

# 基于预训练时空自注意力大模型的交通流量预测

童旭东,周强,顾晶晶,史国梁,崔鸿飞

(南京航空航天大学 计算机科学与技术学院,南京 211106)

E-mail: gujingjing@nuaa.edu.cn

**摘要:**精准的交通流量预测是智能交通系统和智慧城市建设的核心。交通流量数据具有复杂的时空依赖性,呈现出多粒度的特征变化和动态演化规律。现有方法在捕获多粒度时空特征方面存在局限,难以充分挖掘丰富时空关联。预训练大语言模型(Pre-trained Large Language Model, PLM)在特征表示学习方面展现出巨大潜力,但由于其预训练数据主要集中在自然语言领域,与交通流量数据存在显著的领域差异,限制了其直接应用效果。为解决上述问题,本文提出了一种基于预训练时空自注意力模型的交通流量预测框架(Pre-trained Spatio-Temporal Attention Model for Traffic Flow Prediction, PSTAM)。首先通过创新的双路径激活机制解决领域差异问题,实现特征对齐与融合;其次利用预训练策略对多粒度时空特征进行深度建模,增强对复杂时空依赖关系的捕获能力。实验结果表明,PSTAM在多个标准交通数据集上优于现有方法,为智能交通系统的实时决策提供了可靠支持。

**关键词:**交通流量预测;预训练大语言模型;双激活机制;多粒度;领域适配

中图分类号: TP183

文献标识码: A

文章编号: 1000-1220(2026)05-1099-09

## Pre-trained Spatio-temporal Attention Model for Traffic Flow Prediction

TONG Xudong, ZHOU Qiang, GU Jingjing, SHI Guoliang, CUI Hongfei

(College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

**Abstract:** Precise traffic flow prediction is fundamental to the development of intelligent transportation systems and smart city initiatives. Traffic flow data is characterized by complex spatio-temporal dependencies, manifesting multi-scale feature variations and dynamic evolutionary patterns. Existing methodologies exhibit significant limitations in capturing these multi-scale spatio-temporal features, thus failing to adequately exploit the rich spatio-temporal correlations inherent in traffic data. Pre-trained Large Language Models (PLMs) have demonstrated considerable potential in feature representation learning; however, their direct application is substantially constrained by the domain disparity between their pre-training data, which is predominantly concentrated in natural language domains, and the distinctive characteristics of traffic flow data. To address these challenges, this paper proposes a Pre-trained Spatio-Temporal Attention Model for Traffic Flow Prediction (PSTAM). The framework incorporates two primary innovations: firstly, an innovative dual-pathway activation mechanism to resolve domain disparities, facilitating effective feature alignment and fusion; secondly, advanced pre-training strategies for comprehensive modeling of multi-scale spatio-temporal features, substantially enhancing the capacity to capture complex spatio-temporal dependencies. Experimental evaluations demonstrate that PSTAM consistently outperforms state-of-the-art methods across multiple benchmark traffic datasets, providing robust support for real-time decision-making processes in intelligent transportation systems.

**Keywords:** traffic flow prediction; pre-trained large language models; dual activation mechanism; multi-scale; domain adaptation

## 0 引言

在构建智慧城市的过程中,智能交通系统是不可或缺的关键环节,而精准的交通流量预测则是智能交通系统的核心技术之一。高质量的交通流量预测不仅可以为城市交通管理提供科学决策依据,还能优化公共资源分配、改善城市规划,提升城市运行效率和居民生活质量。然而,交通流量数据呈现出复杂的时空依赖关系和多粒度的特征变化,表现为不同时间尺度和空间尺度上的动态演化规律,为精准预测带来了巨

大挑战。

随着深度学习技术的发展,基于神经网络的交通流量预测方法取得了显著进展。其中,预训练大语言模型(Pre-trained Large Language Model, PLM)凭借其强大的特征提取能力和知识迁移能力,在时间序列预测领域展现出巨大潜力<sup>[19]</sup>,被引入交通流量预测任务中。PLM通过自监督学习在大规模数据上进行预训练,学习到了丰富的上下文表示和长距离依赖关系建模能力,这些特性对于捕获交通流量数据中的多粒度时空特征具有重要价值。然而,现有基于PLM的方法忽略了

一个关键问题:PLM 主要基于自然语言语料进行预训练,其表示空间与交通流量序列数据存在显著领域差异.这种领域鸿沟导致预训练知识无法直接有效迁移到交通预测任务中,限制了 PLM 在交通流量预测中的应用效果.

PLM 与交通流量数据之间的领域差异主要体现在数据分布特性、结构关系和模式特征等方面. PLM 处理的是离散符号序列,具有明确的语法结构和上下文关系;而交通流量数据则是连续值序列,呈现出周期性、趋势性和突发性等特征.这些差异使得直接应用 PLM 进行交通流量预测时,难以充分利用预训练获得的知识,甚至可能引入负面干扰.现有研究主要关注如何微调 PLM 以适应交通预测任务,但较少探讨如何解决底层表示空间的领域差异问题,这成为制约 PLM 在交通预测领域发挥潜力的关键瓶颈.

为了解决上述问题,本文提出了一种基于预训练时空自注意力模型的交通流量预测框架 PSTAM(Pre-trained Spatio-Temporal Attention Model for Traffic Flow Prediction).该框架设计了两个核心模块:双激活领域桥接模块(Dual Activation Domain Bridge, DADB)和预训练多粒度时空 Transformer 模块(Pretrained Multi-scale Spatio-Temporal Transformer, PM-STT). DADB 模块通过双路径激活机制实现交通序列数据与预训练模型表示空间的高效对齐与融合,有效克服了领域鸿沟;PMSTT 模块则基于预训练策略,结合时空自注意力机制,增强了对多粒度时空特征的捕获能力.这两个模块通过串联设计,形成了一个端到端的解决方案,不仅有效解决了领域适配问题,还提高了交通流量预测的准确性和鲁棒性.

综上所述,本文的主要贡献如下:

1) 提出了双激活领域桥接模块,通过双路径激活机制解决预训练模型与交通数据的领域差异问题,实现了高效特征对齐与融合,为预训练模型在交通预测任务中的应用提供了新思路.

2) 设计了预训练多粒度时空 Transformer 模块,基于预训练策略对多粒度时空特征进行深度建模,显著增强了对复杂时空依赖关系的捕获能力.

3) 在多个真实交通数据集上进行了广泛实验,验证了所提出模型的有效性,并通过消融实验分析了各关键组件的贡献,为交通流量预测领域提供了新的研究方向.

## 1 相关工作

### 1.1 交通流量预测

交通流量预测需同时处理时间和空间维度的复杂依赖关系.随着深度学习技术的发展,基于神经网络的方法逐渐取代了传统统计模型如 ARIMA<sup>[1]</sup>和 VAR<sup>[2]</sup>.LSTM<sup>[3]</sup> 凭借其捕获长期时序依赖的能力在早期研究中表现出色,而 TCN<sup>[4]</sup> 通过因果卷积有效处理长序列预测问题.然而,这些模型主要关注时间维度,忽略了空间信息的重要性.

