



ELSEVIER

Contents lists available at ScienceDirect

Engineering

journal homepage: www.elsevier.com/locate/eng



Research
Engineering Management—Article

基于物理引导深度学习的突发公共卫生事件期间城市轨道交通短时OD客流预测

章树鑫^a, 张金雷^{a,*}, 杨立兴^a, 陈峰^b, 李树凯^a, 高自友^a

^a School of Systems Science, Beijing Jiaotong University, Beijing 100044, China

^b School of Civil Engineering, Beijing Jiaotong University, Beijing 100044, China

ARTICLE INFO

Article history:

Received 16 November 2023

Revised 8 April 2024

Accepted 30 April 2024

Available online 22 May 2024

关键词

短期起点-终点需求预测

城市轨道交通

疫情

物理引导深度学习

摘要

准确的起点-终点(OD)需求预测对于城市轨道交通(URT)系统的高效运行与管理至关重要,尤其是在疫情期间。然而,该任务面临诸多限制,包括实时可用性差、数据稀疏性、高维性问题以及疫情所带来的影响。因此,本研究提出了一个统一框架,称为物理引导的自适应图时空注意力网络(PAG-STAN),用于疫情背景下的地铁OD需求预测。具体而言,PAG-STAN引入了一个实时OD估计模块,用于估算实时完整的OD需求矩阵。随后,提出了一种新的动态OD需求矩阵压缩模块,用以生成密集的实时OD需求矩阵。此后,PAG-STAN利用多种异构数据来学习疫情期间未来OD客流的演变趋势。最后,掩蔽物理导向损失函数(MPG-loss function)将OD需求与进站流量之间的物理量信息合并到损失函数中,以增强模型的可解释性。疫情和常规情景下PAG-STAN在两个真实世界的地铁OD需求数据集上均表现出良好的性能,凸显了其对地铁OD需求预测的稳健性和敏感性。为验证PAG-STAN各模块的重要性,本研究还开展了一系列消融实验。

© 2024 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. 引言

随着社会经济的快速发展,为满足日益增长的出行需求,城市轨道交通(URT)在一些国家已经实现大规模建设并投入运营。准确的起点-终点(OD)需求预测对于URT系统的高效与安全运行至关重要,因为它能够详细描绘乘客进出URT系统的时间和地点,对于运营方和乘客都具有重要意义。准确的地铁OD需求预测有助于运营方及时调整地铁运行方案,提高运营效率;同时也能帮助乘客提前规划出行路线,减少等车时间,提升出行体验。但是,与进出站客流预测相比,OD需求预测更有难度,

因为不仅需要预测目标车站的进出站客流量,还必须考虑整个URT网络中客流的时空分布。此外,疫情的暴发令地铁OD需求预测变得更加复杂。疫情发生后,人们在出行过程中会考虑更多安全因素,提高了OD需求的随机性和波动性[1–2]。因此,地铁运营方可能难以及时调整列车运行计划,导致运力过剩或不足。鉴于此,准确的地铁OD需求预测是一项关键且极具挑战性的任务,尤其在疫情期间,可能会面临以下问题:

(1) **OD需求数据的实时可用性:** 在地铁客流预测相关研究中[3–5],实时乘客出行数据可通过自动售检票系统(AFC)获取,实现实时客流预测。但是,在地铁OD

* Corresponding author.

E-mail address: zhangjinlei@bjtu.edu.cn (J. Zhang).

需求预测研究中, 每一个OD需求代表一次完整的出行行为, 只有当乘客完成整个旅程后才能获取该信息[6-7]。例如, 在预测08:00-08:30这个时间段的OD需求时, 07:30-08:00期间的完整OD需求数据可能无法获得, 因为仍有部分乘客尚未完成旅程。这种不完整的OD需求信息将会降低预测的准确性。因此, 尽可能获取完整的实时OD需求信息对于提升OD需求预测的准确性至关重要。

(2) **OD需求数据的稀疏性:** 由于城市中不同区域的功能差异, 不同区域之间的OD需求存在显著稀疏性。例如, 住宅区与中心商务区之间通常具有较高的OD需求, 而偏远郊区的OD需求则很低甚至为零。图1展示了南宁地铁单日不同时间段的OD数量分布, 大多数时段的OD需求都非常低或为零。这些OD需求使得OD需求矩阵呈现出高度稀疏性, 进而在训练过程中难以有效捕捉OD需求的时空分布特征[8-9]。此外, 图2展示了新冠病毒肺炎(COVID-19)疫情不同阶段的OD需求矩阵, 可以看到疫情期间OD需求矩阵的稀疏性显著上升, 这是因为人们担心乘坐地铁可能增加感染风险。因此, 有效应对OD需求矩阵的稀疏性问题, 是提升地铁OD需求预测准确性的关键, 尤其是在疫情背景下。

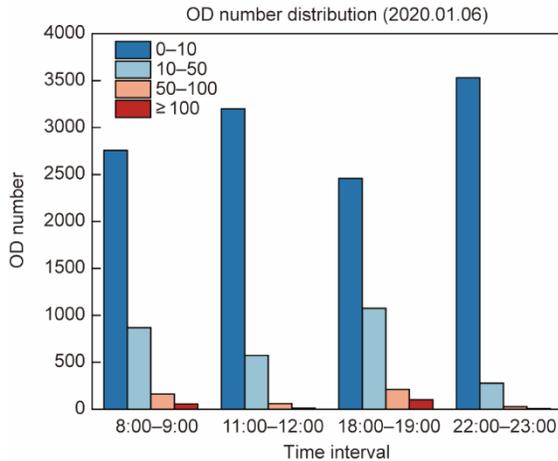


图1. 南宁地铁单日不同时段OD数量可视化图。

(3) **OD需求数据的高维性:** OD需求矩阵的维度为 $N \times N$, 远大于进出站流量矩阵的维度, 其中 N 代表URT系统中的车站数[8,10]。例如, 南宁地铁共有62个车站, 在预测下一个时间段所需的三千多个OD需求时, 要进行极其复杂的计算。因此, OD需求矩阵的高维性问题亟待解决。

(4) **疫情期间外部因素的影响:** 疫情的暴发导致URT系统中的OD需求突发不规律的变化, 给地铁OD需求预测带来巨大挑战。以新冠病毒肺炎疫情为例, 图2展

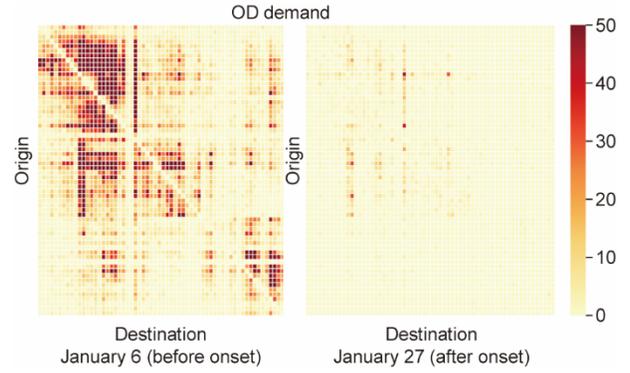


图2. 新冠病毒肺炎疫情不同阶段的OD需求矩阵可视化图。

示了疫情对客流的影响。疫情暴发前的OD需求明显高于疫情期间, 后者几乎接近于零。这种变化主要是因为疫情期间使用城市公共交通的传播风险和感染概率上升。疫情期间OD需求的不规律变化对现有预测模型提出了挑战, 因为传统模型难以有效捕捉多种外部因素对地铁OD需求的影响。因此, 要在疫情期间准确建模OD需求的变化过程, 必须全面理解外部因素对需求分布的影响。

近年来, 短期OD交通需求预测逐渐受到研究者的广泛关注。自Liu等[11]首次提出基于深度学习框架的出租车OD需求预测模型(符合实际环境的时空网络, CSTN)以来, 已有多位学者对OD需求预测展开了深入研究。例如, Noursalehi等[7]提出了一种多分辨率时空特征提取的深度学习模型, 用于地铁OD需求预测; Zou等[12]从起点和终点的角度分别学习OD对的时空相关性, 用于预测长期OD需求。这些方法弥补了交通OD需求预测领域的研究空白。然而, 对疫情期间的交通OD需求预测的研究仍然有限。此外, 这些研究似乎忽略了OD需求与进站客流之间的重要关联性。具体来说, 对于每个车站, 某一时段内的进站客流量应等于该车站在该时段内发出的所有OD需求之和。图3展示了OD需求与进站客流在某一特定时段内的数量关系: 以车站 S_1 为起点的所有OD对的需求总和等于车站 S_1 在该时段的进站流量, 即41人次。理解OD需求与进站流量之间的关系为OD需求预测提供了重要参考。因此, 考虑这一数量关系, 有助于提高模型的可解释性和预测准确性。

本研究通过引入物理引导的自适应图时空注意力网络(PAG-STAN), 解决了疫情期间地铁短期OD需求预测的问题。如前所述, 地铁OD需求的原始实时数据通常存在实时可用性差、数据稀疏性和高维性等问题。因此, PAG-STAN首先利用实时进站流量和长期历史分布比例, 尽可能实现完整的实时OD需求矩阵。随后, 引入了一种新颖的动态OD需求矩阵压缩模块, 该模块能够关注部分高需求OD

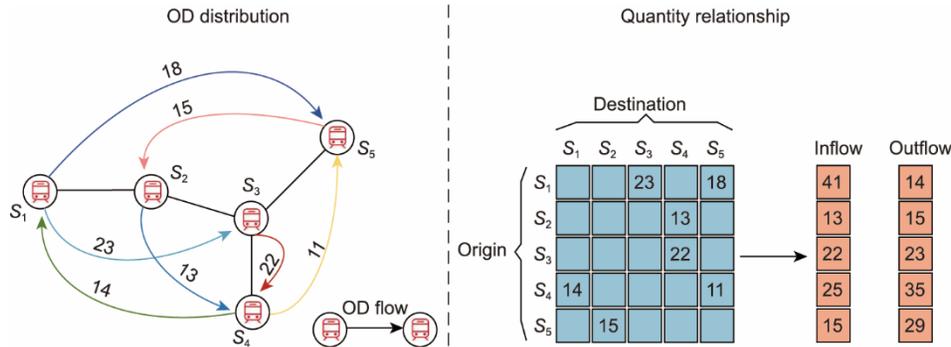


图3. 客流量之间的数量关系。S: 车站。

对, 从而生成稠密的OD需求矩阵。该模块采用编码-解码架构收集疫情期间地铁OD需求的时空关系。在编码器中, 引入了自适应图卷积长短期记忆网络 (AGC-LSTM) 和多周期交叉注意力机制 (MPC-ATTN), 以学习OD需求的周期性时空分布信息。在解码器部分, 采用了多个双向长短期记忆网络 (BiLSTM), 用于建模上下文信息, 从而解码OD需求的全局分布特征。此外, 异构信息融合模块 (HIFB) 引入了疫情相关数据、日期属性数据等异构数据, 用以研究疫情期间外部因素对OD需求的影响。本研究还提出了一种全新的掩码物理引导 (MPG) 损失函数, 将OD需求与进站流量之间的物理量关系嵌入损失函数中, 从而提升训练效率并增强模型的可解释性。最后, 在两组真实的地铁OD需求数据集上进行了大量实验。实验结果表明, PAG-STAN在疫情及常规场景下的地铁OD预测任务中均优于现有方法。因此, 本研究的主要贡献如下:

- 本文提出了一个全新的PAG-STAN模型, 用于提升疫情背景下的地铁OD需求短期预测能力, 有效解决了地铁OD预测中面临的数据实时性差、数据稀疏性、高维性以及疫情冲击等挑战。

- PAG-STAN引入了一种创新的动态OD需求矩阵压缩方法, 能够灵活地为每一个起始站点选择重要的需求特征信息, 从而生成一个稠密的OD需求矩阵。该方法在保留关键OD分布信息的同时, 有效应对了数据稀疏性和高维性问题。

- 为了更准确地预测疫情下的地铁OD需求, PAG-STAN融合了多种异构数据 (多周期OD需求数据、实时进站流量数据、疫情相关数据、日期属性数据等), 以全面挖掘OD分布信息。模型中设计了多个模块, 如AGC-LSTM、MPC-ATTN、BiLSTM以及HIFB, 以更好地捕捉疫情背景下地铁OD需求的时空依赖关系。

- 本研究提出了一种新颖的MPG损失函数, 通过引入OD需求与进站流量之间的物理量关系来引导模型训练, 在提升预测准确性的同时增强了模型的可解释性。

- 本研究在两个真实地铁OD需求数据集上进行了大量实验, 涵盖了疫情及常规场景, 结果显示, PAG-STAN在OD需求预测任务中表现优异。此外, 还开展了一系列消融实验, 以验证各个模块的重要性。

2. 文献综述

城市公共交通系统的短期客流预测在提升服务质量和运营效率方面发挥着关键作用[6]。在实际应用中, 短期预测主要包括两个方面: 客流量预测与OD需求预测。前者为运营方提供进出站/区域的客流信息, 而后者则提供交通网络中客流的时空分布信息。以下两小节将分别回顾短期客流量预测和OD需求预测的相关研究。

2.1. 短期进/出站客流量预测

短期进出站客流量预测在智能交通系统中有巨大应用潜力, 因此已成为研究热点, 其研究方法主要分为两种: 传统的统计方法和基于机器学习的方法。在早期研究中, 大多数学者采用统计方法进行交通流预测。Van der Voort等[13]使用Kohonen映射对交通流进行分类, 并对每类数据采用自回归积分滑动平均模型 (ARIMA) 进行预测。Tan等[14]提出了一种基于ARIMA与指数平滑的集成方法, 用于短期交通流预测。Ni等[15]引入了一种线性回归模型, 利用季节性ARIMA学习地铁在特定事件发生期间的客流变化。然而, 这些统计方法通常将交通流视为平稳时间序列, 忽略了其波动性和复杂性, 因而无法有效捕捉交通流的复杂时空特征。

近年来, 机器学习的快速发展有效弥补了统计模型的不足, 将支持向量回归 (SVR)、K近邻 (KNN) 以及线性回归等方法[16–19]广泛应用于短期客流预测任务中。例如, Li等[20]应用具有高斯损失函数的SVR模型, 以减少城市交通流数据中的随机误差。Hong [21]利用季节性SVR结合混沌模拟退火算法, 处理具有周期性趋势的城

际交通流预测问题。Cai等[22]提出了改进的KNN模型,基于时空状态矩阵之间的时空相关性提高预测准确率。Lin等[23]将带有时延的空间交通序列转化为交通状态向量,并结合SVR与KNN进行交通流预测。Avila和Mezić[24]则采用Koopman模式分解方法分析并预测交通动态,能够有效区分增长或衰减模式,并得出一系列此前未能识别的时间模式。这些基于机器学习的方法考虑了交通流的复杂时间依赖关系,因此在预测性能上有所提升。然而,这些模型难以有效捕捉空间依赖性,通常只能对单一站点的流量进行预测,不适用于整个交通网络的客流预测。

