

基于预训练时空解耦的交通流预测模型

李静, 杜圣东, 史浩琛, 胡节, 杨燕, 李天瑞

引用本文

李静, 杜圣东, 史浩琛, 胡节, 杨燕, 李天瑞. [基于预训练时空解耦的交通流预测模型](#)[J]. 计算机科学, 2026, 53(4): 155-162.

LI Jing, DU Shengdong, SHI Haochen, HU Jie, YANG Yan, LI Tianrui. [Pre-trained Spatio-Temporal Decoupling-based Traffic Flow Prediction Model](#) [J]. Computer Science, 2026, 53(4): 155-162.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[跨模态融合的少样本勒索软件分类器:基于预训练模型的多模态编码](#)

Cross-modal Fusion Few-sample Ransomware Classifier:Multimodal Encoding Based on Pre-trained Models

计算机科学, 2026, 53(4): 435-444. <https://doi.org/10.11896/jsjx.250500078>

[跨模型协同的法律文本相关性无监督表征方法研究](#)

Cross-model Collaborative Unsupervised Representation Method for Legal Texts

计算机科学, 2026, 53(4): 356-365. <https://doi.org/10.11896/jsjx.251100003>

[基于时频域注意力的时间序列异常检测模型](#)

Time-Frequency Attention Based Model for Time Series Anomaly Detection

计算机科学, 2026, 53(2): 161-169. <https://doi.org/10.11896/jsjx.241200106>

[基于双层注意力网络的强化学习方法求解柔性作业车间调度问题](#)

Reinforcement Learning Method for Solving Flexible Job Shop Scheduling Problem Based on Double Layer Attention Network

计算机科学, 2026, 53(1): 231-240. <https://doi.org/10.11896/jsjx.250100088>

[一阶逻辑中一类多线型标准矛盾体的结构](#)

Structures of Multi-line Standard Contradictions in First-order Logic

计算机科学, 2025, 52(12): 200-208. <https://doi.org/10.11896/jsjx.250200060>

基于预训练时空解耦的交通流预测模型

李静 杜圣东 史浩琛 胡节 杨燕 李天瑞

西南交通大学计算机与人工智能学院 成都 611756

(1322601861@my.swjtu.edu.cn)

摘要 交通流预测作为智慧城市动态决策的核心技术,其准确性是影响交通信号控制、路径规划和应急管理的关键。随着城市路网规模的扩大和交通数据的激增,传统方法难以对路网节点间复杂的时空交互特性进行精准建模。预训练模型虽然能进行跨领域知识迁移,但应用于交通流预测任务时,仍面临时空特征耦合所导致的建模瓶颈,以及预训练表征与交通领域特性不匹配的问题。针对上述问题,提出一种基于预训练时空解耦的交通流预测模型(PT-STD)。该方法通过时空分解模块解耦分离空间拓扑关联与多粒度时序模式的深度特征学习,进一步设计分层自适应微调策略,分阶段解冻预训练模型的归一化层与注意力参数,逐步将预训练模型中学习到的通用知识迁移到时空特性建模中。实验表明,该模型在基准数据集上展现出显著优势,而且在数据稀缺场景下的平均绝对误差可降低3.89%。

关键词: 交通流预测; 时空分解; 分层微调; 预训练模型; 城市计算

中图分类号 TP391

Pre-trained Spatio-Temporal Decoupling-based Traffic Flow Prediction Model

LI Jing, DU Shengdong, SHI Haochen, HU Jie, YANG Yan and LI Tianrui

School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu 611756, China

Abstract Traffic flow prediction, as a core technology for dynamic decision-making in smart cities, plays a crucial role in traffic signal control, route planning, and emergency management. With the expansion of urban road networks and the rapid growth of traffic data, traditional methods face challenges in accurately modeling the complex spatio-temporal interactions among road network nodes. Although pre-trained models can transfer knowledge across domains, they still encounter limitations when applied to traffic flow prediction, primarily due to coupled spatio-temporal features and the mismatch between pre-trained representations and traffic-specific characteristics. To address these issues, this paper proposes the pre-trained spatio-temporal decoupling-based traffic flow prediction model (PT-STD). The method employs a spatio-temporal decoupling module to disentangle the deep feature learning of spatial topological relationships and multi-granularity temporal patterns. Furthermore, it designs a hierarchical adaptive fine-tuning strategy that progressively unfreezes the normalization layers and attention parameters of the pre-trained model, gradually transferring the general knowledge learned in the pre-trained model to spatio-temporal feature modeling. Experimental results demonstrate that PT-STD achieves significant improvements on standard benchmark datasets, with a 3.89% reduction in mean absolute error (MAE) under data-scarce scenarios.

Keywords Traffic flow prediction, Spatio-temporal decomposition, Hierarchical fine-tuning, Pretrained model, Urban computing

1 引言

交通流预测作为智慧城市动态决策的核心技术,通过挖掘历史时空数据中隐含的路网演化规律,为交通资源调度优化与拥堵预警提供关键支撑^[1]。其在共享出行工具动态调配^[2]、公共交通运力优化^[3]等场景中具有重要的应用价值。

尽管深度学习模型已取得一定进展,但交通流数据中不同特征的耦合表征会导致模式混淆,而现有方法仍难以有效解耦多尺度时空模式。预训练模型基于离散的文本符号系统

进行建模,具备跨领域知识迁移潜力^[4],而交通流数据属于连续的时空场域,离散的文本符号与交通流数据在特征表征上存在显著差异,因此将预训练模型直接应用于交通流预测会受限于对齐问题^[5],特别是在资源有限的场景下,容易出现预测精度显著波动的问题^[6]。

针对上述挑战,本文提出一种基于预训练时空解耦的交通流预测模型。该模型首先将交通流数据进行时空表征解耦、特征融合与轻量化编码3阶段处理,实现时间与空间的不同特征解析;然后依次进行知识保留、分布适配和动态增强的

到稿日期:2025-06-08 返修日期:2025-09-15

基金项目:四川省重大科技专项(2024ZDZX0012);国家自然科学基金面上项目(62276215);国家自然科学基金联合基金(U2468207)

This work was supported by the Major Science and Technology Special Project of Sichuan Province(2024ZDZX0012), General Program of the National Natural Science Foundation of China(62276215) and Joint Fund of the National Natural Science Foundation of China(U2468207).

