

DOI:10.13232/j.cnki.jnju.2026.01.007

融合趋势分解与自适应图扩散注意力网络的交通流预测

熊章友¹, 李卫军^{1,2*}, 朱晓娟¹, 马馨瑜¹, 杨国梁¹, 丛晓宇¹

(1. 北方民族大学计算机科学与工程学院, 银川, 750021; 2. 图形图像智能处理国家民委重点实验室, 银川, 750021)

摘要: 交通预测在城市交通管理和流量监测中至关重要, 但交通流中复杂的时空关系给准确预测带来了巨大的挑战。时空图神经网络和注意力机制已成为解决交通道路中依赖关系的有效方法, 然而, 大多数基于图神经网络的模型依赖预定义的静态邻接矩阵来建模空间依赖性, 且空间特征的提取依赖于固定图结构中的权重。其次, 现有的注意力机制忽略了交通流数据的特性, 难以捕捉节点之间的相似的交通模式。针对上述问题, 提出 TD-ADGAT 模型, 使用自适应图扩散注意力网络对空间关系进行建模, 无需显式计算预定义图结构的权重, 能自适应生成可训练的邻接矩阵权重, 以显著降低时间复杂度。此外, 针对交通流时间序列数据的特点, 重新设计时间维度的注意力机制, 将交通流数据分解为趋势性和季节性因素, 利用多层感知机捕获趋势性变化和采用傅里叶注意力机制建模季节性变化, 更好地建模交通流的时间关系和节点间的交通模式。最后, 在三个公共数据集上的实验结果证明 TD-ADGAT 模型优于其他基准模型。

关键词: 图注意力网络, 趋势分解, 注意力机制, 交通流预测, 时空特征

中图分类号: TP391

文献标志码: A

Traffic flow prediction based on trend decomposition and adaptive graph diffusion attention network

Xiong Zhangyou¹, Li Weijun^{1,2*}, Zhu Xiaojuan¹, Ma Xinyu¹, Yang Guoliang¹, Cong Xiaoyu¹

(1. School of Computer Science and Engineering, North Minzu University, Yinchuan, 750021, China;

2. Key Laboratory of Images & Graphics Intelligent Processing of State Ethnic Affairs Commission, North Minzu University, Yinchuan, 750021, China)

Abstract: Traffic prediction is crucial in urban traffic management and flow monitoring, but the complex spatial-temporal relationships in traffic flow bring great challenges to accurate prediction. Spatial-temporal graph neural networks and attention mechanisms have become effective methods to solve dependencies in traffic roads. However, most GNN-based models rely on predefined static adjacency matrices to model spatial dependencies, and the extraction of spatial features relies on fixed graph structure weights. Moreover, the existing attention mechanism ignores the characteristics of traffic flow data, and it is difficult to capture similar traffic patterns between nodes. To solve the above problems, this paper proposes the TD-ADGAT model, which uses an adaptive graph diffusion attention network to model spatial relationships. It does not need to explicitly calculate the weights of predefined graph structures, and can adaptively generate trainable adjacency matrix weights, thereby significantly reducing the time complexity. In addition, according to the characteristics of traffic flow time series data, the attention mechanism of time dimension is redesigned, and the traffic flow data is decomposed into trend and seasonal factors. Multilayer Perceptron (MLP) is used to capture the trend change, and Fourier attention is used to model the seasonal changes, so as to better model the temporal relationship of traffic flow and the traffic pattern between nodes. Experimental results on three public datasets demonstrate that TD-ADGAT outperforms existing baseline models in prediction accuracy.

基金项目: 宁夏高等学校科学研究项目(NYG2024086), 银川市科技支撑项目(2025SF49), 北方民族大学研究生创新项目(YCX24363)

收稿日期: 2025-10-25

* 通信联系人, E-mail: lwj@nmu.edu.cn

Keywords: graph attention network, trend decomposition, attention mechanism, traffic flow prediction, spatial-temporal features

早期的交通流预测主要依赖传统的时间序列模型,如支持向量回归(Support Vector Regression, SVR)^[1]、贝叶斯网络(Batch Normalization, BN)^[2]、自回归综合移动平均模型(Autoregressive Integrated Moving Average, ARIMA)^[3]等,但难以有效建模交通数据中的时空特征.随着交通数据集越来越大,深度学习方法被广泛应用于交通流预测任务,例如,引入图神经网络(Graph Neural Network, GNN)来处理具有图结构的交通流数据,如 DCRNN^[4], STGCN^[5], GWNET^[6], AGCRN^[7]和 DSTAGNN^[8]等模型.尽管这些方法在一定程度上缓解了非欧几里得空间下的建模难题,但仍存在显著局限性,它们普遍依赖预定义的静态邻接矩阵,难以捕捉交通流中动态变化的空间关联特性.同时,其图卷积权重通常保持固定,限制了对复杂时空模式的建模能力.最后,随着网络深度增加,这些模型普遍面临过平滑问题.尽管GWNET引入了自适应邻接矩阵来提取空间相关性,但仍未彻底摆脱固定权重的限制.

为了解决上述问题,本文提出一种新的自适应图注意力网络,摒弃传统注意力机制中固定的点积计算方式,引入可学习的参数矩阵来建模节点间的相关性,降低了时间复杂度.同时,引入扩散卷积机制,通过多阶邻居建模远距离节点间的依赖关系,有效缓解了深层网络中的过平滑问题.

此外,为了解决长时间序列依赖问题,注意力机制通过动态分配权重有效突破了RNN (Recurrent Neural Network) 和 CNN (Convolutional Neural Network)在建模远距离依赖方面的局限性,例如,ASTGCN^[9]和GMAN^[10]分别通过多周期建模和转换注意力机制捕捉交通流中的动态时空依赖关系.然而,当前注意力机制在处理具有显著趋势特征的序列时,往往难以有效捕捉时间维度中具有相似日趋势或周趋势节点之间的潜在关联性.如图1a所示,节点150和172的交通流量趋势相似,均在早高峰时段出现流量激增,晚高峰

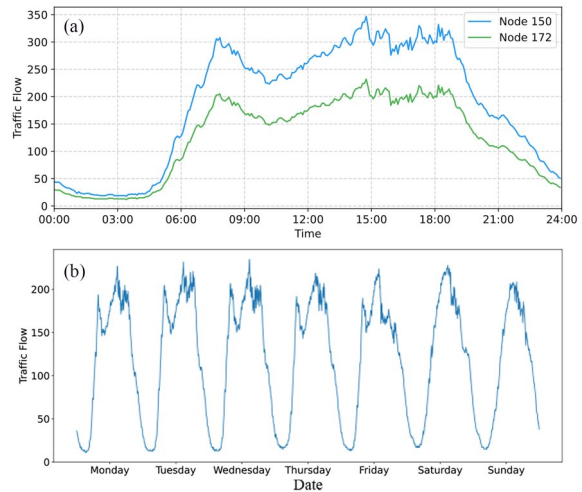


图 1 交通流量的时间相关性分析:(a)日趋势部分流量分析;(b)季节性的周趋势部分流量分析

Fig. 1 Time-related analysis of traffic flow: (a) daily trend component flow analysis, (b) seasonal weekly trend component flow analysis

过后下降,表明两者在实际道路网络中距离较近,具有类似的交通模式.如图1b所示,交通流量在工作日与周末之间存在显著差异,工作日流量上下波动明显,而周末相对平稳,表明两者分别呈现不同的交通模式,并且整体上交通流量随日期变化表现出周期性的波动规律,反映了交通流模式的季节性的周趋势特征.因此,本文将时间序列的数据分为趋势性和季节性两部分,分别使用多层感知机(Multilayer Perceptron, MLP)来捕获趋势性和采用傅里叶注意力机制捕获季节性,更有效地提取时间维度上的相似交通模式相关性.