为解决空间依赖建模问题,图神经网络(Graph Neural Network, GNN)被广泛应用于交通预测领域.DCRNN<sup>[5]</sup> 通过图上的双向随机游走捕获空间依赖,并结合编码器-解码器架构处理时间依赖;STGCN<sup>[6]</sup> 将图卷积神经网络(Graph Convolutional Network, GCN)与时间卷积结合,有效提取节点特征

与时空相关性.ASTGCN<sup>[7]</sup> 引入时空注意力机制处理动态时空相关性,而 GWNet<sup>[8]</sup> 通过自适应依赖矩阵增强对隐藏空间依赖的捕获能力.STSGCN<sup>[9]</sup> 针对远程时空网络的多尺度性设计了同步图卷积结构,STFGNN<sup>[10]</sup> 融合空间和时间图学习隐藏的时空依赖关系并集成门控卷积模块.

随着研究深入,更复杂的架构不断涌现.STGODE<sup>[11]</sup> 基于张量的常微分方程与图神经网络结合,提供了连续时间域上的建模能力;DSTAGNN<sup>[12]</sup> 设计了时空感知距离构建动态图,结合多项式和门控卷积捕获时空依赖关系;AST-GNN<sup>[13]</sup> 使用趋势感知自注意力模块捕获动态的时空信息;AGCRN<sup>[14]</sup> 提出节点自适应参数学习与数据自适应图生成模块,动态捕获交通流量数据的时空相关性;Z-GCNETs<sup>[15]</sup> 结合时间条件拓扑信息捕获时间序列中的重要拓扑特征.PD-Former<sup>[16]</sup> 作为图神经网络模型,通过时空注意力机制有效处理长短期动态时空依赖性;STAEformer<sup>[17]</sup> 则基于 Transformer 架构设计时空自编码器,利用自注意力机制捕获时空数据的长期依赖关系.

尽管这些方法取得了显著进展,但在处理交通数据多粒度特征方面仍存在局限.交通流量数据在不同时间尺度和空间尺度上呈现复杂的依赖关系和变化规律.传统模型往往采用单一结构设计,难以同时有效捕获这些多粒度时空特征,特别是在不同时空尺度上的依赖关系.此外,现有方法较少关注数据领域特性与模型架构的匹配问题,导致模型对交通流量数据特有的时空模式缺乏足够的适应性.

### 1.2 预训练大模型在交通流量预测的应用

预训练大语言模型(PLM) 凭借其强大特征提取和泛化能力,近年来在交通预测领域取得显著进展. PLM 在交通流量预测中的应用主要分为 3 类:基于提示工程(Prompt Engineering)、基于微调(Fine-tuning)和基于预训练的方法.

基于提示工程的方法利用 PLM 已有的知识,无需额外训练.PromptCast<sup>[18]</sup> 等研究尝试通过将时间序列数据转换为文本提示,引导 PLM 直接对交通流量进行预测.这类方法实施简单,计算成本低,但预测精度往往受限于提示设计质量,且难以充分适应交通数据的特殊结构和时空依赖性.

基于微调的方法通过调整 PLM 参数使其适应交通预测任务,在保留模型泛化能力的同时提高预测精度.OFA<sup>[19]</sup> 率先将 PLM 应用于时间序列预测,引入注意力冻结策略保留模型对长序列依赖的捕获能力.ST-LLM<sup>[20]</sup> 进一步探索了 PLM 在交通预测中的应用,提出了嵌入融合技术提取交通流量的时空信息,并采用部分注意力冻结方法优化微调过程.GAT-GPT<sup>[21]</sup> 等研究则结合图注意力网络增强模型对空间依赖的处理能力,但其微调策略仍局限于简单的注意力冻结,未能充分发挥 PLM 的潜力.

基于预训练的方法针对时序数据特点从零构建 PLM.Timer<sup>[22]</sup> 等研究采用类似 GPT 的架构,在大规模时间序列数据上预训练,获得了强大的时间建模能力.STD-MAE<sup>[23]</sup> 提出了基于掩码自编码器的自监督框架,通过时空维度上的重构任务增强模型对交通数据的表征能力.这类方法虽然表现出色,但训练成本极高,需要海量数据支持,在实际应用中面临挑战.

在微调策略方面,现有研究面临多种局限.注意力冻结方

法虽能保留 PLM 的时序建模能力,但难以适应交通数据的多维空间特性;部分注意力冻结虽增强了模型适应性,但易导致预训练知识部分丢失;传统全参数微调则计算成本高且易过拟合。面向上述问题,低秩适应(Low-Rank Adaptation, LoRA)<sup>[24]</sup>作为一种新兴微调策略,通过在原始权重矩阵旁添加低秩分解矩阵实现参数高效的任务适应,同时保留预训练知识,为解决上述问题提供了新的思路。

此外,PLM 与交通预测任务之间存在显著的领域差异也是一个亟待解决的问题。PLM 主要基于自然语言语料预训练,其表示空间与交通流量序列数据存在本质区别。这种领域鸿沟限制了预训练知识向交通预测任务的有效迁移,成为制约 PLM 在交通流量预测中应用效果的关键因素。现有研究往往专注于微调策略的优化,而忽视了解决领域差异的重要性,这使得即使采用先进的微调方法,模型性能仍无法达到理想水平。

## 2 问题定义

交通流量预测是时空序列预测的一种应用,基于历史交通数据以及交通节点网络的空间结构,对未来一段时间内交通流量进行预测。

### 2.1 交通流量数据表示

在本文中,交通流量数据被表示为  $x_T \in \mathbb{R}^{T \times N \times F}$ ,其中  $T$  表示时间步长度, $N$  表示空间节点数量, $F$  表示数据的特征维度。

### 2.2 交通流量预测任务

交通流量预测任务表示为,给定历史交通数据  $x_T = [x^{t-T+1}, x^{t-T+1}, \dots, x^t]$ ,模型通过学习一个映射函数  $f(\cdot; \theta)$ ,将历史数据映射到未来  $T$  个时间步的预测数据  $\hat{x}_T = [x^{t+1}, x^{t+2}, \dots, x^{t+T}]$ ,具体满足如下关系式:

$$f(x_T; \theta) = \hat{x}_T \quad (1)$$

## 3 基于预训练时空自注意力大模型的交通流量预测

本文提出了一种基于预训练时空自注意力模型的交通流量预测框架 PSTAM (Pre-trained Spatio-Temporal Attention Model for Traffic Flow Prediction),其模型架构如图 1 所示。PSTAM 通过创新的双激活机制和预训练策略,在多个层面实现了对交通流量数据特征的深度提取融合并进行建模,从而提供更准确的未来交通流量预测。

交通流量预测任务中的时空特征存在显著的多粒度特性,即数据序列中同时包含不同时间尺度的变化规律和不同空间尺度的依赖关系。传统的单一特征提取方法往往难以同时有效捕获这些不同尺度的特征表达。更具挑战性的是,交通流量数据在不同时空维度上表现出的复杂特征之间存在着线性和非线性的交互关系,这进一步增加了建模的难度。此外,预训练大语言模型(PLM)虽具有强大的特征表示能力,但其预训练数据主要集中在自然语言领域,与交通流量序列数据存在显著的领域差异,限制了其直接应用效果。