基于深度学习的方法是机器学习的新兴分支,已广泛应用于各个领域[25–28],促使研究人员将深度学习技术应用到交通流量预测任务。有研究[29–30]使用循环神经网络(RNN),如LSTM和门控循环单元(GRU),来建模交通流的时间依赖性;还有一些研究[31–32]则采用卷积神经网络(CNN)来建模交通流的空间依赖性。例如,Li等[3]将交通流建模为有向图上的扩散过程,并提出了一种扩散卷积循环神经网络(DCRNN),用以整合空间与时间的依赖性来进行交通流预测。Zhang等[4]提出了一种新的基于深度学习的方法,即时空残差网络(ST-ResNet),用于预测整个城市的人流量,并设计了残差卷积单元来建模空间属性与时间周期性。考虑到URT乘客流的不规则特性,Du等[33]将乘客流建模为多通道矩阵,并提出了一种深度非规则卷积残差网络来进行时空特征学习。然而,交通流预测并不仅仅是一个简单的时间序列处理问题,因为它还涉及复杂的空间交通网络结构[34]。

由于交通网络本质上是具有复杂拓扑特征的图结构,因此越来越多的研究开始采用图卷积网络(GCN)来捕捉交通流的空间依赖性。Yu等[35]将交通网络视为一般图,并开发了一种时空图卷积网络用于交通预测任务。Zhao等[5]提出了一种结合了GCN和GRU的T-GCN新模型,能够同时捕捉交通流的空间和时间依赖关系。然而,由于单一的静态图无法充分反映复杂和动态的空间依赖关系,一些研究人员[36–38]对GCN进行了扩展,以建模多重空间特征。Peng等[39]使用动态交通流概率图来建模交通网络,并使用GCN学习交通流的空间依赖性。Wu等[40]开发了一种新的自适应依赖矩阵,以自动挖掘交通流中的隐藏空间依赖关系。Wang等[41]基于地铁网络拓扑结构和出行模式构建了超图,用于表示客流的高阶空间关系。此外,基于transformer的方法在从多个潜在子空间中建模全局时间依赖性方面具有优势,所以这种方法[42–44]在交通流预测任务中也受到广泛关注。例如,Yan等[45]设计了一种新型的交通transformer模型,用于建模交

通流的动态和分层时空特征。Xie等[46]考虑了人群流动中的非线性时空依赖性,提出了一种新颖的多尺寸块状时空transformer网络,用于融合跨时空和跨尺寸的上下文信息依赖。Xu等[47]设计一种新型的元图transformer模型,用于解决建模复杂交通流动态的问题。上述所有方法都是为预测特定区域/车站的短期进出流量而设计的,仅反映各区域/车站的进出信息,未能全面描述交通网络中交通流的时空分布。为了进一步研究从区域到区域、从车站到车站的交通流动的时空模式,近年来一些学者开始进行OD需求预测研究。

2.2. 短期OD需求预测

OD需求指在特定时间段内,从起点到终点之间的交通流量,能够直接反映区域/站点之间交通流的时空分布。2019年,Liu等[11]提出了一种上下文感知的时空网络,用于预测城市区域间的出租车OD需求,这是解决OD需求预测问题的首次尝试。随后,越来越多的研究提出了基于深度学习的方法来应对OD需求预测的挑战。例如,Chu等[48]提出了一种新型的多尺度卷积LSTM(ConvLSTM)模型,用于在多个尺度上预测出租车需求;Hu等[49]将出行成本建模为一个分布,并提出了一个随机OD矩阵预测问题;为了解决空间OD流缺失的问题,Yao等[50]提出了一种空间交互图卷积网络,用于补充空间OD数据;Zou等[12]设计了一种图深度学习(GDL)模型STGDL,以挖掘时间变化的OD信息中复杂且动态的时空相关性。Ke等[51]将OD对视为节点,用于捕捉OD对之间各种隐藏的时空关系,并开发了一种时空编码器-解码器残差多图卷积网络(ST-EDRMGC),用于预测出租车OD需求。Huang等[52]通过计算预测结果,为多步预测提供未来需求的参考信息,并将预测结果和历史需求引入双向注意力机制中,以探索网约车需求在过去、现在和未来之间的时空相关性;随后,他们利用生成对抗网络(GAN)结构来克服网约车需求数据的高度稀疏性问题[53]。

以上讨论的所有方法主要应用于出租车或网约车的需求预测,而这类任务并不需要考虑OD需求数据的实时可用性。但是,在进行地铁OD需求预测时,实时OD需求数据只有在乘客完成出行后才能获得,因此预测极具挑战性。近年来,学者们尝试解决这一问题。例如,为了解决OD流信息部分可观测的问题,Jiang等[6]提出了一种新的OD流重构机制,以推断完整的OD流信息。Zhang等[8]关注OD数据的可用性问题,提出了一种基于通道的注意力分离的CNN来解决相关限制。Noursalehi等[7]则提出了一种新型多分辨率时空神经网络,通过利用基于出口的

实时OD需求, 预测未来的完整OD需求数据, 从而解决了实时数据可用性的问题。考虑到高维性和实时OD信息不完整的问题, Zhu等[10]提出了一个两阶段OD需求预测方法, 先预测进站量和分离率, 再通过两者相乘得到未来的OD需求。Liu等[9]则通过结合多个历史数据集, 有效解决了OD需求数据的实时可用性问题。

表1 [6–12,48–53]总结了出租车和地铁OD需求预测的主要方法。尽管这些方法在预测性能上取得了一定成效, 但仍存在局限性。首先, 所有方法均未充分考虑前述问题, 如实时可用性、稀疏性、高维性以及客流量关系。其次, 所有方法均只关注常规场景, 忽略了在非正规干扰场景(如疫情)下的OD需求预测需求。这些干扰场景可能

导致OD需求分布发生突发性和不规则变化, 进而引发运力不足或资源浪费等问题。因此, 运营方必须更加关注在非正规场景下的OD需求预测。

本研究重点关注疫情期间URT系统的OD需求预测, 旨在提出一个统一框架, 不仅能有效应对前述问题, 还可作为非正规干扰场景下OD需求预测提供参考。

3. 基础理论

本节首先定义若干基本概念, 以便在疫情背景下对URT系统中的OD需求预测问题进行形式化表述。

定义1 (URT网络): 本研究重点关注URT系统中的

表1 不同方法在整个URT网络上的预测性能

References	Research context	Real-time data availability	Data sparsity	Data high-dimension	Quantity relationship	External factors	Input	Models
Liu et al. [11] (2019)	Taxi demand	/	×	×	×	Weather data	Real-time complete OD	CNN, ConvLSTM
Chu et al. [48] (2020)	Taxi demand	/	×	√	×	Time of day, weather data	Real-time complete OD	ConvLSTM
Hu et al. [49] (2020)	Taxi demand	/	√	×	×	×	Sparse OD stochastic speed tensors	GCN, RNN
Yao et al. [50] (2021)	Taxi demand	/	√	×	×	×	Spatial OD flow	GCN
Zou et al. [12] (2021)	Taxi demand	/	×	√	×	Weather data	Historical/real-time complete OD	CNN, GCN
Ke et al. [51] (2021)	Taxi demand	/	×	×	×	×	Historical/real-time complete OD	GCN, LSTM
Huang et al. [52] (2022)	Taxi demand	/	×	×	×	×	Real-time/pre-predicted OD	Attention mechanism
Huang et al. [53] (2022)	Taxi demand	/	√	×	×	×	Historical/real-time complete OD	GAN, GCN
Zhang et al. [8] (2021)	UTR OD demand	√	√	√	×	×	Historical complete OD, Real-time inflow/outflow	CNN, Attention
Jiang et al. [6] (2022)	URT OD demand	√	×	×	×	×	Historical complete OD, Real-time incomplete OD	LSTM, GCN
Noursalehi et al. [7] (2022)	URT OD demand	√	×	×	×	Time of day, weather data	Real-time complete exit-based OD	Discrete wavelet transform, CNN, ConvLSTM
Zhu et al. [10] (2023)	URT OD Demand	√	×	√	×	×	Real-time inflow, historical separation rate	BiLSTM
Liu et al. [9] (2022)	URT OD Demand	√	√	√	×	×	Real-time incomplete OD, Unfinished order, DO	Graph convolution GRU (GCGRU), Transformer
Our model	URT OD demand	√	√	√	√	Date attribute, pandemic-related data	Historical complete OD, Distribution rate, real-time inflow	AGC-LSTM, Transformer

“/” denotes the issue does not exist in specific scenarios, “×” denotes the issue has not been considered, and “√” denotes the issue has been considered.

OD需求预测。为了表示各站点之间的空间关联性，URT网络被定义为 $G=(S, E, A)$ ，其中： $S=\{s_1, s_2, \dots, s_N\}$ 表示车站的集合， N 是URT网络中车站的数量。 $E \in Z^{N \times N}$ 表示边的集合，其中 Z^+ 表示正整数， $e_{ij} \in E$ 是一个二值变量，用于表示车站 S_i 和车站 S_j 在物理网络中是否相邻。 $A \in R^{N \times N}$ 表示URT网络的加权邻接矩阵，其中 R 表示实数， $a_{ij} \in A$ 是通过将 E 的每一行进行线性归一化后得到的值。通过这种表示方式，我们能够捕捉车站之间的空间关系，并将该信息有效地融入OD需求预测模型中。

定义2 (进站客流矩阵)：基于地铁AFC记录，可以提取历史进站客流序列，并将其按照不同的时间粒度（如10 min或60 min）整合为进站客流矩阵。设 $p_t^{\text{in}}(i)$ 表示在时间间隔 t 内第 S_i 个车站的进站客流，则 $\mathbf{P}_t^{\text{in}} \in Z^{N \times 1}$ 表示一个进站客流向量，反映了所有车站在时间间隔 t 内的进站客流情况。在时间间隔 t ，包含过去 T 个时间步长的进站信息的进站客流矩阵可以表示为： $\mathbf{P}_t^{\text{in}}(T) = [\mathbf{P}_{t-T+1}^{\text{in}}, \mathbf{P}_{t-T+2}^{\text{in}}, \dots, \mathbf{P}_t^{\text{in}}] \in Z^{N \times T}$ ，公式表示如下：

$$\mathbf{P}_t^{\text{in}}(T) = \begin{pmatrix} p_{t-T+1}^{\text{in}}(1) & p_{t-T+2}^{\text{in}}(1) & \cdots & p_t^{\text{in}}(1) \\ p_{t-T+1}^{\text{in}}(2) & p_{t-T+2}^{\text{in}}(2) & \cdots & p_t^{\text{in}}(2) \\ \vdots & \vdots & \vdots & \vdots \\ p_{t-T+1}^{\text{in}}(N) & p_{t-T+2}^{\text{in}}(N) & \cdots & p_t^{\text{in}}(N) \end{pmatrix} \quad (1)$$

定义3 (OD需求矩阵)：设 $x_t(i, j)$ 表示在时间间隔 t 内，从车站 S_i 到车站 S_j 的完整OD需求，其中 $i, j = 1, 2, \dots, N$ 。则完整的OD需求矩阵 $\mathbf{X}_t \in Z^{N \times N}$ 被定义为在时间间隔 t 内所有车站之间的OD需求情况，公式表示如下：

$$\mathbf{X}_t = \begin{pmatrix} x_t(1, 1) & x_t(1, 2) & \cdots & x_t(1, N) \\ x_t(2, 1) & x_t(2, 2) & \cdots & x_t(2, N) \\ \vdots & \vdots & \vdots & \vdots \\ x_t(N, 1) & x_t(N, 2) & \cdots & x_t(N, N) \end{pmatrix} \quad (2)$$

值得注意的是，地铁OD需求与所有车站在单位时间内的进站客流之间存在一个物理量关系，其可形式化表示如下：

$$\mathbf{P}_t^{\text{in}} = \begin{pmatrix} p_t^{\text{in}}(1) \\ p_t^{\text{in}}(2) \\ \vdots \\ p_t^{\text{in}}(N) \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^N x_t(1, j) \\ \sum_{j=1}^N x_t(2, j) \\ \vdots \\ \sum_{j=1}^N x_t(N, j) \end{pmatrix} \quad (3)$$

定义4 (多时间序列OD需求)：本研究考虑了历史多时间序列OD需求数据，包括按周和按日的时间序列OD，用于研究OD需求的长期与短期周期性分布。具体而言，设当前时间间隔为 t ，每天的总时间间隔数为 t_{day} ，则周期性时间序列OD $\mathbf{X}_{t-7 \times t_{\text{day}}}$ 表示的是上周同一时间段的OD需

求。日度时间序列 $\mathbf{X}_{t-t_{\text{day}}}$ 表示的是前一日同一时间段的OD需求。

问题：基于在时间间隔 t 观测到的OD需求与进站客流数据，本研究旨在学习一个函数 $f(\cdot)$ ，用于预测URT系统中下一个时间间隔 \mathbf{X}_{t+1} 的完整OD需求。

$$\mathbf{X}_{t+1} = f(\mathbf{X}_{t-7 \times t_{\text{day}}}, \mathbf{X}_{t-t_{\text{day}}}, \mathbf{X}_{t-T+1}, \dots, \mathbf{X}_t, \mathbf{P}_t^{\text{in}}(T), \mathbf{G}) \quad (4)$$

4. 方法论

为准确预测疫情背景下URT系统中的OD需求，本文提出了一种新颖的PAG-STAN模型。该模型利用多种异构数据（即多尺度的历史时间序列OD需求、实时进站数据、疫情相关数据和日期属性数据）来建模疫情下OD需求的时空演化模式。图4展示了PAG-STAN的架构，由五个模块组成：实时OD估计模块、动态OD需求矩阵压缩模块、编码器、解码器以及MPG损失函数。

实时OD估计模块：为解决实时数据获取的有限性，该模块通过将实时进站数据与历史长短期OD分布率相乘，估计出实时OD需求，从而充分融合了历史OD按周/日分布的信息。

动态OD需求矩阵压缩模块：原始OD需求矩阵的稀疏性和高维性为模型预测带来了巨大挑战。对此，本文提出了一种创新的动态OD需求矩阵压缩模块，用于筛选部分高需求的OD对，从而生成一个稠密的OD需求矩阵。该模块在保留完整OD分布信息的同时，有效降低了OD矩阵的稀疏性与高维性。

编码器：在编码器部分，PAG-STAN引入了AGC-LSTM与MPC-ATTN模块，用于挖掘OD需求中隐藏的周期性时空特征信息。为捕捉OD需求的显著周期性特征，模型引入了周时间序列OD、日时间序列OD以及实时OD需求数据。这些多时间序列OD数据分别输入到三个AGC-LSTM分支中，以提取每种时间序列OD需求数据的隐藏时空分布特征。随后，MPC-ATTN模块对提取到的周期性时空信息进行编码，生成特征表示。最终，这些特征表示被输入一个带有残差操作的卷积网络中，从而增强周期性时空特征表示。

