的特征增强动态图卷积模型 (AFEDGCN) 用于预测交通流量。首先, 模型将空间注意力与动态图卷积融合以增强对节点空间相关性建模, 同时使用交互式卷积增强节点自身特征, 结合 GRU 建模同步时空相关性。其次, 使用多头时间注意力以捕获时间相关性, 并通过全局图卷积对整个时间段内的全局空间相关性进行建模; 最后, 对时空特征进行自适应融合以预测未来交通流状况。通过在多个真实的世界交通数据集上进行实验, 与先进的 GDGCN 和 DGCRN 相比, MAE 指标分别降低了 1.65% 和 4.80%, MAPE 降低了 2.36% 和 3.93%, 结果表明该模型性能与当前先进模型相比更具优势。

关键词: 交通预测; 空间注意力; 多头时间注意力; 特征增强; 动态图卷积

DOI: 10.11907/rjdk.261041

扫描二维码阅读全文:

中图分类号: TP183

文献标识码: A

文章编号: 1672-7800(XXXX)0XX-0001-08

Feature-Enhanced Dynamic Graph Convolutional Neural Model with Attention Fusion for Traffic Flow Prediction

DAI Qingpeng, YANG Hao, WEN Wu

(School of Computer Science, Chengdu University of Information Technology, Chengdu 610225, China)

Abstract: Real-time and accurate traffic flow prediction holds significant practical application value for urban management and governance. Existing methods have limitations in modeling dynamic spatio-temporal dependencies and feature extraction. To further improve prediction accuracy, this study proposes a Feature-Enhanced Dynamic Graph Convolutional Neural Model with Attention Fusion for Traffic Flow Prediction (AFEDGCN). The model integrates spatial attention with dynamic graph convolution to enhance the modeling of node spatial correlations. Meanwhile, interactive convolution is utilized to strengthen the features of individual nodes, and the Gated Recurrent Unit (GRU) is incorporated to model synchronous spatio-temporal dependencies. Multi-head temporal attention is employed to capture temporal correlations, while global graph convolution is used to model the global spatial correlations over the entire time period. Finally, spatio-temporal features are adaptively fused to predict future traffic flow conditions. Experiments conducted on multiple real-world traffic datasets show that, compared with the advanced GDGCN and DGCRN models, the proposed model reduces the MAE (Mean Absolute Error) by 1.65% and 4.80% respectively, and the MAPE (Mean Absolute Percentage Error) by 2.36% and 3.93% respectively. These results indicate that the proposed model has better performance compared with the current state-of-the-art models.

Key Words: traffic prediction; spatial attention; multi-head temporal attention; feature enhancement; dynamic graph convolution

0 引言

随着数字城市智能化进程的快速发展和人口数量的不断增加, 交通拥堵问题日益严重。精准的交通流预测可

以在交通管理部门制定决策时提供科学合理的指导进而有效缓解交通拥堵, 提升城市交通运行效率。因此, 如何有效建模交通流数据中的复杂时空依赖关系, 成为交通智能化领域的核心研究问题之一。

早期的交通流预测方法主要依赖统计学模型和传统

收稿日期: 2026-03-06

基金项目: 四川省科技成果转化示范项目 (2024ZHCG0026)

作者简介: 戴卿鹏 (2001-), 男, 成都信息工程大学计算机学院硕士研究生, 研究方向为交通大数据挖掘; 杨昊 (1981-), 男, 博士, 成都信息工程大学计算机学院副教授, 研究方向为人工智能与大数据挖掘; 文武 (1979-), 男, 博士, 成都信息工程大学计算机学院副教授, 研究方向为人工智能与高性能运算。本文通讯作者: 杨昊。

机器学习模型,如向量自回归模型(VAR)^[1]和支持向量回归(SVR)^[2]等,这类方法在建模复杂非线性关系和高维时空依赖方面能力有限,预测精度受限,预测效果都不佳。随着深度学习的发展,研究者开始引入循环神经网络(RNN)和长短期记忆网络(LSTM)对交通时间序列进行建模^[3],在一定程度上提升了预测性能。然而,这类方法通常忽略交通路网的空间结构信息,难以有效建模节点间的空间相关性。

针对交通路网中节点间复杂的空间依赖关系,图卷积神经网络(GCN)被引入交通流预测领域^[4]。如Yu等^[5]提出时空图卷积网络,通过GCN有效捕捉路网节点间的复杂关系,有效刻画了空间依赖关系;Zhao等^[6]利用GCN和GRU分别提取交通流空间特征与时间特征,通过时空同步建模机制提取交通流量特征的时空一致性。然而,该类方法通常依赖静态邻接矩阵,难以反映交通路网中随时间变化的动态空间关系。

为克服静态图结构的局限性,研究者进一步提出了自适应图和动态图建模方法^[7-8]。Graph WaveNet创新性地 将自适应图与图卷积结合,为固定图结构不能反映真实依赖关系的问题提供了一种解决思路,使模型能够自动学习潜在的空间依赖关系^[9]。Wang等^[10]构建的时空同步图卷积网络,能够有效捕获局部时空相关性并考虑时空数据的异质性。AGCRN模型引入了节点参数学习模块,为每个节点学习一个特定的参数空间并用于图卷积,在减少计算量的同时也增加了可解释性^[11]。Shao等^[12]在动态图学习模块的基础上通过解耦框架构建以捕捉交通数据中两种信号的时空依赖性,有效提升了交通状态预测性能。这类方法在动态空间依赖建模方面取得了显著进展,但多数工作仍侧重于空间结构的动态调整,对节点自身特征的深层次建模关注不足。

注意力机制因其能够动态分配权重、突出关键信息,被广泛应用于交通流预测以增强模型对关键时空特征的 关注能力^[13-14]。GMAN^[15]、ASTGNN^[16]等模型通过引入空间注意力和时间注意力,对不同时刻和不同节点的重要性进行加权,有效提升了模型对复杂时空依赖的建模能力。然而,现有注意力机制多用于特征重加权,较少与动态图卷积深度融合,难以在动态图建模过程中充分发挥其对空间依赖关系的引导作用。