针对上述挑战,PSTAM 框架设计了两个关键模块:双激活领域桥接模块(Dual Activation Domain Bridge, DADB)和预训练多粒度时空 Transformer 模块(Pretrained Multi-scale Spatio-Temporal Transformer, PMSTT)。DADB 模块创新地利用 ReLU 和 GELU 两种不同特性的激活函数来分别捕获时空数据中的线性和非线性特征,并通过自适应机制实现特征的动态融合,有效解决预训练模型与交通数据之间的领域差异问题。PMSTT 模块则巧妙结合了预训练模型的强大特征提取能力和专门设计的时空自注意力机制,通过对多粒度时空特征的深度建模,显著增强了对复杂时空依赖关系的捕获能力。这种串联设计不仅实现了对交通流量数据多粒度特征的有效提取,还解决了预训练模型在交通领域应用的适配问题。模型具体流程如算法 1 所示。

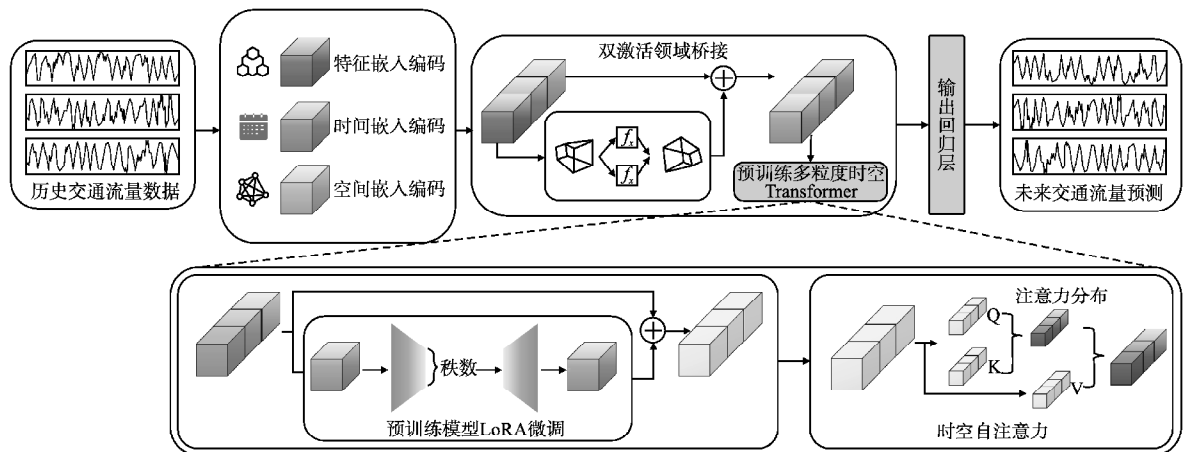


图 1 基于预训练时空自注意力大模型的交通流量预测框架图

Fig. 1 Framework for pre-trained spatio-temporal attention model for traffic flow prediction

### 3.1 双激活领域桥接模块

交通流量数据在不同时空维度上表现出复杂的多粒度特征,而预训练大语言模型(PLM)的表示空间主要基于自然语

言语料构建,二者之间存在显著的领域差异。传统的单一激活函数方法难以同时有效地捕获线性、非线性和异常模式特征,因此无法很好地解决这种领域鸿沟。为此,本文设计了双激活

领域桥接模块(DADB),通过并行的双分支结构和动态门控机制来实现特征的高效对齐与融合。DADB模块的核心创新在于其双分支并行架构。ReLU分支凭借其优秀的稀疏性和梯度传播特性,善于捕获数据中的线性关系和简单非线性模式;而GELU分支通过引入概率分布的思想,能够更好地处理复杂的非线性特征和异常模式,这种互补性使得模块能够全面地提取交通流量数据中的多样化特征,同时更好地适配预训练模型的表示特性。对于输入特征 $X_T$ ,两个分支的处理过程如下:

$$Branch_{ReLU} = (\text{ReLU}(\text{BN}(\text{Conv2d}(X_T))))^2 \quad (2)$$

$$Branch_{GELU} = (\text{GELU}(\text{BN}(\text{Conv2d}(X_T))))^2 \quad (3)$$

其中Conv2d为 $1 \times 1$ 二维卷积,BN为批归一化层。

值得注意的是,中间层的通道设计为输出通道的两倍,这种设置是为了通过扩展特征空间以增强模块的表达能力,同时每个转换操作后配备dropout层来避免过拟合,提高模型的泛化性能。

为了充分利用双分支结构的优势,模块设计了基于 $1 \times 1$ 二维卷积的动态门控机制:

$$W_B = \text{Softmax}(\text{Conv2d}(X_T)) \quad (4)$$

$$\text{output} = w_r \cdot Branch_{ReLU} + w_g \cdot Branch_{GELU} \quad (5)$$

其中 $W_B$ 是可学习的权重矩阵。

动态门控融合机制通过Softmax确保两个分支权重 $w_r$ 和 $w_g$ 之和为1。这种自适应融合机制能够根据输入特征的性质动态调整两个分支的重要性:当数据表现出明显的线性特征时,模块会自动增加ReLU分支的权重;当出现复杂的非线性特征和异常模式时,则会强化GELU分支的影响。这一特性对于实现交通数据与预训练模型表示空间的高效对齐至关重要。

为了适用于PSTAM框架中的不同应用场景,DADB模块设计了3种不同的变体:

1)基础版(DADB):采用两层卷积结构,在保持计算效率的同时提供良好的特征提取能力,有效平衡性能和性能开销。

2)深层版(DADB2):采用四层卷积网络构建更深的特征提取层次。通过逐层扩展和收缩通道数,实现了更强的特征表达能力。

3)线性版(DADBL):将两层卷积结构替换为单层线性结构,并使用层归一化替代批归一化。在保持双激活机制优势的同时,适用于特征投影等场景。

### 3.2 时空特征多重嵌入模块

对交通流量数据进行建模的关键在于对其时间维度和空间维度进行有效的特征表示。为了解决预训练模型主要基于自然语言语料的问题,我们需要将交通流量数据转换为适合预训练模型处理的特征表示。本文设计了多重嵌入模块,通过时间嵌入、节点嵌入和特征嵌入3个部分来捕获交通流量数据的多粒度时空特征。

时间嵌入模块采用了双重时间特征表示。通过两个可学习的嵌入矩阵分别编码日期和星期信息,这种设计能够捕获交通流量在不同时间尺度上的周期性变化:

$$Emb_{Temporal} = W_{Day} + W_{Week} \in \mathbb{R}^{B \times N \times F \times T} \quad (6)$$

空间嵌入采用可学习的方式,为每个交通节点分配独特

的嵌入向量,这使得模型能够学习不同监测站点的固有特性和空间关系:

$$Emb_{Spatial} = w_s \times W_s + b_s \in \mathbb{R}^{B \times N \times F \times T} \quad (7)$$

其中 $W_s$ 为节点嵌入矩阵, $w_s$ 和 $b_s$ 为可学习参数。

输入特征 $X_T$ 通过线性变换映射到统一嵌入空间,确保不同来源的特征可以在相同的表示空间中进行有效融合:

$$Emb_{Feature} = w_f \times X + b_f \in \mathbb{R}^{B \times N \times F \times T} \quad (8)$$