通信作者:杜圣东(sddu@swjtu.edu.cn)

自适应分层微调,实现从通用知识到时空领域的平稳过渡。实验结果表明,PT-STD在4个数据集上的表现都优于主流基线。本文的主要贡献包括:

- 1)提出了一种时空解耦模型设计,通过空间编码与双尺度时序分解实现交通流量特征的解耦表征,同时设计了轻量化MLP编码器提升模型的计算效率;
- 2)设计了自适应分层微调机制,通过知识保留、分布适配和动态增强3阶段优化,有效解决了预训练模型离散文本符号与交通流数据间的维度失配与建模偏差问题;
- 3)基于多场景交通流基准数据集的实验结果表明,PT-STD在预测精度、跨域泛化能力和计算效率方面均显著优于现有方法,尤其在数据稀缺场景下的平均绝对误差可降低3.89%。

2 相关工作

2.1 基于预训练的交通流预测

预训练模型在交通流预测领域经历了显著的技术演进,其核心在于如何更高效地融合时空特征与动态交通模式^[7-9]。早期研究主要依赖LSTM和GRU等循环神经网络处理时序依赖,但受限于局部建模能力和梯度问题^[10-12]。随着Transformer架构的引入,研究者通过自注意力机制突破长距离依赖瓶颈^[13-15]。近期研究进一步向多特征融合和轻量化方向发展,如在模态融合方面,时空预训练框架ST-MFM通过图注意力网络建模环境因素与交通流的动态关联,并利用对比学习对齐跨模态表征,在突发天气场景下的预测误差显著降低^[16]。然而,现有方法仍面临预训练任务与交通流预测任务间存在语义鸿沟、特征融合不充分以及模型计算复杂度高等问题。因此,本文采用HuggingFace Transformers平台的

GPT-2预训练语言模型,加载gpt2-medium权重,并替换其语言生成头以适配交通流序列任务;为缓解文本符号与交通流数据差异带来的迁移障碍,设计了时空自适应分层微调策略,实现从文本特征到交通时空领域特征的渐进式知识迁移。

2.2 基于解耦设计的交通流预测

基于解耦设计的交通流预测近年来涌现出多项创新性研究,主要通过时空特征的分离建模提升预测精度与计算效率^[17]。STWave模型^[18]采用离散小波变换分解交通时序信号,通过双通道时空编码器分别处理趋势分量和事件分量,并利用高效谱图注意力网络捕捉空间相关性。DSTF框架^[19]通过残差分解机制将交通信号解耦为扩散信号和固有信号,结合动态图学习模块建模时变空间依赖关系。ST-TIS模型^[20]采用区域采样策略降低计算复杂度,通过信息融合模块和周期性特征嵌入实现高效的时空依赖性建模。这些方法尽管在特征解耦方面取得了显著进展,但仍存在对不同交通特征耦合表征不足、预训练模型迁移时的特征适配性差等关键问题。为此,本文在预训练模型微调机制的基础上引入了创新性的时空解耦模块,以分离并增强不同时空特征。

3 PT-STD 模型设计

如图1所示,PT-STD模型主要由两个部分组成。第一部分是时空特征解耦,具体是将交通流数据进行时空表征解耦,分解交通流数据的时间特征与空间特征,然后将分解后的数据进行特征融合,再进行深度编码,得到时空解耦编码后的数据。第二部分是自适应分层微调策略,分为3个阶段。首先是将时空解耦编码后的数据输入参数完全冻结的预训练模型中,模型处理后再输入到解冻了中间归一化层的预训练模型中,最后输入到激活了时空注意力层的预训练模型中。

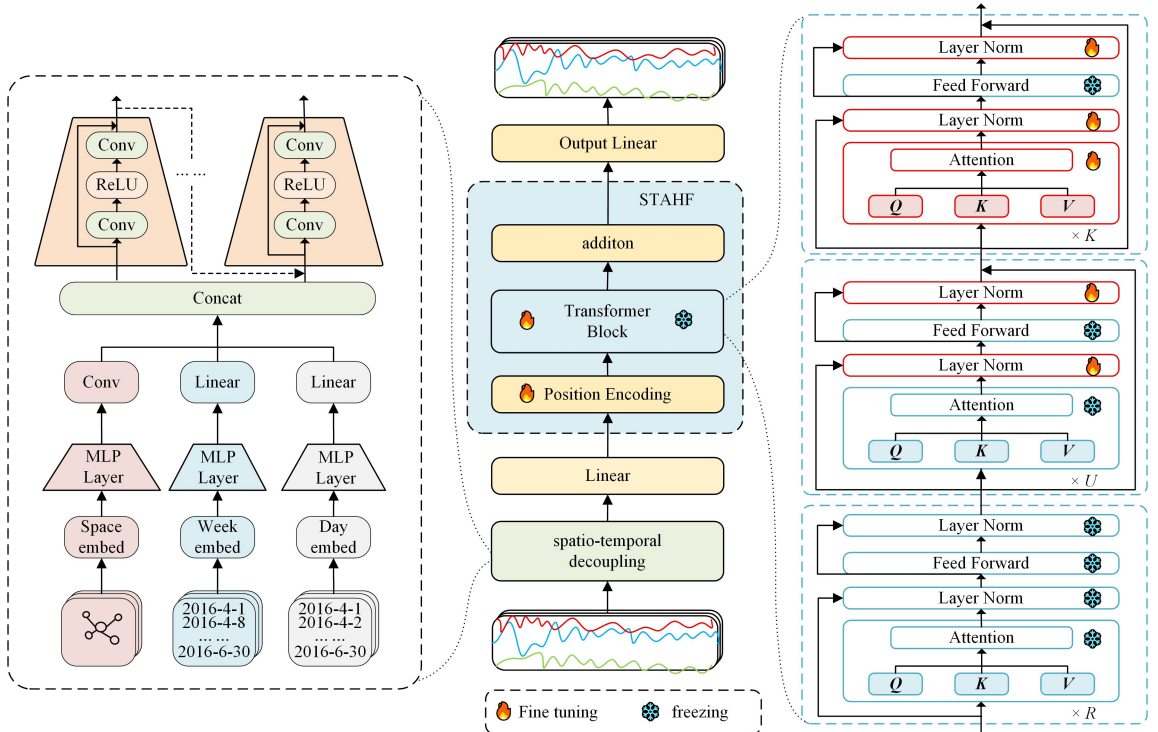


图1 PT-STD模型结构图

Fig. 1 Structure diagram of PT-STD model

3.1 时空解耦模块设计

PT-STD的时空解耦模块采用时空表征解耦、特征融合、深度编码3阶段处理流程。具体地,输入历史流量张量 \mathbf{X}_M 后,首先时空表征解耦阶段将原始数据分解为空间拓扑嵌入 \mathbf{E}_S 、周期性时间嵌入 \mathbf{X}_T 与全局性时间演化嵌入 \mathbf{X}_L 3个分量,再分别通过MLP捕获动态路网关联,提取时间上的周期性波动特征以及时间上的全局长期演化规律;随后,特征融合,阶段将3个分量融合,生成时空耦合表征 \mathbf{X}_Q ;最后,深度编码阶段将耦合表征后的数据通过轻量化级联的MLP结构增强特征表达,同时缓解在训练过程中的梯度消失问题。