解码器：在解码器中，模型使用多个BiLSTM对增强后的周期性时空特征表示进行正向与反向解码。同时，为研究疫情期间外部因素对地铁OD需求的影响，模型引入HIFB模块，用于将OD需求数据与多种外部数据（如疫情相关数据、日期属性数据）进行融合。该模块能够有效学习有意义的信息，从而进一步捕捉OD需求的演化特

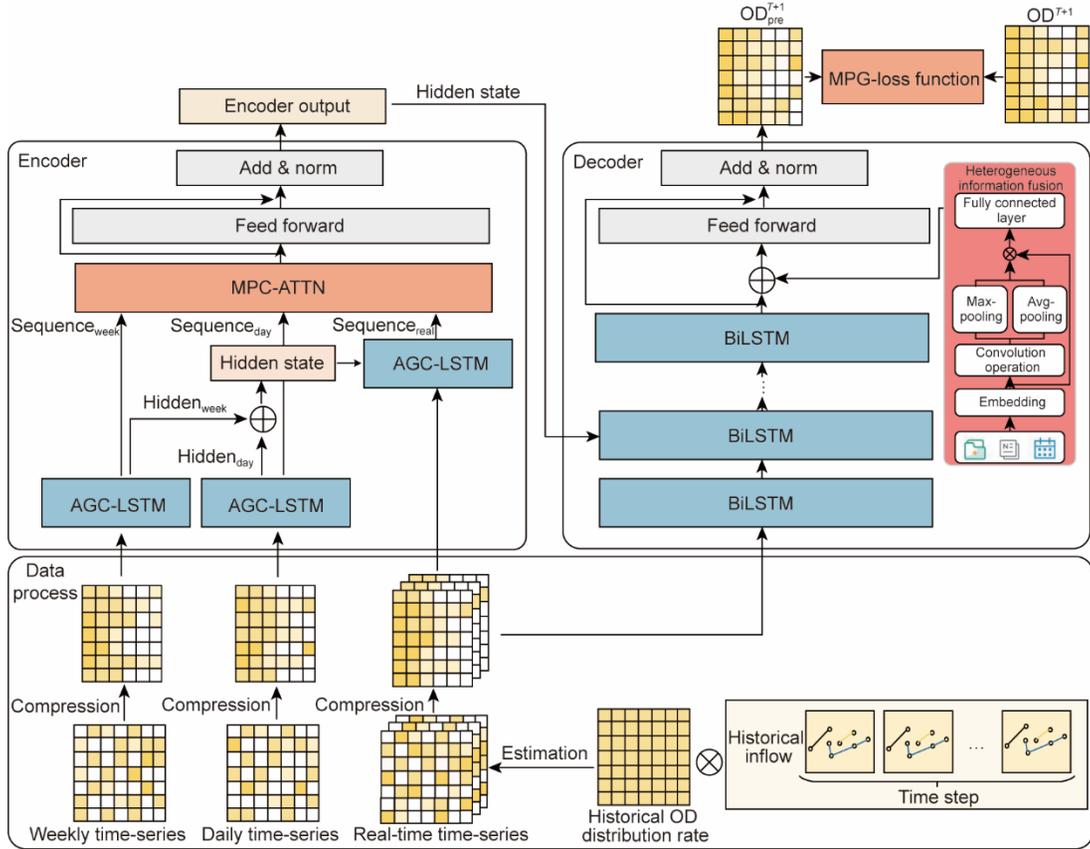


图4. PAG-STAN 框架概览。Max: 最大值; avg: 平均值; add: 加法; norm: 归一化; pre: 前一 (指前一时间步)。

征。最终，通过堆叠若干全连接层，模型预测出下一时间间隔内完整且稠密的OD需求矩阵。

MPG损失函数：为指导模型训练，本文提出了一种新颖的MPG损失函数，该函数考虑了进站客流与OD需求之间的物理量关系。这种方法在保留数据驱动模型高预测准确性的同时，提高了OD需求预测模型的可解释性。此外，MPG损失函数还引入了掩码操作，从而消除动态压缩后的OD需求矩阵中零填充值对训练造成的干扰。

4.1. 实时OD估计模块

由于部分出行在当前时间段内尚未完成，所以直接从AFC记录中提取的实时OD需求通常是不完整的。为此，实时OD估计模块利用实时进站数据和历史目的地分布率来近似生成一个较为完整的OD需求矩阵。这是因为实时进站数据能够全面反映每个出发车站的实时进站情况，而历史终点分布率则包含了OD需求中有意义的周期性分布特征信息。

图5展示了实时OD估计的过程。历史长短期终点分布率分别表示前一周相同时段和前一日相同时段的终点分布，这可以反映由于不同出行活动所引起的出行生成/吸引变化导致的各个OD对之间逐周/日的的需求波动情况

[54]。两类分布率均用于估计在时间间隔 t 时的长/短期潜在OD需求矩阵，分别表示为 $\hat{X}_t^l/\hat{X}_t^s \in R^{N \times K}$ 。考虑历史的长/短期目的地分布率 $DR_t^l/DR_t^s \in R^{N \times K}$ 以及目标出发站的实时进站流量 P_t ，这些潜在OD需求矩阵可按如下方式计算：

$$\hat{X}_t^s(i, j) = P_t(i) \times DR_t^s(i, j) \quad (5)$$

$$\hat{X}_t^l(i, j) = P_t(i) \times DR_t^l(i, j) \quad (6)$$

获得这两个潜在的OD需求矩阵后，将其拼接在一起，并输入到一个 1×1 卷积单元中，从而生成一个近似完整的周期性OD分布的实时OD需求矩阵。

$$\hat{X}_t = \text{Conv}(\text{Concat}[\hat{X}_t^s, \hat{X}_t^l]) \quad (7)$$

4.2. 动态OD需求矩阵压缩模块

原始OD需求矩阵具有高度稀疏性和高维性，给OD需求预测带来巨大挑战。一方面，稀疏的OD需求矩阵会干扰OD需求的时空分布建模；另一方面，在处理高维OD需求矩阵时，模型也会产生较高的计算成本[55]。因此，本文提出了一种动态OD需求矩阵压缩模块，用于筛选部分高需求的OD对，以生成稠密的OD需求矩阵。如图6所示，对于每个起始站点 S_i ，首先计算其通向所有终

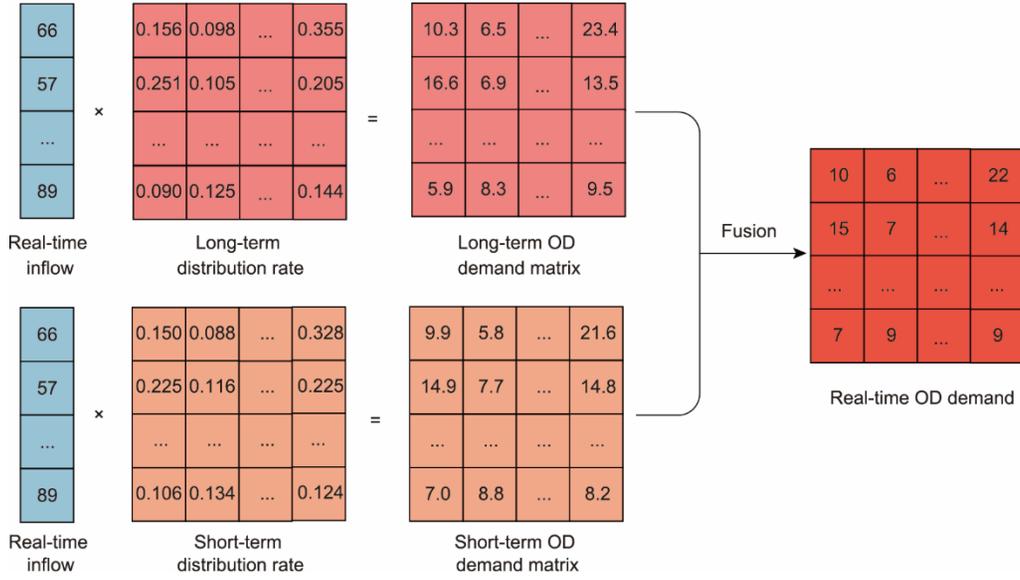


图5. 实时OD估计流程图。

点站的终点分布率 $\mathbf{DR}(i) = [DR_1(i), DR_2(i), \dots, DR_N(i)]$

$$\mathbf{DR}(i) = \begin{pmatrix} DR_1(i) \\ DR_2(i) \\ \vdots \\ DR_N(i) \end{pmatrix}^T = \begin{pmatrix} \sum_t x_t(i, 1) / \sum_t p_t^{\text{in}}(i) \\ \sum_t x_t(i, 2) / \sum_t p_t^{\text{in}}(i) \\ \vdots \\ \sum_t x_t(i, N) / \sum_t p_t^{\text{in}}(i) \end{pmatrix}^T \quad (8)$$

根据终点分布率对终点站进行排序，并依据预设的客流比例阈值（PFP）选择第 $k_i - 1$ 个具有较高分布率的终点站。需要注意的是，可以预设该阈值 PFP，从而控制压缩后矩阵的稀疏程度。本研究中，将该阈值设定为 70%，用于地铁 OD 需求预测。后续的实验中也将通过调整 PFP 来探究模型在不同稀疏程度下的预测性能。终点筛选过程可近似表示为以下最小化问题：

$$\begin{aligned} & \text{Minimize } |j| \\ & \text{s.t. } \sum_{j \in N} DR_j(i) \geq \text{PFP} \end{aligned} \quad (9)$$

式中， $|j| = k_i - 1$ 表示被选中的终点站数量； k_i 表示起点站 S_i 所对应的 OD 特征数量，不同的起点站 S_i 具有不同的 k_i 。在筛选出高需求的终点站后，目标站点与终点站之间的

OD 需求被视为 $k_i - 1$ 个 OD 特征。其余站点的 OD 需求总和和被归为第 k_i^{th} 个 OD 特征。由于不同起点站的 OD 特征数量不同，作者使用 0 来填充 OD 特征向量的空缺，从而确保每个 OD 特征向量的维度均为 K ，其中 $K = \text{Max}(k_1, k_2, \dots, k_N)$ 。由此，可以在时间间隔 t 生成一个稠密的 OD 矩阵 $\hat{\mathbf{X}}_t \in R^{N \times K}$ 。值得注意的是，由于填充的 0 值会干扰预测，因此在模型训练过程中对其进行了掩码处理，防止误差反向传播。因此，该模块在保留 OD 分布信息的同时，有效地降低了 OD 需求矩阵的稀疏性与高维性。

4.3. 编码器

在获得稠密的 OD 需求矩阵后，本文提出了一个通用的编码器-解码器框架。其中，编码器旨在从多重时间序列 OD 需求中挖掘隐藏的周期性时空分布信息。为此，本文提出了一种新的 AGC-LSTM 结构和创新的 MPC-ATTN 机制，用于融合多重时间序列 OD 需求，从而对其内在的周期性时空分布特征进行编码。

4.3.1. AGC-LSTM

URT 系统本质上是一个具有复杂空间拓扑结构的动

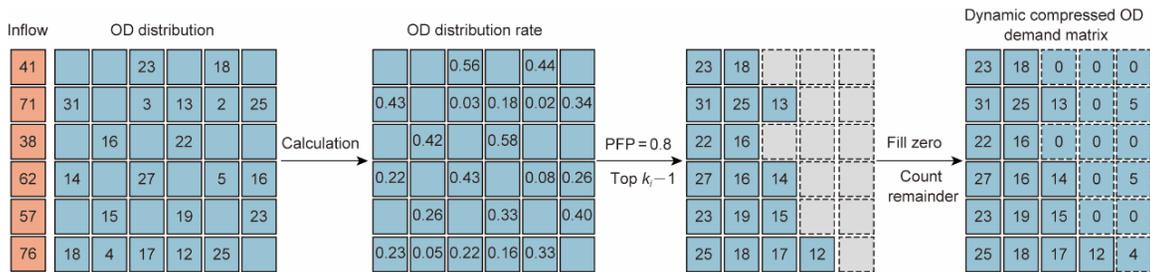


图6. 动态OD需求矩阵压缩模块概览图。PFP: 预设客流比例阈值。

态交通系统，因此使用GCN来捕捉交通流的复杂空间依赖关系合乎情理[5,39,56]。由于Kipf和Welling [57]提出的图卷积操作在结构上简单且高效，本文采用了该版本的GCN进行图卷积操作。

$$H_{l+1} = \sigma \left[\left(D^{1/2} \hat{A} D^{-1/2} \right) \cdot H_l \cdot W_l \right] \quad (10)$$

式中， H_l 表示第 l 层的隐藏特征； σ 是激活函数，如 $\text{ReLU}(\cdot)$ 或 $\text{tanh}(\cdot)$ 。 $\hat{A} = A + I_N$ 表示添加了自连接的邻接矩阵，以便在卷积操作中保留节点自身的的信息，其中 I_N 为单位矩阵。 D 是度矩阵， $D_{ii} = \sum_j A_{ij}$ ，而 W_l 是第 l 层图卷积的可训练权重矩阵。

虽然传统的GCN能够有效建模网络的拓扑空间特征，但最近的研究[40]发现，预定义的静态邻接矩阵可能会限制其在交通流预测中的表现，因为交通节点之间的空间依赖关系可能会随时间动态变化。因此，本文在GCN中引入了自适应邻接矩阵 \tilde{A}_{adp} ，用于动态建模随时间变化的空间依赖。自适应邻接矩阵 \tilde{A}_{adp} 不依赖于先验知识，而是通过随机梯度下降算法进行学习优化。具体来说，首先为源节点嵌入向量 $N_s \in R^{N \times d_{\text{model}}}$ 和目标节点嵌入向量 $N_t \in R^{N \times d_{\text{model}}}$ 随机初始化可学习参数，随后生成自适应邻接矩阵如下：

$$\tilde{A}_{\text{adp}} = \text{Softmax} \left(\text{ReLU} \left(N_t \cdot N_s^T \right) \right) \in R^{N \times N} \quad (11)$$

在该过程中， ReLU 激活函数用于忽略较弱的依赖关系， Softmax 激活函数用于对自适应邻接矩阵进行归一化处理。值得一提的是，自适应邻接矩阵能够补充节点（站点）之间的不确定关系，并在训练过程中动态建模隐藏的空间依赖关系。自适应图卷积操作的计算公式如下：

$$H_{l+1} = \sigma \left[\left(D^{1/2} \hat{A} D^{-1/2} \oplus \tilde{A}_{\text{adp}} \right) \cdot H_l \cdot W_l \right] \quad (12)$$