其中 $w_f$ 和 $b_f$ 为可学习参数。

最后将3种嵌入特征进行Concat拼接,得到完整的时空特征表示:

$$Emb_{Fusion} = \text{Concat}(Emb_{Feature}, Temporal, Spatial) \quad (9)$$

为了进一步增强特征的表达能力并促进领域适配,本文使用DADB模块对拼接后的特征进行增强处理,并通过残差连接保留原始特征信息:

$$X_{Feature} = Emb_{Fusion} + \text{DADB}(Emb_{Fusion}) \quad (10)$$

这种多重嵌入和增强策略不仅能够充分捕获交通流量数据的多粒度时空特征,还有效地将交通数据转换为更适合预训练模型处理的表示形式,为后续模块提供了高质量的输入特征,从而缓解了预训练模型与交通数据的领域差异问题。

### 3.3 预训练多粒度时空Transformer模块

交通流量预测任务中,准确捕获时间序列的长短期依赖关系至关重要。为此,本文设计了专门的时空自注意力机制,通过可学习的时空编码和差异化的特征投影策略,增强了模型对时空依赖关系的建模能力。与此同时,OFA模型<sup>[19]</sup>证明了预训练大语言模型(PLM)在时序预测的效果,后续的一系列研究<sup>[20-23]</sup>也验证了PLM在交通流量预测领域的有效性。基于这些发现,本文提出将专门设计的时空自注意力机制与PLM相结合,并采用注意力冻结方法和LoRA<sup>[24]</sup>策略进行高效微调,形成了预训练多粒度时空Transformer模块(PM-STT)。该模块通过预训练知识迁移和增强的时空建模能力,实现了对交通流量多粒度时空特征的深入表达。

对于预训练特征提取部分,首先使用DADBL进行维度对齐,将特征 $X_{Feature}$ 映射到PLM的维度空间:

$$P_{ToPLM} = \text{DADBL}(X_{Feature}) \in \mathbb{R}^{B \times N \times F_d \times T} \quad (11)$$

其中 $F_d$ 为预训练大模型的特征维度。

本文选择GPT-2<sup>[25]</sup>作为预训练大模型PLM。PLM采用了L层Transformer<sup>[26]</sup>结构,每层的处理过程为:

$$H_i^{mha} = \text{LN}(H_{i-1} + \text{MHA}(H_{i-1})) \quad (12)$$

$$H_i = \text{LN}(H_i^{mha} + \text{FFN}(H_i^{mha})) \quad (13)$$

其中 $H_i$ 为第 $i$ 层的隐藏状态,MHA为多头自注意力操作,FFN为前馈神经网络,LN为层归一化。

为了在保持预训练知识的同时实现高效微调,我们引入了LoRA策略,通过在注意力层中添加低秩矩阵来实现参数高效的任务适应。

具体而言,对于每一层,其参数更新策略如下:

前层L-U,仅更新LoRA参数和层归一化参数;后U层,允许包括LoRA参数在内的注意力机制参数更新,但保持前馈网络层冻结。

这种渐进式的微调策略既保留了预训练知识,又实现了与交通流量预测任务的高效对齐。除此之外,为了进行完整的跨时间步特征融合,在PLM特征提取完成后,我们将输入 $X_T$

的原始嵌入  $X_{Feature}$  作为残差项与投影回原始维度的  $X_{FromPLM}$  相加并输入 TSA 模块:

$$X_{FromPLM} = \text{DADBL}(P_{FromPLM}) \quad (14)$$

$$X_{output} = \text{TSA}(Emb_{TSA}, TSA_{Q,K,V}) \quad (15)$$

通过引入可学习的时空编码  $Emb_{TSA}$  并使用 DADB 模块对特征进行查询、键、值的紧凑投影, TSA 模块能够感知时间步和空间节点间的相对位置关系,挖掘深层次的多粒度时空特征:

$$TSA_{Q,K,V} = \text{DADB}(X_{FromPLM} + X_{Feature}) \quad (16)$$

$$Emb_{TSA} \in \mathbb{R}^{B \times N \times F \times T} \quad (17)$$

PMSTT 模块的设计有效解决了两个关键问题:一方面,通过 LoRA 微调策略和残差连接机制,保留了 PLM 的预训练知识,同时实现了与交通领域数据的有效对齐;另一方面,通过专门设计的时空自注意力机制,增强了对多粒度时空依赖关系的建模能力.这种创新设计使得模型能够充分挖掘交通流量数据在不同时间尺度和空间尺度上的复杂依赖关系,提高预测准确性.

### 3.4 输出回归层

为了将提取的特征映射到目标预测空间,本文设计了专门的输出预测层.该层采用了两阶段的特征转换策略,首先通过 DADB 模块增强特征表达,然后使用卷积层生成最终的预测结果.具体而言,预测过程表示为:

$$X_{enhanced} = \text{DADB}(X_{output}) \quad (18)$$

$$\hat{x}_T = \text{Conv } 2d(X_{enhanced}) \quad (19)$$

DADB 模块通过双分支的协同作用,增强了特征的表达能力.这种设计使得模型能够同时捕获线性和非线性模式,为后续的预测提供更丰富的特征表示.同时面对交通流量数据的非平稳性, DADB 的动态门控机制可以表现出较好的泛化性能.

最终输出为  $\hat{x}_T = [x^{t+1}, x^{t+2}, \dots, x^{t+T}] \in \mathbb{R}^{B \times T \times N \times F}$ , 即模型的交通流量预测结果.

模型采用 MAE 作为 Loss 损失函数进行反向传播梯度更新,对模型参数进行优化.

#### 算法 1. PSTAM 算法

输入:历史交通流量数据

$$X_T = [x^{t-T+1}, x^{t-T+1}, \dots, x^t] \in \mathbb{R}^{B \times T \times N \times F}$$

输出:未来交通流量预测

$$\hat{x}_T = [x^{t+1}, x^{t+2}, \dots, x^{t+T}] \in \mathbb{R}^{B \times T \times N \times F}$$

For 每个 epoch Do

  打乱训练数据

  For 每个批次  $X_T$  在训练数据中 Do

$Emb_{Fusion} =$  嵌入编码融合( $X_T$ )

$X_{Feature} = \text{DADB}(Emb_{Fusion})$

$P_{ToPLM} = \text{DADBL}(X_{Feature})$

    For PLM 的每个 layer Do

$P_{ToPLM} = \text{PLM\_LoRA}(P_{ToPLM})$

    End for

$P_{FromPLM} = P_{ToPLM}$

$X_{FromPLM} = \text{DADBL}(P_{FromPLM})$

$TSA_{Q,K,V} = \text{DADB}(X_{FromPLM} + X_{Feature})$

$Emb_{TSA} =$  随机初始化

$X_{output} = \text{TSA}(Emb_{TSA}, TSA_{Q,K,V})$

$X_{enhanced} = \text{DADB}(X_{output})$

$\hat{x}_T =$  输出回归( $X_{enhanced}$ )

  更新模型参数

  End for

End for

## 4 实验与分析

### 4.1 实验设置

本文实验基于 NVIDIA A100 显卡进行,显存大小为 80G,模型使用 Pytorch 框架搭建.

实验使用前 12 个时间步长预测后 12 个时间步长.初始学习率设置为 0.01,学习率会随着实验进行逐渐降低,批大小为 64.