时空表征解耦:时空表征解耦模块采用多路并行处理架构,空间嵌入路径通过 1×1 卷积 $\theta_S \in \mathbf{R}^{T \times D}$ 将输入 $\mathbf{X}_S \in \mathbf{R}^{N \times T}$ 映射为 \mathbf{E}_S ,再经残差连接的MLP进行非线性增强。其数学表达为:

$$\mathbf{E}_S = \mathbf{F}_{\text{conv}}(\mathbf{X}_S; \theta_S) \in \mathbf{R}^{N \times D} \quad (1)$$

$$\begin{aligned} \mathbf{E}_S' &= \text{MLP}_S(\mathbf{E}_S) \\ &= \mathbf{F}_{\text{conv}}^2(\sigma(\mathbf{F}_{\text{conv}}^1(\mathbf{E}_S; \theta_{s1})); \theta_{s2}) + \mathbf{E}_S \in \mathbf{R}^{N \times D} \end{aligned} \quad (2)$$

其中, T 是时间步长; N 是节点数; D 是特征维度; θ_{s1}, θ_{s2} 为可学习参数; σ 为非线性激活函数ReLU。 1×1 卷积的主要功能是对输入特征进行通道维度上的线性映射,实现与后续MLP模块维度的对齐与适配。

时间嵌入路径则是构建了双尺度架构。对于周粒度的周期性模式,通过可学习矩阵 $\mathbf{W}_T \in \mathbf{R}^{T \times D}$,将原始时序 $\mathbf{X}_T \in \mathbf{R}^{N \times T}$ 映射为周期嵌入 \mathbf{E}_T ,并引入周期MLP进行特征细化。其数学表达为:

$$\mathbf{E}_T = \mathbf{X}_T \mathbf{W}_T \in \mathbf{R}^{N \times D} \quad (3)$$

$$\mathbf{E}_T' = \text{MLP}_T(\mathbf{E}_T) \in \mathbf{R}^{N \times D} \quad (4)$$

对于日粒度全局演化趋势,采用独立参数矩阵 $\mathbf{W}_L \in \mathbf{R}^{L \times D}$ 生成全局时间嵌入 \mathbf{E}_L ,其中 L 为预测步长,并通过趋势MLP捕获长期演化规律。其数学表达为:

$$\mathbf{E}_L = \mathbf{X}_L \mathbf{W}_L \in \mathbf{R}^{N \times D} \quad (5)$$

$$\mathbf{E}_L' = \text{MLP}_L(\mathbf{E}_L) \in \mathbf{R}^{N \times D} \quad (6)$$

双路时间嵌入的构建既可以捕获到全局的时间信息,也可以捕获到周期性的时间信息,确保了不同时间尺度信息的完整性。

特征融合:特征融合模块采用了多特征协同架构。将空间拓扑嵌入 $\mathbf{E}_S' \in \mathbf{R}^{N \times D}$ 、周期时间嵌入 $\mathbf{E}_T' \in \mathbf{R}^{N \times D}$ 与全局演化嵌入 $\mathbf{E}_L' \in \mathbf{R}^{N \times D}$ 分别经 1×1 卷积核 $\theta_S, \theta_T, \theta_L \in \mathbf{R}^{T \times D}$ 进行特征变换后,沿通道维度拼接构建联合表征。

$$\mathbf{E}_{\text{STL}} = \mathbf{F}_{\text{conv}}(\mathbf{E}_S'; \theta_S) \oplus \mathbf{F}_{\text{conv}}(\mathbf{E}_T'; \theta_T) \oplus \mathbf{F}_{\text{conv}}(\mathbf{E}_L'; \theta_L) \in \mathbf{R}^{N \times 3D} \quad (7)$$

其中, \oplus 表示拼接操作。该过程将不同的特征在通道维度上进行集成,最大程度地保留了不同特征的原始信息,为后续模块实现深度特征交互奠定了基础。

深度编码:深度编码模块采用残差连接的层级MLP架构。每层编码过程表示为:

$$\mathbf{E}_Q^{n+1} = \mathbf{F}_{\text{conv}}^{(n,2)}(\sigma(\mathbf{F}_{\text{conv}}^{(n,1)}(\mathbf{E}_Q^n))) + \mathbf{E}_Q^n \quad (8)$$

其中, $\mathbf{F}_{\text{conv}}^{(n,1)}$ 与 $\mathbf{F}_{\text{conv}}^{(n,2)}$ 分别为该层的前、后卷积操作, σ 为ReLU激活函数。通过具有自适应权重调节能力的MLP架构,模型可以逐步提取到数据的时空交互特征。在这个过程中采用

了残差连接结构^[21],主要的目的是增强模型的训练稳定性与特征表达能力,通过显式构建映射路径,实现梯度的直接传播,有效缓解了训练过程中的梯度衰减问题,还保留了关键的数据结构信息与拓扑连续性。

3.2 时空自适应分层微调设计

为实现预训练语言模型在交通流预测任务中的有效迁移,以GPT-2为基础,构建了时空自适应分层微调机制(Spatio-Temporal Adaptive Hierarchical Fine-tuning, STAHF)。其中,GPT-2是由OpenAI提出的,采用了12层堆叠的Transformer解码器结构,具有强大的上下文处理能力与长距离依赖学习能力。与编码器-解码器架构相比,GPT-2更适用于自回归任务,非常契合交通流预测任务中用历史时序预测未来趋势的场景。

模型具体实现方面,通过HuggingFace Transformers平台加载gpt2-medium预训练权重,保留其主干参数结构,并将语言生成头部替换为交通序列预测的回归模块。同时,为缓解模型迁移过程中知识遗忘与表征偏移的问题,设计了分阶段渐进式微调策略,对预训练模型进行逐层解冻与结构适配。

STAHF机制采用3阶段渐进式优化策略:在知识保留阶段,冻结全部预训练参数来保持基础语义表征能力;在分布适配阶段,解冻归一化层参数来实现交通流数据特征的分布校准;在动态增强阶段,激活注意力机制参数来强化时空动态依赖建模。最终,通过回归卷积层将隐空间特征映射至预测空间,输出未来 L 步的流量张量 $\mathbf{Y}_L \in \mathbf{R}^{L \times N \times S}$ 。这种方式通过知识继承与领域适应的协同优化,在保持预训练模型泛化能力的同时,实现了交通时空特性的针对性建模。具体的操作路径如下。