综上,本文的贡献如下.

(1)通过两个模块捕获时空相关性,趋势分解模块将时间序列分解为趋势性和季节性来捕获时间相关性的相似交通模式.自适应图扩散注意力模块融合了自适应图注意力网络与扩散卷积,用于提取局部和多阶节点间的空间相关性.

(2)提出的趋势分解模块从解耦的角度分析交通流数据,使用线性模型 MLP 提取趋势部分,并通过基于频域的傅里叶注意力机制建模季节性

交通模式,更有效地表示时间序列中的长期趋势与周期性变化.

(3)提出一种新的注意力机制,称为自适应图注意力网络,能从数据中自适应地学习相邻节点的权重,无需显式计算注意力权重矩阵,有效降低了模型的计算复杂度.

(4)在三个交通数据集上进行实验,实验结果表明,提出的模型优于其他的基准方法.

1 相关工作

1.1 基于图神经网络的交通流预测 交通网络数据具有显著的时空特性,既包含动态的时间变化又具有复杂的空间拓扑结构.RNN难以建模长期依赖,CNN受限于欧氏空间假设,难以处理非规则的交通路网.而GCN更适用于非欧几里得结构,能有效提取交通流的时空特征.例如,崔建勋等^[11]将交通流划分为网格化与拓扑化两类,系统分析了深度学习的1D因果卷积、2D和3D图像卷积的优缺点.Song et al^[12]的STSGCN模型通过设计时空同步建模模块直接捕获局部时空关联,并引入多模块结构应对时空异质性问题.Fang et al^[13]的STGODE模型采用张量形式的常微分方程(Ordinary Differential Equation, ODE)来构建深度连续网络,同时考虑道路节点的空间邻域和语义邻域关系,以实现交通流时空特征的深度建模.赵文竹等^[14]的MVSTGCN模型采用静态结构与动态模式构建多视角图,利用动态图卷积提取空间依赖特征.Jiang et al^[15]提出一种学习时空图的新型的元图卷积循环网络Mega-CRN模型,有效解决了时空异质性问题.针对可扩展性不足和过度平滑等问题,Zhang et al^[16]的LightST模型通过时空知识蒸馏与自适应对比机制,提高模型在复杂场景下的鲁棒性与泛化能力.然而,上述模型大都依赖固定的图结构,难以建模节点间的潜在关联.因此,本文提出了自适应图扩散注意力网络,将自适应注意力机制与扩散卷积结合,动态建模交通网络中的空间依赖关系.

1.2 基于注意力机制的交通流预测 注意力机制因其能动态捕捉交通流的时空依赖,被广泛应用于交通预测.例如,Zhang et al^[17]的ASTCG模

型融合CNN,GRU与注意力机制提取时空特征.Huang et al^[18]设计了局部流注意力机制,通过流感知的定向注意力建模交通流动态,并利用局部空间注意力替代全局计算.Chen et al^[19]的TFM-GCAM模型以交通流矩阵作为图卷积网络的邻接矩阵来表征节点交通流特性,并设计了动态特征提取器和注意力特征融合机制来有效捕捉交通数据的时空动态特征.Transformer的自注意力机制作为注意力机制的一种,擅长捕捉序列内部相似性.例如,Chen et al^[20]的Bi-STAT模型利用循环Transformer与双向预测机制来提升时空建模效率,但Bi-STAT没有考虑外部因素对交通流的影响,因此,ST-MAN^[21]引入位置感知的时空记忆网络,并结合多尺度注意力机制与条件外部因素融合模块,实现对长期时空依赖性与外部干扰因素的有效建模.Li et al^[22]的SSL-STFormer模型结合自监督任务与纠缠感知机制,建模时空耦合特性.高榕等^[23]提出改进的时空Transformer模型ISTTM,通过编码器结合稀疏空间注意力和层次扩散卷积捕捉时空特征,解码器引入双重交叉注意力建模历史与未来的细粒度和全局关系.Wang et al^[24]的STGformer模型结合GCN和Transformer的优势,在单层中有效地捕获全局和局部模式的高阶时空相互作用.然而上述模型普遍忽略了交通流的趋势性与季节性特征,且常用的空间注意力机制依赖全局计算,计算开销较大,难以捕捉节点间的相似交通模式.针对这些不足,本文深入分析了交通流数据的时空特征.

2 模型方法

2.1 问题定义

定义1 道路网络 将道路网络建模为一个交通拓扑图 $G=(\nu, \epsilon, A)$,其中, $\nu=\{v_1, v_2, \dots, v_N\}$ 表示包含 N 个交通节点的集合($|\nu|=N$), $\epsilon \subseteq \nu \times \nu$ 为节点间连接关系的集合, A 为图 G 的邻接矩阵, N 代表节点的总数.

定义2 交通流量张量 T 时间步长的交通流量张量为 $X=\{X_1, \dots, X_t, \dots, X_T\} \in R^{N \times C \times T}$,其中, $X_t \in R^{N \times C}$ 表示在时间步长 t 时道路网络中各

节点的交通流特征观测值, C 是每个节点对应的特征数量.

交通流量预测任务旨在根据历史交通观测数据来推断未来一段时间内的交通状况. 具体地, 给定过去 T 步交通流张量 X , 通过学习一个映射函数 f , 从历史观测中推测未来 T' 步的交通流量. 其形式化表示如式(1)所示:

$$[X_{(t-T+1)}, \dots, X_t; G] \xrightarrow{f} [X_{(t+1)}, \dots, X_{(t+T)}] \quad (1)$$

TD-ADGAT 模型整体框架如图 2 所示, 主要

包括趋势分解模块和自适应图扩散注意力模块. 时间维度上, 模型将交通流序列分解为趋势性和季节性部分, 分别通过 MLP 提取长期趋势特征和利用傅里叶时间注意力提取季节性周期特征, 并通过时空嵌入层融合时间和空间信息, 生成带时间嵌入的交通流张量. 空间维度上, 将交通流张量与路网节点输入自适应图扩散注意力模块, 结合自适应图注意力和扩散卷积, 前者降低复杂度, 后者捕捉多阶节点空间依赖. 最终融合各特征输入输出层, 生成预测结果.