实验使用平均绝对误差 MAE、均方根误差 RMSE、平均绝对百分比误差 MAPE 作为评价指标.各指标的计算公式如下所示:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (20)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (21)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100 \quad (22)$$

实验采用 3 种指标,可以衡量模型预测偏差的多个方面, MAE 直观反映平均误差大小, RMSE 关注较大的误差对整体的影响, MAPE 则提供误差相对实际值的比例,全方位地展现模型的交通流量预测效果.在这些性能评价指标之中,本实验采用 MAE 作为指导模型进行更新时的 Loss 损失函数.

### 4.2 数据集

为了证明本文提出的 PSTAM 模型的性能,本文采用 4 个在真实世界上采集的交通流量数据集进行实验.数据集详情描述如表 1 所示.

表 1 数据集详情

Table 1 Details of datasets

数据集	PEMS03	PEMS04	PEMS07	PEMS08
节点数量	358	307	883	170
时间步数	26208	16992	28224	17856
时间间隔	5min	5min	5min	5min
时间跨度	3 个月	2 个月	4 个月	2 个月

PEMS03:数据集采集 2018 年 9 月 1 日~11 月 30 日总计 3 个月内的洛杉矶地区高速公路上的车辆和速度数据,以 5 分钟为时间间隔,共包含 26208 个时间步,其中包含 358 个交通检测器作为空间节点.

PEMS04:数据集采集 2018 年 1 月 1 日~2 月 28 日总计 2 个月内的旧金山湾区的交通流量数据,以 5 分钟为时间间隔,共包含 16992 个时间步,其中包含 307 个传感器作为空间节点.

PEMS07:数据集采集 2017 年 5 月 1 日~8 月 31 日总计 4 个月内的加州第七区的交通数据,以 5 分钟为时间间隔,共包含 28224 个时间步,其中包含 883 个传感器作为空间节点.

PEMS08:数据集采集自 2016 年 7 月 1 日至 8 月 31 日总计 2 个月内的圣贝纳迪诺地区的交通数据,以 5 分钟为时间间隔,共包含 17856 个时间步,其中包含 170 个传感器作为空

间节点.

所有数据集均按照 6:2:2 的比例划分为训练集、验证集、测试集.

### 4.3 对比实验

为了对 PSTAM 模型的性能进行评估,本文采取了 20 个模型作为基线与本文模型进行比较:

1) 传统统计模型:包括经典的 ARIMA<sup>[1]</sup> 和 VAR<sup>[2]</sup>, 这些方法基于数学统计原理,分别用于单变量和多变量时间序列的预测,虽然模型简单但在某些场景下仍具有参考价值.

2) 深度学习基础模型:包括擅长捕获长期时序依赖关系的 LSTM<sup>[3]</sup> 和利用因果卷积处理长序列预测问题的 TCN<sup>[4]</sup>, 这些模型奠定了深度学习在时间序列预测中的基础,但只关注时间维度而忽略了空间信息.

3) 图神经网络模型:这类模型专注于时空依赖关系的联合建模,包括 DCRNN<sup>[5]</sup> 通过图上随机游走捕获空间依赖, STGCN<sup>[6]</sup> 将图卷积与时空卷积结合, GWNNet<sup>[8]</sup> 利用自适应依赖矩阵, STSGCN<sup>[9]</sup> 和 STFGNN<sup>[10]</sup> 分别针对多尺度性和隐藏时空依赖设计, STGODE<sup>[11]</sup> 结合常微分方程, DSTAGNN<sup>[12]</sup> 构建动态图, AGCRN<sup>[14]</sup> 提出自适应参数学习,以及 Z-GC-

NETs<sup>[15]</sup> 融合时间条件拓扑信息.

4) 基于 Transformer 的模型:包括 ASTGCN<sup>[7]</sup> 引入时空注意力机制, ASTGNN<sup>[13]</sup> 使用趋势感知自注意力, PDFormer<sup>[16]</sup> 处理长短期依赖性, STAEformer<sup>[17]</sup> 设计的时空自编码器,这些模型基于 Transformer 架构或自注意力机制,在捕获长距离时空依赖关系方面展现出优势.

5) 基于预训练大模型的方法:这类方法是当前研究热点,包括 OFA<sup>[19]</sup> 将预训练大模型应用于时间序列预测并使用注意力冻结, ST-LLM<sup>[20]</sup> 提出嵌入融合方法提取交通流量序列的时空信息,以及 STD-MAE<sup>[23]</sup> 采用自监督框架训练互为解耦的掩码自编码器,在时空维度上重构序列,这些方法探索了预训练大模型在交通流量预测中的潜力.

对比实验结果如表 2 所示,其中最优的实验结果用加粗标记,次优的实验结果用下划线标记.本文中的部分基线实验结果引用了 STD-MAE 中的数据.从实验结果可以观察到, PSTAM 模型在 PEMS03 数据集上均取得了最优的效果,在 PEMS04、PEMS07、PEMS083 个数据集上仅有 MAPE 参数略差于 STD-MAE 和 ST-LLM,而其他两种指标相比于其他基线都得到了显著的提升.

表 2 交通流量预测结果对比

Table 2 Comparison of traffic flow prediction results

Dataset	PEMS03			PEMS04			PEMS07			PEMS08		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
ARIMA	35.31	47.59	33.78	33.73	48.80	24.18	38.17	59.27	19.46	31.09	44.32	22.73
VAR	23.65	38.26	24.51	23.75	36.66	18.09	75.63	115.24	32.22	23.46	36.33	15.42
LSTM	21.33	35.11	23.33	27.14	41.59	18.20	29.98	45.84	13.20	22.20	34.06	14.20
TCN	19.31	33.24	19.86	31.11	37.25	15.48	32.68	42.23	14.22	22.69	35.79	14.04
DCRNN	18.18	30.31	18.91	24.70	38.12	17.12	25.30	38.58	11.66	17.86	27.83	11.45
STGCN	17.49	30.12	17.15	22.70	35.55	14.59	25.38	38.78	11.08	18.02	27.83	11.40
ASTGCN	17.69	29.66	19.40	22.93	35.22	16.56	28.05	42.57	13.92	18.61	28.16	13.08
GWNNet	19.85	32.94	19.31	25.45	39.70	17.29	26.85	42.78	12.12	19.13	31.05	12.68
STSGCN	17.48	29.21	16.78	21.19	33.65	13.90	24.26	39.03	10.21	17.13	26.80	10.96
STFGNN	16.77	28.34	16.30	19.83	31.88	13.02	22.07	35.80	9.21	16.64	26.22	10.60
STGODE	16.50	27.84	16.69	20.84	32.82	13.77	22.99	37.54	10.14	16.81	25.97	10.62
DSTAGNN	15.57	27.21	14.68	19.30	31.46	12.70	21.42	34.51	9.01	15.67	24.77	9.94
ASTGNN	15.07	26.88	15.80	19.26	31.16	12.65	22.23	35.95	9.25	15.98	25.67	9.97
AGCRN	16.06	28.49	15.85	19.83	32.26	12.97	21.29	35.12	8.97	15.95	25.22	10.09
Z-GCNETs	16.64	28.15	16.39	19.50	31.61	12.78	21.77	35.17	9.25	15.76	25.11	10.01
PDFormer	14.94	25.39	15.82	18.32	29.97	12.10	19.83	32.87	8.53	13.58	23.51	9.05
STAEformer	15.35	27.55	15.18	18.22	30.18	<u>11.98</u>	19.14	32.60	<u>8.01</u>	13.46	23.25	8.88
OFA	18.43	29.73	17.11	23.37	36.37	14.93	25.53	37.51	11.35	18.89	28.67	11.24
ST-LLM	14.19	24.76	15.23	<u>17.65</u>	29.70	12.37	18.89	<u>31.32</u>	8.15	13.64	<u>22.45</u>	<b>8.70</b>
STD-MAE	<u>13.80</u>	<u>24.43</u>	<u>13.96</u>	17.80	<u>29.25</u>	<b>11.97</b>	<u>18.65</u>	31.44	<b>7.84</b>	<u>13.44</u>	22.47	8.76
PSTAM	<b>13.41</b>	<b>24.04</b>	<b>13.82</b>	<b>17.52</b>	<b>29.20</b>	12.57	<b>18.58</b>	<b>31.01</b>	8.31	<b>13.21</b>	<b>22.28</b>	<u>8.73</u>