以往研究中已证实ConvLSTM [58]在建模序列数据的长期时空关系方面具有良好的稳定性和强大的建模能力[59–61]。但是，传统的ConvLSTM中的卷积操作仅适用于规则的欧几里得结构，无法有效提取URT网络拓扑中的非欧几里得空间信息。为了同时捕捉OD需求的复杂空间拓扑特征与时间演化特征，本文提出了一种新型AGC-LSTM模型（图7）。该模型通过在ConvLSTM的状态到状态和输入到状态转移中，将传统卷积操作替换为自适应图卷积操作。其计算方式如下所示：

$$in_t = \sigma \left(W_{\text{seq}}^{\text{in}} * \text{Seq}_t + W_{\text{h}}^{\text{in}} * H_{t-1} + W_{\text{c}}^{\text{in}} \circ C_{t-1} + b_{\text{in}} \right) \quad (13)$$

$$F_t = \sigma \left(W_{\text{seq}}^{\text{f}} * \text{Seq}_t + W_{\text{h}}^{\text{f}} * H_{t-1} + W_{\text{c}}^{\text{f}} \circ C_{t-1} + b_{\text{f}} \right) \quad (14)$$

$$C_t = F_t \circ C_{t-1} + in_t \cdot \tanh \left(W_{\text{seq}}^{\text{c}} * \text{Seq}_t + W_{\text{h}}^{\text{c}} * H_{t-1} + b_{\text{c}} \right) \quad (15)$$

$$O_t = \sigma \left(W_{\text{seq}}^{\text{o}} * \text{Seq}_t + W_{\text{h}}^{\text{o}} * H_{t-1} + W_{\text{c}}^{\text{o}} \circ C_t + b_{\text{o}} \right) \quad (16)$$

$$H_t = O_t \circ \tanh \left(C_t \right), \quad (17)$$

式中， $*$ 表示自适应图卷积操作； \circ 表示Hadamard积； $\{W_{\text{seq}}^{\text{in}}, \dots, W_{\text{c}}^{\text{o}}\}$ 表示相关的权重矩阵； $\{b_{\text{in}}, \dots, b_{\text{o}}\}$ 表示相关偏置项； $\text{Seq}_1, \dots, \text{Seq}_t$ 表示输入序列； in_1, \dots, in_t 表示输入门； F_1, \dots, F_t 表示遗忘门； C_1, \dots, C_t 表示记忆单元的输出； O_1, \dots, O_t 表示输出门； H_1, \dots, H_t 表示隐藏状态。AGC-LSTM采用自适应图卷积操作来编码时空信息，因此可以通过自适应地考虑目标节点在拓扑结构中的邻居节点的输入和隐藏状态，来确定该目标节点的未来状态。

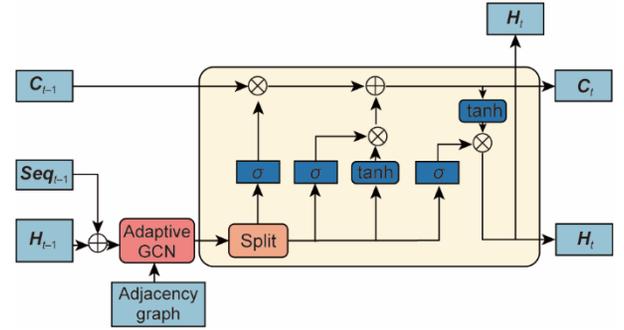


图7. AGC-LSTM可视化图。

得益于AGC-LSTM单元，PAG-STAN能够自适应地有效学习复杂的时空依赖关系。具体来说，按周与按日的OD模式首先分别输入到两个独立的AGC-LSTM分支中，从而生成长短期的周期性隐藏状态：

$$O_t^w, H_t^w = \text{AGC-LSTM} \left(X_{t-7:t_{\text{day}}}, H_{t-1} \right) \quad (18)$$

$$O_t^d, H_t^d = \text{AGC-LSTM} \left(X_{t-t_{\text{day}}}, H_{t-1} \right) \quad (19)$$

隐藏状态 H_t^w 和 H_t^d 通过一个 1×1 的卷积层进行融合，从而构建稠密的周期性隐藏状态 H_t^p 。视该稠密周期性隐藏状态 H_t^p 为初始状态，并与估算出的实时OD模式一起输入到一个独立的AGC-LSTM分支中，从而生成具有长短期周期信息的实时特征。

$$H_t^p = \text{Conv} \left(H_t^w \oplus H_t^d \right) \quad (20)$$

$$O_t^r, H_t^r = \text{AGC-LSTM} \left(X_t, H_t^p \right) \quad (21)$$

4.3.2. MPC-ATTN

由于出行行为具有周期性，多种OD需求模式之间可能存在较强的相关性。例如，周度OD需求模式能够反映工作日/周末的长期OD分布规律，而日度OD需求模式则能反映短期内OD需求的波动趋势。多头注意力机制能够联合关注来自不同表示子空间的信息，从而提高特征表示能力[62]。为了进一步挖掘OD需求中显著的周期性时空分布特征，本文提出了一种新颖的MPC-ATTN机制

(图8), 用于从多个角度建模多种OD需求模式之间的内在关联。该机制能够有效编码来自三个AGC-LSTM分支的多个OD特征输出, 从而生成一个潜在表示。

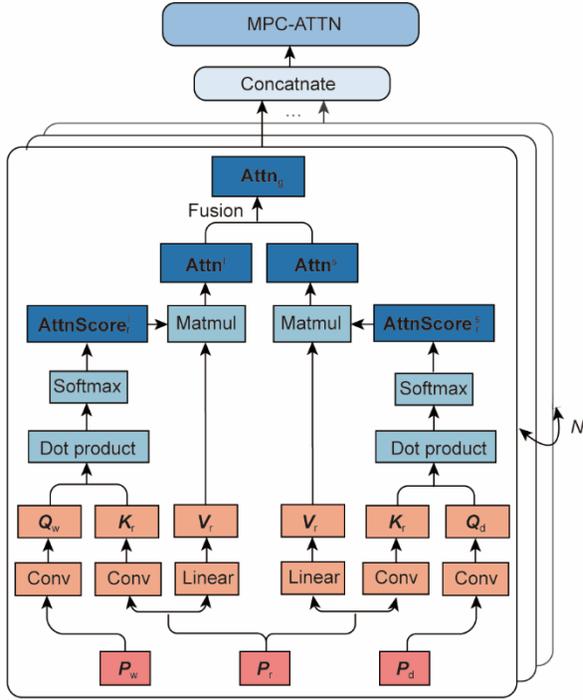


图8. MPC-ATTN 概览图。Attn: 注意力; Matmul: 矩阵乘法; V : 值; Q : 查询; K : 键; w : 周期; r : 实时; d : 日度。

具体来说, 三个独立AGC-LSTM分支的特征输出首先投影到相同维度 d_{model} , 即 $Q, K, V \in R^{d_{\text{model}} \times N \times K}$ 。在MPC-ATTN中, 传统注意力机制中对查询和键的线性投影由卷积操作取代, 以便考虑OD需求的局部分布信息。同时计算了长短期的交叉注意力分数, 用于探索多个特征输出之间的长短期周期相关性。

$$\text{AttnScore}_r^l = \text{Softmax} \left(\frac{Q_w \cdot K_r^T}{\sqrt{d_{K_r}}} \right) \quad (22)$$

$$\text{AttnScore}_r^s = \text{Softmax} \left(\frac{Q_d \cdot K_r^T}{\sqrt{d_{K_r}}} \right) \quad (23)$$

式中, AttnScore_r^l 表示长周期分布信息的相关性权重; AttnScore_r^s 表示短周期分布信息的相关性权重; d_{K_r} 为 K_r 的维度, 等于 d_{model} 。随后, 长期与短期的相关性权重分别与实时值 V_r 进行逐元素乘法, 以传播具有高注意力权重的周期信息。

$$\text{Attn}^l = \text{Matmul}(\text{AttnScore}_r^l, V_r) \quad (24)$$

$$\text{Attn}^s = \text{Matmul}(\text{AttnScore}_r^s, V_r) \quad (25)$$

在信息传播后, 长期交叉注意力 Attn^l 与短期交叉注

意力 Attn^s 拼接到一起, 通过一个 1×1 的卷积层进行投影, 从而生成具有长期与短期周期分布信息的全局交叉注意力表示。

$$\text{Attn}_g = \text{Conv}(\text{Attn}^l \cup \text{Attn}^s), \quad (26)$$

式中, \cup 表示对注意力得分的拼接。

具体来说, 本研究进一步采用多头注意力机制[62]来从不同的表示子空间获取信息, 从而对AGC-LSTM模块生成的多个OD特征输出进行编码, 进而生成一个包含全球长短周期OD分布信息的潜在表示。

$$\text{MultiHead-Attn}_g =$$

$$\text{Concat}(\text{Attn}_g^1, \text{Attn}_g^2, \dots, \text{Attn}_g^h) W^{\text{latent}} \quad (27)$$

$$\text{Attn}_g^h = \text{Conv}(\text{Attn}^l \cup \text{Attn}^s) \quad (28)$$

式中, h 是注意力头的数量; $W^{\text{latent}} \in R^{h \cdot d_v \times d_{\text{model}}}$ 为潜在注意力的权重矩阵。

4.4. 解码器

在编码器中充分学习了长短期的周期性时空分布信息之后, 解码器部分利用多个BiLSTM对这些分布信息进行正向与反向的解码。这些BiLSTM模型能够有效地捕捉上下文信息, 并学习OD需求的全局演变模式。此外, 地铁OD需求很容易受到多种外部因素(如疫情和日期属性等)的影响, 这使得其时空分布表现出高度的不规律性。这种不规律性进一步增加了地铁OD需求预测的难度。为研究疫情期间外部因素对地铁OD需求的影响, 本文提出了一种新型的HIFB, 用于融合不同来源的数据, 包括疫情相关信息和日期属性数据。该模块能够挖掘外部因素与OD需求之间的相互信息, 从而增强OD需求随时间演化的能力。最终, 通过多个全连接层预测下一时间区间内完整的OD需求。

4.4.1. BiLSTM

由于疫情背景下OD需求演化具有突发性和不确定性, 传统的单向建模方法可能无法充分捕捉OD需求在疫情期间的全局时空分布信息。因此, 本文采用多个BiLSTM模型对有意义的历史周期分布信息进行解码, 从前向和后向两个方向提供互补信息, 从而全面理解OD需求复杂的全局演变模式。BiLSTM是对LSTM的改进版本, 结合了前向LSTM和后向LSTM, 如图9所示。具体来说, BiLSTM首先按正向和反向顺序处理输入数据, 分别获得前向和后向的时间依赖关系。随后拼接这两个时间依赖关系, 并输入到一个全连接层中, 从而获得最终的隐藏状态。

具体来说, 首先由一个独立的BiLSTM处理估计的实

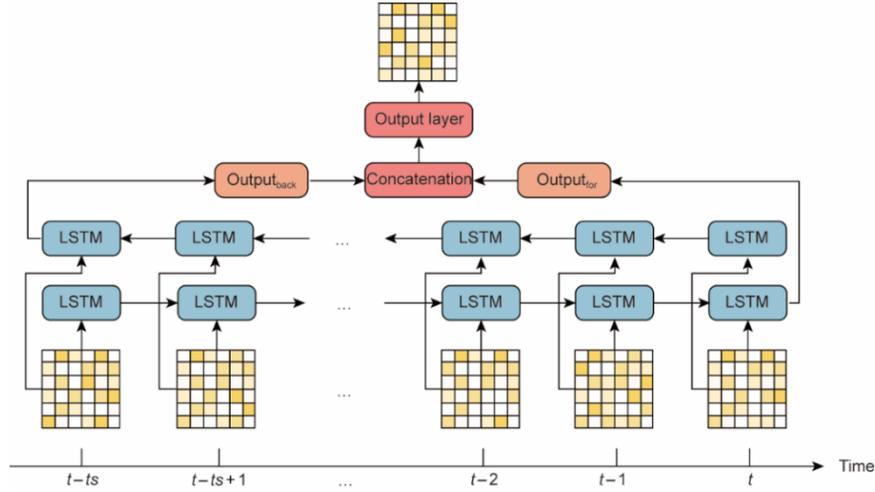


图9. BiLSTM 概览图。Back: 反向; for: 正向; ts : 历史时间步长。

时OD需求, 捕捉解码器序列。随后, 视编码器的潜在表示为初始状态, 与解码器序列一同输入到另一个独立的BiLSTM中。最终, 通过堆叠多个BiLSTM网络, 有效解码历史周期性时空分布信息。这样, 解码器序列中的每个位置都能与编码器提取的周期性分布特征进行融合, 有助于更好地理解OD需求在前向和后向方向上的全局演变模式。

4.4.2. HIFB

疫情期间, 地铁OD需求容易受到多种外部因素的影响。疫情对人们的出行构成安全威胁, 导致客流量出现异常波动。然而, 除了疫情之外, 日期属性等外部因素也会影响客流的动态变化。因此, 为研究疫情期间外部因素对地铁OD需求的影响, 本文提出了一种新型的HIFB, 用于充分整合与疫情相关的数据及日期属性数据。

将异构特征矩阵记作 $F_t \in R^{f_{num} \times ts}$, 其中 f_{num} 表示异构特征的数量; ts 表示历史时间步长, 主要包含多个时间区间 t 的异构数据 (如疫情相关数据、日期属性数据等)。由于这些异构数据是按天收集的, 因此时间区间 t 的异构特征矩阵 F_t 共享日度记录的数据。图10展示了整个异构信息融合的过程。具体来说, HIFB首先通过学习得到的嵌入将异构特征矩阵 F_t 转换为潜在特征矩阵 $\hat{F}_t \in R^{d_{model} \times f_{num} \times ts}$,

实现高维特征映射。随后, 潜在特征矩阵 \hat{F}_t 被送入 1×1 卷积单元处理, 并在时间维度上执行池化操作 (包括最大池化和平均池化)。通过这种方式, 可以有效地从异构数据中提取疫情期间的全局演化信息。接着, 将提取出的有意义的演化信息矩阵 M_t 与潜在特征矩阵 \hat{F}_t 通过逐元素相乘进行融合, 得到增强后的特征矩阵 $\hat{F}_t^e \in R^{d_{model} \times f_{num} \times ts}$ 。

$$M_t = \sigma \left(\text{AvgPool} \left[\text{Conv}(\hat{F}_t) \right] + \text{MaxPool} \left[\text{Conv}(\hat{F}_t) \right] \right) \quad (29)$$

$$\hat{F}_t^e = \hat{F}_t \otimes M_t \quad (30)$$

式中, σ 表示 Sigmoid 函数。在获得增强后的特征矩阵 \hat{F}_t^e 后, 应用若干全连接层将 \hat{F}_t^e 维度转换为与稠密 OD 需求矩阵 $\hat{X}_t \in R^{ts \times N \times K}$ 一致。最终, 通过按元素相加的方式将 \hat{F}_t^e 与 \hat{X}_t 直接融合, 从而有效增强了疫情期间 OD 需求的全局演化特征。因此, HIFB 模块通过整合多种异构数据, 深入探讨其对 OD 需求时空分布的影响, 为建模疫情期间不规则 OD 分布提供了有价值的辅助信息。

4.5. MPG 损失函数

如前所述, 在 URT 网络中, OD 需求与进站客流之间存在明显的数量关联, 这种物理规律可作为 OD 分布演化过程的约束, 有助于提升模型的可解释性。因此, 将客流量规律嵌入损失函数中, 从而有效引导模型训练。此