综上所述,现有研究在动态空间依赖建模、注意力机制引入和节点特征增强等方面均取得了一定进展。但仍存在以下不足:①动态图卷积与注意力机制多为独立设计,缺乏深度融合;②节点自身特征通常作为图卷积的附属项进行简单增强,未被显式建模;③多数方法侧重于单一时间尺度的建模,难以同时刻画局部动态与时间段内的全局空间相关性。

为解决上述问题,本文提出了一种注意力融合的特征增强动态图卷积交通流预测模型(Feature-Enhanced Dy-

namic Graph Convolutional Neural Model with Attention Fusion for Traffic Flow Prediction, AFEDGCN)。与现有的交通流量预测方法相比,本文方法的核心差异体现在以下3个方面:①在动态图构建过程中引入空间注意力机制,使动态图卷积能够根据节点间的相关性动态分配权重,而非仅依赖隐式的参数更新;②通过交互式特征增强模块对节点自身特征进行显式建模,避免将节点特征简单视为图卷积的残差项;③采用分层时空建模策略,在刻画局部动态时空依赖的基础上,引入全局图卷积以捕获时间段内的长期空间相关性,从而实现多尺度时空特征的协同建模。

1 问题定义

在交通预测问题中,通常将交通路网定义为图 $G = (V, E, A)$ 以描述道路网间的拓扑结构,其中 V 表示交通路网的一组节点, N 为节点数, E 是交通路网的边缘。 $A \in R^{N \times N}$ 表示道路之间的连接关系。

使用 $X_t \in R^{N \times C}$ 表示 t 时刻下路网中节点的流量观测值,其中 C 是交通流量的维度。过去 P 个时间片的历史交通流量数据表示为如式(1)所示。

$$\mathcal{X} = (X_{t-P+1}, \dots, X_t) \in R^{P \times N \times C} \quad (1)$$

交通流的预测问题可以看作是通过学习一个函数 $g(\cdot)$,通过历史数据 $\mathcal{X}$ 预测未来 Q 步的交通流量,如式(2)所示。

$$[X_{t-P+1}, \dots, X_t; G] \xrightarrow{g(\cdot)} [X_{t+1}, \dots, X_{t+Q}] \quad (2)$$

2 方法模型

模型总体结构如图1所示,它由4个主要模块组成,即数据嵌入层、注意力融合的增强动态图卷积递归神经网络框架(EDG-GRU)、全局图卷积模块(GGCN)和多头时间注意模块(MHTA)。

首先,该方法借助数据嵌入层,将交通数据中的已知信息融入原始交通数据,增强交通特征的表达能力。其次,使用EDG-GRU框架捕捉交通节点在不同时刻下的动态依赖关系。在高效提取动态空间依赖关系的过程中,同步完成节点特征的深度增强,进而优化模型对复杂交通流数据的表征能力。再次,将通过EDG-GRU框架的交通信息再使用全局图卷积模块与多头时间注意力模块分别进一步地捕获全局尺度下的空间特征与时间特征;最后,对时空特征进行自适应融合后叠加输入到预测层,经过预测层输出最终预测的未来交通流状态。

2.1 数据嵌入层

交通数据具有显著周期性与复杂动态时空特性,直接使用原始数据会增加模型建模难度^[17]。为此,设计数据嵌入层,通过嵌入关键信息引导模型捕捉交通数据的动态时空规律。