在基于 PLM 的方法中, PSTAM 与 OFA 相比, 在 4 个数据集上的 MAE 指标分别提升了 27.24%, 25.03%, 27.22%, 30.07%; 与 STLLM 相比, 在 4 个数据集上的 MAE 指标分别提升了 5.50%, 0.74%, 1.64%, 3.15%; 与 STD-MAE 相比, 在 4 个数据集上的 MAE 指标分别提升了 2.83%, 1.57%, 0.38%, 1.71%. 这表明 PSTAM 通过双激活领域桥接模块有效解决了预训练模型与交通数据的领域差异问题, 同时通过预训练多粒度时空 Transformer 模块增强了对复杂时空依赖

关系的捕获能力. 相比之下, OFA 没有考虑预训练模型与交通数据的领域差异问题, 也未对节点的空间信息进行编码嵌入, 因此无法充分挖掘数据中的时空信息; STLLM 虽然对节点间的空间信息进行建模, 但未解决领域适配问题, 也忽视了时空特征在不同尺度上的多粒度特性; STD-MAE 使用预训练时空掩码学习时空特征, 但其自监督策略难以针对领域差异提供有效解决方案.

在使用图神经网络和自注意力机制的模型中, PSTAM 与

STAEformer 相比,在 4 个数据集上的 MAE 指标分别提升了 12.64%、3.84%、2.93%、1.86%;与 PDFormer 相比,在 4 个数据集上的 MAE 指标分别提升了 10.24%、4.37%、6.30%、2.72%。这表明 PSTAM 结合了预训练大模型的知识迁移能力和专门设计的时空自注意力机制,能够更有效地捕获和建模多粒度时空特征。PDFormer 虽然使用时空注意力处理长短期依赖,但缺乏预训练知识的支持,且未能解决特征领域适配问题;STAEformer 基于 Transformer 架构设计时空自编码器,但没有充分考虑不同时间尺度和空间尺度上的特征变化规律,导致在多粒度特征提取方面存在局限性。

总体而言,实验结果充分证明了本文提出的 PSTAM 框架在解决领域差异问题和多粒度时空特征建模方面的有效性。通过双激活领域桥接模块和预训练多粒度时空 Transformer 模块的协同作用,PSTAM 能够更加全面地捕获交通流量数据在不同时空尺度上的复杂依赖关系,从而获得更准确的预测结果。

#### 4.4 消融实验

##### 4.4.1 模块消融实验

为了验证 PSTAM 各个模块的有效性,本文设计了 8 个变体模型。PSTAM-D 表示在 PSTAM 的基础上去掉所有阶段的双激活领域桥接模块,PSTAM-DE 表示仅去掉编码阶段,PSTAM-DP 表示仅去掉预训练微调阶段,PSTAM-DS 表示仅去掉时空自注意力阶段,PSTAM-DR 表示仅去掉回归预测阶段;PSTAM-PS 表示去掉预训练多粒度时空 Transformer 模块,PSTAM-P 表示去掉预训练多粒度时空 Transformer 模块中的预训练微调部分,PSTAM-S 表示去掉预训练多粒度时空 Transformer 模块中的时空自注意力部分。这些变体模型充分讨论了双激活领域桥接模块和预训练多粒度时空 Transformer 模块在模型中的作用。消融实验所采用的数据集与对比实验保持一致,实验中除了上述变体模型更改的部分,其余实验设置均与对比实验相同,消融实验结果如表 3 所示,其中最优的实验结果采用加粗标记。

表 3 交通流量预测消融结果对比

Table 3 Comparison of traffic flow prediction ablation results

Dataset	PEMS03			PEMS04			PEMS07			PEMS08		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
PSTAM-D	14.31	26.34	14.12	18.61	31.01	13.78	20.07	32.93	9.33	13.78	23.79	9.83
PSTAM-DE	13.73	25.53	14.01	17.87	29.88	13.41	19.47	32.27	9.17	13.57	23.58	9.57
PSTAM-DP	13.61	25.37	13.93	17.74	29.73	13.37	19.43	32.53	9.14	13.53	23.55	9.53
PSTAM-DS	13.57	25.29	13.90	17.72	29.81	13.30	19.43	32.40	9.13	13.68	23.64	9.67
PSTAM-DR	14.01	25.97	14.07	18.03	30.32	13.53	19.79	32.51	9.23	13.71	23.73	9.70
PSTAM-PS	14.43	25.45	14.31	18.78	31.32	13.81	20.13	33.32	9.57	13.89	23.95	10.08
PSTAM-P	14.26	25.38	14.26	18.55	30.89	13.88	20.01	33.02	9.23	13.66	23.68	9.80
PSTAM-S	14.06	26.03	14.09	18.07	30.04	13.51	19.81	32.67	9.21	13.58	23.53	9.73
PSTAM	<b>13.41</b>	<b>24.04</b>	<b>13.82</b>	<b>17.52</b>	<b>29.20</b>	<b>12.57</b>	<b>18.58</b>	<b>31.01</b>	<b>8.31</b>	<b>13.21</b>	<b>22.28</b>	<b>8.73</b>

对于双激活领域桥接模块,对比 PSTAM-D 与 PSTAM 可以观察到,PSTAM-D 在去掉双激活领域桥接模块后各个指标显著下降,预测精度大大降低。这表明双激活领域桥接模块成功解决了预训练模型与交通数据的领域差异问题,通过双路径激活机制实现了交通序列数据与预训练模型表示空间的高效对齐与融合。进一步对比 PSTAM-DE、PSTAM-DP、PSTAM-DS、PSTAM-DR 与 PSTAM 的结果表明,无论在哪个阶段移除双激活领域桥接模块,模型的预测性能都会有所下降,其中在回归预测阶段该模块的影响最大,编码阶段次之,预训练微调和时空自注意力阶段的影响较小。具体而言,在编码阶段使用该模块可以有效解决输入特征与预训练模型的初始领域差异;在预训练微调阶段使用有助于保持特征转换过程中的领域一致性;而在回归预测阶段的应用则确保了最终特征表示能够有效映射回交通预测空间。缺少双激活领域桥接模块,预训练模型难以适应交通数据的分布特性,导致模型在面对复杂交通流量数据时预测能力下降。