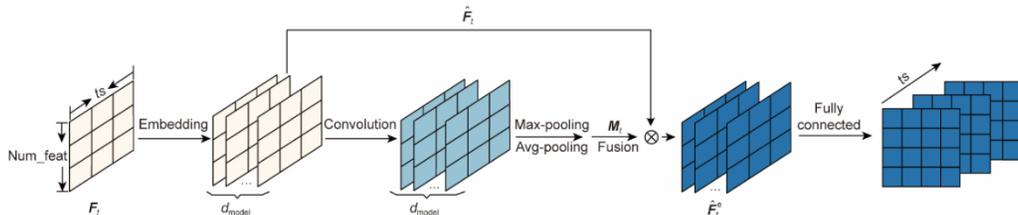


图10. HIFB 概览图。Num_feat: 外部特征数量。

外, 为构建动态压缩 OD 矩阵 $\hat{\mathbf{X}}$, 引入了 0 值填补策略, 但这些填补的 0 值可能对预测性能产生负面影响。因此, 损失函数中引入了掩码操作, 以消除由压缩 OD 矩阵填 0 值所带来的训练误差。新型 MPG 损失函数公式如下所示:

$$\begin{aligned} \text{MPG-Loss} = & W_{\text{OD}} \cdot \frac{1}{(N \times K)_{\text{masked}}} \sum_{i \in N} \sum_{j \in K} \|y_{ij} - \hat{y}_{ij}\|^2 + \\ & W_{\text{in}} \cdot \frac{1}{N} \sum_{i \in N} \left\| \sum_{j \in K} y_{ij} - \mathbf{P}_i^{\text{in}} \right\|^2 \end{aligned} \quad (31)$$

式中, W_{OD} 和 W_{in} 分别表示损失函数中不同项的权重, 用于平衡损失函数中各项之间的权衡关系, 并可动态调节不同类型信息对模型训练的影响; $(N \times K)_{\text{masked}}$ 表示在反向传播过程中未被遮盖的 OD 对数量; y_{ij} 表示从车站 S_i 到车站 S_j 的真实 OD 需求; \hat{y}_{ij} 表示对应的预测值; \mathbf{P}_i^{in} 表示车站 S_i 的进站客流量。

具体来说, MPG 损失函数的第一项评估了模型学习 OD 需求分布特征的能力, 第二项则评估模型学习物理数量关系的能力。该函数通过将乘客数量规律嵌入损失函数中, 在模型训练过程中回传有意义的物理信息, 从而在保证模型预测准确率的同时提升模型的可解释性。此外, 在训练过程中, 采用掩码操作对压缩 OD 矩阵中填补的 0 进行遮盖, 消除因填 0 值所引起的训练干扰。

5. 实验

本节以新冠病毒肺炎疫情期间的地铁 OD 需求为例, 验证所提出模型在疫情下进行短期地铁 OD 需求预测的有

效性。为了进一步验证 PAG-STAN 模型在不同场景下的实际复现能力, 还使用另一个真实 OD 需求数据集, 在日常场景下进行了地铁 OD 需求预测研究。

5.1. 实验设置

5.1.1. 数据设置

OD 需求数据。本研究引入了两个用于短期地铁 OD 需求预测的真实世界数据集, 这些数据集由数百万 AFC 记录构成, 分别采集自南宁地铁和北京地铁, 简称为 NNMOD 和 BJMOD。NNMOD 包含了 URT 网络中 62 个车站的 OD 需求数据, 时间范围为 2020 年 1 月 31 日至 2020 年 4 月 30 日, 覆盖了新冠病毒肺炎疫情的关键阶段, 包括最初的暴发、控制措施实施和后续的稳定阶段。BJMOD 包含北京地铁 276 个车站在工作日的 OD 需求数据, 数据时间范围为 2016 年 2 月 29 日至 2016 年 4 月 1 日。图 11 展示了这两个数据集中的 OD 流动趋势。可明显看出, 随着疫情的稳定, NNMOD 的 OD 流量逐步上升, 而 BJMOD 则维持了正常的通勤特征。为了检验数据分布的一致性, 采用 Cramér-von Mises (CvM) 检验分析两个数据集。NNMOD 和 BJMOD 均按照 7:2:1 的比例被划分为训练集、验证集和测试集, 且 CvM 检验中的所有 P 值均大于 0.05, 说明在 0.05 的显著性水平下, 三个子集的数据分布保持一致。

新冠病毒肺炎疫情相关数据。日度的新冠病毒肺炎疫情相关数据来自官方病例数据和相关的社交媒体数据。日度确诊病例数据来源于中国国家卫生健康委员会官网, 可直接反映疫情的发展趋势。社交媒体的兴起为理解日常活

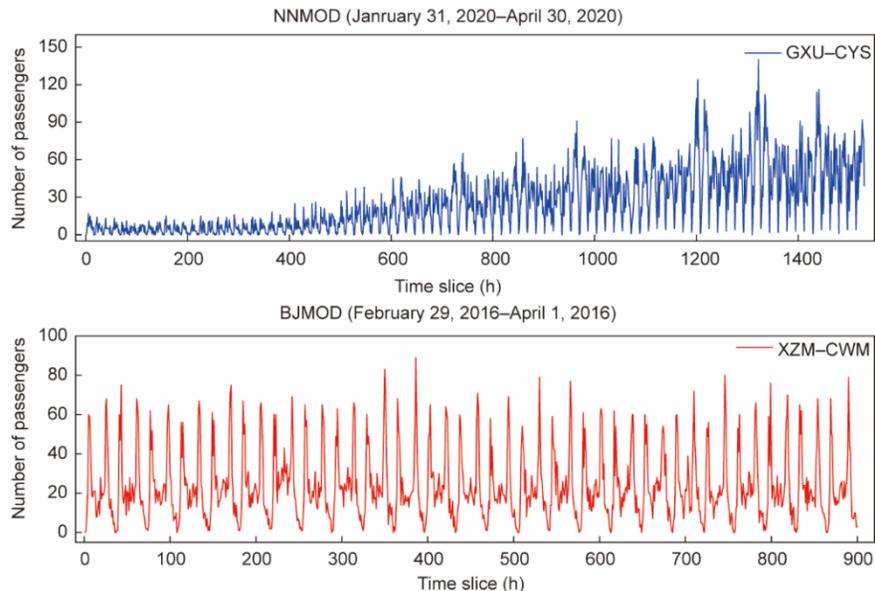


图 11. 不同 OD 对之间的 OD 时间序列数据可视化图。GXU-CYS: 广西大学站到朝阳广场站; XZM-CWM: 西直门站到崇文门站。

动与地铁OD需求之间的关系提供了可行途径[15,63]。通过新浪微博流式应用程序接口(API)配合地理位置过滤,收集关键词为“新冠肺炎”“南宁”“地铁”的相关内容。由于疫情相关数据通常按日发布,因此每天的数据作为该日各时间段的异构特征进行共享。但由于实时性限制,当日的数据往往难以及时获取。已有研究表明,前一天的外部信息会对第二天的交通流量产生影响[64]。因此,本研究使用前一天完整的疫情相关数据作为当前时段的异构信息。为判断新冠肺炎疫情相关数据与OD需求数据之间是否存在相关性,对两者之间的皮尔逊相关系数进行了计算。结果显示,其绝对值均大于0.6,表明两者之间具有显著相关性。

数据属性。日期属性数据包括星期几、日期类型(工作日/周末)和节假日信息。通过独热(one-hot)编码将这些非数值型外部特征转换为二进制向量,最终在模型中选用了共11个变量来表示所有外部因素,即外部特征数量为 $\text{num_feat} = 11$ 。表2列出了所有数据属性的详细信息。

表2 异构数据描述

Heterogeneous data	Type of value	Description
Day of week	Dummy variables	One-hot vector with 7 dummy variables
Date type	Dummy variables	0: non-weekend; 1: weekend
Holiday	Dummy variables	0: non-weekend; 1: weekend
Confirmed case data	Continuous value	Daily confirmed case volumes
Social media data	Continuous value	Daily post volumes

5.1.2. 模型配置

本研究中,所有模型均使用PyTorch框架,并在配备Intel® Core™ i9-10900X CPU、64 GB内存和NVIDIA GeForce RTX3060 GPU的台式机上运行。本文提出的模型包含两个编码器层和两个解码器层。为了优化训练效果,在编码器和解码器中嵌入了丢弃率为0.1的随机失活(Drop-out)层。批量大小设为32,优化器采用Adam,学习率为0.0001。此外,所有实验结果均进行了显著性检验(P 值<0.01的 t 检验)。

为进一步分析多个关键超参数对PAG-STAN的影响,本研究对PAG-STAN中的四个超参数进行了详细研究,分别是模型维度 d_{model} 、多头注意力机制的头数 Head_{num} 、时间步数 TS_{num} 、乘客流量保留比例PFP。具体来说,设置的超参数取值区间为: d_{model} (64, 128, 256, 512), Head_{num} (3, 4, 6, 8), TS_{num} (8, 10, 12, 14), PFP(50%, 60%, 70%, 80%)。上述四个超参数对PAG-STAN的影响如图12所示。

总体来看,随着 d_{model} 的增加,均方根误差(RMSE)和平均绝对百分比误差(MAPE)逐渐降低。然而,当 d_{model} 超过256后,进一步增加时,性能提升不明显。综合计算成本,最终将 d_{model} 设置为256。超参数 Head_{num} 和 TS_{num} 预测结果表现相似,故分别将其设置为4和12。在乘客流量比例方面,当PFP小于70%时,模型预测性能基本保持稳定;但当PFP超过70%时,预测性能明显下降。这是因为较大的PFP降低了压缩矩阵稀疏性的削减效果,进而影响模型预测能力。较小的PFP有助于提升预测精度,但也会限制模型对部分OD对需求的预测能力。最终经过综合考量,将PFP设置为70%,以在预测性能和可预测OD对数量之间取得平衡。

在模型训练过程中,采用模型检查点(Model Checkpoint)和早停(Early Stopping)技术,从而保存最佳模型并防止过拟合。在触发早停之前,训练损失和验证损失如图13所示。可明显看出,训练损失收敛迅速且稳定,而验证损失在第400个周期后逐渐收敛并在大约第600个周期趋于稳定,接近训练损失。该现象可归因为NNMOD数据集的组成,此数据集涵盖了疫情期间南宁地铁复杂的OD需求数据。这些OD需求数据存在不规则波动,需要多轮训练迭代才能有效捕捉其复杂模式。本研究进一步探讨了PAG-STAN模型在地铁OD需求预测中的运行效率。具体来说,该模型的平均训练时间为每轮11.83 s,符合实际应用的接受范围。值得注意的是,如果不采用动态OD需求压缩操作,模型的执行时间将增加至每轮18.21 s。这一观察结果表明,动态OD需求压缩操作能有效缓解高维问题,降低计算成本。因此,本文提出的模型在运行时表现出良好的实时性,其运行效率不会成为任务执行的瓶颈。由此可见,应重点关注提高疫情期间地铁OD需求预测的准确性。

5.1.3. 评估指标

本研究选取了RMSE、MAE和加权平均绝对百分比误差(WMAPE)作为评估指标。

$$\text{RMSE} = \sqrt{\frac{1}{(N \times K)_{\text{masked}}} \sum_{i \in N} \sum_{j \in K} \|y_{ij} - \hat{y}_{ij}\|^2} \quad (32)$$

$$\text{MAE} = \frac{1}{(N \times K)_{\text{masked}}} \sum_{i \in N} \sum_{j \in K} |y_{ij} - \hat{y}_{ij}| \quad (33)$$

$$\text{WMAPE} = \sum_{ij} \left(\frac{y_{ij}}{\sum_{ij} y_{ij}} \left| \frac{y_{ij} - \hat{y}_{ij}}{y_{ij}} \right| \right) \quad (34)$$

式中, y_{ij} 表示从 S_i 车站到 S_j 车站的实际OD需求值; \hat{y}_{ij} 表示对应的预测值。

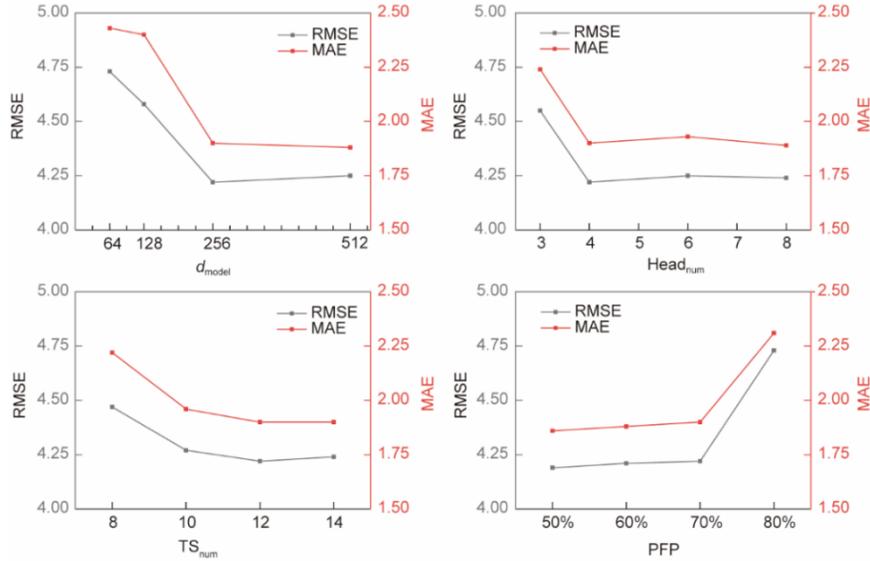


图12. PAG-STAN模型中四个超参数对NNMODD（60 min粒度）的影响。RMSE：均方根误差；MAE：平均绝对误差。

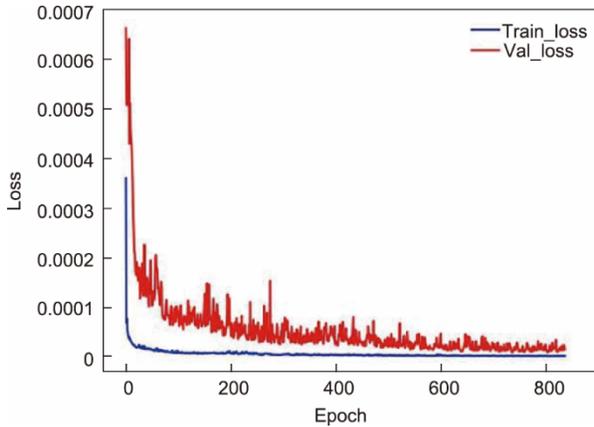


图13. 训练损失和验证损失的变化。

5.2. 前沿成果对比分析

为了全面评估PAG-STAN的预测性能，本文选取以下基准模型，与PAG-STAN的预测效果进行比较。本研究根据官方代码搭建运行了各基准模型。值得注意的是，在以下实验中，所有基线模型均使用估计的实时OD需求数据进行地铁OD需求预测。

