

基于时空感知 Transformer 的交通流预测模型^{*}鲁思源¹ 沈琴琴¹ 包银鑫² 高锐锋¹ 施 隼¹

(1. 南通大学交通与土木工程学院 南通 226019; 2. 南通大学信息科学技术学院 南通 226019)

摘要: 交通流预测是智能交通系统的一个热点研究领域,其根本挑战是对交通数据中复杂的时空相关性进行有效建模。针对大部分现有时空 Transformer 模型在构建时空相关性矩阵时忽略了时间趋势性和空间异质性的重要影响的问题,提出一种基于时空感知 Transformer 的交通流预测模型。首先,采用改进的时空感知自注意力机制挖掘交通流数据中潜在的时间趋势性和空间异质性特征,建立精确的时空相关性矩阵以获取全局时空特征;然后,使用多尺度扩散卷积模拟交通流在路网中的多阶扩散过程,捕获节点多邻域范围的局部空间特征;最后,采用多元特征融合模块对捕获的时空特征进行自适应融合并输出预测结果。在 PeMS04 和 PeMS08 两个真实交通数据集上进行实验,结果表明,与最近提出的 RPConvformer、ASTGNN、PDFormer 等基于 Transformer 的基线模型相比,新模型的平均绝对误差分别降低了 8.0%、6.5% 和 2.0%。

关键词: 交通流预测;时空相关性;Transformer;自注意力机制;扩散卷积

中图分类号: U491.14;TN953.1 **文献标识码:** A **国家标准学科分类代码:** 510.4030

Traffic flow prediction model based on spatial-temporal aware Transformer

Lu Siyuan¹ Shen Qinqin¹ Bao Yinxin² Gao Ruifeng¹ Shi Quan¹

(1. School of Transportation and Civil Engineering, Nantong University, Nantong 226019, China;

2. School of Information Science and Technology, Nantong University, Nantong 226019, China)

Abstract: Traffic flow prediction is a hot research area in intelligent transportation systems, and the fundamental challenge is to effectively model the complex spatial-temporal correlations in traffic data. To address the problem that most existing spatial-temporal Transformer models ignore the important effects of temporal trend and spatial heterogeneity when constructing spatial-temporal correlation matrices, a traffic flow prediction model based on Spatial-Temporal Aware Transformer (STAFORMER) is proposed. First, an improved spatial-temporal aware self-attention mechanism is used to mine potential temporal trend and spatial heterogeneity features in traffic flow data, establishing an accurate spatial-temporal correlation matrix to obtain global spatial-temporal features. Then, the multi-range diffusion convolution is used to simulate the multi-order diffusion process of traffic flow in the road network to capture the local spatial features. Finally, the multivariate feature fusion module is used to adaptively fuse the captured spatial-temporal features and output the prediction results. Experiments are conducted on two real traffic datasets, i. e. PeMS04 and PeMS08, and the results show that, compared with the recently proposed Transformer-based models such as RPConvformer, ASTGNN, and PDFormer, the mean absolute errors of STAFORMER are reduced by 8.0%, 6.5%, and 2.0%, respectively.

Keywords: traffic flow prediction; spatial-temporal correlation; Transformer; self-attention mechanism; diffusion convolution

0 引言

经济社会的高速发展使得汽车保有量不断增加,现有的道路基础设施已经难以满足日益增长的交通需求,导致交通拥堵频繁发生,这也成为困扰城市高质量发展的难题。

作为智能交通系统的关键技术之一,实时准确的交通流预测可以为交通管理部门提供数据支持,帮助规划路线,指导车辆调度,达到缓解交通拥堵的目的^[1]。

早期的交通流预测模型主要采用统计学方法和机器学习方法^[2-3]。自回归差分移动平均和向量自回归等统计学

方法^[3]受线性和平稳性假设的限制,在实际应用中表现不佳。为了处理交通流预测的非线性问题,支持向量机^[4-5]和 K 近邻^[6]等传统的机器学习方法受到关注。但随着交通数据量的爆发式增长,这些传统的机器学习方法的浅层结构制约了其对复杂特征的提取能力。

近年来,深度学习的快速发展极大促进了交通流预测的研究。长短期记忆网络^[7]和门控递归单元(gated recurrent unit, GRU)^[8]作为常用的时序数据处理方法,能够有效捕获交通流数据中复杂的时间特征。但由于相邻道路之间的连接关系,空间信息在一定程度上也会影响交通流预测的精度。针对该问题,一些研究工作将路网划分为标准网格,引入卷积神经网络(convolution neural network, CNN)^[9-10]以提取不同区域间的空间相关性。然而基于 CNN 的空间建模方法忽略了交通路网底层的道路拓扑结构,无法从本质上描述交通数据的空间相关性。为此,研究人员进一步将交通路网抽象为图结构,使用图神经网络进行空间信息交互^[11],并融合时间序列建模方法,提出一系列基于时空融合的图神经网络交通流预测模型^[12-17]。如, Zhao 等^[14]分别使用 GRU 和图卷积网络(graph convolution network, GCN)提取交通流数据中的时间和空间特征,提出了时间图卷积网络(temporal graph convolution network, T-GCN)。