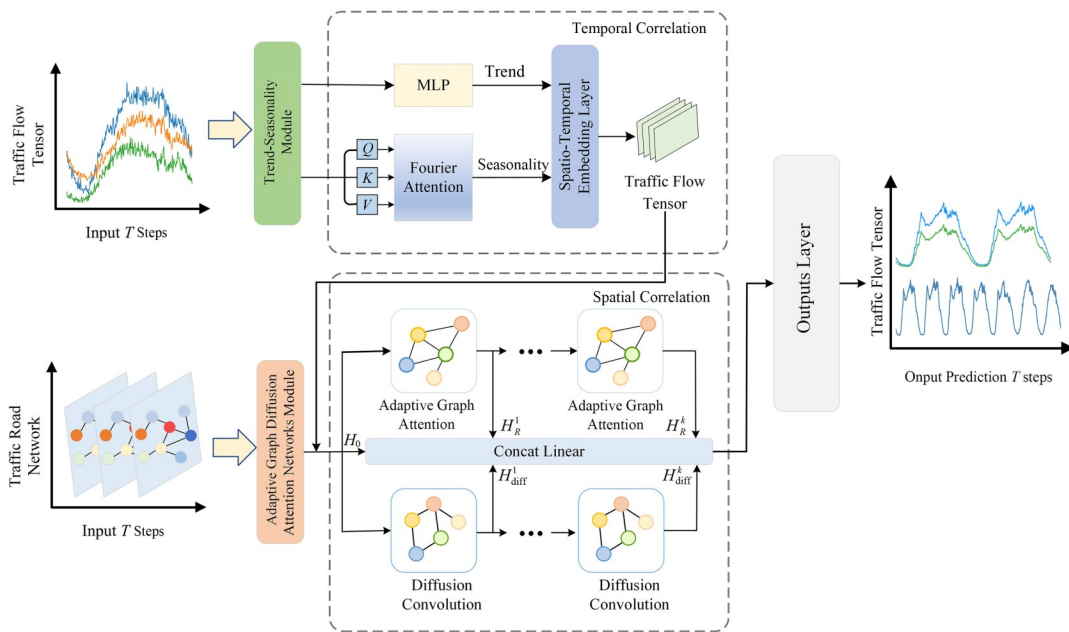


图 2 TD-ADGAT 模型图

Fig. 2 The diagram of TD-ADGAT model

2.2 趋势分解的时间关系分析模块 交通流时间序列通常包含短期和长期趋势、季节性变化以及外部事件引起的非周期波动, 本文重点讨论趋势性与季节性的变化. 虽然有研究如 FEDformer^[25] 考虑了交通流的时间变化问题, 但在分解模式建模上仍显不足, 难以同时捕捉复杂的趋势性与季节性波动特征. TDformer^[26] 在此基础上进一步分析了注意力机制在建模历史趋势时存在显著误差, 并通过趋势分解提升预测性能. 然而, 该方法对趋势分解结果高度依赖, 使用简单线性模型难以捕捉非线性变化, 且在应对复杂模式时稳健性不足. 而 Shao et al^[27] 提出的 STID 模型通过引入空间和时间标识信息, 证明了 MLP 处理时间序列

数据的有效性.

因此, 本文的趋势分解模块将交通流时间序列输入数据分解为趋势性和季节性两部分. 首先, 采用 MLP 来提取交通流趋势部分, 并通过从原始时间序列中减去趋势部分来得到季节性分量. 为了提取输入序列中的趋势信息, 采用滑动平均的方法对原始输入数据进行平滑处理. 对于任意时间步 t , 趋势特征 X_{tre} 定义为该时刻及其前 $m - 1$ 时间步内输入序列的平均值, 对于长度为 T 的输入数据 $X_{input} \in R^{N \times C \times T}$ 进行移动平均处理. 其计算过程如式(2)所示, 分别得到趋势部分 X_{tre} 和季节性分量 X_{sea} .

$$X_{\text{tre}} = \frac{1}{m} \sum_{i=0}^{m-1} X_{\text{input}}(t-i), X_{\text{sea}} = X_{\text{input}} - X_{\text{tre}} \quad (2)$$

其中, $X_{\text{input}}(t-i)$ 是原始输入序列在时间步 $t-i$ 的数值, m 是滑动窗口大小.

2.2.1 趋势部分 在传统的 Transformer 框架中, 自注意力机制主要关注输入序列中不同位置之间的语义相关性, 而对时间关系的建模通常依赖位置编码进行隐式处理. 虽然位置编码在一定程度上保留了时间顺序的信息, 但其本质是静态的, 无法有效表达时间间隔、趋势变化和季节性等时间序列中的关键特征. 然后, 自注意力本质上是对历史信息的加权整合, 缺乏对时间演化过程的建模能力, 所以对时间依赖关系的表达不充分.

为了解决上述问题, 本文采用具有 l 层的 MLP 来预测未来交通流量的趋势, 首先将前一层 ($l-1$ 层) 的输出特征 X_{tre}^{l-1} 与 l 层的可学习权重矩阵 $W_L^{(l)}$ 进行逐元素相乘, 再对所有 d 个特征维度求和并叠加偏置项 $b_L^{(l)}$, 最后通过 $ReLU$ 激活函数引入非线性变换, 这种层级式设计通过 l 层网络的逐级特征提取, 显式地捕捉交通流量中的时序动态和演化趋势. 具体如式(3)所示.

$$\hat{X}_{\text{tre}} = ReLU \left(\sum_{i=1}^d W_L^{(l)} \odot X_{\text{tre}}^{(l-1)} + b_L^{(l)} \right) \quad (3)$$

2.2.2 季节性部分 时间序列数据中的季节性特征通常表现为明显的周期性波动模式. 与常规的时域分析方法相比, 频域建模具有显著优势, 一方面, 频谱分析直接揭示时间序列中的主要周期成分, 避免了时域中烦琐的滑动窗口操作; 另一方面, 频域中的能量分布往往具有稀疏性, 少数关键频率便可携带主要信号信息, 显著提升了特征提取的效率. 因此, 本文引入频域的傅里叶注意力机制, 并结合 MLP 捕获交通流的时间相关性.

如图3所示, 傅里叶注意力机制采用了一种新的处理流程, 首先, 通过快速傅里叶变换(Fast Fourier Transform, FFT)将原始季节性输入从时域转换到频域; 接着, 在频域中引入三个可训练的参数矩阵, 分别对查询(Q)、键(K)和值(V)进行线性变换, 并通过矩阵乘法计算 Q 与 K 的相似度, 经缩放(Scale)操作对 V 进行加权融合, 以提取频域特征; 最后, 使用逆快速傅里叶变换

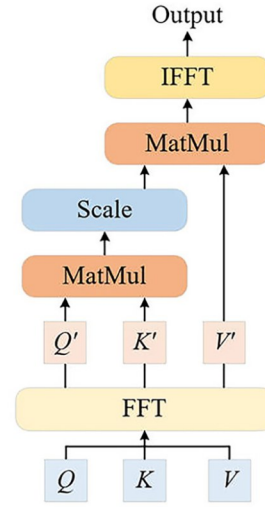


图3 傅里叶变化的时间注意力机制框架图

Fig. 3 Temporal attention mechanism with Fourier changes

(Inverse Fast Fourier Transform, IFFT)将融合后的特征还原回时域. 该机制的引入使算法复杂度从 $O(n^2)$ 降低至 $O(n \lg n)$. 其得到的查询(Q)、键(K)和值(V)如式(4)所示, 得到的季节性预测值如式(5)所示.

$$Q'_j = FFT(X_{\text{sea}} W_q), K'_j = FFT(X_{\text{sea}} W_k), \quad (4)$$

$$V'_j = FFT(X_{\text{sea}} W_v)$$

$$\hat{X}_{\text{sea}} = IFFT \left(\text{Softmax} \left(\frac{Q'_j (K'_j)^T}{\sqrt{d}} \right) V'_j \right) \quad (5)$$

其中, FFT 代表快速傅里叶变换, $IFFT$ 是 FFT 的逆变化操作. W_q , W_k 和 W_v 都是可学习的线性变换矩阵, \hat{X}_{sea} 是输出的增强后季节性分量.