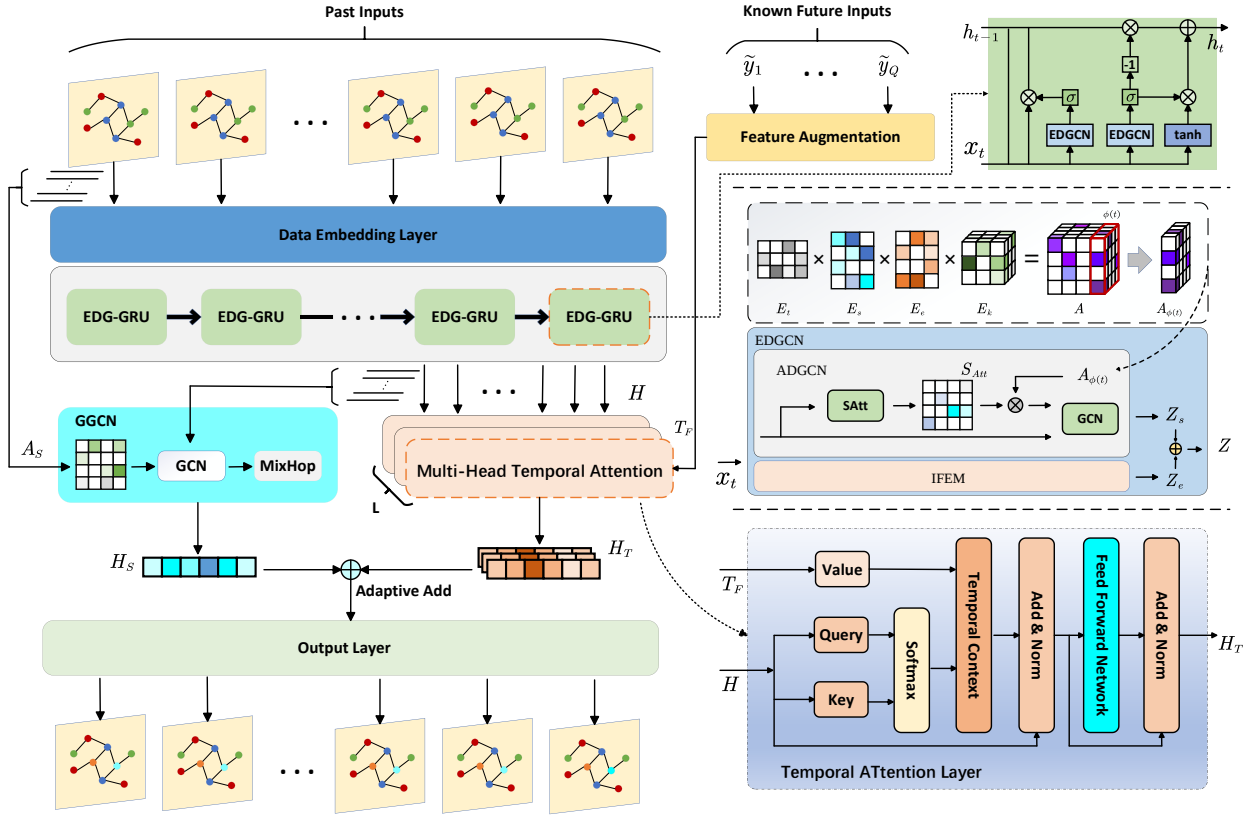


Fig. 1 Model structure

图1 模型结构

利用全连接层将原始数据映射到高维特征空间, 获得特征嵌入 $X_{data} \in R^{T \times N \times d_e}$ 。

定义周嵌入词典和日嵌入词典, 并使用数据中的星期信息和时间戳信息从其中提取相应的嵌入向量, 通过拼接和广播获得周期嵌入 $X_p \in R^{T \times N \times 2d_e}$ 。

为传感器的每个时间步嵌入专属的时空自适应嵌入 $X_a \in R^{T \times N \times d_a}$, 模型可以通过自适应地学习时空关联, 增强对时空特征的表达能。

通过拼接上述嵌入, 得到隐藏的时空表征 $X_{emb} \in R^{T \times N \times D_{hid}}$, 如式(3)所示。

$$X_{emb} = X_{data} \parallel X_p \parallel X_a \quad (3)$$

其中, d_e, d_p, d_a 为嵌入维度, D_{hid} 为隐藏维度, $\parallel$ 是向量拼接操作。

2.2 增强型动态图卷积递归网络

在交通流预测中, 常使用空间域图卷积神经网络捕获空间相关性, 定义为如式(4)所示。

$$\begin{aligned} GCN(A, X) &= \left(I_N + D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \right) XW + b \\ &= I_N XW + D^{-\frac{1}{2}} A D^{-\frac{1}{2}} XW + b \end{aligned} \quad (4)$$

其中, A 是交通图的邻接矩阵, D 是 A 的度矩阵, W 和 b 是可学习参数。将图卷积嵌入 GRU 得到图卷积递归神经网络以聚合时空特征。

但现有研究中, 模型更侧重对时空上的特征进行提取, 对节点自身特征的提取关注甚少。式(4)中, $I_N XW$ 一

般被视为是增强自身相关性, 也认为是捕捉节点自身特征。但这种方式实质上仅是一种相对简单的叠加操作, 未深入挖掘节点自身的深层特征。

同时, 现有的大部分图卷积方法多采用静态权重机制构建邻接矩阵, 无法根据实时交通场景变化, 动态调整权重分配策略, 限制了模型对复杂路网结构的表达能力, 无法有效建模交通数据的动态空间依赖。

因此, 本文提出了一种增强型动态图卷积模块。使用交互式特征增强模块, 通过不同尺度的特征交互以挖掘节点自身特征的高阶表示。并且, 使用融合空间注意力的动态图卷积模块, 通过动态图动态捕捉路网拓扑的时变特性, 并结合空间注意力机制增强关键节点的特征权重, 从而精准刻画交通路网节点间的空间依赖关系。进一步地, 将交互式特征增强模块和空间注意力融合动态图卷积模块的结果相叠加, 在提取动态空间依赖关系的同时, 实现对节点特征的深度增强, 从而提升模型对复杂交通流数据的表征能力。

最后, 将上述的增强型动态图卷积嵌入 GRU 单元, 构建增强型动态图卷积递归网络, 以捕捉交通数据的时空特征。与其他典型的动态图卷积网络仅依赖参数隐式生成动态图的方法不同, 本文的增强型动态图卷积通过显式融合空间注意力以量化节点间的动态关联, 同时引入交互式卷积以深度挖掘节点自身特征, 从而更精准、更鲁棒地建模交通流中复杂的时空依赖, 最终实现时空建模与特征增

强的协同优化。

2.2.1 交互式特征增强模块

传统卷积模型在特征提取时使用固定卷积核,难以兼顾细粒度细节与全局特征。靠堆叠层数扩大感受野,会导致计算冗余与梯度衰减;在特征融合时,通过拼接或求和融合易稀释信息,难以挖掘特征关联性。因此,设计交互式特征增强模块,使用双尺度并行卷积,用小核(如 1×1)提取局部细节,用大核(如 1×3)扩大感受野捕获特征。通过元素乘法融合特征,动态强化节点有用特征。交互式特征增强模块通过不同尺度特征之间的元素级交互,建模特征之间的相互影响关系,实现特征间的协同增强^[18]。相比于通道注意力模块主要通过通道重标对特征进行加权,以及门控机制侧重于控制信息的通过与抑制,交互式特征增强模块不仅关注特征的重要性分配,更强调特征之间的交互表达能力。具体计算如式(5)一式(7)所示。