相较于传统的时空图神经网络,Transformer 凭借高效的全局性和动态性建模能力,近年来在交通流预测领域取得了显著的研究进展。自注意力机制作为 Transformer 的核心组件,通过动态学习输入特征之间的相关性,为交通流数据的空间和时间特征提取提供了强大的支持。Zheng 等^[18]利用自注意力机制提取交通流数据的时空特征,提出了图多头注意力网络(graph multi-attention network, GMAN),但并未充分考虑路网结构、交通流的周期性特征等,导致其时空特征提取能力有限。为此,一些研究工作通过在交通流数据中引入自适应空间位置嵌入^[19]和自适应周期性特征嵌入^[20]来增强数据的时空信息。此外,大量工作还从时间和空间两个维度改进自注意力机制,以提升模型对时空特征的提取能力。在空间维度上,为解决复杂路网中的注意力分散问题,研究者基于路网结构^[21]、交通流序列相似性^[21]和节点上下游关系^[22]设计掩码矩阵,引导模型关注起主导作用的空间依赖,忽略相关性较低的节点对。在此基础上, Jiang 等^[23]通过空间信息延迟感知模块,学习交通流在不同节点间传播的时间延迟,进一步提高空间建模的准确性,提出了传播延迟感知 Transformer(propagation delay-aware dynamic long-range transformer, PDFormer)。然而,这些方法在所有时刻共享相同的注意力参数,限制了模型构建随时间动态变化的复杂空间关系的能力。刘起东等^[24]针对这一问题,提出了时间感知的自注意力机制,能够根据预测时间动态调整注意力参数,以建模不同时刻下的动态空间关系。在时间维度上, Wen 等^[25]

针对自注意力机制对输入序列的顺序信息不敏感的问题,在键矩阵和值矩阵中加入自适应位置偏差向量,以学习各时间步间的相对位置关系,提出改进相对位置编码和一维卷积层的 Transformer(Transformer with relative positional encoding and 1D convolutional layer improvements, RPConvformer)。同时,为了增强模型对局部时间特征的关注, Guo 等^[26]引入卷积自注意力机制,通过一维卷积捕获交通流序列的局部变化趋势,提出了基于注意力的时空图神经网络(attention based spatial temporal graph neural network, ASTGNN); Wang 等^[27]通过掩码矩阵将注意力的关注范围限制在邻近时间步,但可能导致模型丢失长期时间信息; Geng 等^[28]提出门控时间自注意力模块,结合门控一维卷积和自注意力机制,分别提取局部和全局时间特征。

当前研究从多个角度对自注意力机制进行了改进,并取得了不错的效果。但是,现有自注意力机制在计算时空相关性的过程中仅关注单一节点或时间步的交通特征,未深入挖掘交通流数据中潜在的时间趋势性和空间异质性特征,因此难以获取准确的全局时空关系。此外,大部分现有方法基于路网结构等先验知识设计掩码矩阵,从而限制空间注意力的关注范围,捕获起主导作用的空间关系,但这种固定的掩码矩阵难以适应动态变化的交通模式,制约了模型挖掘潜在空间关系的能力。

针对上述问题,本文提出一种通过 3 个并行的特征编码模块进行时空特征提取的时空感知 Transformer 交通流预测模型(spatial-temporal aware transformer, STAFORMER)。首先,采用空间感知自注意力(spatial aware self-attention, SASA)模块和时间感知自注意力(temporal aware self-attention, TASA)模块,将图卷积和时域一维卷积与自注意力机制相结合,充分挖掘交通流数据中潜在的时间趋势性和空间异质性特征,以获取精确的全局时空特征;然后,通过模拟交通流的多阶扩散过程,利用多尺度扩散卷积(multi-range diffusion convolution, MDC)模块对局部空间特征进行提取;最后,将 3 个特征编码模块提取的时空特征输入多元特征融合模块进行自适应融合,实现准确的交通流预测。在两个实际的交通数据集 PeMS04 和 PeMS08 上进行验证,结果表明,与最近提出的 RPConvformer、ASTGNN、PDFormer 等基于 Transformer 的基线模型相比, STAFORMER 的平均绝对误差分别降低了 8.0%、6.5% 和 2.0%。

1 问题概述

交通流预测即使用历史交通流数据对未来一段时间内的交通流进行预测,是一个典型的时空预测问题。给定路网结构 $G = (S, E, A)$, 其中 S 是交通传感器的集合, E 是连接交通传感器的道路的集合, $A \in \mathbf{R}^{N \times N}$ 是邻接矩阵(N 表示传感器的数量), 反映各传感器间的连接关系, 若两个

传感器间有道路直接相连则为1,否则为0。

交通流历史数据由布置在路网中的传感器节点采集,每隔固定时间记录一次该时间间隔内通过传感器节点的车辆数。本文以 X_t^n 表示第 n 个传感器在 t 时间步所记录的交通流数据,则所有传感器在 t 时间步记录的交通流数据表示为 $\mathbf{X}_t = (X_t^1, X_t^2, \dots, X_t^N) \in \mathbf{R}^{N \times 1}$ 。交通流预测任务的目标是:对于当前时间步 t ,根据路网结构 $\mathbf{G}$ 和历史 T 个时间步的交通流历史值学习一个函数 f 来预测未来 τ 个时间步的交通流,即:

$$\{\mathbf{X}_{t+1}, \mathbf{X}_{t+2}, \dots, \mathbf{X}_{t+\tau}\} = f(\mathbf{X}_t, \mathbf{X}_{t-1}, \dots, \mathbf{X}_{t-T+1}; \mathbf{G}) \quad (1)$$

2 方法

STAFormer 的模型架构如图1所示。输入数据首先经过数据嵌入层进行时空信息嵌入;然后由 L 个堆叠的编码器依次提取时空特征;最后,输出层对每个编码器的输出进行聚合,并通过双层全连接网络生成预测结果。