对于预训练多粒度时空 Transformer 模块,对比 PSTAM-PS 与 PSTAM 可以观察到,PSTAM 在去掉预训练多粒度时空 Transformer 模块后各个指标下降明显,预测精度下降幅度较大。这表明预训练多粒度时空 Transformer 模块通过结合预训练大模型的知识迁移能力和专门设计的时空自注意力机

制,显著增强了对多粒度时空特征的捕获能力。进一步对比 PSTAM-P、PSTAM-S 与 PSTAM 的结果证明,预训练微调部分通过 LoRA 策略在保留预训练知识的同时实现了与交通预测任务的高效对齐;而时空自注意力部分则通过多头注意力机制捕获了不同时间尺度和空间尺度上的依赖关系,两者的协同作用显著提高了模型对复杂时空依赖关系的建模能力,从而提高了 PSTAM 的预测精度。

消融实验结果充分验证了双激活领域桥接模块在解决领域差异问题方面的关键作用,以及预训练多粒度时空 Transformer 模块在增强多粒度时空特征捕获能力方面的重要贡献。这两个核心模块的协同作用是 PSTAM 框架取得优异性能的关键因素。

##### 4.4.2 参数敏感性实验

在预训练多粒度时空 Transformer 模块中,通过对 PLM 进行 LoRA 高效微调,可以在保留预训练知识的同时实现与交通预测任务的有效对齐。其中 PLM 的层数,进行全量微调的层数以及 LoRA 采用的秩数大小对于解决领域差异和多粒度特征建模的效果至关重要,因此这是 3 个需要谨慎选择的参数。对于上述 3 个参数的调节结果如图 2~图 4 所示,其中实验设置除对应调节参数外保持与对比实验相同,数据集采用 PEMS04。

从实验结果图 2 可以发现, PLM 层数过小过大时, 模型效果都出现显著下降. 具体而言, PLM 层数较少时随着层数增加模型效果有所提升并在层数 9 时达到最优, 但是在层数超过 9 时模型效果显著下降并趋于平稳. 对于这种现象, 推测是在层数较低时, 模型的表示能力有限, 难以充分建模交通数据的多粒度时空特征, 因此对解决领域差异的帮助有限; 而在层数较高时, 过多的参数导致模型可能过度关注预训练语料中的语言特性而非交通数据的时空特征, 从而影响领域适配效果, 导致模型的预测精度下降.

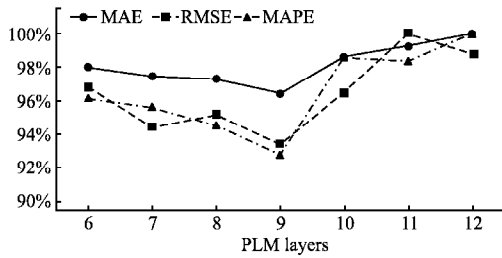


图 2 PLM 层数敏感性分析

Fig. 2 Sensitivity analysis of PLM layers

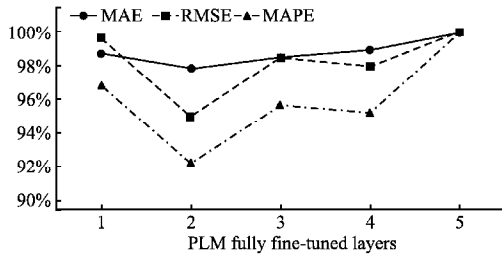


图 3 PLM 完全微调层数敏感性分析

Fig. 3 Sensitivity analysis of PLM fully fine-tuned layers

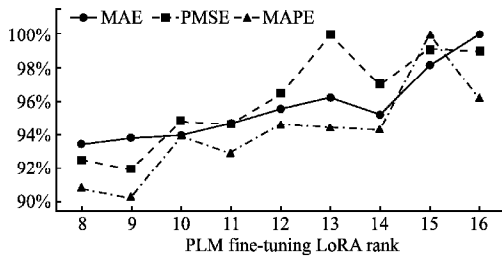


图 4 PLM 微调 LoRA 秩数敏感性分析

Fig. 4 Sensitivity analysis of PLM fine-tuning LoRA rank

从实验结果图 3 可以发现, PLM 的完全微调层数同样是过小过大时都会影响模型的预测效果, 但总体对于各项指标的影响较小. 这是因为非完全微调层通过保持原有注意力参数, 可以更好地保留 PLM 在预训练阶段学习到的长距离依赖关系建模能力, 这对捕获交通流量数据中的多粒度时间特征十分重要. 但是在完全微调层数较小时, 参与梯度更新的参数量占比过低, 模型难以有效适应交通领域的特定特征分布; 而在完全微调层数较大时, 过度微调会导致模型丢失原有的预训练知识, 减弱其对多粒度时空特征的建模能力, 增加过拟合风险.

从实验结果图 4 可以发现, LoRA 采用的秩数大小在较小时, 模型效果随着秩数增大有所提高, LoRA 秩数处于中间

值时, 模型效果最优, 随着 LoRA 秩数的继续增大, 模型的性能开始出现下降. 这种现象的出现是因为 LoRA 秩数控制低秩矩阵的维度, 直接影响模型在领域适配过程中的灵活性. 在秩数较小时, 参与更新的训练参数过少, 模型难以充分调整以适应交通数据的分布特性, 导致领域适配不足; LoRA 秩数不断增大提供了更强的表示能力, 有助于捕获交通数据中的多粒度特征, 但是超过某个阈值后, 过多的参数更新会破坏 PLM 原有的知识结构, 导致预训练优势的丧失. 此外, 过大的 LoRA 秩数也可能导致模型过度关注训练数据中的局部特征而忽略全局模式, 影响对不同时间尺度和空间尺度特征的均衡捕获, 最终导致模型预测精度的下降.

通过这些参数调节分析, 我们确定了 PSTAM 框架中 PLM 层数、完全微调层数以及 LoRA 秩数的最优配置, 为解决预训练模型与交通数据的领域差异问题以及增强多粒度时空特征的捕获能力提供了有效支持

### 4.5 案例研究

为了更直观地展示 PSTAM 模型的性能和工作机制, 本节选择了 PEMS04 数据集中的典型交通场景进行分析. 我们主要关注高峰期交通流量预测的准确性和模式转换期的预测能力. 这些场景能够有效验证模型在捕获多粒度时空特征和处理领域差异方面的优势.