知识保留阶段:知识保留阶段采用参数冻结策略继承预训练模型的通用表征能力,模型前 R 层参数完全冻结。该阶段可分解为以下4个步骤。

1)冻结的多头注意力提取全局特征

$$\mathbf{H}_{\text{attn}}^{(r)} = \text{MultiHeadAttn}_{\theta_{\text{attn}}}^{\text{frozen}}(\mathbf{H}_{\text{out}}^{(r-1)}) \quad (9)$$

其中, θ_{attn} 为冻结的注意力参数矩阵, $\mathbf{H}_{\text{out}}^{(r-1)} \in \mathbf{R}^{N \times 3D}$ 为第 r 层的输入特征。该操作通过预训练阶段学习到的注意力分布模式,捕获输入序列中潜在的全局依赖关系。

2)冻结的层归一化处理残差连接

$$\mathbf{H}_{\text{layer}}^{(r)} = \text{LayerNorm}_{\gamma_1, \beta_1}^{\text{frozen}}(\mathbf{H}_{\text{attn}}^{(r)} + \mathbf{H}_{\text{out}}^{(r-1)}) \quad (10)$$

其中, $\gamma_1, \beta_1 \in \mathbf{R}^{3D}$ 为冻结参数,用于稳定注意力输出分布,保留预训练的领域不变性。

3)冻结的前馈网络进行非线性变换

$$\mathbf{H}_{\text{ffn}}^{(r)} = \text{FFN}_{\theta_{\text{ffn}}}^{\text{frozen}}(\mathbf{H}_{\text{layer}}^{(r)}) \quad (11)$$

其中, θ_{ffn} 为冻结的前馈网络参数。

4)输出经过冻结层归一化

$$\mathbf{H}_{\text{out}}^{(r)} = \text{LayerNorm}_{\gamma_2, \beta_2}^{\text{frozen}}(\mathbf{H}_{\text{ffn}}^{(r)} + \mathbf{H}_{\text{layer}}^{(r)}) \quad (12)$$

其中,冻结参数 $\gamma_2, \beta_2 \in \mathbf{R}^{3D}$ 抑制微调初期参数偏移,保持隐空间结构化投影。这个阶段是利用预训练获得的通用能力对输入特征进行结构化投影,形成具有交通流领域不变性的初始隐空间,有效抑制微调初期的参数偏移风险。

分布适配阶段:在中间 U 层中解冻层归一化参数 γ 和 β ,以实现交通数据分布的动态校准。该阶段可描述为:

$$\mathbf{H}_{\text{layer}}^{(u)} = \text{LayerNorm}_{\gamma_3, \beta_3}^{\text{train}} (\text{MultiHeadAttn}_{\theta_{\text{attn}}}^{\text{frozen}} (\mathbf{H}_{\text{out}}^{(u-1)}) + \mathbf{H}_{\text{out}}^{(u-1)}) \quad (13)$$

$$\mathbf{H}_{\text{out}}^{(u)} = \text{LayerNorm}_{\gamma_4, \beta_4}^{\text{train}} (\text{FFN}_{\theta_{\text{fin}}}^{\text{frozen}} (\mathbf{H}_{\text{layer}}^{(u)}) + \mathbf{H}_{\text{layer}}^{(u)}) + \mathbf{H}_{\text{out}}^{(u-1)} \quad (14)$$

其中, $\gamma_3, \beta_3, \gamma_4, \beta_4$ 为可训练参数, 主要是用于对特征分布进行自适应校准, 有效缓解交通流数据非平稳性导致的领域漂移问题。这个阶段在保持注意力机制冻结的同时, 通过动态归一化实现交通流领域数据的分布适配。

在 3 阶段优化过程中, 优先解冻归一化层参数, 主要是基于迁移学习中的两个关键理论: 第一是归一化层在神经网络中承担特征标准化与分布调节功能, 是特征对齐的关键位置; 第二是归一化层参数较少且梯度更新稳定, 优先解冻可在不破坏主干语义结构的基础上平滑调整模型对交通流数据的处理, 为后续注意力机制的解冻提供稳定梯度基础, 避免梯度振荡与训练不稳定的问题^[22]。因此, STAHF 以先解冻归一化后解冻注意力的策略, 确保了通用语言知识向交通时空特征的平滑迁移。相较于传统的固定冻结策略 (FPT), STAHF 在提升模型迁移稳定性的同时, 更适应交通流中的非平稳性与多尺度周期特征。

动态增强阶段: 在末端 K 层中解冻注意力参数 θ_{attn} , 以激活时空依赖建模能力。该阶段的核心计算流程为:

$$\mathbf{H}_{\text{layer}}^{(k)} = \text{LayerNorm}_{\gamma_5, \beta_5}^{\text{train}} (\text{MultiHeadAttn}_{\theta_{\text{attn}}}^{\text{train}} (\mathbf{H}_{\text{out}}^{(k-1)}) + \mathbf{H}_{\text{out}}^{(k-1)}) \quad (15)$$

$$\mathbf{H}_{\text{out}}^{(k)} = \text{LayerNorm}_{\gamma_6, \beta_6}^{\text{train}} (\text{FFN}_{\theta_{\text{fin}}}^{\text{frozen}} (\mathbf{H}_{\text{layer}}^{(k)}) + \mathbf{H}_{\text{layer}}^{(k)}) + \mathbf{H}_{\text{out}}^{(k-1)} \quad (16)$$

其中, θ_{attn} 为可训练注意力权重矩阵。这个阶段激活多头注意力的时空模式挖掘能力, 在预训练的先验基础上, 可以学习到节点间的动态空间关联与多粒度时间演化规律。通过注意力权重矩阵的更新, 模型能够自适应地聚焦于重要的时空依赖关系, 从而提升了模型的泛化能力。

为保证微调过程的稳定性与迁移效果, STAHF 3 阶段策略中的关键参数 (R, U, K) 由验证集性能调优确定。在具体设置中, GPT-2 总层数为 $L = 12$, 在知识保留阶段, 冻结前 R 层, 其中 $R \in \{2, 4, 6, 8\}$, 通过实验调试最终在 $R = 4$ 时取得最优性能。在分布适配阶段, 解冻第 $R+1$ 至 $R+U$ 层中的归一化层参数, 经调试设置 $U = 3$; 在动态增强阶段, 进一步解冻后续 $K = 5$ 层的多头注意力模块。该分层配置在不同交通任务中均展现出良好的训练稳定性与泛化能力。

跨阶段特征融合: 设计全局残差通路可与学习权重, 实现多尺度特征融合。

$$\mathbf{H}_{\text{final}} = \mathbf{H}_{\text{out}}^{(k)} + \mathbf{E}_Q \odot \mathbf{G} \quad (17)$$