2.3 自适应图扩散注意力网络模块

2.3.1 自适应图注意力网络 以往的交通流预测方法普遍依赖预定义的邻接矩阵权重, 难以准确捕捉道路间随时间变化的空间相关性. 如图4a传统的 GAT 所示, 首先, 原始节点特征通过前馈神经网络进行非线性变换, 生成高阶特征表示, 并通过共享线性变换 W 将节点特征映射为投影, 得到节点 i 和节点 j 的投影特征 Wh_i 和 Wh_j ; 随后, 节点间的特征通过交互操作(如拼接或点积)计算注意力分数; 最后, 结合预定义的邻接矩阵作为掩码, 在局部邻域内根据注意力分数对节点特征进行加权求和. 尽管该机制能有效建模节点间的重

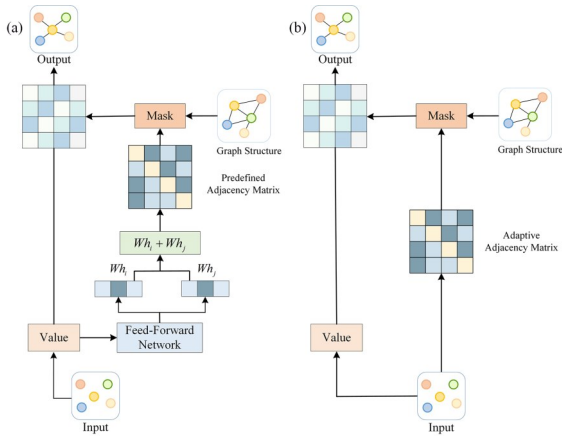


图 4 GAT (a)和 Adaptive GAT (b)对比

Fig. 4 Comparison between GAT (a) and Adaptive GAT (b)

要性差异,但在实际应用中仍存在显著局限性. 第一,注意力权重的计算需要对所有节点 (i, j) 进行组合,导致整体计算复杂度达到 $O(n^2 + n^2)$;第二,由于每次都需要显式地计算 $Wh_i + Wh_j$,导致训练和推理效率较低.

为了解决这些问题,本文提出了自适应图注意力网络,如图4b所示,不再对所有节点显式计算注意力得分,而是通过随机自适应生成可训练的邻接矩阵,结合图结构掩码,有选择性地建立节点连接. 首先,这些权重值是可训练的或是固定的,通过生成一系列自适应矩阵 $R = \{R_1, R_2, \dots, R_T\}$,其对应不同的时间步长的注意力矩阵,其中, R_t 表示时间步 t 生成的注意力矩阵. 最后,得到自适应GAT的注意力权重矩阵如式(6)所示,具体计算过程如式(7)所示.

$$A_i = \text{Softmax}(\text{mask}(R_t)) \quad (6)$$

$$\alpha_{i,j} = \text{Softmax}(S_{i,j}) = \frac{\exp(S_{i,j})}{\sum_{r \in V_{(i)}} \exp(S_{i,r})} \quad (7)$$

其中, $\alpha_{i,j}$ 表示节点 i 对节点 j 的注意力权重, $\text{mask}(R^{(t)})$ 表示仅保留预定义图结构中存在的节点的权重, $S_{i,j}$ 表示学习到的权重相关性, $V_{(i)}$ 代表邻居节点的集合. 注意力网络的隐藏状态如式(8)所示.

$$H_{R_t} = \sigma(A_i f(H_i)) \quad (8)$$

其中, H_{t_i} 表示时间步长 t_i 的隐藏状态.

为了稳定学习过程,将自适应图注意力网络进一步扩展为多头机制. 具体来说,使用参数 h 将并行的注意力机制拼接,如式(9)所示.

$$H_{R_t} = \parallel_{k=1}^h \sigma(A_i^k f^k(H_i^k)) \quad (9)$$

其中, \parallel 代表拼接操作,合并多头输出, h 是注意力头的数量, A_i^k 是第 k 个头的注意力权重矩阵, $f(\cdot)$ 是第 k 个头的独立非线性变换和输出 $d = D/k$ 维度特征. $H_{R_t} \in R^{N \times D}$ 是通过时间步长 t_i 学习到的隐藏状态,然后将所有时间步长的隐藏状态合并到 $H_R \in R^{P \times N \times D}$.

2.3.2 自适应图扩散注意力网络 为了更有效地捕捉交通网络中多阶节点之间的空间依赖关系,本文在自适应图注意力网络中引入扩散卷积机制,构建自适应图扩散注意力模块. 该模块结合了局部注意力建模和多阶扩散特征提取,增强了空间特征表达的丰富性与灵活性. 扩散卷积能建模节点与多阶邻居之间的复杂关系,提升空间感知能力,并在一定程度上缓解过平滑问题.

在实际交通网络中,节点(如传感器或路口)之间的影响往往超越直接邻居,存在更远距离的依赖. 因此,本模型采用扩散卷积模拟特征在图结构上的多阶传播过程,其定义如式(10)所示.

$$H_{\text{diff}}^k = \tilde{A}^k H^{k-1} \quad (10)$$

其中, \tilde{A} 为归一化的邻接矩阵, H^{k-1} 为第 $k-1$ 层节点隐藏状态,通过对不同阶数扩散结果进行加权叠加,选择较大的 K 阶扩散卷积可以捕捉更远距离的依赖关系,并缓解过度平滑问题. 最后,得到扩散卷积的整体输出,其表达式如(11)所示.

$$\text{DiffConv}(H) = \sum_{k=0}^K H_{\text{diff}}^k W_k \quad (11)$$

其中, W_k 为第 k 阶扩散对应的可学习权重矩阵, K 为最大扩散阶数.

根据自适应GAT的权重矩阵的计算式,节点特征在自适应邻居权重下进行更新,其更新式如(12)所示. 为了同时利用局部交互特征与多阶扩散特征,图扩散注意力模块采用双分支结构进行特征提取. 最终输出通过融合自适应GAT分支与扩散卷积分支的结果实现,自适应图扩散注意力网络融合模块如式(13)所示.

$$H_R^k = \sigma(A_R^k f(H_R^{k-1})) \quad (12)$$

$$H_s^K = \sum_{k=0}^K (H_R^k W_R^k + H_{\text{diff}}^k W_{\text{diff}}^k) \quad (13)$$

其中, A_R^k 为第 k 阶自适应注意力邻接矩阵, $f(\cdot)$ 为特征变换函数, $\sigma(\cdot)$ 为激活函数, W_R^k 和 W_{diff}^k 为可学习的参数矩阵, H_{diff} 是扩散卷积, H_R 是自适应 GAT.

2.4 时空嵌入层 在时空嵌入层中, 本文构建了一种融合时空特征的嵌入机制, 以增强模型对交通流动态变化的建模能力.

首先, 借助一个全连接网络将原始输入数据映射到高维特征空间, 得到高维特征表示 $X_t \in R^{N \times D \times T}$. 其中, N 是节点数量, D 表示特征投影维度, T 表示时间步长. 随后, 为了建模交通流中的趋势性与季节性特征, 构造了周期性时间嵌入 $TE(X)$. 其中, $X_{\text{tre}} \in R^{N \times D \times T_d}$ 表示前一天的历史流量序列, 用于刻画短期趋势. $X_{\text{sea}} \in R^{N \times D \times T_w}$ 表示前一周同类型工作日和周末的流量序列, 用于建模周周期的季节性规律.

然后, 将 X_{tre} 和 X_{sea} 拼接后形成的时间嵌入特征表示为 $TE(X) = \text{Concat}(X_{\text{tre}}, X_{\text{sea}})$. 引入自适应图扩散注意力机制建模节点间随时间演化的空间依赖关系, 得到空间结构嵌入 $SE(X)$, 并借助位置编码嵌入 $PE(X)$ 加强对时序顺序信息的感知能力.