$$S_1 = \sigma(\Theta_1(x_t)) \odot \Theta_3(x_t) \quad (5)$$

$$S_2 = \sigma(\Theta_3(x_t)) \odot \Theta_1(x_t) \quad (6)$$

$$Z_e = \Theta_1(S_1 + S_2) \quad (7)$$

其中, σ 为激活函数, Θ_1 、 Θ_3 表示卷积核为1、3的一维卷积模块, x_t 为 t 时刻嵌入的数据。

2.2.2 空间注意力融合动态图卷积模块

由于路段间的隐式空间关系会随时间动态变化,传统静态和单个邻接矩阵难以表达真实情况下的交通状态,因而难以建模动态的空间关系^[19]。因此,设计一个将时间步长和节点都纳入考虑范围的动态关系图 $A \in R^{N_t \times N \times N}$,其中 N_t 表示一天含有的时间段数量, $A_{\phi(t)}$ 为 t 时刻的邻接矩阵。通过张量分解由4个可学习向量简化动态图的构造:① $E^k \in R^{d_h \times d_h \times d_h}$ 核心张量;② $E^t \in R^{N_t \times d_h}$ 时隙嵌入;③ $E^s \in R^{N \times d_h}$ 起始节点信息嵌入;④ $E^e \in R^{N \times d_h}$ 结束节点信息嵌入(d_h 为节点隐藏维度)。反向传播更新可学习向量—学习空间关系的动态表示,确保模型在不同时间步下能够根据节点状态和交通模式学习当前时刻的空间依赖关系,提升模型对复杂交通场景下的适应性。具体计算如式(8)所示。

$$A = \text{soft max}(\text{ReLU}(E^k \times E^t \times E^s \times E^e)) \quad (8)$$

交通网络中的某些节点由于其地理位置(在城市中心)或特性可能对其他多个节点产生影响,表现出与多个节点的广泛连接关系,从而形成以这些节点为中心、多节点交互的复杂时空网络^[20]。直观地表现是不同节点间的空间相关性程度有着不同的差异。为了进一步精准量化节点间动态相关性的影响关系,通过融合注意力机制动态计算节点间的空间相关性程度,深度捕捉路网中节点之间的动态空间关联。相关计算公式如式(9)所示。

$$e_{i,j} = a(Wx_i^t \parallel Wx_j^t) \quad (9)$$

$$\alpha_{i,j} = \text{soft max}(e_{i,j}) = \frac{\exp(\text{LeakReLU}(e_{i,j}))}{\sum_{k \in N_i} \exp(\text{LeakReLU}(e_{i,k}))} \quad (10)$$

$$S_{Att} = (\alpha_{i,j}), S_{Att} \in R^{N \times N} \quad (11)$$

其中, x_i^t 为路网节点 i 在 t 时刻的嵌入数据, $\alpha_{i,j}$ 表示 t 时刻节点 i 和 j 之间的特征相关性强度。最后,通过空间注意力相关性权重矩阵 S_{Att} 和动态图 A ,得到动态图与空间注意力融合图卷积模块的输出,如式(12)所示。

$$Z_s = \text{GCN}(A_{(t)}, S_{Att}, x_t) = \sigma((A_{(t)} \odot S_{Att})x_t W + b) \quad (12)$$

通过并联架构将交互式特征增强模块和空间注意力融合动态图卷积模块结合,如式(13)所示。

$$Z = Z_s + Z_e \quad (13)$$

将增强型动态图卷积组件嵌入GRU单元,初步对时间依赖进行建模,得到增强动态图卷积递归网络(EDG-GRU)。给定先前隐藏的表示 h_{t-1} 和时间步 t 处的输入数据,通过GRU单元控制信息的记忆和遗忘,从而得到时间步 t 处的输出 h_t 。最终将所有时刻的输出结果加以聚合,得到时空路网信息 $H = \{h_1, h_2, \dots, h_p\}$, $H \in R^{P \times N \times D_{hid}}$, H 包含了 p 时间段内交通路网的时空特征和节点自身特征。

2.3 全局图卷积模块

由于递归神经网络按时序步依次计算,因而EDG-GRU所聚合的空间关系主要为某一时刻或相邻时刻下的路网节点空间依赖,对整个时间段内节点间潜在的长期全局空间相关性建模有所不足。对此,本文提出了一种全局图卷积模块以进一步提取时空路网信息中的全局空间相关性。

该模块从整个时间段的角度出发,旨在捕获路网中不同节点在整个时间段内的全局空间相关性。计算公式如式(14)所示。

$$e_{ij} = \frac{\sum_{t=1}^P (X_{i,t} X_{j,t}^T) w_{ij}}{\sqrt{\sum_{t=1}^P (X_{i,t})^2} \sqrt{\sum_{t=1}^P (X_{j,t})^2}} \quad (14)$$

$$A_s = (\text{sigmoid}(e_{i,j})), A_s \in R^{N \times N} \quad (15)$$

其中, $X \in R^{N \times P}$ 为路网的流量特征输入, $X_i \in R^P$ 表示节点 i 在 p 时间段内的流量特征, $W = (w_{ij}) \in R^{N \times N}$ 是一个可训练的参数矩阵。 e_{ij} 通过计算 P 时间步长期间节点 i 处的交通流与在 j 处的交通流之间的相似性而获得。

将全局相似邻接矩阵 A_s 与时空路网信息 H 进行全局图卷积,并借助MixHop传播加以实现,在保留一定比例的节点原始特征下进行更深邻域的探索。相关公式如式(16)、式(17)所示。

$$H_\psi = \beta H + (1 - \beta) A_s H_{\psi-1} \quad (16)$$

$$H_s = \sum_{\psi=0}^{\Psi} H_\psi H_\psi \quad (17)$$

其中, H 表示图卷积层的输入(时空路网信息), Ψ 表

示传播深度, H_ψ 是第 ψ 层的表示, $H_0 = H$ 。 β 是控制原始输入与来自不同层信息之间比率的超参数。最终得到全局图卷积的计算结果 $H_s \in R^{P \times N \times D_{hid}}$ 。

在计算复杂度方面, 全局相似邻接矩阵构建主要涉及节点间特征相似度计算, 其时间复杂度为 $O(N^2d)$, 其中 N 为节点数量, d 为特征维度。需要指出的是, 该相似度矩阵在每个时间段内统一构建, 并使用原始低维数据以构建相似性邻接矩阵, 而非对映射到高维的时间序列进行计算, 这有效降低了计算与存储开销。在本文所使用的数据规模下, 该计算开销处于可接受范围。此外, 在更大规模路网场景下, 可通过稀疏化近邻筛选、低秩分解或分块计算等方式降低计算与存储成本。同时, 引入 MixHop 传播机制, 在保持浅层信息的同时避免深层邻域的过度扩展, 使得该模块在节点规模较大时仍具有良好的可扩展性。