其中, $\mathbf{G} \in \mathbf{R}^{N \times 3D}$ 为可学习权重。该通路将深层输出与早期的特征进行结合, 为模型提供了跨层级的特征表示, 有助于提高模型在训练时的稳定性, 还可以提升特征的代表能力。

最终, 回归卷积层将融合后的特征映射为预测结果。

$$\hat{\mathbf{Y}}_L = \mathbf{F}_{\text{conv}} (\mathbf{H}_{\text{final}}; \boldsymbol{\theta}_L) \quad (18)$$

其中, $\boldsymbol{\theta}_L \in \mathbf{R}^{K \times 3D \times S}$ 为卷积核参数, 用于将维度为 3D 的融合特征投影到一维输出, 从而生成最终的预测结果。该框架通过分层知识迁移、动态分布校准与时空特征增强的协同优化,

突破了传统冻结策略的适应性瓶颈。

PT-STD 的具体过程如算法 1 所示。

算法 1 基于预训练时空解耦的交通流预测模型

输入: 历史 M 时间步长的流量特征 \mathbf{X}_M 和所有超参数

输出: 训练完成的 PT-STD 模型

1. 初始化模型参数, 设置训练轮数 T
2. for epoch = 1 to T do
3. 随机打乱训练数据顺序
4. for 每个数据批次 \mathbf{X}_M in 训练数据集 do
 - // 时空表征解耦阶段
 - 5. 空间特征 $\mathbf{E}_s' =$ 空间编码 (\mathbf{X}_M) // 式(1)和式(2)
 - 6. 时间特征 $\mathbf{E}_T' \mathbf{E}_L' =$ 时间编码 (\mathbf{X}_M) // 式(3)–式(6)
 - // 特征融合与深度编码
 - 7. 特征融合 $\mathbf{E}_{\text{STL}} =$ 融合 ($\mathbf{E}_s' \mathbf{E}_T' \mathbf{E}_L'$) // 式(7)
 - 8. 深度编码 $\mathbf{E}_Q^n =$ 编码 (\mathbf{E}_{STL}) // 式(8)
 - // 时空自适应分层微调机制
 - 9. for $i = 1$ to $R+U+K$ do
 - 10. if $i \leq R$ then
 - 11. $\mathbf{H}_{\text{out}}^i =$ 知识保留训练 // 式(9)–式(12)
 - 12. else if $i \leq U$ then
 - 13. $\mathbf{H}_{\text{out}}^i =$ 分布适配训练 // 式(13)和式(14)
 - 14. else
 - 15. $\mathbf{H}_{\text{out}}^i =$ 动态增强训练 // 式(15)和式(16)
 - 16. end if
 - 17. end for
 - // 预测与损失计算
 - 18. $\mathbf{Y}_{\text{pred}} =$ 预测 (\mathbf{H}) // 式(18)
 - 19. Loss = 损失 ($\mathbf{Y}_{\text{pred}}, \mathbf{Y}_{\text{true}}$)
 - 20. end for
 - 21. end for

4 实验及结果分析

4.1 数据集

为构建具有时空泛化能力的交通预测模型, 选取纽约市出租车 (NYC Taxi) 与芝加哥共享单车 (CHBike) 两类场景下的交通系统构建基准数据集^[24]。两类场景的时间跨度为 2016 年 4 月至 6 月, 时间分辨率为 30 min, 共 4368 个时间步。其中, 每个类别包含两个数据集, 分别是下车数据集 Drop-off 与上车数据集 Pick-up, 所以共使用 4 个数据集进行实验。

NYC Taxi 场景下的数据将行程记录经空间聚类处理成 266 个站点, 反映跨区域机动出行特征。CHBike 场景下的数据则是根据骑行记录筛选出 250 个高频使用站点, 刻画短途的区域分布规律。两类场景下的 4 个数据集在空间尺度和出行模式上形成互补, 为模型学习多尺度交通特征提供了数据支撑。数据集的具体统计特征如表 1 所列。

表 1 数据集类别描述

Table 1 Dataset category description

数据集类别	NYC Taxi ^[24]	CHBike ^[24]
总行程/km	35	3.6
站数	266	250
时间跨度	2016-04-01–2016-06-30	2016-04-01–2016-06-30
时间步数	4368	4368
时间间隔/min	30	30

4.2 基准模型与实验设置

为全面评估 PT-STD 的性能优势,选取了 10 个具有代表性的基线模型进行对比实验,这些模型涵盖了当前时空预测领域的主要技术路线。1)传统时空图卷积网络,包括基于固定图结构的 STGCN^[25]与采用扩散卷积的 DCRNN^[26];2)动态图演化模型,涵盖自适应邻接矩阵构建的 AGCRN^[27]和 DGCRN^[28]以及集成注意力机制的 ASTGCN^[29]和 ASTGNN^[30];3)预训练架构,选取融合图卷积与语言模型的 GCNGPT 和 GATGPT^[23];4)通用预测框架 OFA^[31]与 ST-LLM。其中,STGCN 和 DCRNN 分别采用时空卷积和扩散过程来捕捉局部时空特征;AGCRN 和 DGCRN 通过自适应邻接矩阵学习动态时空关联;ASTGCN 和 ASTGNN 引入多头注意力机制增强长期依赖建模能力;GCNGPT 和 GATGPT 探索了预训练语言模型在时空任务中的迁移潜力;ST-LLM 和 OFA 则分别代表了时空注意力、通用预测的最新进展。这一基准集合全面覆盖了静态与动态图建模、传统与预训练架构、卷积与注意力机制等关键技术路线,为验证 PT-STD 的综合性能提供了多维度的对比参照。

本研究采用严格的时间序列划分策略,将 NYC Taxi 和 CHBike 场景下的 4 个数据集按 6:2:2 的比例划分为连续的训练集、验证集和测试集。实验设置历史时间窗口为 12,预测步长为 12,以评估模型的多步预测能力。

实验在 NVIDIA RTX 4090 GPU 平台上运行,基于 PyTorch 2.1.2 框架和 CUDA 12.6 加速环境实现。训练过程中设置批量大小为 64,最大训练轮次为 500,并采用早停机制

(验证集 MAE 连续 30 轮未改善即终止训练)。这一标准化实验配置,确保了研究结果的可重复性和可比性。

4.3 评估指标

本研究采用了 4 类指标全面评估模型性能:1)MAE,衡量预测值与真实值的平均绝对偏差;2)RMSE,反映对数据异常值的敏感程度;3)MAPE,表示相对误差百分比;4)WAPE,通过流量加权消除低值区域的影响。

$$MAE = \frac{1}{m} \sum_{i=1}^m |\hat{Y}_i - Y_i| \quad (19)$$

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (\hat{Y}_i - Y_i)^2} \quad (20)$$

$$MAPE = \frac{1}{m} \sum_{i=1}^m \left| \frac{\hat{Y}_i - Y_i}{Y_i} \right| \times 100\% \quad (21)$$

$$WAPE = \frac{\sum_{i=1}^m |\hat{Y}_i - Y_i|}{\sum_{i=1}^m |Y_i|} \times 100\% \quad (22)$$