最终, 融合上述各类信息, 获得综合的时空嵌入表示如式(14)所示.

$$X_{\text{emb}} = X_t + TE(X) + SE(X) + PE(X) \quad (14)$$

2.5 输出层 在输出层, 模型将对应的趋势特征 X_{tre} 、季节性特征 X_{sea} 和空间特征 X_{spatial} 在通道维度拼接形成融合特征 $X_{\text{fusion}} = X_{\text{tre}} \parallel X_{\text{sea}} \parallel X_{\text{spatial}}$. 融合后特征经过第一个 1×1 卷积层 $\text{Conv1}(\cdot)$ 映射到跳跃维度, 得到跳跃连接维度 X_{sc} . 然后对所有跳跃连接输出进行求和, 得到隐藏状态为 $X_h = \sum \text{Conv1}(X_{\text{fusion}})$. 最后, 通过第二个 1×1 卷积层 $\text{Conv2}(\cdot)$ 映射为目标输出维度, 得到最终预测结果如式(15)所示.

$$\hat{X} = \text{Conv2}(\sigma(X_h)) \quad (15)$$

3 实验设计与分析

3.1 实验设置

3.1.1 实验配置和数据集分析 为了保证实验的有效性, 模型采用 Python 语言, 基于 PyTorch 框架实现模型, 并在 NVIDIA GeForce RTX 4090 GPU 上完成实验. 实验在配置有 PyTorch 1.10.1+cu111 和 Python 3.7 的虚拟环境中运行. 模型的超参数通过在验证集上的实验进行调整, 并以验证集性能为依据选取最优模型. 最终选定的超参数包括特征维度 D 为 64, 注意力头数为 8, 编码器层数为 $[3, 4, 4]$. 模型训练使用 Adam 优化器, 初始学习率为 0.001, 批处理大小为 16, 聚类大小为 4, 训练总轮数为 100 个 epoch.

为了评估提出的 TD-ADGAT 模型的预测性能, 实验选用交通流预测领域广泛使用的标准数据集 PeMS. PeMS 是一个重要的交通监测数据库, 主要用于分析和评估加利福尼亚州的高速公路交通状况. 该系统通过遍布全州超过 39000 个探测器实时采集数据, 这些探测器主要分布于高速公路和大都市区域, 其采集的信息包括车速、车流量以及道路拥堵情况等, 为交通管理、城市规划和学术研究提供了可靠的数据支持. 目前公开的 PeMS 数据集主要包括 PeMS04, PeMS07 和 PeMS08 等数据集, 其统计信息如表 1 所示.

表 1 数据集信息

Table 1 Datasets description

Datasets	Node	Time step	Time range
PeMS04	307	16992	2018/01/01—2018/02/28
PeMS07	883	28224	2017/05/01—2017/08/31
PeMS08	170	17856	2016/07/01—2016/08/31

3.1.2 基准模型和评价指标 为了验证 TD-ADGAT 模型的有效性, 选取三类代表性基线方法对 TD-ADGAT 模型进行评估, 这些方法虽然在模型架构上存在差异, 但都致力于解决交通流预测中的关键挑战. 传统机器学习方法 (如 SVR^[1], ARIMA^[3]) 着重于时间序列建模, 但未充分考虑交通数据的空间特性; 基于图神经网络的方法 (如 DCRNN^[4], STGCN^[5], GWNET^[6], AGCRN^[7], DSTAGNN^[8], STSGCN^[12], ST-

GODE^[13], STFGCN^[28])通过引入图卷积操作显式建模路网空间结构,但大多采用静态或预定义的图结构;基于注意力和Transformer的方法(如GMAN^[10], STID^[27], Bi-STAT^[20], PDFormer^[29])利用自注意力机制动态捕捉时空依赖关系. 本实验采用短时交通流预测中常用的评价指标,包括平均绝对误差(MAE)、均方根误差(RMSE)和平均绝对百分比误差(MAPE),分别如式(16),式(17)和式(18)所示.

$$MAE = \frac{1}{J} \sum_{i=1}^J |x_i - y_i| \quad (16)$$

$$RMSE = \sqrt{\frac{1}{J} \sum_{i=1}^J (x_i - y_i)^2} \quad (17)$$

$$MAPE = \frac{1}{J} \sum_{i=1}^J \left| \frac{x_i - y_i}{x_i} \right| * 100\% \quad (18)$$

其中, x_i 和 y_i 分别表示时间间隔 i 的交通流量的实际观测的数值和预测值, J 表示交通预测测试样本的数量.

3.2 模型对比实验与分析 本实验将提出的模型与三种基线方法进行性能对比,结果如表2所示. 其中,黑体表示该指标结果最优,下划线表示次优结果. 从实验结果可以看出,TD-ADGAT在PeMS04, PeMS07和PeMS08三类数据集上的表

现均优于其他基准模型.

如表2所示,传统预测方法例如SVR和ARIMA仅考虑了时间相关性,忽略了交通数据中复杂的空间相关性,所以整体效果最差. 基于图结构的深度学习模型优于传统预测方法,主要原因在于其使用了预定义网络结构来捕获道路网络的空间性,例如,DCRNN和STGCN. 为了进一步提升对动态空间相关性的建模能力,一些研究引入了自适应嵌入方法,例如,GWNET和AGCRN. 然而这些模型的图结构在训练前已固定,难以根据交通数据的实时变化进行调整,因此后续提出了一些数据驱动的动态图构建方法,例如,STSGCN,GMAN和DSTAGNN.

图结构模型虽在时空关系表征能力上具有优势,但其复杂的图结构设计和图卷积操作导致了较高的时间复杂度. TD-ADGAT省略了邻接矩阵的显式计算,有效降低了模型的时间复杂度. 因此,在整体性能上,TD-ADGAT优于上述图神经网络模型. 此外,由于Transformer引入的自注意力机制可以动态关注关键的时空信息,带有Transformer架构的模型如Bi-STAT和PDFormer在预测性能上比传统GNN模型更具有有一定优

表2 模型对比实验与分析

Table 2 Model comparison experiment and analysis

Model	PeMS04			PeMS07			PeMS08		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE
SVR	28.70	19.20%	44.56	32.49	14.26%	50.22	23.25	14.64%	36.16
ARIMA	28.55	19.55%	40.36	33.89	17.60%	46.38	31.23	19.25%	33.47
DCRNN	22.74	14.75%	36.58	23.63	12.28%	36.51	18.18	11.23%	28.18
STGCN	21.76	13.87%	34.77	22.89	11.98%	35.44	17.84	11.21%	27.12
GWNET	19.36	13.30%	31.72	21.22	9.08%	34.12	15.06	9.51%	24.86
AGCRN	19.83	12.97%	32.26	22.37	9.12%	36.55	15.95	10.09%	25.22
STSGCN	21.18	13.88%	33.65	24.26	10.20%	39.03	17.13	10.96%	26.78
GMAN	19.25	13.00%	30.85	20.68	9.31%	33.56	14.87	9.77%	24.06
STGODE	20.85	13.78%	32.82	22.97	10.14%	36.19	16.82	10.62%	26.78
STID	19.58	13.38%	31.79	21.52	9.15%	36.29	15.58	10.33%	25.89
DSTAGNN	19.28	12.81%	31.30	21.42	9.01%	34.51	15.81	10.00%	25.01
Bi-STAT	19.01	12.93%	30.55	20.28	<u>8.55%</u>	33.24	14.08	9.33%	23.58
PDFormer	<u>18.51</u>	12.31%	<u>30.02</u>	19.93	8.68%	<u>32.88</u>	<u>13.66</u>	9.10%	23.56
STFGCN	18.60	<u>12.27%</u>	30.34	<u>19.72</u>	8.64%	33.24	13.68	9.03%	<u>23.45</u>
TD-ADGAT	18.31	12.07%	29.81	19.61	8.50%	32.74	13.03	8.92%	22.83

势. Bi-STAT通过Transformer模块和复杂的交叉注意力机制捕获时空相关性,但其动态注意力机制本身需要额外的参数和计算步骤,效果不及TD-ADGAT. 与最近的STFGCN模型相比,虽然STFGCN考虑了交通流的语义和多尺度动态时间依赖关系,但忽略了交通流的趋势性和季节性因素,导致其在PeMS08数据集上的各项指标分别比本模型高0.65,0.11%和0.62.