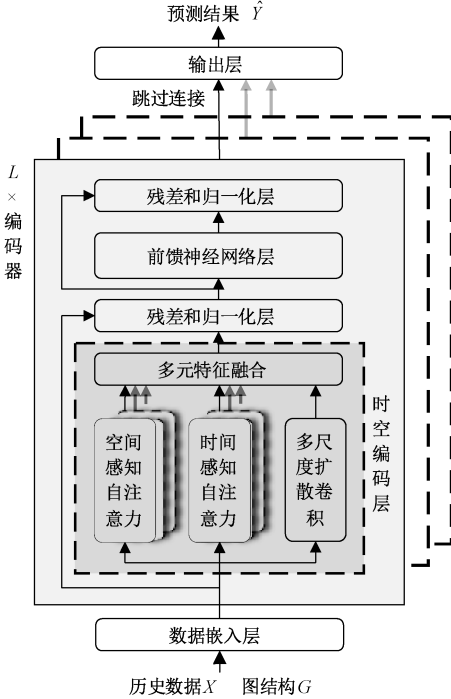


图1 STAFormer 整体架构

2.1 数据嵌入层

数据嵌入层首先通过全连接网络将原始输入 $\mathbf{X} \in \mathbf{R}^{T \times N \times 1}$ 映射到高维空间 $\mathbf{X}_{FC} \in \mathbf{R}^{T \times N \times d}$, d 是嵌入维度。然后分别计算时间位置嵌入 $\mathbf{E}_{TP} \in \mathbf{R}^{T \times d}$ 、空间位置嵌入 $\mathbf{E}_{SP} \in \mathbf{R}^{N \times d}$ 和时间周期嵌入 $\mathbf{E}_P \in \mathbf{R}^{T \times d}$, 最后将3个嵌入向量与 $\mathbf{X}_{FC} \in \mathbf{R}^{T \times N \times d}$ 相加得到数据嵌入层输出 $\mathbf{X}_E \in \mathbf{R}^{T \times N \times d}$, 具体流程如图2所示。

1) 时间位置嵌入

模型基于自注意力机制捕获交通流数据中的时间相关性,其本身不具备位置信息的概念,无法直接理解序列中每

个时间步的相对位置。因此模型通过正余弦位置编码进行时间位置嵌入,计算公式如下:

$$\mathbf{E}_{TP}(t, 2i) = \sin(t/10\,000^{2i/d}) \quad (2)$$

$$\mathbf{E}_{TP}(t, 2i+1) = \cos(t/10\,000^{2i/d}) \quad (3)$$

其中, t 是时间步, $2i$ 和 $2i+1$ 是当前嵌入的特征维度。

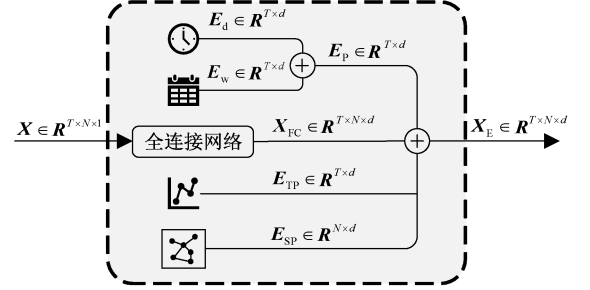


图2 数据嵌入层结构

2) 空间位置嵌入

为了表示交通路网的结构信息,模型使用图拉普拉斯位置编码获取空间位置嵌入。首先根据路网的邻接矩阵计算图拉普拉斯矩阵,并进行特征分解,计算过程如下:

$$\mathbf{L} = \mathbf{I} - \mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2} = \mathbf{U}^T \mathbf{A} \mathbf{U} \quad (4)$$

其中, $\mathbf{L}$ 是拉普拉斯矩阵, $\mathbf{D}$ 是邻接矩阵 $\mathbf{A}$ 的度矩阵, $\mathbf{A}$ 是特征值矩阵, $\mathbf{U}$ 是特征向量矩阵。然后,选取 q 个最小的非零特征值对应的特征向量,输入全连接网络进行维度变换,得到空间位置嵌入 $\mathbf{E}_{SP} \in \mathbf{R}^{N \times d}$ 。

3) 周期嵌入

交通流数据具有明显的周期性,如每日的高峰期都出现在相似的时间。本文通过构建可学习的周周期矩阵 $\mathbf{P}_w \in \mathbf{R}^{7 \times d}$ 和日周期矩阵 $\mathbf{P}_d \in \mathbf{R}^{1440 \times d}$ 来建模交通流数据的周期性特征。具体而言,周一至周日和每天的1440 min 都对应一个 d 维嵌入向量,对于输入模型的历史 T 个时间步的交通流数据,根据每个数据的采集时间从周期矩阵中获取日周期向量 $\mathbf{E}_d \in \mathbf{R}^{T \times d}$ 和周周期向量 $\mathbf{E}_w \in \mathbf{R}^{T \times d}$, 两者相加得到周期嵌入向量 $\mathbf{E}_P \in \mathbf{R}^{T \times d}$ 。

2.2 时空编码层

时空编码层是模型的主体部分,由空间感知自注意力模块、时间感知自注意力模块和多尺度扩散卷积模块3个主要模块构成,用于建模复杂的动态时空相关性。

1) 空间感知自注意力模块

空间异质性是指受道路类型、宽度及路网局部拓扑结构等空间属性的影响,不同区域的交通流分布存在差异。识别具有相似空间属性的节点并构建联系,有助于提高交通流预测的准确性。然而,节点的详细空间属性通常难以获取,限制了模型的泛化能力。为此,本文设计了如图3所示的空间感知自注意力机制。不同于传统空间自注意力机制仅考虑单个节点的交通特征,新方法利用GCN聚合节点局部范围内的交通特征,从而匹配具有相似局部交通模式

的节点,如同处于交通枢纽的节点。空间感知自注意力机制的计算过程如下:首先,对于输入 $\mathbf{X}_E$,使用 GCN 生成查询矩阵 $\mathbf{Q}_S$ 和键矩阵 $\mathbf{K}_S$,使用线性变换生成值矩阵 $\mathbf{V}_S$:

$$\text{GCN}(\mathbf{X}_E) = \text{ReLU}(\tilde{\mathbf{D}}^{-1/2} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-1/2} \mathbf{X}_E \mathbf{W}_G) \quad (5)$$

$$\mathbf{Q}_S = \text{GCN}(\mathbf{X}_E), \mathbf{K}_S = \text{GCN}(\mathbf{X}_E), \mathbf{V}_S = \mathbf{X}_E \mathbf{W}_S \quad (6)$$