其中, m 是样本数; \hat{Y}_i 与 Y_i 分别表示第 i 个样本的预测值与真实值。该指标体系从绝对误差(MAE, RMSE)和相对误差(MAPE, WAPE)两个维度,兼顾了精度一致性和异常敏感性,特别适合评估交通数据中的非平稳特性。所有指标均为正向指标,数值越小表示预测性能越好。

4.4 结果与分析

为系统评估模型的能力,在 NYC Taxi 与 CHBike 场景下的 4 个数据集上开展对比实验。实验结果如表 2 所列,PT-STD 在 4 项基准任务中均展现出显著的性能优势,其预测精度、不同场景稳定性与计算效率均超越了现有方法。

表 2 模型在 NYC Taxi 和 CHBike 数据集上的比较

Table 2 Comparison of models on NYC Taxi and CHBike datasets

Metric	NYC Taxi Drop-off				NYC Taxi Pick-up				CHBike Drop-off				CHBike Pick-up			
	MAE	RMSE	MAPE/%	WAPE/%	MAE	RMSE	MAPE/%	WAPE/%	MAE	RMSE	MAPE/%	WAPE/%	MAE	RMSE	MAPE/%	WAPE/%
ASTGCN	6.98	14.70	45.48	26.60	7.43	13.84	47.96	28.04	2.79	4.20	69.88	56.49	2.76	4.55	64.23	55.71
DGCRN	<u>5.14</u>	9.39	35.09	<u>19.64</u>	5.44	9.82	35.78	20.58	1.96	2.93	51.99	39.70	<u>2.06</u>	3.21	54.06	<u>41.51</u>
DCRNN	5.19	9.63	37.78	19.82	<u>5.40</u>	9.71	35.09	<u>20.43</u>	1.96	2.94	51.42	39.61	2.09	3.30	54.22	42.26
STGCN	5.38	9.60	39.12	20.55	5.71	10.22	36.51	21.62	2.01	3.07	50.45	40.62	2.08	3.31	53.63	42.08
ASTGNN	6.28	12.00	49.78	23.91	5.90	10.71	40.15	22.32	2.24	3.35	57.21	45.27	2.37	3.67	60.08	47.81
AGCRN	5.45	9.56	40.67	20.81	5.79	10.11	40.40	21.93	2.06	3.19	51.91	41.78	2.16	3.46	56.35	43.69
ST-LLM	5.17	<u>9.10</u>	34.09	19.75	5.41	<u>9.37</u>	<u>34.86</u>	20.48	<u>1.93</u>	<u>2.90</u>	51.65	<u>39.14</u>	2.03	3.16	54.66	47.66
GCNGPT	6.64	12.24	42.46	25.32	6.58	12.23	40.19	24.88	2.24	3.48	51.05	45.37	2.37	3.80	56.24	47.66
GATGPT	5.66	10.39	37.36	21.60	5.92	10.55	37.83	22.39	1.95	2.94	49.26	39.43	2.07	3.23	<u>52.54</u>	41.70
OFA	5.60	10.14	37.39	21.36	5.82	10.42	36.67	22.00	1.96	2.97	49.64	39.68	2.06	3.21	53.55	41.70
PT-STD	4.97	8.59	<u>34.80</u>	18.98	5.15	8.93	33.61	19.48	1.91	2.85	50.12	38.63	2.03	<u>3.17</u>	52.16	40.95

具体实验结果表明,PT-STD 在不同场景的交通预测任务中展现出全面领先的优势。在 NYC Taxi 场景下,模型在 Drop-off 和 Pick-up 任务中的 MAE 均取得最优,MAE 分别达到 4.97 与 5.15,较基线模型 ST-LLM 分别降低 3.9% 和 4.8%;在最具挑战性的 Drop-off 任务中,模型的 WAPE 指标显著优于其他方法,较传统时空模型 DCRNN 优化 4.3%。在 CHBike 场景下,PT-STD 同样保持优势,Drop-off 任务的 RMSE 较次优模型 ST-LLM 降低了 1.7%,这充分验证了其在小规模交通系统中的适用性。与基于图神经网络的基线模型 DGCRN 和 AGCRN 相比,PT-STD 在 4 个预测任务中的平均误差降低了 8%,这些都充分证明了 PT-STD 创新性的

时空建模机制的有效性。

进一步分析表明,PT-STD 在 MAE 等误差指标上的显著改进,主要得益于模型结构与训练策略的协同优化:一方面,时空解耦模块通过独立处理空间拓扑、周期性变化与全局时间趋势演化特征,提升了特征的分离性与表达能力;另一方面,分层微调机制通过渐进式解冻策略有效缓解了预训练模型迁移过程中的表征漂移问题,显著增强了适配能力与训练稳定性。两者联合作用,全面提升了模型对复杂时空依赖结构的建模能力,有效降低了模型的整体预测误差。

少样本预测:为探究 PT-STD 在数据稀缺场景下的领域适应能力,本研究构建了基于 20% 训练样本的少样本学习

实验,具体结果如表 3 所列。

由实验结果可知,PT-STD 在少样本学习场景下展现出卓越的领域适应能力。在 NYC Taxi 场景下,PT-STD 相较于最优基线模型实现了显著提升,在 Drop-off 任务中的平均绝对误差从 ST-LLM 模型的 8.50 降低至 5.46;在 Pick-up 任务中,其 WAPE 从 30.64% 优化至 22.04%。PT-STD 在 CHBike 场景下同样保持优势,Drop-off 任务的 RMSE 从 GCNGPT 的 4.30 降低至 3.65;而且在较难的 Pick-up 任务

中,PT-STD 比最优基线模型 ST-LLM 的平均绝对误差降低了 3.89%,充分验证了模型在少样本条件下的稳定泛化能力。跨领域实验结果对比显示,PT-STD 在复杂出租车任务中的性能提升幅度显著高于相对简单的共享单车场景,这表明模型的时空自适应机制能有效应对不同领域的分布差异。这些结果充分证明,通过时空解耦与分层微调策略,PT-STD 可以有效解决数据稀缺场景下的领域适应问题。