为了进一步验证模型在不同预测时间步下的性能表现,本文在PeMS04和PeMS08数据集上进行了12个时间步的对比实验,如图5所示.其

中,DSTAGNN在两个数据集上的表现最差,Bi-STAT由于使用自注意力机制,预测效果优于DSTAGNN. PDFormer同时建模了语义信息和地理特征,其在PeMS04数据集上进行12个时间步预测时的表现优于Bi-STAT. 与其他基准模型相比,TD-ADGAT在所有时间步上整体表现最优. 尽管在PeMS04数据集的RMSE指标上,TD-ADGAT在第一至第四时间步的预测效果略低于STFGCN,但在后续时间步中表现出更优的预测能力,这表明TD-ADGAT在中长期预测任务中具有更强的稳定性和准确性.

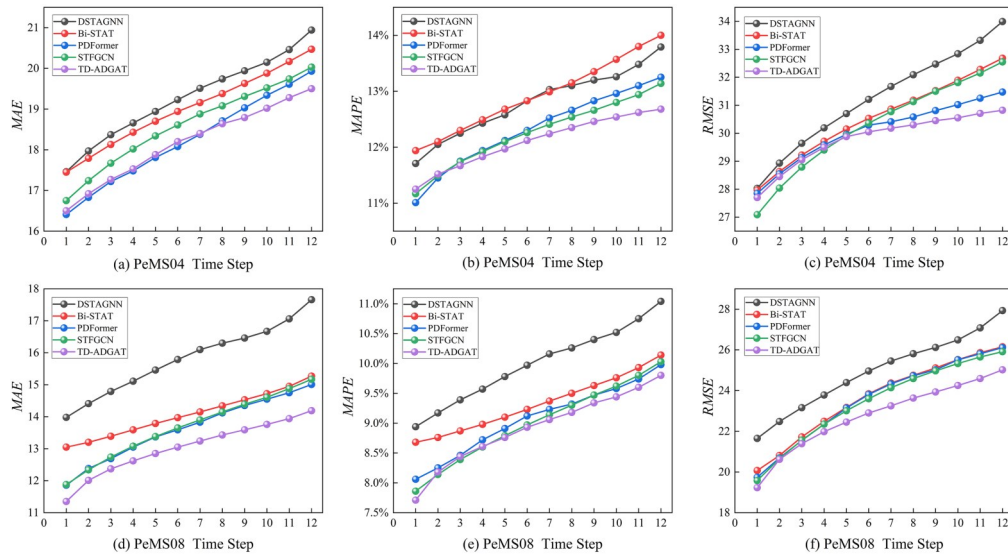


图5 PeMS04和PeMS08数据集的不同时间步对比实验

Fig. 5 Comparison experiments of PeMS04 and PeMS08 datasets at different time steps

3.3 消融实验

3.3.1 模型消融实验 为了证明每个模块的有效性以及对模型的影响,将模型与以下变体在PeMS04和PeMS08数据集上进行消融实验比较.

(1) DEC-TD: 移除趋势分解模块,仅使用线性模型MLP提取时间特征.

(2) DEC-ADG: 移除自适应图扩散注意力模块,仅使用静态图卷积捕获空间相关性.

(3) DEC-FFT: 移除趋势分解模块中的傅里叶注意力,使用时域注意力.

(4) DEC-TSE: 移除时空嵌入,交通流不会对趋势性和季节性成分数据进行处理.

整体模型的消融实验结果如表3所示,表中黑体字表示结果最优.

由表可见,本文提出的TD-ADGAT模型与其他变体相比,预测效果最优,进一步验证了模型

表3 模型消融实验结果

Table 3 Model ablation experiments

Model	PeMS04			PeMS08		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE
DEC-TD	18.61	12.43%	30.63	13.38	9.23%	23.41
DEC-ADG	18.75	12.55%	30.68	13.54	9.22%	23.59
DEC-FFT	18.47	12.52%	31.11	13.34	9.13%	23.44
DEC-TSE	19.30	12.78%	31.76	14.23	9.58%	23.65
TD-ADGAT	18.31	12.07%	29.81	13.03	8.92%	22.83

的有效性. 去除趋势分解模块后, DEC-TD 仅使用 MLP 提取时间特征存在明显的缺陷, MLP 只能捕捉局部的时间模式, 难以建模交通流数据中复杂的周期性规律, 导致在 PeMS04 数据集上的预测效果不佳. 去除自适应图扩散注意力模块后, DEC-ADG 在 PeMS04 和 PeMS08 数据集上的预测效果显著下降, 这说明仅使用传统的固定的邻接矩阵的静态图卷积方法, 难以捕获交通流的动态空间特征. 而自适应图扩散注意力通过可学习的自适应权重以及聚合多节点的特征信息, 更有效地捕获交通流的空间相关性. 使用时域注意力替换傅里叶注意力后, DEC-FFT 预测效果明显下降, 特别是在 PeMS04 数据集上的 RMSE 激增变大, 比 TD-ADGAT 误差高 1.30, 这证明傅里叶注意力比时域注意力更适合捕获交通流的季节性. 去除时空嵌入模块后, DEC-TSE 预测效果最差, 交通流无法通过时空嵌入层处理趋势性和季节性的数据, 无法捕获时空信息, 导致处理交通流数据的时间复杂度偏高.

3.3.2 不同图结构消融实验 为了进一步验证自适应图扩散注意力模块的有效性, 本文设计了几个变体进行实验对比.

(1) PG: 仅使用预定义邻接矩阵提取空间特征.

(2) GAT: 使用传统的 GAT 替换本文模型的自适应 GAT.

(3) DGCN: 仅使用扩散卷积捕获空间相关性.

(4) AGAT: 仅使用自适应图注意力捕获空间相关性.

实验结果如表 4 所示, 表中黑体字表示结果最优. 可以得到如下结论.

(1) 不同图结构消融方式在 PeMS04 和 PeMS08 数据集上均表现出明显差异. PG 变体在两个数据集上的 MAE, MAPE 和 RMSE 指标均较高, 整体效果普遍不如其他变体, 表明单纯依赖固定的预定义邻接矩阵无法充分捕获复杂的空间相关性.

(2) 虽然 GAT 变体大部分效果优于 PG, 但 GAT 每次计算时需要对节点进行交互以获取注意力矩阵, 并通过 *Softmax* 函数进行归一化操作,

表 4 不同图结构消融实验结果

Table 4 Ablation experiments with different graph structures

Model	PeMS04			PeMS08		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE
PG	18.53	12.41%	31.24	13.30	9.48%	23.25
GAT	18.37	12.66%	30.72	13.33	9.24%	23.11
DGCN	18.45	12.56%	30.47	13.35	9.19%	23.09
AGAT	18.38	12.25%	30.18	13.06	9.11%	23.07
TD-ADGAT	18.31	12.07%	29.81	13.03	8.92%	22.83

导致时间复杂度偏高, 有大量的计算成本. 而 AGAT 通过自适应生成邻接矩阵, 避免了对所有节点显式计算注意力得分, 其权重通过训练过程学习获得, 消除了自注意力机制的交互过程, 显著降低了时间复杂度.

(3) DGCN 引入了扩散卷积, 可以聚合多阶邻域的特征信息, 并缓解了图神经网络中过度平滑问题, 因此, 在捕获空间依赖性方面比 PG 和 GAT 更具优势.

(4) TD-ADGAT 结合了 DGCN 和 AGAT 的优点, 在所有变体中取得的效果最好, 进一步提升了模型对复杂空间依赖关系的建模能力.