2.4 多头时间注意力模块

由于 EDG-GRU 依赖时序顺序传递信息, 故难以学习交通数据中长时间序列依赖, 难区分长时序下的语义差异^[21]。因此, 使用多头时间注意力模块以进一步提取时空路网信息 H 中的全局上下文时间信息, 并在引入预测时刻之前即可获得, 不依赖于未来交通状态观测的外生信息, 通过这些未来已知信息调整输出, 缓解误差传播。

通过可提前获得的未来流量所对应的星期信息和时间戳信息, 分别提取相应的特征嵌入向量, 以获得未来时间嵌入 Y_p 。同时, 使用未来时空自适应嵌入 Y_a , 通过可学习嵌入矩阵映射以刻画不同未来时间步的时空差异。拼接未来时间嵌入和未来时空自适应嵌入, 得到未来隐藏时空表征 $Y_{emb} \in R^{T \times N \times D_{hid}}$, 如式(18)所示。

$$Y_{emb} = Y_p \parallel Y_a \quad (18)$$

将 Y_{emb} 通过线性层进行特征增强, 得到未来已知流量信息特征 $T_F \in R^{T \times N \times D_{hid}}$; 将 T_F 和时空路网信息 H 用以生成查询、关键字和值矩阵; 将未来已知信息 T_F 仅用于生成值, 并结合历史隐藏表示(时空路网信息 H)生成查询和关键字, 模型能够在不违反预测设定的前提下, 引导时间注意力机制更合理地分配时间权重。同时, 采用多头时间注意力模块以丰富信息的表示, 学习交通流量不同角度的时空相关性。最后, 将结果连接进行线性变换, 得到输出 H_{Att} 。相关公式如式(19)一式(22)所示。

$$Q = HW_q^T, K = HW_k^T, V = T_F W_v^T \quad (19)$$

$$Attention(Q, K, V) = \text{soft max} \left(\frac{QK^T}{\sqrt{d_k}} \right) V \quad (20)$$

$$head_i = Attention(HW_i^Q, HW_i^K, T_F W_i^V) \quad (21)$$

$$H_{Att}(Q, K, V) = Concat(head_1, head_2, \dots, head_h) W_o \quad (22)$$

其中, W_q 、 W_k 和 W_v 是可学习参数矩阵, d_k 是权重缩放因子。 $head_i$ 代表每个头的注意力, W_o 是输出投影矩阵。

最后, H_{Att} 被传递到残差网络得到多头时间注意力模块输出 $H_T \in R^{P \times N \times D_{hid}}$ 。

2.5 预测层

将全局图卷积模块和多头时间注意力模块提取到的空间特征 H_s 和时间特征 H_T , 经自适应融合后映射到输出空间, 从而得到最后预测结果。形式如式(23)所示。

$$\hat{X}_{t+1:t+Q} = Conv(\rho H_s + (1 - \rho) H_T) \quad (23)$$

其中, $\hat{X} \in R^{Q \times N \times 1}$ 是最终预测结果, ρ 是时空特征融合的自适应参数。

3 实验与结果分析

3.1 数据集

本文使用的数据集是由美国加利福尼亚州运输部性能测量系统(PeMS)收集的3个真实地区的高速公路数据(PEMS04、PEMS07 和 PEMS08)和腾讯地图导航数据中的北京市交通数据 NE-BJ。数据集以 5 min 为间隔进行记录, 一天的交通数据流包含 288 个数据。详细信息如表 1 所示。

Table 1 Dataset description

表 1 数据集描述

数据集	传感器数量/个	边数量/条	时间跨度	时间步长
PEMS04	307	340	01/01/2018–02/28/2018	16 992
PEMS07	883	866	05/01/2017–08/31/2017	28 224
PEMS08	170	295	07/01/2016–08/31/2016	17 856
NE-BJ	500	110734	01/07/2020–31/07/2020	6 509

3.2 实验设置

与大多数方法保持一致, 本文按照 6:2:2 的比例将所有数据集划分为训练集、验证集和测试集, 使用过去 1 h 的交通流量预测未来 1 h 的交通流量(历史步长 P 和预测步长 Q 都设置为 12)。模型的隐藏层维度 D_{hid} 设置为 64, 特征嵌入维度 d_e 、周嵌入和时间戳嵌入维度 d_t 、时空自适应嵌入维度 d_a 之比为 4:1:2。节点隐藏维度 d_N 设置为 4, 全局图卷积模块中信息传播比率 β 设置为 0.05。本文所有实验均在 NVIDIA GeForce RTX 3080 上进行。使用 Huber 损失函数, 它能够平衡 MSE 和绝对误差之间的权衡, 具有良好的鲁棒性和可微性。

所使用的基线模型均采用公开实现版本, 并在相同的数据划分、输入特征、训练轮数和评估指标设置下进行实验。模型超参数遵循原文推荐设置, 以确保实验对比公平性与结果可复现性。

3.3 基准模型

本文选择了 8 种基线模型以作对比。①VAR: 向量自回归模型, 能够捕捉多个时间序列之间的成对关系; ②SVR: 支持向量回归模型, 利用线性支持向量机实现预测任务; ③STGCN: 将图卷积和门控时间卷积结合, 捕获全面的时空相关性; ④GWNEN: 结合自适应图卷积和扩张的随机卷积捕获时空依赖关系; ⑤GMAN: 将空间、时间和变换的注意力叠加在一起, 使用时空注意力层捕获动态时空依赖, 并通过转换注意力层缓解误差传播; ⑥ASTGNN: 使用

GCN结合自注意力模块以聚合邻居信息;⑦DGCRN^[22]:通过超网络构建动态邻接矩阵,将GNN与RNN结合,捕捉时空依赖关系;⑧GDGCN^[23]:基于图的视图将历史时隙视为节点以处理时间关系,通过图卷积处理时间动态。