其中, $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$ 是图 $\mathbf{G}$ 添加了自环的邻接矩阵, $\mathbf{I}$ 是单位矩阵, $\tilde{\mathbf{D}}$ 是 $\tilde{\mathbf{A}}$ 对应的度矩阵, $\mathbf{W}_G$ 和 $\mathbf{W}_S \in \mathbf{R}^{d \times d_h}$ 是可学习的参数矩阵, d_h 是自注意力头的特征维度, ReLU 是激活函数。

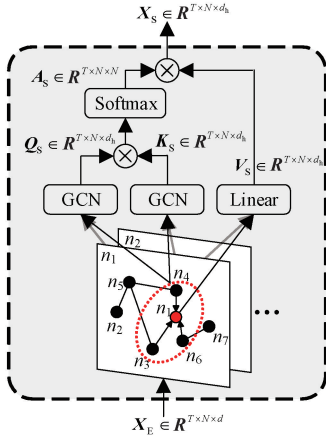


图 3 空间感知自注意力模块

然后,对查询矩阵和键矩阵执行缩放点积注意力计算,得到空间注意力权重矩阵,并对值矩阵进行加权求和,得到单个空间感知自注意力头的输出 $\mathbf{X}_S \in \mathbf{R}^{N \times T \times d_h}$:

$$\mathbf{X}_S = \text{Softmax}\left(\frac{\mathbf{Q}_S (\mathbf{K}_S)^T}{\sqrt{d_h}}\right) \mathbf{V}_S \quad (7)$$

其中, Softmax 函数表示对矩阵进行归一化操作。

2) 时间感知自注意力模块

自注意力机制最初用于处理单词等离散的数据,并没有考虑序列数据中重要的时间趋势信息,直接将其用于交通流序列的时间相关性建模将导致注意力分配不准确的问题。如图 4 所示,在传统时间自注意力机制中,时间步 D 对 A、B、C 的注意力是相似的,因为它们具有相似的特征值。然而,时间步 B、D 的时间相关性显然高于其他时间步,因为它们具有相似的局部变化趋势。

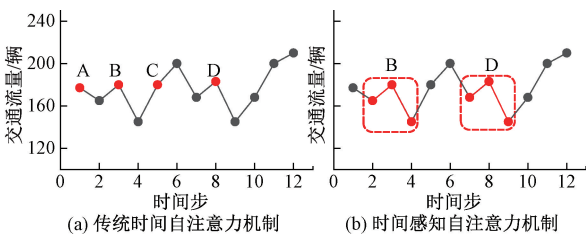


图 4 两种自注意力机制的区别

为了捕获准确的时间依赖关系,本模型采用时间感知自注意力机制进行时间相关性建模。该机制利用如图 5 所示的时域一维卷积取代传统自注意力中生成查询矩阵和键矩阵的线性变换操作。时域一维卷积以当前及相邻时间步的交通特征作为输入,经卷积运算生成新的特征表示,使模型能够感知交通序列的局部趋势信息,从而匹配具有相似局部变化趋势的时间步。时间感知自注意力机制的计算公式定义如下:

$$\mathbf{Q}_T = \Phi_Q * \mathbf{X}_E, \mathbf{K}_T = \Phi_K * \mathbf{X}_E, \mathbf{V}_T = \mathbf{X}_E \mathbf{W}_T \quad (8)$$

$$\mathbf{X}_T = \text{Softmax}\left(\frac{\mathbf{Q}_T (\mathbf{K}_T)^T}{\sqrt{d_h}}\right) \mathbf{V}_T \quad (9)$$

其中, $*$ 表示卷积运算,卷积核大小为 3×1 , Φ_Q 和 Φ_K 是卷积核参数, $\mathbf{W}_T \in \mathbf{R}^{d \times d_h}$ 是可学习的参数矩阵, $\mathbf{X}_T \in \mathbf{R}^{N \times T \times d_h}$ 是单个时间感知自注意力头的输出。

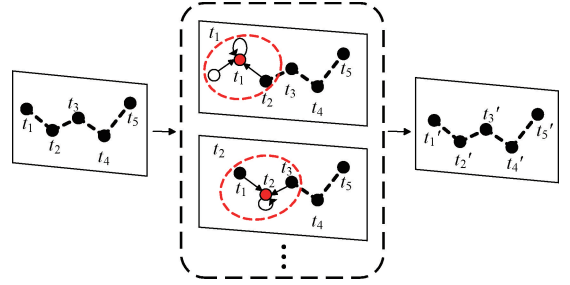


图 5 时域一维卷积

3) 多尺度扩散卷积模块

空间感知自注意力机制通过匹配具有相似交通模式的节点,实现全局空间相关性建模。在此基础上,模型使用多尺度扩散卷积加强对节点重要邻接区域的局部空间相关性建模,从全局和局部两个视角充分提取交通流数据的空间特征。扩散卷积^[12]将交通流与扩散过程结合起来,用 K 个有限步骤对交通流在路网中的扩散过程进行建模。 k 阶扩散用于模拟交通流沿路网进行 k 次扩散后对可达节点造成的影响,计算公式为:

$$\mathbf{X}_{\text{adj}}^k = (\mathbf{C})^k \mathbf{X}_E \mathbf{W}_{\text{adj}}^k \quad (10)$$

其中, $\mathbf{W}_{\text{adj}}^k \in \mathbf{R}^{d \times d_{\text{DC}}}$ 是可学习的参数矩阵, d_{DC} 是扩散卷积模块的特征维度, $\mathbf{C} = \mathbf{D}^{-1} \mathbf{A}$ 是扩散过程的转移矩阵, $(\mathbf{C})^k$ 是 $\mathbf{C}$ 的 k 次方,表示交通流沿路网进行 k 次扩散后可到达的节点。上述扩散过程是基于邻接矩阵所定义的显示空间相关性,即只有两个节点在邻接矩阵中相连时,才能进行空间信息的交互。为了挖掘交通路网中潜在的空间相关性,引入可学习的自适应矩阵,其计算方法如下:

$$\mathbf{A}_{\text{adp}} = \text{Softmax}(\text{ReLU}(\mathbf{E}_1 (\mathbf{E}_2)^T)) \quad (11)$$