3.4 时间复杂度分析 为了验证 TD-ADGAT 的计算效率, 本文将 TD-ADGAT 与四个主流基准模型进行了参数实验, 其训练和推理的平均时间如表 5 所示.

由表 5 可见, TD-ADGAT 在计算效率方面具有显著优势. 这得益于 TD-ADGAT 采用自适应 GAT, 减少了节点间的冗余交互, 时间复杂度从 $O(n^2 + n^2)$ 降低到 $O(n)$. 而 DSTAGNN 模型结构复杂, 导致计算复杂度最高, TD-ADGAT 在训练和推理时间上分别比其减少了约 65% 和 70%.

表 5 PeMS04 数据集时间复杂度分析

Table 5 Time complexity analysis of the PeMS04 dataset

Model	Training time (s·epoch ⁻¹)	Inference time (s)
DSTAGNN	350.37	43.93
Bi-STAT	247.96	25.32
PDFormer	195.55	19.92
STFGCN	143.78	17.57
TD-ADGAT	121.70	13.10

PDFormer因其复杂的全局自注意力机制导致计算量随时间步数呈平方级增长,而TD-ADGAT在两项指标上比其降低37.76%和34.24%。尽管STFGCN在预测性能上表现较好,但由于其包含复杂的时空交互机制及动态图构建模块,训练和推理时间仍明显高于TD-ADGAT。

3.5 参数分析

3.5.1 注意力头数参数分析 为了验证本模型中的注意力机制的有效性,在PeMS04和PeMS08数据集上对比了不同注意力头数(4,6,8,10,12)对模型性能的影响。如图6所示,模型

在不同注意力头数设置下的MAE与RMSE表现出明显差异。实验结果表明,适当的注意力头数有助于模型有效提取时空特征,而过多的注意力头数可能带来计算开销增加与预测精度下降的问题。例如,在PeMS04数据集上,随着注意力头数从4增加至8,模型的MAE和RMSE误差显著下降,其中,当注意力头数为8时,达到最优。继续增加至10和12后,误差反而上升,说明过多的注意力头数可能引入冗余信息,削弱模型的泛化能力。PeMS08数据集也呈现出类似趋势,模型在注意力头数为8时,同样取得最佳性能。

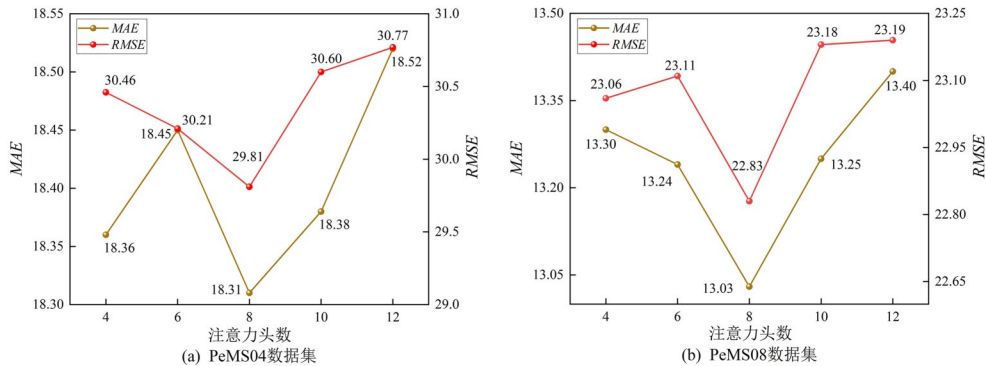


图6 注意力头数参数分析

Fig. 6 Attention head count analysis

3.5.2 聚类大小参数分析 由表6中不同聚类大小参数在PeMS04和PeMS08数据集上的对比结果可以看出,聚类数量对模型空间关系具有显著影响。实验结果表明,聚类大小数为4时,模型效果最优。在PeMS04数据集中,其MAE,MAPE和RMSE分别比聚类为3时降低了0.14,0.24%和1.24,比聚类为6时分别降低了0.56,0.43%和1.99,在PeMS08数据集中也表现如此。由此可见,随着聚类数量的增加,误差指标整体呈上升趋势,尤其从4增加到6时,误差增长更显著。这说明聚类数过多可能引入冗余,干扰模型对关键空间结构的学习,而聚类数过少可能导致簇内区域过大,难以捕捉细粒度的局部动态特征。

表6 聚类大小参数分析

Table 6 Analysis of cluster size parameters

数据集	评价指标	3	4	5	6
PeMS04	MAE	18.45	18.31	18.66	18.87
	MAPE	12.31%	12.07%	12.45%	12.50%
	RMSE	30.95	29.81	31.39	31.80
PeMS08	MAE	13.09	13.03	13.31	13.46
	MAPE	9.28%	8.92%	9.17%	9.29%
	RMSE	23.04	22.83	23.20	23.44

3.6 可视化分析 为了更好地验证趋势分解模块中的趋势性和季节性的交通模式,采用不同趋势的节点进行可视化分析,如图7a所示,节点97和203趋势一致且重叠。如图7b所示,节点21和

127趋势完全不同,还选取图1a趋势一致的节点150和172。

采用t-SNE方法将节点嵌入维度降至二维,对某一天工作日和周末的不同时刻的节点分布进行可视化分析,如图8所示,黑框代表节点150和172,蓝框代表节点21和127,红框代表节点97和203。

在趋势性方面,由图8可见,无论工作日还是周末,在不同时间段,部分节点如150与170因趋

势相似,在嵌入空间中距离较近.节点 97 与 203 因趋势紧密重叠,在嵌入空间中距离很近,说明其交通模式高度相似.相反,节点 21 与 127 因图 7b 中流量趋势差异明显,因此,在图 8a 和图 8b 中不同时刻的嵌入距离始终较远,表明趋势相似度很低.这从侧面验证了本模型引入的扩散卷积能有效捕捉多阶节点之间的空间依赖关系.在季节性

方面,图 8a 工作日的不同时刻节点分布较为分散,且离散度在 7:00,12:00 和 18:00 等典型高峰时刻及其邻近时间段达到峰值.而图 8b 周末的不同时刻节点分布更集中,不同时间段的节点基本无交集,特别高峰期时段.由此可见,本模型可以有效捕捉趋势相同的节点交通模式,并能清晰区分工作日与周末的季节性差异.

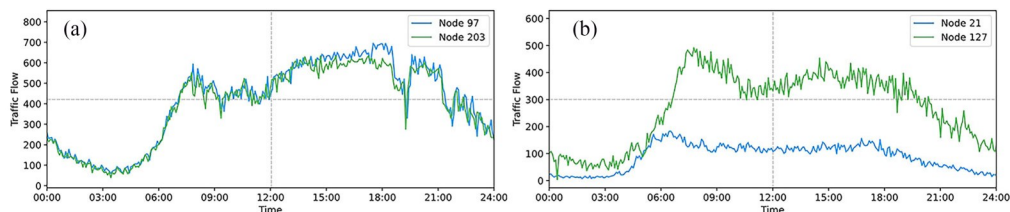


图 7 节点可视化分析: (a)节点 97 和节点 203 交通流量可视化; (b)节点 21 和节点 127 交通流量可视化

Fig. 7 Node visual analysis: (a) visualization of traffic flow at Node 97 and Node 203, (b) visualization of traffic flow at Node 21 and Node 127