#### 4.5.1 高峰期交通流量预测分析

图 5 展示了 PSTAM 模型与次优模型 STD-MAE 在高峰期流量预测对比. 从图中可以观察到两个模型的预测曲线整体上都能较好地跟踪真实交通流量的主要趋势, 在大部分时间段内表现相近. 不过 PSTAM 在捕捉交通流量的突变点和异常模式方面展现出明显优势, 图中左上角放大区域显示当

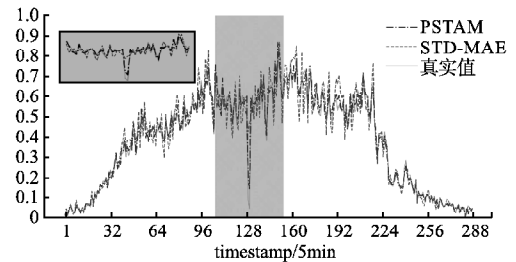


图 5 高峰期交通流量预测分析

Fig. 5 Analysis of peak period traffic flow prediction

交通流量出现短暂下降后快速回升的异常模式时, PSTAM 能够准确捕捉这种变化, 而 STD-MAE 模型则表现出一定的滞后性, 无法及时反映这种突变特征. 这表明双激活领域桥接模块通过有效处理领域差异, 增强了模型对异常交通模式的敏感度. 此外, 对于交通流量出现的多次短暂尖峰和谷值, PSTAM 模型能够更精确地识别这些异常点, 而 STD-MAE 的预测则相对平滑, 错过了关键的短期变化特征. 这一差异说明预训练多粒度时空 Transformer 模块能够有效捕获多尺度时间依赖, 对异常模式具有更强的辨识能力.

#### 4.5.2 交通模式转换期预测分析

图 6 展示了 PSTAM 模型与次优模型 STD-MAE 在交通模式转换期的预测表现对比. 从图中可以明显看出, 尽管两个模型在整体流量趋势预测上效果接近, 但 PSTAM 在处理交

通模式快速转换和异常点识别方面具有显著优势. 图中左上角放大区域清晰地展示了一个典型的流量波动周期, 此时交通流量呈现出非常规的“谷-峰-谷”模式. PSTAM 能够准确追踪这种复杂变化, 尤其是在流量急剧下降和随后快速回升的转换点上, 预测曲线与真实值几乎重合; 而 STD-MAE 则明显表现出“平滑化”效应, 无法捕捉这些关键的转换特征.

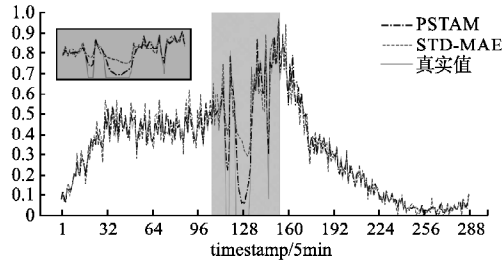


图 6 高峰期交通流量预测分析

Fig. 6 Analysis of transition period traffic flow prediction

特别值得注意的是时间步 128 附近的显著异常点, 真实交通流量在此处出现了一个急剧的下降, PSTAM 成功识别并准确预测了这一异常现象, 而 STD-MAE 则完全错过了这一重要特征, 预测曲线与实际值存在较大偏差. 类似地, 在时间步 140~160 区间的多次波动中, PSTAM 能够更精确地捕捉流量的起伏变化, 而 STD-MAE 则倾向于预测更为平滑的曲线, 丢失了许多关键的短期异常特征.

## 5 结论

本文提出了一种基于预训练时空自注意力模型的交通流量预测框架 PSTAM, 针对预训练大语言模型与交通数据的领域差异问题, 以及交通流量数据多粒度时空特征的有效建模挑战. 我们设计了双激活领域桥接模块, 通过双路径激活机制实现交通数据与预训练模型表示空间的高效对齐与融合; 同时提出预训练多粒度时空 Transformer 模块, 基于预训练策略对多粒度时空特征进行深度建模, 显著增强对复杂时空依赖关系的捕获能力. 在 4 个标准数据集上的实验结果表明, PSTAM 框架在多数评估指标上优于现有最先进的方法. 消融实验证明了两个核心模块的重要性及其协同作用, 参数调节分析揭示了模型设计中的关键考量. 未来将继续优化 PSTAM 模型, 进一步提高对多粒度时空特征的捕获能力, 提升交通流量预测的准确性和鲁棒性.

## References:

- [1] Box G E P, Pierce D A. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models[J]. Journal of the American statistical Association, 1970, 65(332): 1509-1526.
- [2] Sims C A. Macroeconomics and reality[J]. Econometrica: Journal of the Econometric Society, 1980, 48(1): 1-48.
- [3] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [4] Bai S, Kolter J Z, Koltun V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling[J]. arXiv preprint arXiv:1803.01271, 2018.
- [5] Li Y, Yu R, Shahabi C, et al. Diffusion convolutional recurrent neural network: data-driven traffic forecasting[J]. arXiv preprint arXiv:1707.01926, 2017.

- [6] Yu B, Yin H, Zhu Z. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting[J]. arXiv preprint arXiv:1709.04875, 2017.
- [7] 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.
- [8] Wu Z, Pan S, Long G, et al. Graph wavenet for deep spatial-temporal graph modeling[J]. arXiv preprint arXiv:1906.00121, 2019.
- [9] 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 AAAI Conference on Artificial Intelligence, 2020: 914-921.
- [10] Li M, Zhu Z. Spatial-temporal fusion graph neural networks for traffic flow forecasting[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2021: 4189-4196.
- [11] Fang Z, Long Q, Song G, et al. Spatial-temporal graph ode networks for traffic flow forecasting[C]//Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, 2021: 364-373.
- [12] Lan S, Ma Y, Huang W, et al. Dstagnn: dynamic spatial temporal aware graph neural network for traffic flow forecasting[C]//International Conference on Machine Learning, PMLR, 2022: 11906-11917.
- [13] 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, 2021, 34(11): 5415-5428.
- [14] 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, doi: 10.48550/arXiv.2007.02842.
- [15] Chen Y, Segovia I, Gel Y R. Z-GCNets: time zigzags at graph convolutional networks for time series forecasting[C]//International Conference on Machine Learning, PMLR, 2021: 1684-1694.
- [16] Jiang J, Han C, Zhao W X, et al. Pdformer: propagation delay-aware dynamic long-range transformer for traffic flow prediction[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2023: 4365-4373.
- [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] Xue H, Salim F D. Promptcast: a new prompt-based learning paradigm for time series forecasting[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 36(11): 6851-6864.
- [19] Zhou T, Niu P, Sun L, et al. One fits all: power general time series analysis by pretrained Im[J]. Advances in Neural Information Processing Systems, 2023, 36: 43322-43355, doi: 10.48550/arXiv.2302.11939.
- [20] Liu C, Yang S, Xu Q, et al. Spatial-temporal large language model for traffic prediction[C]//25th IEEE International Conference on Mobile Data Management (MDM), 2024: 31-40.
- [21] Chen Y, Wang X, Xu G. Gtgpt: a pre-trained large language model with graph attention network for spatiotemporal imputation[J]. arXiv preprint arXiv:2311.14332, 2023.
- [22] Liu Y, Zhang H, Li C, et al. Timer: generative pre-trained transformers are large time series models[J]. arXiv preprint arXiv:2402.02368, 2024.
- [23] Gao H, Jiang R, Dong Z, et al. Spatial-temporal-decoupled masked pre-training for spatiotemporal forecasting[J]. arXiv preprint arXiv:2312.00516, 2023.
- [24] Hu E J, Shen Y, Wallis P, et al. Lora: low-rank adaptation of large language models[C]//Proceedings of the International Conference on Learning Representations (ICLR), 2022.
- [25] Radford A, Wu J, Child R, et al. Language models are unsupervised multitask learners[J]. OpenAI Blog, 2019, 1(8): 9.
- [26] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Advances in Neural Information Processing Systems, 2017: 5998-6008.