其中, $\mathbf{E}_1$ 和 $\mathbf{E}_2 \in \mathbf{R}^{N \times 10}$ 是随机初始化的可学习参数矩阵, ReLU 激活函数用于消除自适应矩阵中的弱连接。基于自适应矩阵的 k 阶扩散过程定义为:

$$\mathbf{X}_{\text{adp}}^k = (\mathbf{C}_{\text{adp}})^k \mathbf{X}_E \mathbf{W}_{\text{adp}}^k \quad (12)$$

其中, $\mathbf{W}_{\text{adp}}^k \in \mathbf{R}^{d \times d_{\text{DC}}}$ 是可学习的参数矩阵。为了获取

来自多个邻域范围的空间特征的集成表示,将原始输入、显式扩散结果和隐式扩散结果拼接后输入全连接网络进行自适应融合,得到多尺度扩散卷积模块的输出:

$$\mathbf{X}_{\text{DC}} = \text{Concat}(\mathbf{X}_{\text{E}}, \mathbf{X}_{\text{adj}}^1, \dots, \mathbf{X}_{\text{adj}}^K, \mathbf{X}_{\text{adp}}^1, \dots, \mathbf{X}_{\text{adp}}^K) \mathbf{W}_{\text{DC}} \quad (13)$$

其中, Concat 表示矩阵拼接操作, $\mathbf{W}_{\text{DC}} \in \mathbf{R}^{(2K+1)d_{\text{DC}} \times d_{\text{DC}}}$ 是可学习的参数矩阵。

4) 多元特征融合

时空编码层集成了上述 3 种特征编码模块以并行捕获时空特征,各模块的输出经过拼接后输入全连接网络进行自适应融合,得到时空编码层的输出 $\mathbf{X}_{\text{ST}} \in \mathbf{R}^{T \times N \times d}$:

$$\mathbf{X}_{\text{ST}} = \text{Concat}(\mathbf{X}_{\text{DC}}, \mathbf{X}_{\text{S}}^1, \dots, \mathbf{X}_{\text{S}}^{h_{\text{S}}}, \mathbf{X}_{\text{T}}^1, \dots, \mathbf{X}_{\text{T}}^{h_{\text{T}}}) \mathbf{W}_{\text{ST}} \quad (14)$$

其中, h_{S} 和 h_{T} 是空间和时间感知自注意力模块的注意力头数, $\mathbf{W}_{\text{ST}} \in \mathbf{R}^{d \times d}$ 是可学习的参数矩阵。

2.3 输出层

输出层首先通过跳过连接聚合每个编码器捕获的时空特征,得到 $\mathbf{X}_{\text{out}} \in \mathbf{R}^{T \times N \times d_{\text{sk}}}$:

$$\mathbf{X}_{\text{out}} = \mathbf{X}^1 \mathbf{W}_{\text{sk}}^1 + \dots + \mathbf{X}^l \mathbf{W}_{\text{sk}}^l \quad (15)$$

其中, $\mathbf{X}^l \in \mathbf{R}^{T \times N \times d}$ 是第 l 个编码器的输出, $\mathbf{W}_{\text{sk}}^l \in \mathbf{R}^{d \times d_{\text{sk}}}$ 是可学习的参数矩阵。然后,使用双层全连接网络进行预测,一次输出多步预测结果,其计算公式如下:

$$\hat{\mathbf{Y}} = \text{ReLU}(\text{ReLU}(\mathbf{X}_{\text{out}} \mathbf{W}_{\text{O}}^1) \mathbf{W}_{\text{O}}^2) \quad (16)$$

其中, $\mathbf{W}_{\text{O}}^1 \in \mathbf{R}^{d_{\text{sk}} \times 1}$ 和 $\mathbf{W}_{\text{O}}^2 \in \mathbf{R}^{T \times r}$ 是可学习的参数矩阵, $\hat{\mathbf{Y}} \in \mathbf{R}^{T \times N \times 1}$ 是模型最终输出的预测结果。

3 实验

3.1 数据集

本文的实验数据来源于美国加州交通运输局性能评估系统公开的交通流数据,其中 PeMS04 数据集^[26]采集自旧金山地区,PeMS08 数据集^[26]采集自圣贝纳迪诺地区,数据集详细信息如表 1 所示。按照 6:2:2 的比例将数据集划分为训练集、验证集和测试集。

表 1 数据集详细信息

数据集	节点数	边数	本间隔/ min	训练 集样 本数	验证 集样 本数	测试 集样 本数
PeMS04	307	340	5	10 181	3 394	3 394
PeMS08	170	295	5	10 700	3 566	3 567

3.2 实验参数设置

实验基于深度学习框架 Pytorch 开发,使用 RTX3070 GPU 进行训练。预测目标为使用前 1 h($T=12$)的历史交通流预测未来 1 h($\tau=12$)的交通流。模型优化器选用

AdamW,学习率调度器选用 CosineLRScheduler,初始学习率设为 0.001,batch size 设为 16,模型迭代次数设为 100。模型结构中的其余参数如表 2 所示。

表 2 模型结构参数

参数	取值
时空编码块个数 L	6
扩散步骤 K	2
时空嵌入层特征维度 d	64
多尺度扩散卷积模块特征维度 d_{DC}	16
空间感知自注意力模块特征维度 d_{S}	32
时间感知自注意力模块特征维度 d_{T}	16
自注意力头特征维度 d_{h}	8
输出层跳过连接特征维度 d_{sk}	256