3.4 评估指标

实验使用均绝对误差(MAE)、均方根误差(RMSE)和平均绝对百分比误差(MAPE)这3个交通流量预测的常用指标作为衡量模型性能的标准,具体计算公式如式(24)~式(26)所示。

$$MAE = \frac{1}{N} \left(\sum_{i=1}^N |y_i - \hat{y}_i| \right) \quad (24)$$

$$RMSE = \sqrt{\frac{1}{N} \left(\sum_{i=1}^N (y_i - \hat{y}_i)^2 \right)} \quad (25)$$

$$MAPE = \frac{1}{N} \left(\sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \right) \quad (26)$$

其中, N 为样本数据总数量, y_i 和 $\hat{y}_i$ 为预测值和真实值。

3.5 实验结果分析

基准模型与本文所提模型AFEDGCN在高速公路数据集(PEMS03、PEMS04、PEMS08)和北京市交通数据集上未来1 h的平均预测结果如表2所示。可以清晰地看出,本文AFEDGCN模型在数据集的多数评估指标上优于其他基线模型,这一结果充分证明了该模型在交通流预测任务中的有效性和优越性。

在众多基线模型中,传统机器学习模型SVR与VAR模型受限于模型架构的固有特性,因而难以处理交通流数

据的高维度、非线性及动态相关性,预测结果不理想。在处理复杂交通流预测问题时,深度学习方法相较于传统方法展现出了显著优势。

时空图卷积模型STGCN同时考虑了时间与空间相关性,预测性能良好;GWNET将扩散卷积集成到TCN中,有效地学习交通数据的时间依赖性,并通过自适应邻接矩阵对空间依赖关系进行建模,预测性能进一步提升。但由于采用的是静态权重机制,因此STGCN和GWNET对时空关联刻画不全面,难以建模交通数据复杂的动态时空相关性。ASTGNN借助注意力机制并同时考虑了周期性和空间异质性,使预测性能得到进一步提升。但ASTGNN主要聚焦于节点级别的特征权重分配,对于路网局部拓扑结构的细节刻画不足,因而在复杂路网场景中预测误差较大,预测效果不足。GMAN基于注意力机制,通过多个注意力模块对全局时空相关性进行建模,但在捕捉路网局部结构信息和空间动态变化方面存在短板,因此对动态变化的时空关联刻画不准确。DGCRN通过捕获交通数据间的动态空间依赖性,与使用预定义或自适应矩阵的模型相比,预测精度显著提高,这凸显了在模型中对动态空间关系进行建模的必要性。然而,DGCRN对时间维度动态变化的刻画能力较弱,时间特征提取不够全面,预测精度还有待提升。GDGCN则分别在空间和时间维度上,采用动态图卷积方法捕捉空间和时间动态相关性,并通过在空间、时间、特征等所有维度上跨层组合参数共享和独立块,提取出了各种稳定的模式。但模块间的融合方式较为简单,导致模型对复杂交通流场景的适配性仍有提升空间。

Table 2 The average prediction results of the next 12 time steps

表2 未来12个时间步的平均预测结果

Module	PEMS04			PEMS07			PEMS08			NE-BJ		
	MAE	RMSE	MAPE/%	MAE	RMSE	MAPE/%	MAE	RMSE	MAPE/%	MAE	RMSE	MAPE/%
VAR	23.75	36.66	18.09	101.20	155.14	39.69	22.32	33.83	14.47	6.14	9.65	23.33
SVR	28.66	44.59	19.15	32.97	50.15	15.43	23.25	36.15	14.71	6.04	9.61	23.30
STGCN	21.76	34.77	13.87	22.90	35.44	11.98	17.84	27.12	11.21	5.39	9.09	22.14
GWNET	19.35	31.72	13.30	21.22	33.11	9.07	15.06	24.85	9.51	4.99	9.20	19.45
GMAN	19.14	31.60	13.19	20.97	34.10	9.05	15.31	24.92	10.13	4.80	9.18	18.36
ASTGNN	18.60	31.02	12.63	20.61	34.01	8.86	14.97	24.71	9.49	4.92	9.41	18.51
DGCRN	18.82	30.57	13.03	21.45	34.53	9.21	14.95	23.77	9.69	4.79	9.23	17.98
GDGCN	18.57	29.99	12.57	20.22	33.38	8.55	14.58	23.98	9.87	4.72	9.09	18.12
AFEDGCN	18.28	30.39	12.18	20.12	33.32	8.52	14.15	23.86	9.41	4.59	8.68	17.84

从具体数值来看,在4个数据集上,AFEDGCN的平均MAE、RMSE和MAPE值分别为14.28、24.06和11.99%;GDGCN的相应平均值为14.52、24.11和12.28%;DGCRN的平均性能为15.00、24.52和12.48%。具体来看,相较于最优的GDGCN和DGCRN两个模型,AFEDGCN在MAE、RMSE和MAPE指标上的增益有一定的增幅。与GDGCN相比MAE、RMSE和MAPE指标分别有1.65%、0.21%和2.36%的预测性能提升。与DGCRN相比,MAE、RMSE和MAPE指标分别有4.80%、1.87%和3.93%的预测性能提

升。这进一步表明,AFEDGCN模型在交通流量预测任务中,具备更优预测效果。

为了更好地描述预测结果,将AFEDGCN与整体效果最好的基线模型GDGCN在一天内进行结果可视化比较,如图2—图5所示。可以观察到,AFEDGCN在部分时间步的预测比GDGCN效果更好。