- **ARIMA**：一种典型的传统数理统计模型。本研究使用国际商业机器公司（International Business Machines Corporation，美国）推出的SPSS®软件中的专家建模器（Expert Modeler）自动获得最优ARIMA结果。

- **三维卷积神经网络（3DCNN）**：3DCNN通过卷积操作提取丰富的信息特征。本研究在模型中使用了两个 3×3 的3DCNN卷积层和一个包含256个神经元的全连接层来预测OD需求。

- **LSTM**：LSTM能够有效捕捉时间特征。本文采用了两层全连接的LSTM层来预测未来的OD需求，隐藏状

态的维度设置为256。

- **ConvLSTM**：ConvLSTM [58]用卷积操作替代了LSTM中的全连接操作，从而能够同时建模空间和时间特征。ConvLSTM的模型配置与LSTM相同。

- **Graph WaveNet (GWN)**：GWN [40]引入了一个自适应邻接矩阵来建模隐藏的空间依赖关系，并开发了堆叠扩张一维（1D）卷积单元来建模长期时间依赖。

- **CSTN**：CSTN [11]是首个基于深度学习研究交通OD需求预测的模型，融合了局部空间上下文、时间演化上下文和全局相关上下文，从而预测区域间交通需求。

- **ST-ResNet**：该模型[4]使用残差卷积单元建模时空特征。本研究使用三个分支的残差卷积单元来学习多重时空依赖关系。

- **ST-ED-RMGC**：ST-ED-RMGC [51]是用于地铁OD需求预测的模型，该模型中引入了多种图结构来学习OD需求的复杂模式。

- **Transformer**：Transformer [62]是一种基于注意力机制的模型，利用多头注意力机制来学习序列数据的注意力权重。其注意力头数设置为8，嵌入维度 d_{model} 设置为512。

- **Informer**：Informer [65]是一种基于Transformer的长序列时间序列高效预测模型，其超参数设置与Transformer一致。

- **异构信息聚合机 (HIAM)**：HIAM [9]充分利用历史数据中的异构信息，以联合学习OD和DO（终点-起点）客流的演化模式，从而预测未来的跨站点客流。

(1) NNMOD的性能。首先，对PAG-STAN在

NNMOD测试集上的预测性能与其他方法进行了比较。所有方法的性能总结如表3所示。可以观察到,在所有比较模型中,传统的统计模型ARIMA的性能最差,因为ARIMA很难捕捉到OD需求中复杂的时空特征。基础深度学习模型CNN和LSTM通过显式建模OD需求的空间或时间依赖关系,实现了一定程度的性能提升。复合模型(如ConvLSTM、GWN和ST-ResNet)同时建模OD需求的时空分布,优于基础的深度学习模型,说明必须充分挖掘复杂时空依赖关系。当前主流的OD需求预测方法CSTN和ST-ED-RMGC在NNMOD数据集上的预测效果并不理想,这可能是因为这些方法主要针对出租车OD需求的预测,其时空依赖关系与地铁OD需求有显著差异。因此,这两种方法不适合用于地铁OD需求预测。基于注意力机制的Transformer模型和Informer模型在所有时间区间上均表现出色,它们能有效建模序列中任意距离的时间关系。尽管未考虑空间特征,但南宁地铁的空间拓扑结构较为简单,因此这两种模型也能实现较好的预测性能。HIAM的预测

性能在所有时间区间中都排第二,因为它充分利用了多种信息(如不完整OD客流、未完成订单和DO客流)来建模地铁OD分布。但由于未考虑OD需求与进站客流之间的数量关系,HIAM模型在预测结果的可解释性方面略有不足。

尽管现有模型在预测性能方面取得了一定进展,但上述所有模型仅关注历史OD需求信息,忽视了新冠病毒肺炎疫情期间影响OD需求分布的外部因素。PAG-STAN不仅深度融合多周期的OD需求模式,用于捕捉OD需求的周期性时空分布特征,还引入异构数据源(如疫情相关数据、日期属性数据)以学习反映疫情影响的有意义信息,从而增强OD需求的演化特征。因此,PAG-STAN在所有时间区间的表现均优于其他方法,分别在RMSE、MAE和WMAPE方面平均提升了7.77%、11.50%和6.33%。

(2) BJMOD的性能。为了全面评估PAG-STAN的稳健性和敏感性,本研究在另一个真实数据集BJMOD上进行了地铁OD需求预测任务。需注意,在该任务中,日期

表3 不同方法在NNMOD/BJMOD数据集上的预测性能表现

Time interval	Metrics	ARI-MA	3DCNN	LSTM	ConvLSTM	GWN	CSTN	ST-ED-RMGC	ST-ResNet	Transformer	Informer	HIAM	PAG-STAN	
NNMOD														
15 min	RMSE	5.18	3.22 ± 0.19	2.76 ± 0.14	2.74 ± 0.10	2.60 ± 0.15	3.04 ± 0.08	2.98 ± 0.15	2.39 ± 0.17	2.34 ± 0.14	2.32 ± 0.11	2.19 ± 0.17	1.90 ± 0.20*	
		MAE	5.09	1.34 ± 0.11	1.17 ± 0.13	1.21 ± 0.09	1.14 ± 0.12	1.27 ± 0.10	1.26 ± 0.17	0.97 ± 0.15	0.94 ± 0.15	0.93 ± 0.13	0.89 ± 0.21	0.80 ± 0.11*
	WMAPE	0.712	0.644 ± 0.023	0.594 ± 0.019	0.561 ± 0.021	0.552 ± 0.030	0.595 ± 0.025	0.591 ± 0.019	0.536 ± 0.025	0.531 ± 0.018	0.527 ± 0.020	0.516 ± 0.018	0.491 ± 0.022*	
		MAE	7.46	2.45 ± 0.18	1.98 ± 0.22	1.98 ± 0.15	1.96 ± 0.23	2.47 ± 0.16	2.35 ± 0.24	1.66 ± 0.27	1.53 ± 0.20	1.50 ± 0.17	1.46 ± 0.19	1.33 ± 0.13*
	30 min	RMSE	8.28	5.68 ± 0.23	4.47 ± 0.26	4.42 ± 0.19	4.40 ± 0.25	5.68 ± 0.11	5.25 ± 0.21	3.41 ± 0.22	3.41 ± 0.19	3.39 ± 0.15	2.98 ± 0.16	2.82 ± 0.17*
			MAE	0.693	0.609 ± 0.021	0.506 ± 0.018	0.503 ± 0.032	0.495 ± 0.016	0.547 ± 0.025	0.557 ± 0.026	0.392 ± 0.022	0.390 ± 0.019	0.393 ± 0.028	0.384 ± 0.015
60 min	RMSE	15.27	10.17 ± 0.28	7.95 ± 0.34	7.00 ± 0.25	7.39 ± 0.19	8.17 ± 0.22	8.13 ± 0.14	5.68 ± 0.17	5.14 ± 0.21	4.96 ± 0.23	4.43 ± 0.19	4.22 ± 0.26*	
		MAE	14.30	4.46 ± 0.23	3.64 ± 0.19	3.18 ± 0.22	3.32 ± 0.17	3.88 ± 0.12	3.85 ± 0.20	2.64 ± 0.15	2.41 ± 0.25	2.40 ± 0.20	2.19 ± 0.16	1.90 ± 0.19*
	WMAPE	0.669	0.570 ± 0.026	0.467 ± 0.029	0.421 ± 0.031	0.432 ± 0.019	0.476 ± 0.011	0.495 ± 0.015	0.331 ± 0.027	0.329 ± 0.018	0.322 ± 0.024	0.319 ± 0.026	0.293 ± 0.020*	
		MAE	7.46	4.55 ± 0.18	3.94 ± 0.23	3.85 ± 0.17	2.62 ± 0.22	3.56 ± 0.21	3.53 ± 0.19	2.08 ± 0.24	1.96 ± 0.22	1.89 ± 0.25	1.79 ± 0.26	1.66 ± 0.19*
	30 min	RMSE	5.82	1.76 ± 0.21	1.59 ± 0.16	1.49 ± 0.18	1.04 ± 0.26	1.29 ± 0.21	1.32 ± 0.22	1.05 ± 0.32	0.93 ± 0.24	0.88 ± 0.14	0.82 ± 0.25	0.73 ± 0.23*
			MAE	0.769	0.696 ± 0.020	0.684 ± 0.014	0.678 ± 0.030	0.533 ± 0.013	0.607 ± 0.022	0.597 ± 0.024	0.516 ± 0.021	0.490 ± 0.014	0.452 ± 0.018	0.424 ± 0.019
BJMOD														
30 min	RMSE	7.46	4.55 ± 0.18	3.94 ± 0.23	3.85 ± 0.17	2.62 ± 0.22	3.56 ± 0.21	3.53 ± 0.19	2.08 ± 0.24	1.96 ± 0.22	1.89 ± 0.25	1.79 ± 0.26	1.66 ± 0.19*	
		MAE	5.82	1.76 ± 0.21	1.59 ± 0.16	1.49 ± 0.18	1.04 ± 0.26	1.29 ± 0.21	1.32 ± 0.22	1.05 ± 0.32	0.93 ± 0.24	0.88 ± 0.14	0.82 ± 0.25	0.73 ± 0.23*
	WMAPE	0.769	0.696 ± 0.020	0.684 ± 0.014	0.678 ± 0.030	0.533 ± 0.013	0.607 ± 0.022	0.597 ± 0.024	0.516 ± 0.021	0.490 ± 0.014	0.452 ± 0.018	0.424 ± 0.019	0.382 ± 0.020*	

Data in the columns are mean ± standard deviation; the best and second-best prediction performances are highlighted in bold and underlined, respectively; numbers marked with * indicate that the improvement is statistically significant compared with the best baseline (t -test with P -value < 0.01)

属性数据仅作为辅助信息用于探究外部因素对OD需求的影响。所有方法的预测性能如表3所示。考虑到北京地铁相比南宁地铁具有更复杂的空间结构,那些忽略OD需求空间依赖关系的模型在BJMOD数据集上表现较差。尤其是广泛使用的OD需求预测方法CSTN和ST-ED-RMGC在BJMOD数据集上表现不佳,凸显了它们在地铁OD需求预测中的鲁棒性较差。相反,HIAM在BJMOD数据集上依旧表现出竞争力,表明其鲁棒性良好。PAG-STAN在所有基准模型中表现最佳,这是因为该模型利用多种异构数据捕捉未来OD客流的演化趋势并在损失函数中引入OD需求与进站客流之间的数量关系,从而增强了模型预测结果的可解释性。这些结果有力地说明了PAG-STAN的鲁棒性和敏感性,并为其在常规场景下的实际应用提供了有价值的参考。

(3) 高需求站点的性能。地铁OD分布具有非均匀性,大多数OD需求空间上集中在居住区和商业区之间。因此,准确预测高需求站点之间的OD需求对于地铁运营方提前制定时刻表具有重要意义。本节评估了PAG-STAN在高起点需求站点的OD需求预测性能。本研究选择了15个高起点需求站点进行实验,覆盖了南宁地铁大约65%的OD需求。不同方法在高需求站点的预测性能如表4所示。具体来说,PAG-STAN相较于现有最优模型在RMSE上降低了5.74%、MAE上降低了5.48%、WMAPE上降低了6.45%。同时,随着时间间隔的增加,PAG-STAN相比其他模型的优势更加明显。这些预测结果表明,PAG-STAN在高需求站点之间的OD需求预测中表现

优异。

(4) OD需求预测结果的可视化。为了更直观地展示PAG-STAN的预测性能,本研究通过热力图对不同时段(8:00–8:30、12:00–12:30和19:00–19:30)的预测OD矩阵进行了可视化。列表示起点站,行表示终点站,如图14所示。由于新冠病毒肺炎疫情的影响,即使在早晚高峰时段,大多数OD对的OD需求值仍然较小甚至为零,只有少数OD对具有较高的OD需求。如此显著的OD需求矩阵稀疏性使得疫情期间准确预测OD需求具有挑战性。为了实现准确的地铁OD需求预测,PAG-STAN整合了多个历史时序的OD需求数据,以挖掘OD分布的周期性特征。因此,对于低需求和高需求的OD对,预测值均接近真实值。PAG-STAN在早晚高峰时段的预测性能优于中午时段,这是因为早晚高峰期间OD需求的分布特征更为显著。此外,本研究进一步可视化了从NNMOD和BJMOD两个数据集中各选取的一个OD对的时间序列需求数据,分别为“广西大学站到朝阳广场站(GXU–CYS)”和“西直门站到崇文门站(XZM–CWM)”。如图15所示,无论是在NNMOD还是BJMOD数据集上,PAG-STAN都实现良好的预测性能。尽管GXU–CYS的需求时间序列数据受疫情影响出现了不规则波动,但该模型能有效从多种历史异构数据(如疫情相关数据和日期属性数据)中学习OD需求的时空分布信息。因此,其预测结果能够较好地捕捉需求数据的演化趋势。XZM–CWM的需求时间序列数据反映了工作日期间OD需求的变化,表现出明显的早高峰特征和规律性波动。通过利用历史周需求和日需求时

表4 不同方法在高需求车站上的预测性能表现

Time interval	Metrics	ARIMA	3DCNN	LSTM	ConvLSTM	GWN	CSTN	ST-ED-RMGC	ST-ResNet	Transformer	Informer	HIAM	PAG-STAN
15 min	RMSE	7.38	5.22 ± 0.27	3.83 ± 0.18	4.17 ± 0.22	4.69 ± 0.19	4.11 ± 0.16	5.19 ± 0.20	3.66 ± 0.15	3.72 ± 0.25	3.68 ± 0.17	<u>3.45 ± 0.23</u>	3.24 ± 0.26*
	MAE	6.45	3.24 ± 0.17	2.44 ± 0.10	2.63 ± 0.13	2.92 ± 0.09	2.60 ± 0.14	3.20 ± 0.26	2.30 ± 0.15	2.39 ± 0.11	2.36 ± 0.20	<u>2.17 ± 0.19</u>	2.02 ± 0.21*
	WMAPE	0.735	0.554 ± 0.035	0.449 ± 0.026	0.476 ± 0.028	0.505 ± 0.024	0.467 ± 0.031	0.565 ± 0.024	0.409 ± 0.019	0.430 ± 0.017	0.403 ± 0.014	<u>0.397 ± 0.008</u>	0.368 ± 0.012
30 min	RMSE	11.26	9.30 ± 0.38	6.38 ± 0.28	6.95 ± 0.31	9.09 ± 0.24	7.62 ± 0.26	7.42 ± 0.21	5.89 ± 0.17	5.88 ± 0.23	5.66 ± 0.29	<u>5.54 ± 0.10</u>	5.21 ± 0.15*
	MAE	9.37	5.91 ± 0.31	3.99 ± 0.26	4.57 ± 0.22	5.90 ± 0.28	4.98 ± 0.19	4.64 ± 0.23	4.00 ± 0.34	3.75 ± 0.25	3.68 ± 0.20	<u>3.57 ± 0.20</u>	3.38 ± 0.19*
	WMAPE	0.721	0.567 ± 0.034	0.430 ± 0.031	0.442 ± 0.027	0.511 ± 0.040	0.474 ± 0.029	0.428 ± 0.021	0.373 ± 0.033	0.366 ± 0.024	0.353 ± 0.027	<u>0.341 ± 0.028</u>	0.317 ± 0.031*
60 min	RMSE	22.30	19.01 ± 0.53	15.69 ± 0.60	16.28 ± 0.66	18.25 ± 0.59	16.41 ± 0.44	16.35 ± 0.47	13.36 ± 0.38	17.22 ± 0.51	16.91 ± 0.40	<u>14.67 ± 0.27</u>	13.91 ± 0.33*
	MAE	16.18	12.30 ± 0.28	9.91 ± 0.22	10.50 ± 0.30	11.68 ± 0.25	10.49 ± 0.21	10.29 ± 0.19	8.43 ± 0.22	10.13 ± 0.36	10.00 ± 0.23	<u>8.29 ± 0.20</u>	7.94 ± 0.21*
	WMAPE	0.768	0.716 ± 0.033	0.537 ± 0.024	0.632 ± 0.030	0.599 ± 0.029	0.707 ± 0.037	0.620 ± 0.022	0.485 ± 0.017	0.598 ± 0.031	0.596 ± 0.023	<u>0.461 ± 0.019</u>	0.438 ± 0.028*