3.3 评价指标及对比模型

实验选取 3 种指标对预测结果进行分析以评估模型性能,分别为平均绝对误差(mean absolute error,MAE)、均方根误差(root mean square error,RMSE)和平均绝对百分比误差(mean absolute percentage error,MAPE)。计算公式如下:

$$\text{MAE} = \frac{1}{m} \sum_{i=1}^m |Y_i - \hat{Y}_i| \quad (17)$$

$$\text{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^m (Y_i - \hat{Y}_i)^2} \quad (18)$$

$$\text{MAPE} = \frac{1}{m} \sum_{i=1}^m \frac{|Y_i - \hat{Y}_i|}{Y_i} \times 100\% \quad (19)$$

其中, Y_i 为真实观测值, $\hat{Y}_i$ 为预测值, m 为样本数量。为验证所提出的模型在交通流预测中的有效性,本文选取了以下两类主流的交通流预测模型进行对比分析:

第 1 类是基于时空图神经网络的模型:GWN^[12]、STSGCN^[13]、T-GCN^[14];

第 2 类是基于 Transformer 的模型:GMAN^[18]、PDFormer^[23]、RPConvformer^[25]、ASTGNN^[26]、S²TAT^[27]。

3.4 模型训练

交通流数据由布置在路网中的大量传感器进行采集,受传感器故障的影响,数据中存在部分缺失值与异常值。因此,模型使用 *Huber Loss* 作为损失函数。该函数在误差较大时使用绝对误差,能够增强模型对异常值及噪声的鲁棒性;在误差较小时使用连续可导的平方误差,有助于加快损失函数的收敛。其计算公式如下:

$$\text{Huber Loss} = \begin{cases} \frac{1}{m} \sum_{i=1}^m \frac{1}{2} (Y_i - \hat{Y}_i)^2, & |Y_i - \hat{Y}_i| \leq \delta \\ \frac{1}{m} \sum_{i=1}^m (\delta |Y_i - \hat{Y}_i| - \frac{1}{2} \delta^2), & \text{其他} \end{cases} \quad (20)$$

其中, δ 是控制损失函数切换的阈值,本文设置为 2。

为验证 Huber Loss 对于解决均方误差(mean square error, MSE)对异常值过度敏感问题的有效性,本文分别使用 Huber Loss 和 MSE 作为损失函数对模型进行训练,以对比两种损失函数的收敛性能。MSE 的计算公式如下:

$$MSE = \frac{1}{m} \sum_{i=1}^m (Y_i - \hat{Y}_i)^2 \tag{21}$$

图 6 给出了使用不同损失函数进行训练时,训练集和验证集 MAE 在模型训练过程中的下降曲线。可以看出,在训练集和验证集上,Huber Loss 的收敛性能均优于 MSE,且在验证集上具有更大的优势。这是因为采用 MSE 作为损失函数时,模型过度拟合数据中的异常值及噪声,导致其在验证集上的泛化能力下降。

3.5 实验结果及分析

本文所提出的 STAFormer 与对比模型的实验结果如表 3 所示。STAFormer 在两个数据集上的误差指标均优

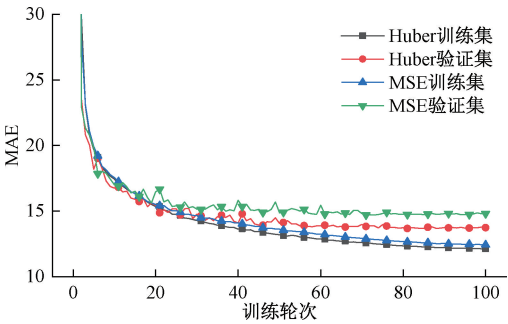


图 6 模型训练过程中训练集和验证集的误差下降曲线

于所有对比模型,与预测精度最高的 PDFormer 相比,在 PeMS04 数据集上,MAE、MAPE、RMSE 分别降低 2.1%、3.2% 和 1.0%。在 PeMS08 数据集上分别降低 1.9%、3.1% 和 1.0%。

表 3 不同模型性能对比

数据集	评价指标	T-GCN	STSGCN	GWNET	GMAN	S ² TAT	RPCnvformer	ASTGNN	PDFormer	STAFormer
PeMS04	MAE	20.91	21.67	19.09	19.14	19.24	19.09	18.81	18.44	18.05
	MAPE/%	14.80	14.37	13.34	13.19	12.83	12.63	12.47	12.44	12.04
	RMSE	33.38	34.39	31.73	31.60	30.97	30.91	30.76	29.96	29.66
PeMS08	MAE	16.26	17.89	15.27	15.52	15.57	15.17	14.93	13.85	13.58
	MAPE/%	11.86	11.78	9.99	11.08	9.97	9.80	9.73	9.30	9.01
	RMSE	25.65	27.59	24.48	24.31	24.62	24.38	24.41	23.18	22.94

为进一步对比分析各模型长短期预测性能的差异,本文选取 3 个最具代表性的 Transformer 基线模型,绘制 MAE 和 RMSE 误差在不同预测时间步长下的变化曲线,如图 7 所示。从图中可以看出,PDFormer 的短期预测能力优于 STAFormer,但 STAFormer 具有更加稳定的长期预测能力。ASTGNN 由于采用自回归预测,存在严重的误差累积,其长期预测能力显著下降。

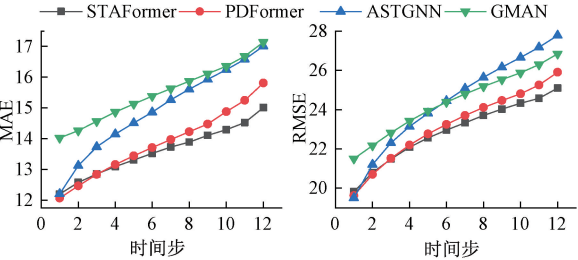


图 7 不同模型在 PeMS08 数据集上的多步预测性能对比

为直观的展示模型的预测效果,本文选取 PeMS08 数据集第 25 号节点,绘制其一天内 288 个时间步的交通流真实值与各模型预测值的对比图,如图 8 所示。并使用 MAE 量化各模型预测值与真实值的误差,计算结果表明 STAFormer 表现最佳,其 MAE 仅为 2.13,明显优于 PDFormer 和 ASTGNN 的 2.86 和 3.03。