3.6 消融实验分析

为了进一步评估所提模型中各组成部分对模型预测性能的影响,在PEMS04和PEMS07数据集上的未来1 h预

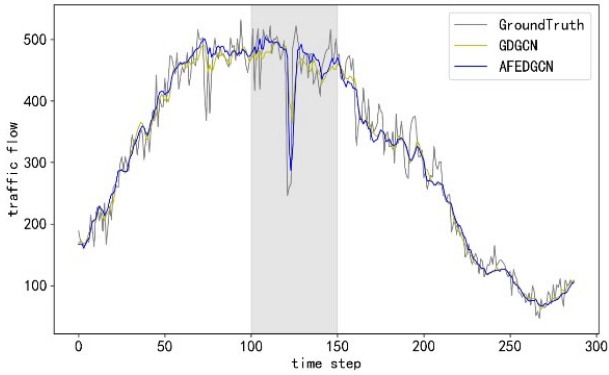


Fig. 2 Node 45 on PEMS04

图2 PEMS04上的45节点

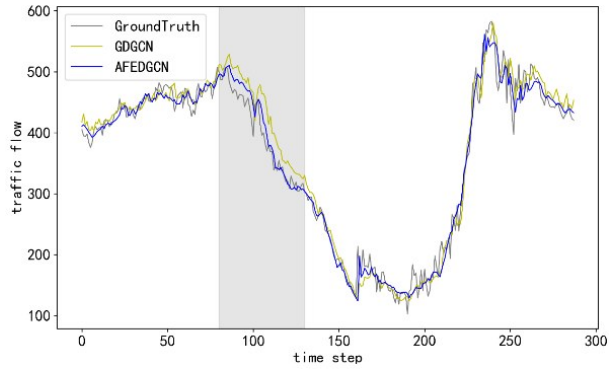


Fig. 3 Node 563 on PEMS07

图3 PEMS07上的563节点

测进行消融实验(见表3),设计了以下5种AFEDGCN的变体。①w/o GGCN:移除全局图卷积模块;②w/o IFEM:使用残差链接替换交互式特征增强模块;③w/o ADGCN:移除空间注意力融合图卷积模块;④w/o EDGCN:使用传统静态图卷积模块替换增强动态图卷积模块;⑤w/o MHTA:移除多头时间注意力模块。从结果可以看出,各模块对预测结果均起到了积极作用。模型在替换增强型动态图卷积后性能出现了明显下降,这表明本文所提出的增强型动态图卷积与传统静态图卷积相比,能够更好地捕捉交通路网中的

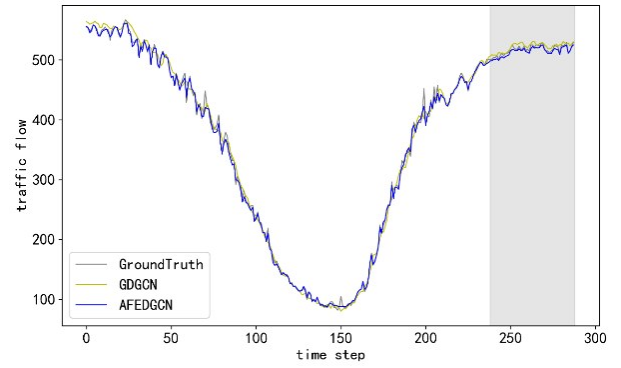


Fig. 4 Node 121 on PEMS08

图4 PEMS08上的121节点

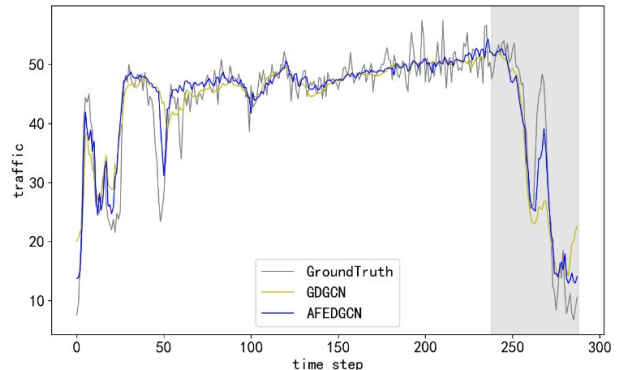


Fig. 5 Node 54 on NE-BJ

图5 NE-BJ上的54节点

动态空间特性;在移除多头时间注意力模块后,模型难以提取序列的上下文时间信息,从而导致性能明显下降。而二者结合可以有效地捕捉交通数据中复杂的时空特征。

通过结果还可以发现,全局图卷积模块通过捕获路网节点在整个时间段内潜在的、长期的全局空间相关性,能够进一步提升模型预测性能。交互式特征增强模块能够对节点的特征进行有效提取增强,比简单的残差链接效果更好。

Table 3 Ablation experiment result

表3 消融实验结果

Module	PEMS04			PEMS07			NE-BJ		
	MAE	RMSE	MAPE/%	MAE	RMSE	MAPE/%	MAE	RMSE	MAPE/%
AFEDGCN	18.28	30.39	12.18	20.12	33.32	8.52	4.59	8.68	17.84
w/o GGCN	18.46	30.45	12.47	20.27	33.55	8.77	4.61	8.69	17.92
w/o IFEM	18.50	30.52	12.36	20.47	34.52	8.71	4.61	8.73	18.09
w/o ADGCN	18.57	30.45	12.74	20.60	33.95	9.24	4.77	9.28	19.98
w/o EDGCN	19.89	32.71	13.50	22.04	36.32	9.81	5.06	9.67	20.86
w/o MHTA	19.69	33.04	13.09	22.87	38.19	10.44	5.09	9.65	19.60

3.7 模型参数影响分析

为了进一步研究网络超参数设置对预测性能的影响,模型在PEMS08数据集上进行了不同的网络核心超参数设置实验。选择的超参数包括数据隐藏维度 D_{hid} 、时间注意力头数量 h 和节点隐藏维度 d_N 。具体实验结果如图6所示。

可以看出,对于数据隐藏维度,当隐藏维度为64时,模型达到了更优预测结果。对于较小的维度,可能无法提取复杂特征,导致预测结果较差。而较大的维度则增加了模型复杂度,导致过拟合或难以学习有效表示,影响了性能。类似地,节点隐藏维度设置过小会导致无法全面表示路网的节点信息;将维度设置过大则会导致信息冗余,结