表 3 少样本实验预测结果

Table 3 Prediction results of few-shot experiments

Metric	NYC Taxi Drop-off				NYC Taxi Pick-up				CHBike Drop-off				CHBike Pick-up			
	MAE	RMSE	MAPE/%	WAPE/%	MAE	RMSE	MAPE/%	WAPE/%	MAE	RMSE	MAPE/%	WAPE/%	MAE	RMSE	MAPE/%	WAPE/%
ST-LLM	8.50	18.61	70.20	32.44	8.09	15.39	62.09	30.64	2.49	4.11	57.07	50.28	2.57	4.49	56.62	51.79
GATGPT	11.26	24.36	92.13	43.09	9.67	18.30	76.23	36.60	2.88	4.82	59.81	58.33	2.94	5.16	59.44	59.25
GCNGPT	11.05	23.49	73.30	42.18	10.52	20.15	66.32	39.81	2.64	4.30	62.42	53.44	2.67	4.62	66.74	53.88
PT-STD	5.46	9.43	38.37	20.84	5.83	9.99	38.54	22.04	2.28	3.65	52.17	46.14	2.47	4.19	54.18	49.83

4.5 消融对比分析

1) 消融研究

为验证 PT-STD 架构各核心组件的必要性,采用系统性消融策略,分别在 CHBike 和 NYC Taxi 场景下评估了时

空表征解耦模块、轻量化 MLP 编码器以及 STAHF 模块的贡献。消融实验结果如图 2 和表 4 所示。通过对比完整模型与各模块缺失配置的性能差异,深入揭示了模型架构设计的有效性。

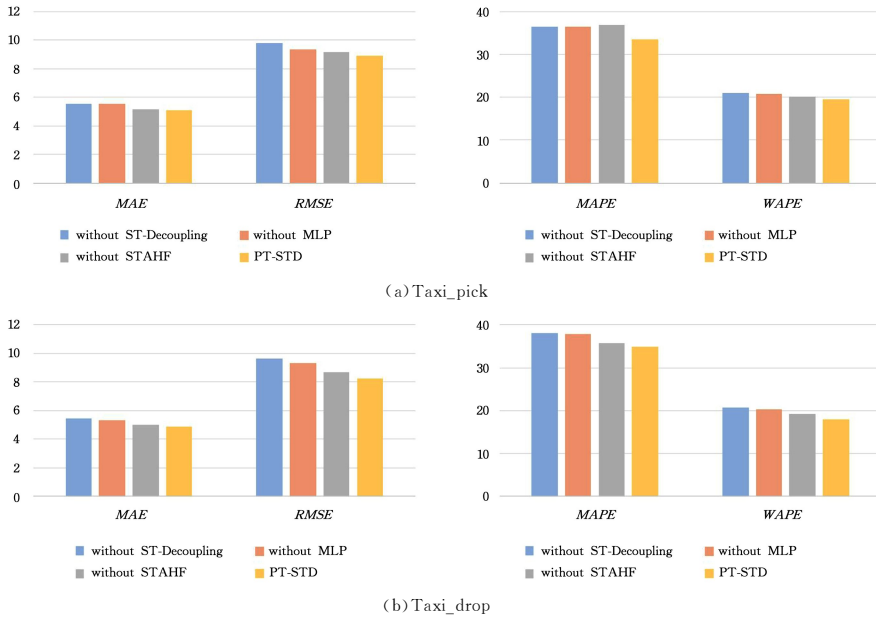


图 2 模型在 NYC Taxi 任务下的消融结果

Fig. 2 Ablation results of the model under NYC Taxi task

表 4 模型在 CHBike 任务下的消融结果

Table 4 Ablation results of the model under the CHBike task

时空表征 解耦	MLP 模块	STAHF 模块	Bike_pick				Bike_drop			
			MAE	RMSE	MAPE/%	WAPE/%	MAE	RMSE	MAPE/%	WAPE/%
—	✓	✓	2.04	3.19	53.55	41.26	1.94	2.94	51.10	39.34
✓	—	✓	2.05	3.20	53.96	41.33	1.92	2.87	50.27	38.93
✓	✓	—	2.08	3.32	52.74	42.09	1.97	2.99	50.23	39.83
✓	✓	✓	2.03	3.17	52.16	40.95	1.91	2.85	50.12	38.63

实验结果系统地验证了 PT-STD 架构中各功能模块的协同优化机制。在共享单车场景中,完整模型在 Pick-up 任务上取得最优性能指标,其中时空表征解耦模块对预测精度的提升较为显著,使 RMSE 指标降低 4.7%。在出租车场景中,

完整模型在 Drop-off 任务中的 MAE 较缺失 STAHF 模块的配置降低 2.8%,且 WAPE 实现了 1.7% 的优化。同时,STAHF 模块在跨任务场景中具有不可替代的作用,其缺失会导致 Pick-up 的 MAPE 指标恶化 3.38%,这一结果充分验

证了 STAHF 模块的有效性。相比之下,MLP 模块则主要影响特征非线性转换能力,这个模块的缺失会使 Drop-off 任务的 RMSE 指标上升 4.9%。实验表明,3 个模块相互影响,形成互补,时空表征解耦模块实现宏观时空特征的分离,MLP 实现微观特征增强与非线性映射,而 STAHF 则负责跨域知识的自适应迁移与融合,三者缺一不可。

2) 性能分析

图 3 给出了 PT-STD 模型与其他模型的效率对比。实验结果表明,PT-STD 在计算效能维度上均展现出显著优势。在训练时间和推理效率维度上,PT-STD 模型在 4 个基准任务中的表现均显著优于 GATGPT。在共享单车场景的下车任务中,PT-STD 的训练与推理效率较 ST-LLM 模型具有明显提升;在上车任务中,PT-STD 的训练效率略低于 ST-LLM 的训练效率。在出租车场景中的下车与上车任务中,PT-STD 的训练效率略低于 GCNGPT 和 ST-LLM 的训练效率。此外,PT-STD 在不同交通特征和流向任务中均保持稳定的时间消耗波动,这充分验证了其架构设计的鲁棒性。