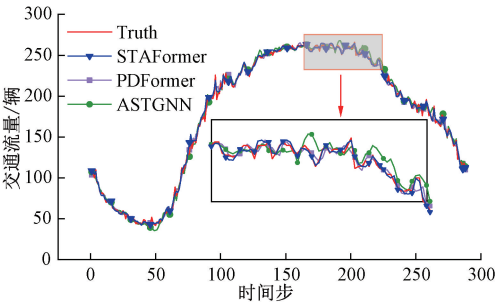


图 8 PeMS08 数据集 25 号节点预测值与真实值对比

3.6 消融实验

1) 时空编码层消融实验

为验证时空编码层各模块的有效性,本文在 STAFormer 的基础上替换或移除部分组件,形成如下变体模型:(1) w/o GCN: 去除多尺度扩散卷积;(2) w/o MDC: 使用普通图卷积代替多尺度扩散卷积;(3) w/o TASA: 使用传统时间自注意力机制代替时间感知自注意力机制;(4) w/o SASA: 使用传统空间自注意力机制代替空间感知自注意力机制。上述变体模型在 PeMS04 数据集上的消融实验结果如图 9 所示。可以看出,STAFormer 在 3 个指标上均优于所有变体模型。使用传统自注意力机制代替 SASA 和 TASA 后,MAE 分别上升 0.18 和

0.27,说明考虑交通流的空间异质性及时间趋势性,能够有效提高模型的预测精度。将多尺度扩散卷积替换为普通图卷积后,MAE 上升 0.86,说明在多个尺度上执行扩散卷积后进行特征融合,有助于模型获取更丰富的局部空间信息。完全去除多尺度扩散卷积模块后,模型的预测精度大幅下降,表明本文所采用的全局和局部融合空间建模方法能够有效提升模型预测精度。消融实验结果表明本文所提出的模型结构设计合理,各模块的改进均对预测性能有独立贡献。

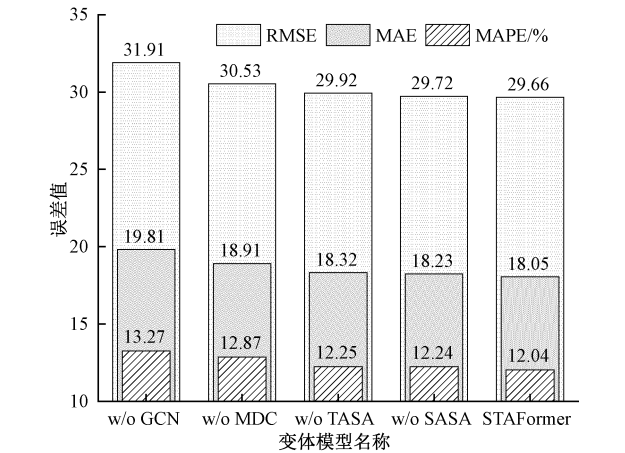


图 9 时空编码层消融实验

2)数据嵌入层消融实验

为验证时空嵌入方法的有效性,本文设计了以下几种变体模型:(1)w/o SPE:去除空间位置嵌入;(2)w/o TPE:去除时间位置嵌入;(3)w/o Period:去除时间周期嵌入。上述变体在 PeMS04 数据集上的消融实验结果如表 4 所示,去除任意一种时空嵌入,都会导致模型丢失必要的时空特征,进而影响预测精度。

表 4 数据嵌入层消融实验

模型名称	MAE	RMSE	MAPE/%
STAFormer	18.05	12.04	29.66
w/o SPE	18.21	12.27	29.93
w/o TPE	18.41	12.28	30.16
w/o Period	18.69	12.56	30.13

3.7 模型复杂度分析

本文在 PeMS08 数据集上对各模型的复杂度进行对比分析,结果如表 5 所示。GMAN 和 ASTGNN 均采用编码器-解码器架构,且 GMAN 在空间和时间维度都使用全局注意力,因此训练时间和参数量最高。ASTGNN 采用自回归预测方法,每次仅输出单个时间步的预测结果,导致推理效率较低。PDFormer 和 STAFormer 采用了相似的网络架构,使用双层全连接网络代替解码器生成预测结果,有效降低了模型的计算复杂度。

表 5 不同模型在 PeMS08 数据集上的复杂度对比

模型名称	训练时间/ (s/epoch)	推理时间/ s	参数量	MAE
GMAN	283.1	21.8	9.0×10^5	15.50
ASTGNN	146.2	46.9	4.6×10^5	14.93
PDFormer	116.4	7.6	5.3×10^5	13.85
STAFormer	121.6	8.3	5.3×10^5	13.58

4 结 论

本文通过对传统自注意力机制进行改进,提出了一种基于时空感知 Transformer(STAFormer)的交通流预测模型。通过将时域一维卷积和图卷积与自注意力机制相结合,时空感知自注意力模块能够充分考虑时间趋势性和空间异质性对时空关系的影响,捕获精确的全局时空特征。此外,模型采用一种全局-局部空间建模方法,在空间感知自注意力模块捕获全局空间特征的基础上,采用多尺度扩散卷积模块强化对节点重要邻接区域的局部空间特征的提取。最后,将捕获的多元时空特征进行自适应融合,生成预测结果。两个数据集上的实验结果表明,STAFormer 优于所有对比模型,取得了最好的预测效果。与最优的基线模型 PDFormer 相比,STAFormer 具有更加稳定的长期预测性能,在模型复杂度相似的情况下提高了预测精度。综上所述,本文所提出的模型能够进行实时、准确的交通流量预测,为交通管理部门提供可靠的数据支持。

参考文献

[1] YIN X, WU G, WEI J, et al. Deep learning on traffic prediction: Methods, analysis, and future directions[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(6): 4927-4943.