果显示设置为4时预测结果更好。

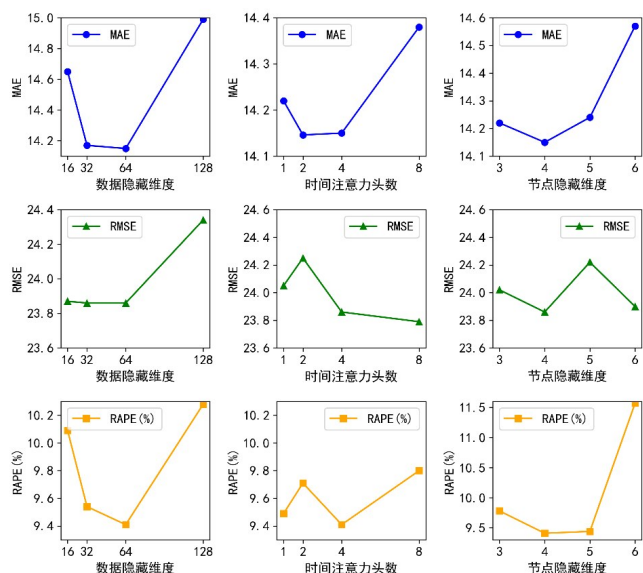


Fig. 6 Model parameter analysis on PEMS08

图6 PEMS08的模型参数分析

时间注意力头 h 的数量能够关注输入序列的不同方面,从而提取出更全面的特征。从结果可以看出,当注意力头数量为4时,模型整体性能更优。

4 结语

为更好地对交通流复杂的动态时空关系进行建模,提出了一种用于交通流量预测的注意力融合的增强动态图卷积模型AFEDGCN。该方法使用空间注意力以增强动态图卷积,同时结合交互式卷积增强节点自身特征;将改进的增强动态GCN与GRU相结合,捕获动态空间关系时同步地对时间相关性进行初步提取;使用多头时间注意力对时间相关性进行更精准的建模。为了进一步捕捉整个时间段内的空间特征,使用一种全局图卷积对整个时间段内的空间特征进行深层次建模。实验结果表明,AFEDGCN模型在不同的预测性能上均优于基线模型,验证了该模型能够有效捕获交通流动态复杂的时空依赖性,提高预测精度。

参考文献:

- [1] ZIVOT E, WANG J. Vector autoregressive models for multivariate time series[M]. New York: Springer New York, 2006: 369-413.
- [2] XU H, JIANG C. Deep belief network-based support vector regression method for traffic flow forecasting[J]. Neural Computing and Applications, 2020, 32(7): 2027-2036.
- [3] 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.
- [4] WU Z, PAN S, LONG G, et al. Connecting the dots: multivariate time series forecasting with graph neural networks[C]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2020: 753-763.
- [5] YU B, YIN H, ZHU Z. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting[DB/OL]. <https://arxiv.org/abs/1709.04875>.
- [6] ZHAO L, SONG Y, ZHANG C, et al. T-GCN: a temporal graph convolutional network for traffic prediction[J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 21(9): 3848-3858.
- [7] LIN J, LI Z, LI Z, et al. Dynamic causal graph convolutional network for traffic prediction[C]//2023 IEEE 19th International Conference on Automation Science and Engineering(CASE), 2023: 1-8.
- [8] JU X, XIE X, YU Y. Spatial-temporal dynamic fusion graph convolutional network for traffic forecasting[C]//2023 5th International Academic Exchange Conference on Science and Technology Innovation (IAECST), 2023: 709-713.
- [9] WU Z, PAN S, LONG G, et al. Graph WaveNet for deep spatial-temporal graph modeling[BD/OL]. <https://arxiv.org/abs/1906.00121>.
- [10] WANG X, MA Y, WANG Y, et al. Traffic flow prediction via spatial temporal graph neural network[C]//Proceedings of the Web Conference 2020, 2020: 1082-1092.
- [11] 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.
- [12] SHAO Z, ZHANG Z, WEI W, et al. Decoupled dynamic spatial-temporal graph neural network for traffic forecasting[DB/OL]. <https://arxiv.org/abs/2206.09112>.
- [13] SONG Y, LUO R, ZHOU T, et al. Graph attention informer for long-term traffic flow prediction under the impact of sports events[J]. Sensors, 2024, 24(15): 4796.
- [14] 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.
- [15] 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.
- [16] ZHOU H, REN D, XIA H, et al. AST-GNN: an attention-based spatio-temporal graph neural network for interaction-aware pedestrian trajectory prediction[J]. Neurocomputing, 2021, 445: 298-308.
- [17] LIU H, DONG Z, JIANG R, et al. Spatio-temporal adaptive embedding makes vanilla transformer sota for traffic forecasting[C]//Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, 2023: 4125-4129.
- [18] ELDELE E, CHEN Z, WU M, et al. TSLANet: rethinking transformers for time series representation learning[M]. Florida: CRC Press, 2024: 35-57.
- [19] CAO L, WANG B, JIANG G, et al. Spatiotemporal-aware trend-seasonality decomposition network for traffic flow forecasting[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2025: 11463-11471.
- [20] KONG W, GUO Z, LIU Y. Spatio-temporal pivotal graph neural networks for traffic flow forecasting[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2024: 8627-8635.
- [21] LUO X, ZHU C, ZHANG D, et al. Dynamic graph convolutional network with attention fusion for traffic flow prediction[DB/OL]. <https://arxiv.org/abs/2302.12598>.
- [22] LI F, FENG J, YAN H, et al. Dynamic graph convolutional recurrent network for traffic prediction: benchmark and solution[J]. ACM Transactions on Knowledge Discovery from Data, 2023, 17(1): 1-21.
- [23] XU Y, HAN L, ZHU T, et al. Generic dynamic graph convolutional network for traffic flow forecasting[J]. Information Fusion, 2023, 100: 101946.

(责任编辑:孙娟)