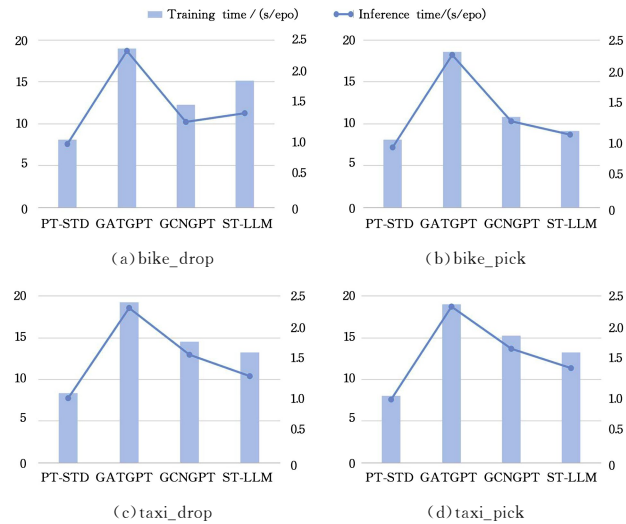


图 3 模型性能对比

Fig. 3 Comparison of model performance

结束语 本研究提出了一种基于预训练时空解耦的交通流预测模型 PT-STD,通过时空解耦架构与自适应分层微调机制的协同设计,实现了在交通流场景下的深度适配。实验结果表明,PT-STD 模型在 4 个基础数据集需求预测任务中,相较于对比模型都有显著提升,同时在少样本场景下仍能保持较高的预测精度。未来将进一步探索预训练模型在多元交通场景下的泛化能力,同时结合实时数据流与增量学习机制,拓展在城市交通动态管控中的应用。

参考文献

[1] AOUEDI O, LE V A, PIAMRAT K, et al. Deep learning on network traffic prediction: Recent advances, analysis, and future directions[J]. *ACM Computing Surveys*, 2025, 57(6): 1-37.

[2] LI J, LIN F, HAN G, et al. PAG-TSN: Ridership demand forecasting model for shared travel services of smart transportation [J]. *IEEE Transactions on Intelligent Transportation Systems*,

2023, 24(12): 15876-15889.

[3] ULVI H, YERLIKAYA M A, YILDIZ K. Urban traffic mobility optimization model: A novel mathematical approach for predictive urban traffic analysis[J]. *Applied Sciences*, 2024, 14(13): 5873.

[4] LI K, ZHAN T, FU K, et al. MergeNet: Knowledge Migration across Heterogeneous Models, Tasks, and Modalities[C]// *Proceedings of the AAAI Conference on Artificial Intelligence*. 2025: 4824-4832.

[5] BOMMASANI R, HUDSON D A, ADELI E, et al. On the opportunities and risks of foundation models[J]. *arXiv*: 2108.07258, 2021.

[6] SU J C, MAJI S, HARIHARAN B. When does self-supervision improve few-shot learning? [C]// *European Conference on Computer Vision*. Cham: Springer, 2020: 645-666.

[7] MIN B, ROSS H, SULEM E, et al. Recent advances in natural language processing via large pre-trained language models: A survey[J]. *ACM Computing Surveys*, 2023, 56(2): 1-40.

[8] LIU Y, OTT M, GOYAL N, et al. Roberta: A robustly optimized bert pretraining approach[J]. *arXiv*: 1907.11692, 2019.

[9] QIU X, SUN T, XU Y, et al. Pre-trained models for natural language processing: A survey[J]. *Science China Technological Sciences*, 2020, 63(10): 1872-1897.

[10] HUSSAIN B, AFZAL M K, AHMAD S, et al. Intelligent traffic flow prediction using optimized GRU model[J]. *IEEE Access*, 2021, 9: 100736-100746.

[11] DAI G, MA C, XU X. Short-term traffic flow prediction method for urban road sections based on space-time analysis and GRU [J]. *IEEE Access*, 2019, 7: 143025-143035.

[12] HUSSAINA H A, TAHER M A, MAHMOOD O A, et al. Urban traffic flow estimation system based on gated recurrent unit deep learning methodology for Internet of Vehicles [J]. *IEEE Access*, 2023, 11: 58516-58531.

[13] WEN Y, XU P, LI Z, et al. RPCConvformer: A novel Transformer-based deep neural networks for traffic flow prediction[J]. *Expert Systems with Applications*, 2023, 218: 119587.

[14] REZA S, FERREIRA M C, MACHADO J J M, et al. A multi-head attention-based transformer model for traffic flow forecasting with a comparative analysis to recurrent neural networks [J]. *Expert Systems with Applications*, 2022, 202: 117275.

[15] LUO Q, HE S, HAN X, et al. LSTTN: A long-short term transformer-based spatiotemporal neural network for traffic flow forecasting[J]. *Knowledge-Based Systems*, 2024, 293: 111637.

[16] ZHOU B, LIU J, CUI S, et al. A large-scale spatio-temporal multimodal fusion framework for traffic prediction[J]. *Big Data Mining and Analytics*, 2024, 7(3): 621-636.

[17] GUO C, HWANG F J, CHEN C H, et al. Signal Decoupling Optimization for Robust Graph-Based Traffic Forecasting[J]. *IEEE Transactions on Industrial Informatics*, 2025, 21(8): 6158-6168.

[18] FANG Y, QIN Y, LUO H, et al. When spatio-temporal meet wavelets: Disentangled traffic forecasting via efficient spectral graph attention networks[C]// *2023 IEEE 39th International*

- Conference on Data Engineering(ICDE). IEEE,2023:517-529.
- [19] SHAO Z,ZHANG Z,WEI W,et al. Decoupled dynamic spatial-temporal graph neural network for traffic forecasting[J]. arXiv:2206.09112,2022.
- [20] LI G,ZHONG S,DENG X,et al. A lightweight and accurate spatial-temporal transformer for traffic forecasting [J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 35(11):10967-10980.
- [21] HE K,ZHANG X,REN S,et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016:770-778.
- [22] LI Y,WANG N,SHI J,et al. Revisiting batch normalization for practical domain adaptation[J]. arXiv:1603.04779,2016.
- [23] CHEN Y,WANG X,XU G. Gatgpt:A pre-trained large language model with graph attention network for spatiotemporal imputation[J]. arXiv:2311.14332,2023.
- [24] LIU C,YANG S,XU Q,et al. Spatial-temporal large language model for traffic prediction[C]//2024 25th IEEE International Conference on Mobile Data Management(MDM). IEEE,2024:31-40.
- [25] YU B,YIN H,ZHU Z. Spatio-temporal graph convolutional networks:A deep learning framework for traffic forecasting[J]. arXiv:1709.04875,2017.
- [26] LI Y,YU R,SHAHABI C,et al. Diffusion convolutional recurrent neural network:Data-driven traffic forecasting[J]. arXiv:1707.01926,2017.
- [27] 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.
- [28] 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.
- [29] 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.
- [30] 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.
- [31] ZHOU T,NIU P,SUN L,et al. One fits all:Power general time series analysis by pretrained lm[J]. Advances in Neural Information Processing Systems,2023,36:43322-43355.



LI Jing, born in 2000, postgraduate. Her main research interests include artificial intelligence, deep learning and traffic flow prediction.



DU Shengdong, born in 1981, Ph.D, associate professor, Ph.D supervisor, is a member of CCF (No. 73290M). His main research interests include artificial intelligence, machine learning and knowledge engineering.

(责任编辑:柯颖)