间序列数据，PAG-STAN能够捕捉到需求数据的周期性信息，从而实现预测结果与实际需求数据之间的良好拟合。这些结果展示了PAG-STAN在疫情期间和常规场景下进行地铁OD需求预测的优越性。

5.3. 消融研究

本节进行了大量的消融研究，以验证PAG-STAN各个组成部分的有效性。

5.3.1. 多个历史时间序列OD需求的有效性

PAG-STAN考虑了多个历史时间序列OD需求，以研究OD分布的长短期周期性。本节探索了来自周度和日度时间序列中OD需求的长短期OD分布信息的影响。引入了如下几个变体：

- **仅实时OD网络模型 (Net-real-time OD)**：该变体直接使用估计的实时OD需求来预测OD需求。
- **实时OD+周度OD网络模型 (Net-real-time OD + weekly OD)**：该变体利用周度时间序列OD需求，探索长期历史OD信息对OD需求预测的影响。
- **实时OD+日度OD网络模型 (Net-real-time OD + daily OD)**：该变体使用日度时间序列OD需求，研究短期历史OD信息对地铁OD需求预测的影响。
- **多周期OD网络模型 (Net-multi-periodic OD)**：该网络融合了周度和日度时间序列OD需求，以捕捉长期和短期历史OD分布信息，从而预测地铁OD需求。

如表5总结所示，仅实时OD网络模型在所有时间区

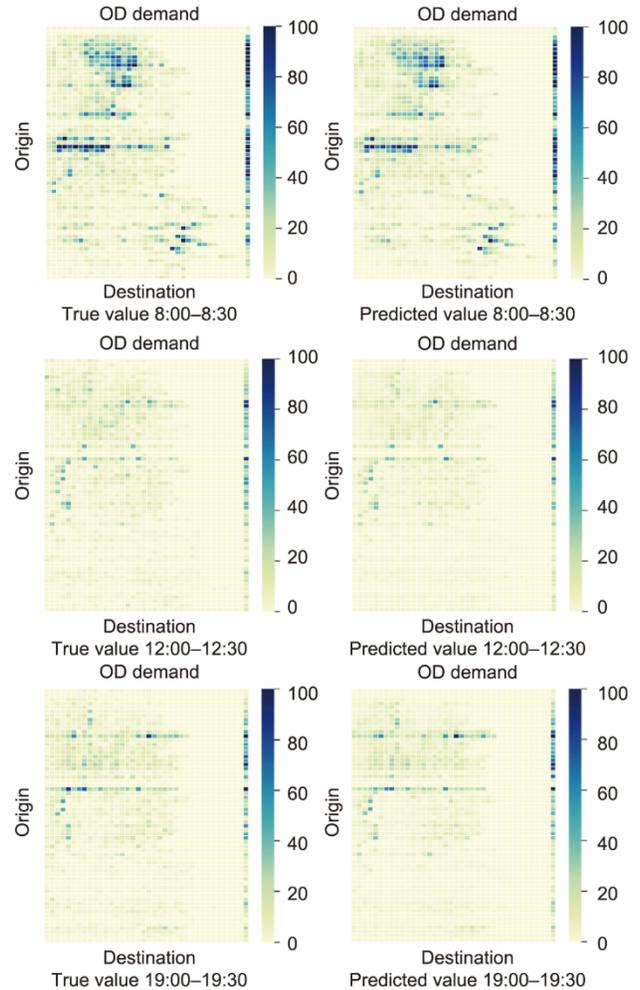


图14. 使用PAG-STAN模型对比OD实际值与预测值。

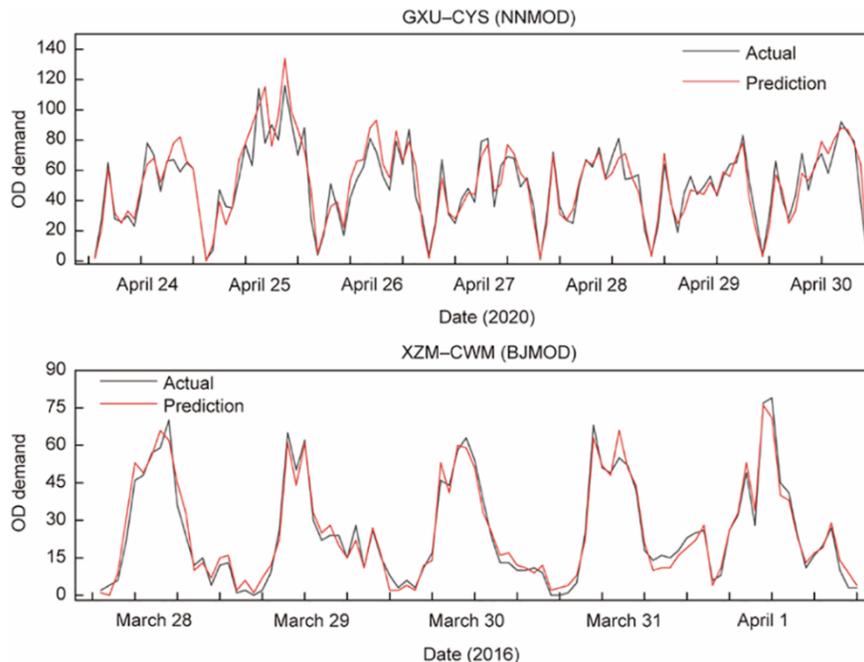


图15. 使用PAG-STAN模型对比OD时间序列真值与预测值。

间内的预测性能都较差，这是因为该变体仅考虑了有限的实时信息来进行OD需求预测。当引入周度或日度时间序列OD时，PAG-STAN的预测性能有了小幅提升，表明长期或短期的OD分布信息对OD需求预测是有益的。此外，实时OD+日度OD网络模型在大多数情况下优于实时OD+周度OD网络模型，这说明短期的时空信息更能准确刻画OD的周期演化模式。相比之下，多周期OD网络模型通过挖掘来自多个时间序列OD需求中的长短期周期分布信息，实现了最佳预测性能。在平均RMSE、MAE和WMAPE三项指标上，分别相较于仅实时OD网络模型提高了10.95%、11.02%和8.62%。因此，来自多个历史时间序列OD需求中的长短期OD分布信息能够显著提升OD需求预测的准确性。

5.3.2. 实时OD需求估计的有效性

本研究根据实时客流量和历史长短期OD分布率来估计实时OD需求。为了研究实时OD需求估计对地铁OD需求预测的影响，提出了几个变体。

- **不完整OD预测 (Net-incomplete OD)**: 该变体直接使用不完整的实时OD需求来预测未来的地铁OD需求。

- **仅长期估计OD预测 (Net-estimated OD, long-**

term): 该变体利用完整的长期OD分布率来估计实时OD需求，用于进行地铁OD需求预测。

- **仅短期估计OD预测 (Net-estimated OD, short-term)**: 该变体利用完整的短期OD分布率来估计实时OD需求，用于进行地铁OD需求预测。

- **长短期估计OD预测 (Net-estimated OD, long- and short-term)**: 该变体使用了含长期和短期的完整OD的需求分布信息，用于进行地铁OD需求预测。

所有变体的预测表现总结见表6。可以观察到，直接使用不完整的实时OD需求矩阵时，不完整OD预测变体的预测性能最差。这是因为不完整的矩阵没有记录未完成的出行信息，干扰了OD分布信息的学习。同时，可以观察到，仅长期估计OD预测和仅短期估计OD预测变体的预测性能有所提高，因为它们都考虑了完整的OD需求信息。与之前的实验结果相似，仅短期估计OD预测在大多数时间间隔内优于仅长期估计OD预测，证明短期性的OD分布数据能为OD需求预测提供有用信息。相比之下，长短期估计OD预测在所有时间间隔内均实现了理想的预测性能，表明实时OD估计模块有效估计了长短期性分布信息的实时完整OD需求。

表5 多周期OD需求在OD预测中的表现

Time interval	Metrics	Net-real-time OD	Net-real-time and weekly OD	Net-real-time and daily OD	Net-multi-periodic OD
15 min	RMSE	2.202	2.136	2.002	1.898
	MAE	0.886	0.865	0.842	0.803
	WMAPE	0.531	0.520	0.514	0.491
30 min	RMSE	3.152	3.041	2.933	2.821
	MAE	1.457	1.403	1.372	1.328
	WMAPE	0.391	0.380	0.373	0.361
60 min	RMSE	4.612	4.523	4.408	4.218
	MAE	2.235	2.126	2.043	1.903
	WMAPE	0.328	0.315	0.306	0.293

表6 实时OD估计在OD预测中的表现

Time interval	Metrics	Net-incomplete OD	Net-estimated OD (long-term)	Net-estimated OD (short-term)	Net-estimated OD (long- and short-term)
15 min	RMSE	2.321	2.098	1.977	1.898
	MAE	1.016	0.863	0.857	0.803
	WMAPE	0.537	0.511	0.502	0.491
30 min	RMSE	3.240	2.989	2.927	2.821
	MAE	1.471	1.365	1.341	1.328
	WMAPE	0.422	0.386	0.375	0.361
60 min	RMSE	4.643	4.431	4.327	4.218
	MAE	2.428	2.156	2.043	1.903
	WMAPE	0.331	0.320	0.307	0.293

5.3.3. 外部因素的有效性

本研究结合了多种异构数据（如疫情相关数据、日期属性数据）来研究外部因素对疫情期间地铁OD分布的影响。本小节探讨了多种外部因素对模型预测性能的影响。以下是几种变体的介绍：

- **无外部因素 (Net-no external factors)**：此变体不使用任何异构数据来考虑外部因素对疫情期间地铁OD分布的影响。

- **仅疫情信息 (Net-pandemic information)**：此变体仅使用与疫情相关的数据来研究疫情对地铁OD分布的影响。

- **仅日期属性信息 (Net-date attribute information)**：此变体仅使用日期属性信息来研究日期属性对疫情期间地铁OD分布的影响。

- **异构信息 (Net-heterogeneous information)**：该方法结合了疫情相关信息和日期属性信息，全面研究外部因素对疫情期间地铁OD分布的影响。

如表7所示，当结合所有外部因素时，PAG-STAN在所有时间区间实现最佳预测性能。疫情相关数据使得模型能够从全局角度全面理解OD分布的演变模式，而日期属性数据则帮助模型从局部角度捕捉OD分布的周期性（工作日/周末）。将这两种异构信息结合到PAG-STAN中，能够有效地研究外部因素对地铁OD分布的影响，同时从全局和局部角度促进对疫情期间的演变特征的学习。

值得注意的是，仅将疫情相关数据与OD需求数据融合时，模型在所有时间区间的表现都优于仅将日期属性数据与OD需求数据融合的情景。这个结果表明，在疫情期间捕捉OD分布的全局演变特征比仅关注OD分布的局部周期性更有价值。因此，理解外部因素对地铁OD分布的影响对于预测疫情期间的地铁OD需求至关重要。

5.3.4. 动态OD需求矩阵压缩的有效性

本研究将原始的 $N \times N$ 维度OD需求矩阵转化为稠密

的 $N \times K$ 维度OD需求矩阵，从而解决了数据稀疏性和高维性的问题。为了研究OD需求矩阵压缩操作是否能改善OD需求预测，本研究利用具有 $N \times N$ 维度的原始完整OD需求矩阵来预测地铁OD需求。在得到预测的OD需求矩阵 $\hat{X}_t \in R^{N \times N}$ 后，将预测的OD需求矩阵的形状转化为密集OD需求矩阵 $\hat{X}_t \in R^{N \times K}$ ，从而比较预测性能。

如表8所示，应用原始OD需求矩阵时，其他站点的WMAPE超过了60%。这个高误差是因为原始OD需求矩阵中的值较小，甚至为零，导致了稀疏矩阵，进而妨碍了特征学习。然而，当使用稠密OD需求矩阵时，PAG-STAN展示的预测性能有所改善，在所有时间区间的其他站点的平均WMAPE为27%。这是因为压缩操作让模型能够捕捉到OD需求中的重要分布特征。压缩操作改善了前 $K-1$ 个站点的预测性能。在压缩OD需求矩阵中，所有时间区间的前 $K-1$ 个站点的平均WMAPE为40.33%，而原始OD矩阵为53.83%。这是因为原始OD矩阵的稀疏性在研究重要的OD分布信息时对模型产生了负面影响。因此，压缩操作有效克服了数据稀疏性问题，从而提高了地铁OD需求的可预测性。

5.3.5. MPG损失函数的有效性

本研究提出了一种新的MPG损失函数，将物理客流量规律嵌入损失函数中，用于指导模型训练。此外，采用了掩蔽操作，消除了稠密OD矩阵中填充0值带来的训练干扰。本小节探讨了MPG损失函数对地铁OD需求预测的影响。

- **一般损失函数 (Net-general loss function)**：该变体根据一般均方误差（MSE）损失函数训练PAG-STAN。

- **信息嵌入损失函数 (Net-information embedding loss function)**：该变体仅将客流量规律嵌入损失函数，以增强模型训练。

- **掩蔽损失函数 (Net-masked loss function)**：该变体

表7 不同外部因素对OD预测性能的影响

Time interval	Metrics	Net-no external factors	Net-pandemic information	Net-date attribute information	Net-heterogeneous information
15 min	RMSE	2.088	1.969	2.010	1.898
	MAE	0.843	0.821	0.837	0.803
	WMAPE	0.529	0.509	0.517	0.491
30 min	RMSE	3.117	2.913	3.039	2.821
	MAE	1.403	1.351	1.387	1.328
	WMAPE	0.384	0.371	0.375	0.361
60 min	RMSE	4.550	4.307	4.393	4.218
	MAE	2.247	2.019	2.057	1.903
	WMAPE	0.320	0.304	0.313	0.293