[2] 姚俊峰,何瑞,史童童,等. 基于机器学习的交通流预测方法综述[J]. 交通运输工程学报, 2023, 23(3): 44-67.

[3] XU D, WANG Y, JIA L, et al. Real-time road traffic state prediction based on ARIMA and Kalman filter[J]. Frontiers of Information Technology & Electronic Engineering, 2017, 18: 287-302.

[4] 屈展,赵鑫,屈伸. 引入果蝇优化算法的最小二乘支持向量机交通流量预测[J]. 电子测量技术, 2018, 41(16): 18-22.

[5] 王逸文,王维莉,陈怡霏,等. 基于奇异谱分析的长时交通流混合预测模型[J]. 电子测量与仪器学报, 2022, 36(11): 98-106.

[6] CAI P, WANG Y, LU G, et al. A spatiotemporal correlative k-nearest neighbor model for short-term traffic multistep forecasting [J]. Transportation Research Part C: Emerging Technologies, 2016, 62: 21-34.

- [7] YANG B, SUN S, LI J, et al. Traffic flow prediction using LSTM with feature enhancement [J]. Neurocomputing, 2019, 332: 320-327.
- [8] SUN P, BOUKERCHE A, TAO Y. SSGRU: A novel hybrid stacked GRU-based traffic volume prediction approach in a road network[J]. Computer Communications, 2020, 160: 502-511.
- [9] 马永杰, 程时升, 马芸婷, 等. 卷积神经网络及其在智能交通系统中的应用综述[J]. 交通运输工程学报, 2021, 21(4): 48-71.
- [10] ZHANG J, ZHENG Y, QI D, et al. Predicting citywide crowd flows using deep spatio-temporal residual networks[J]. Artificial Intelligence, 2018, 259: 147-166.
- [11] JIANG W, LUO J. Graph neural network for traffic forecasting: A survey [J]. Expert Systems with Applications, 2022, 207: 117921.
- [12] WU Z, PAN S, LONG G, et al. Graph wavenet for deep spatial-temporal graph modeling[C]. Proceedings of the 28th International Joint Conference on Artificial Intelligence, 2019: 1907-1913.
- [13] SONG C, LIN Y, GUO S, et al. Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting [C]. Proceedings of the 34th AAAI Conference on Artificial Intelligence, 2020, 34(1): 914-921.
- [14] 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, 2020, 21(9): 3848-3858.
- [15] 殷礼胜, 吴洋洋. 基于改进 VMD-GAT-GRU 的交通流量组合预测模型[J]. 电子测量与仪器学报, 2022, 36(7): 62-72.
- [16] 谷振宇, 陈聪, 郑家佳, 等. 考虑时空相似性的动态图卷积神经网络交通流预测[J]. 控制与决策, 2023, 38(12): 3399-3408.
- [17] 曹阳, 朱镕琦, 沈琴琴, 等. 基于时域图卷积神经网络的交通流预测模型[J]. 计算机工程与设计, 2023, 44(12): 3700-3706.
- [18] ZHENG C, FAN X, WANG C, et al. GMAN: A graph multi-attention network for traffic prediction [C]. Proceedings of the 34th AAAI Conference on Artificial Intelligence, 2020, 34(1): 1234-1241.
- [19] 张安勤, 秦添. 基于多通道时空编码器的交通流量预测模型[J]. 计算机应用研究, 2024, 41(1): 83-87.
- [20] SUN W, CHENG R, JIAO Y, et al. Transformer network with decoupled spatial-temporal embedding for traffic flow forecasting[J]. Applied Intelligence, 2023, 53(24): 30148-30168.
- [21] YE X, FANG S, SUN F, et al. Meta graph transformer: A novel framework for spatial-temporal traffic prediction[J]. Neurocomputing, 2022, 491: 544-563.
- [22] 李昕光, 王琬, 曲大义. 基于特征选择的时空融合交通流预测[J]. 电子测量技术, 2023, 46(22): 87-93.
- [23] JIANG J, HAN C, ZHAO W, et al. PDFormer: Propagation delay-aware dynamic long-range transformer for traffic flow prediction [C]. Proceedings of the 37th AAAI Conference on Artificial Intelligence, 2023, 37(4): 4365-4373.
- [24] 刘起东, 刘超越, 邱紫鑫, 等. 基于时间感知 Transformer 的交通流预测方法[J]. 计算机科学, 2023, 50(11): 88-96.
- [25] WEN Y, XU P, LI Z, et al. RPConvformer: A novel transformer-based deep neural networks for traffic flow prediction [J]. Expert Systems with Applications, 2023, 218: 119587.
- [26] GUO S, LIN Y, WAN H, et al. Learning dynamics and heterogeneity of spatial-temporal graph data for traffic forecasting [J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 34(11): 5415-5428.
- [27] WANG T, CHEN J, LYU J, et al. Synchronous spatiotemporal graph transformer: A new framework for traffic data prediction[J]. IEEE Transactions on Neural Networks and Learning Systems, 2023, 34(12): 10589-10599.
- [28] GENG Z, XU J, WU R, et al. STGAFormer: Spatial-temporal gated attention transformer based graph neural network for traffic flow forecasting[J]. Information Fusion, 2024, 105: 102228.

作者简介

鲁思源, 硕士研究生, 主要研究方向为智能交通。

E-mail: lsy@stmail.ntu.edu.cn

沈琴琴(通信作者), 博士, 副教授, 硕士生导师, 主要研究方向为智能交通。

E-mail: shenqq@ntu.edu.cn

包银鑫, 博士研究生, 主要研究方向为智能交通。

E-mail: 835761833@qq.com

高锐锋, 博士, 副教授, 硕士生导师, 主要研究方向为无线通信系统、智能信息处理。

E-mail: grf@ntu.edu.cn

施佳, 博士, 教授, 博士生导师, 主要研究方向为智能信息处理。

E-mail: sq@ntu.edu.cn