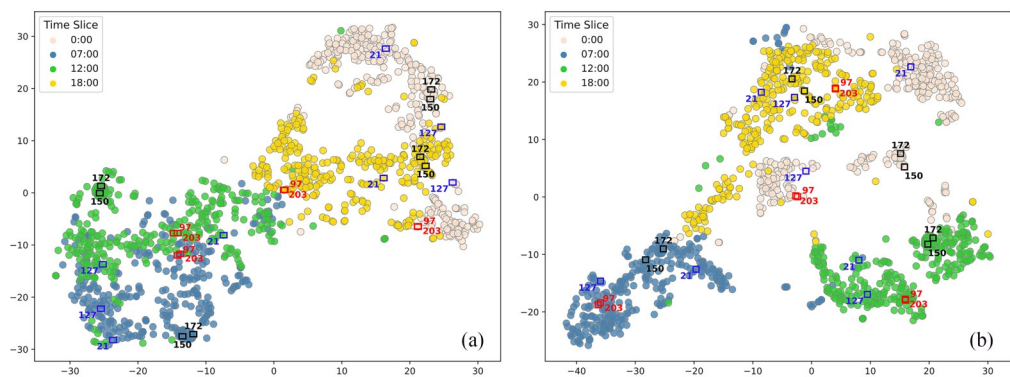


图 8 节点可视化分析: (a)工作日节点可视化; (b)周末节点可视化

Fig. 8 Node visualization analysis: (a) node visualization on working days, (b) node visualization on weekends

4 结论

为了更有效地捕捉交通流中的时空相关性,本文提出了一种融合趋势分解模块与自适应图扩散注意力网络的交通流预测模型.该模型通过趋势分解模块将交通流时序数据分解为趋势性和季节性成分,利用多层感知机提取趋势性特征,并引入傅里叶注意力机制建模季节性变化.为了捕捉空间相关性,设计了自适应图扩散注意力模块.自适应图注意力在传统图注意力的基础上进行改进,简化了节点间的复杂交互过程,更真实地学习图结构中的权重.同时,引入扩散卷积机制,有效捕捉多阶节点之间的空间依赖关系.

本文存在一定的局限性,尚未考虑外部因素对交通流的影响.未来将引入事件、天气、温度等外部信息,以进一步提升交通流预测的准确性.

参考文献

[1] Smola A J, Schölkopf B. A tutorial on support vector regression. *Statistics and Computing*, 2004, 14(3): 199–222.

[2] Sun S L, Zhang C S, Yu G Q. A bayesian network approach to traffic flow forecasting. *IEEE Transactions on Intelligent Transportation Systems*, 2006, 7(1): 124–132.

[3] Williams B M, Hoel L A. Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process:

- Theoretical basis and empirical results. *Journal of Transportation Engineering*, 2003, 129(6): 664–672.
- [4] Li Y G, Yu R, Shahabi C, et al. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. <https://arxiv.org/abs/1707.01926>, 2018–02–22.
- [5] Yu B, Yin H T, Zhu Z X, et al. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting//*Proceedings of the 27th International Joint Conference on Artificial Intelligence*. Menlo Park: AAAI Press, 2018: 3634–3640.
- [6] Wu Z H, Pan S R, Long G D, et al. Graph wavenet for deep spatial-temporal graph modeling//*Proceedings of the 28th International Joint Conference on Artificial Intelligence*. Menlo Park: AAAI Press, 2019: 1907–1913.
- [7] Bai L, Yao L N, Li C, et al. Adaptive graph convolutional recurrent network for traffic forecasting. *Advances in Neural Information Processing Systems*, 2020, 33: 17804–17815.
- [8] Lan S Y, Ma Y T, Huang W K, et al. Dstagnn: Dynamic spatial-temporal aware graph neural network for traffic flow forecasting//*Proceedings of the 39th International Conference on Machine Learning*. New York: PMLR, 2022: 11906–11917.
- [9] Guo S N, Lin Y F, Feng N, et al. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, 33(1): 922–929.
- [10] Zheng C P, Fan X L, Wang C, et al. Gman: A graph multi-attention network for traffic prediction. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, 34(1): 1234–1241.
- [11] 崔建勋, 要甲, 赵泊媛. 基于深度学习的短期交通流预测方法综述. *交通运输工程学报*, 2024, 24(2): 50–64.
- [12] Song C, Lin Y F, Guo S N, et al. Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, 34(1): 914–921.
- [13] Fang Z, Long Q Q, Song G J, et al. Spatial-temporal graph ODE networks for traffic flow forecasting//*Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. New York: Association for Computing Machinery, 2021: 364–373.
- [14] 赵文竹, 袁冠, 张艳梅, 等. 多视角融合的时空动态GCN城市交通流量预测. *软件学报*, 2024, 35(4): 1751–1773.
- [15] Jiang R H, Wang Z N, Yong J W, et al. Spatio-temporal meta-graph learning for traffic forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023, 37(7): 8078–8086.
- [16] Zhang Q R, Gao X Y, Wang H X, et al. Efficient traffic prediction through spatio-temporal distillation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2025, 39(1): 1093–1101.
- [17] Zhang Q Y, Chang W F, Yin C H, et al. Attention-based spatial-temporal convolution gated recurrent unit for traffic flow forecasting. *Entropy*, 2023, 25(6): 938.
- [18] Huang X, Zhang B W, Feng S S, et al. Interpretable local flow attention for multi-step traffic flow prediction. *Neural Networks*, 2023, 161: 25–38.
- [19] Chen J, Zheng L, Hu Y Z, et al. Traffic flow matrix-based graph neural network with attention mechanism for traffic flow prediction. *Information Fusion*, 2024, 104: 102146.
- [20] Chen C L, Liu Y B, Chen L, et al. Bidirectional spatial-temporal adaptive transformer for urban traffic flow forecasting. *IEEE Transactions on Neural Networks and Learning Systems*, 2023, 34(10): 6913–6925.
- [21] Liu Y, Guo B, Meng J X, et al. Spatio-temporal memory augmented multi-level attention network for traffic prediction. *IEEE Transactions on Knowledge and Data Engineering*, 2024, 36(6): 2643–2658.
- [22] Li Z T, Hu Z, Han P, et al. SSL-STMFormer self-supervised learning spatio-temporal entanglement transformer for traffic flow prediction. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2025, 39(11): 12130–12138.
- [23] 高榕, 万以亮, 邵雄凯, 等. 面向改进的时空Transformer的交通流量预测模型. *计算机工程与应用*, 2023, 59(7): 250–260.
- [24] Wang H J, Chen J Y, Pan T, et al. STGformer: Efficient spatiotemporal graph transformer for traffic

- forecasting. (2024-10-15) [2025-10-01]. <https://arxiv.org/abs/2410.00385>.
- [25] Zhou T, Ma Z Q, Wen Q S, et al. Fedformer: frequency enhanced decomposed transformer for long-term series forecasting//Proceedings of the 39th International Conference on Machine Learning. New York: PMLR, 2022: 27268-27286.
- [26] Zhang X Y, Jin X Y, Gopalswamy K, et al. First de-trend then attend: Rethinking attention for time-series forecasting. (2022-12-15) [2025-10-01]. <https://arxiv.org/abs/2212.08151>.
- [27] Shao Z Z, Zhang Z, Wang F, et al. Spatial-temporal identity: A simple yet effective baseline for multivariate time series forecasting//Proceedings of the 31st ACM International Conference on Information & Knowledge Management. New York: Association for Computing Machinery, 2022: 4454-4458.
- [28] Li H, Liu J, Han S Y, et al. STFGCN: Spatial-temporal fusion graph convolutional network for traffic prediction. Expert Systems With Applications, 2024, 255(Part B): 124648.
- [29] Jiang J W, Han C K, Zhao W X, et al. PDFformer: Propagation delay-aware dynamic long-range transformer for traffic flow prediction. Proceedings of the AAAI Conference on Artificial Intelligence, 2023, 37(4): 4365-4373.

(责任编辑 高善露)