表8 压缩/原始OD需求预测性能

Time interval	Metrics	With compression		Without compression	
		Top $K-1$ stations	Remaining stations	Top $K-1$ stations	Remaining stations
15 min	RMSE	1.578	6.025	2.875	10.104
	MAE	0.676	3.681	1.043	6.582
	WMAPE	0.591	0.366	0.834	0.612
30 min	RMSE	2.186	8.593	2.192	17.326
	MAE	1.042	5.654	1.137	13.793
	WMAPE	0.418	0.206	0.436	0.629
60 min	RMSE	5.369	21.015	7.938	55.261
	MAE	2.953	15.518	4.547	38.639
	WMAPE	0.201	0.199	0.345	0.618

仅对由于压缩操作而填充的零值进行掩蔽，避免在模型训练过程中误差反向传播。

• **MPG 损失函数 (Net-MPG-loss function)**: 该变体通过 MPG 损失函数训练 PAG-STAN。

表9总结了不同损失函数下 PAG-STAN 的预测性能。结果表明，使用一般 MSE 损失函数时，PAG-STAN 的表现最差，因为它忽略了 OD 需求与进站流量之间的物理量信息以及填充零值带来的干扰。当采用掩蔽操作消除填充零值带来的误差时，性能有所改善。此外，将 OD 需求和客流量之间的物理量关系嵌入损失函数中，增强了预测性能，突出了物理信息对模型训练和可解释性的重要性。值得注意的是，在使用 MPG 损失函数训练 PAG-STAN 时，MPG 损失函数在所有时间间隔中始终表现出优越的预测性能，验证了其在增强模型训练和可解释性方面的有效性。

表9 不同损失函数对 OD 预测的性能

Time	Metrics	Net-general loss	Net-informa- tion loss	Net-masked loss	Net-MPG-loss
15 min	RMSE	2.149	1.991	2.014	1.898
	MAE	0.869	0.823	0.843	0.803
	WMAPE	0.508	0.497	0.503	0.491
30 min	RMSE	2.942	2.895	2.904	2.821
	MAE	1.431	1.373	1.394	1.328
	WMAPE	0.380	0.368	0.372	0.361
60 min	RMSE	4.393	4.291	5.317	4.218
	MAE	2.172	2.007	2.109	1.903
	WMAPE	0.314	0.299	0.306	0.293

5.3.6. 多模块共同效果

为研究单一模块的有效性，展开了上述消融研究。然而，不同的模块可能会共同影响预测性能。因此，本小节讨论了多个模块（多周期 OD 模式、实时 OD 估计模块和异构信息融合模块）对模型的共同影响，进一步帮助理解每个模块的贡献。

• **仅实时估计 OD (Net-only real-time estimated OD)**: 该变体仅利用实时估计的 OD 需求预测未来的 OD 需求。

• **多周期历史 OD 和不完整实时 OD (Net-multi-periodic historical OD and incomplete real-time OD)**: 该变体使用多周期 OD 模式（周度/日度历史 OD 模式）和不完整的实时 OD 需求数据来预测未来的 OD 需求。

• **多周期历史 OD 和实时估计 OD (Net-multi-periodic historical OD and real-time estimated OD)**: 该变体使用多周期 OD 模式（周度/日度历史 OD 模式）和实时估计的 OD 需求来预测未来的 OD 需求。

• **实时估计 OD 和异构信息 (Net-real-time estimated OD and heterogeneous information)**: 该变体利用实时估计的 OD 需求和异构信息来预测未来的 OD 需求。

• **不完整实时 OD 和异构信息 (Net-incomplete real-time OD and heterogeneous information)**: 该变体利用不完整的实时 OD 需求和异构信息来预测未来的 OD 需求。

表10展示了不同模块组合的评估指标。需注意，模型的性能显著受到实时 OD 需求完整性的影响。即使加入了周度或日度时间序列 OD 需求信息或异构信息，不完整的 OD 需求输入仍导致预测性能较差。当结合周度和日度时间序列 OD 需求信息时，相关变体的表现优于考虑异构信息的变体，表明历史长周期性信息在提升预测性能方面的重要性。此外，多周期历史 OD 和不完整实时 OD 的预测性能位居第二，这得益于估算的实时 OD 需求数据和历史周度/日度时间序列 OD 需求数据所提供的丰富信息。这些数据源从多个角度提供了 OD 需求的时空分布的全面信息。这表明，估算的实时 OD 需求数据和历史周度/日度时间序列 OD 需求数据共同提升了预测性能。因此，估算的实时 OD 需求数据和历史周度及日度时间序列 OD 需求数据的共同效果显著增强了地铁 OD 预测，异构数据在一定程度上提供了辅助的演化信息。

表10 不同组合模块的性能

Time interval	Metrics	Net-real-time estimated OD	Net-estimated OD + multi-periodic	Net-estimated OD + heterogeneous data	Net-incomplete OD + multi-periodic	Net-incomplete OD + heterogeneous data	PAG-STAN
15 min	RMSE	2.283	2.088	2.202	2.345	2.377	1.898
	MAE	0.934	0.843	0.886	1.117	1.165	0.803
	WMAPE	0.551	0.529	0.531	0.568	0.583	0.491
30 min	RMSE	3.201	3.117	3.152	3.311	3.329	2.821
	MAE	1.486	1.403	1.457	1.493	1.501	1.328
	WMAPE	0.411	0.384	0.391	0.454	0.468	0.361
60 min	RMSE	4.662	4.550	4.612	4.685	4.699	4.218
	MAE	2.274	2.247	2.235	2.434	2.453	1.903
	WMAPE	0.348	0.320	0.328	0.352	0.366	0.293

6. 结论

本研究聚焦于一个具有挑战性的任务，即在疫情背景下的URT系统短期OD需求预测，研究的贡献如下：

- 本文全面总结了地铁OD需求预测中存在的问题，包括OD需求数据实时可用性、稀疏性、高维性以及疫情下外部因素的影响。

- 此外，本文提出了一个新的PAG-STAN框架来解决上述问题。该框架首先得到近似完整的实时稠密OD需求矩阵。随后，通过编码器-解码器框架捕捉OD地铁需求的多周期时空依赖性。最后，通过将客流量关系嵌入损失函数中，增强了模型的训练和可解释性。

- 在大规模新冠病毒肺炎疫情地铁OD需求数据集上的实验表明，PAG-STAN在疫情期间的地铁OD需求预测上超越了当天的前沿方法。与此同时，模型的通用性也在北京地铁的另一个非疫情期间的数据集上得到了验证。然而，本研究也面临一些限制。

- 尽管动态OD需求矩阵压缩模块有效解决了稀疏性和高维性问题，但它导致了OD对之间相对位置信息的丢失。未来的工作将探索更佳方案，在保留完整相对位置信息的同时解决稀疏性和高维性问题。

- 在其他突发事件（如大型体育赛事）下的URT系统OD需求预测也是具有重要研究价值的课题。未来的工作将应用PAG-STAN预测其他突发事件下的OD需求，增强PAG-STAN的普适性。

致谢

本研究受国家自然科学基金项目(72288101、72201029、72322022)资助。

Compliance with ethics guidelines

Shuxin Zhang, Jinlei Zhang, Lixing Yang, Feng Chen, Shukai Li, and Ziyou Gao declare that they have no conflicts of interest or financial conflicts to disclose.

References

- [1] Yao W, Yu J, Yang Y, Chen N, Jin S, Hu Y, et al. Understanding travel behavior adjustment under COVID-19. *Commun Transp Res* 2022;2:100068.
- [2] Borjigin SG, He Q, Niemeier DA. COVID-19 transmission in U. S. transit buses: a scenario-based approach with agent-based simulation modeling (ABSM). *Commun Transp Res* 2023;3:100090.
- [3] Li Y, Yu R, Shahabi C, Liu Y. Diffusion convolutional recurrent neural network: data-driven traffic forecasting. 2017. arXiv:1707.01926.
- [4] Zhang J, Zheng Y, Qi D. Deep spatio-temporal residual networks for citywide crowd flows prediction. In: *Proceedings of the AAAI-17: 31st AAAI conference on artificial intelligence*; 2017 Feb 4–9; San Francisco, CA, USA. Washington, DC: Association for the Advancement of Artificial Intelligence (AAAI); 2017.
- [5] Zhao L, Song Y, Zhang C, Liu Y, Wang P, Lin T, et al. T-GCN: a temporal graph convolutional network for traffic prediction. *IEEE Trans Intell Transp Syst* 2020;21(9):3848–58.
- [6] Jiang W, Ma Z, Koutsopoulos HN. Deep learning for short-term origin-destination passenger flow prediction under partial observability in urban railway systems. *Neural Comput Appl* 2022;34:4813–30.
- [7] Noursalehi P, Koutsopoulos HN, Zhao JH. Dynamic origin-destination prediction in urban rail systems: a multi-resolution spatio-temporal deep learning approach. *IEEE Trans Intell Transp Syst* 2022;23(6):5106–15.
- [8] Zhang J, Che H, Chen F, Ma W, He Z. Short-term origin-destination demand prediction in urban rail transit systems: a channel-wise attentive split-convolutional neural network method. *Transp Res Part C Emerg Technol* 2021; 124:102928.
- [9] Liu L, Zhu Y, Li G, Wu Z, Bai L, Lin L. Online metro origin-destination prediction via heterogeneous information aggregation. 2022. arXiv: 2107.00946v5.
- [10] Zhu G, Ding J, Wei Y, Yi Y, Xu SSD, Wu EQ. Two-stage OD flow prediction for emergency in urban rail transit. *IEEE Trans Intell Transp Syst* 2023;25(1):920–8.
- [11] Liu L, Qiu Z, Li G, Wang Q, Ouyang W, Lin L. Contextualized spatial-temporal network for taxi origin-destination demand prediction. *IEEE Trans Intell Transp Syst* 2019;20(10):3875–87.
- [12] Zou XX, Zhang SY, Zhang CH, Yu JJQ, Chung E. Long-term origin-destination demand prediction with graph deep learning. *IEEE Trans Big Data* 2021;8(6):1481–95.
- [13] Van der Voort M, Dougherty M, Watson S. Combining kohonen maps with arima time series models to forecast traffic flow. *Transp Res Part C Emerg Technol* 1996;4(5):307–18.
- [14] Tan MC, Wong SC, Xu JM, Guan ZR, Zhang P. An aggregation approach to short-term traffic flow prediction. *IEEE Trans Intell Transp Syst* 2009; 10(1):

- 60–9.
- [15] Ni M, He Q, Gao J. Forecasting the subway passenger flow under event occurrences with social media. *IEEE Trans Intell Transp Syst* 2016;18(6):1623–32.
 - [16] Castro-Neto M, Jeong YS, Jeong MK, Han LD. Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions. *Expert Syst Appl* 2009;36(3):6164–73.
 - [17] Högborg P. Estimation of parameters in models for traffic prediction: a nonlinear regression approach. *Transp Res* 1976;10(4):263–5.
 - [18] Sun H, Liu HX, Xiao H, He RR, Ran B. Use of local linear regression model for short-term traffic forecasting. *Transp Res Rec* 2003;1836(1):143–50.
 - [19] Zheng Z, Su D. Short-term traffic volume forecasting: a k -nearest neighbor approach enhanced by constrained linearly sewing principle component algorithm. *Transp Res Part C Emerg Technol* 2014;43:143–57.
 - [20] Li MW, Hong WC, Kang HG. Urban traffic flow forecasting using Gauss–SVR with cat mapping, cloud model and PSO hybrid algorithm. *Neurocomputing* 2013;99:230–40.
 - [21] Hong WC. Traffic flow forecasting by seasonal SVR with chaotic simulated annealing algorithm. *Neurocomputing* 2011;74(12–13):2096–107.
 - [22] Cai P, Wang Y, Lu G, Chen P, Ding C, Sun J. A spatiotemporal correlative k -nearest neighbor model for short-term traffic multistep forecasting. *Transp Res Part C Emerg Technol* 2016;62:21–34.
 - [23] Lin G, Lin A, Gu D. Using support vector regression and k -nearest neighbors for short-term traffic flow prediction based on maximal information coefficient. *Inf Sci* 2022;608:517–31.
 - [24] Avila AM, Mezic' I. Data-driven analysis and forecasting of highway traffic dynamics. *Nat Commun* 2020;11:2090.
 - [25] Lv Y, Duan Y, Kang W, Li Z, Wang FY. Traffic flow prediction with big data: a deep learning approach. *IEEE Trans Intell Transp Syst* 2014;16(2):865–73.
 - [26] Liu Y, Liu Z, Jia R. DeepPF: a deep learning based architecture for metro passenger flow prediction. *Transp Res Part C Emerg Technol* 2019;101:18–34.
 - [27] Liu L, Chen RC. A novel passenger flow prediction model using deep learning methods. *Transp Res Part C Emerg Technol* 2017;84:74–91.
 - [28] Polson NG, Sokolov VO. Deep learning for short-term traffic flow prediction. *Transp Res Part C Emerg Technol* 2017;79:1–17.
 - [29] Guo J, Xie Z, Qin Y, Jia L, Wang Y. Short-term abnormal passenger flow prediction based on the fusion of SVR and LSTM. *IEEE Access* 2019;7:42946–55.
 - [30] Jing Y, Hu H, Guo S, Wang X, Chen F. Short-term prediction of urban rail transit passenger flow in external passenger transport hub based on LSTM-LGB-DRS. *IEEE Trans Intell Transp Syst* 2021;22(7):4611–21.
 - [31] An J, Fu L, Hu M, Chen W, Zhan J. A novel fuzzy-based convolutional neural network method to traffic flow prediction with uncertain traffic accident information. *IEEE Access* 2019;7:20708–22.
 - [32] Liu Y, Lyu C, Liu X, Liu Z. Automatic feature engineering for bus passenger flow prediction based on modular convolutional neural network. *IEEE Trans Intell Transp Syst* 2021;22(4):2349–58.
 - [33] Du B, Peng H, Wang S, Bhuiyan MZA, Wang L, Gong Q, et al. Deep irregular convolutional residual LSTM for urban traffic passenger flows prediction. *IEEE Trans Intell Transp Syst* 2020;21(3):972–85.
 - [34] Chen C, Liu Y, Chen L, Zhang C. Bidirectional spatial–temporal adaptive transformer for urban traffic flow forecasting. *IEEE Trans Neural Netw Learn Syst* 2022;34(10):6913–25.
 - [35] Yu B, Yin H, Zhu Z. Spatio–temporal graph convolutional networks: a deep learning framework for traffic forecasting. 2017. arXiv:1709.04875.
 - [36] Ali A, Zhu Y, Zakarya M. Exploiting dynamic spatio–temporal graph convolutional neural networks for citywide traffic flows prediction. *Neural Netw* 2022;145:233–47.
 - [37] Lv M, Hong Z, Chen L, Chen T, Zhu T, Ji S. Temporal multi-graph convolutional network for traffic flow prediction. *IEEE Trans Intell Transp Syst* 2021;22(6):3337–48.
 - [38] Yu B, Lee Y, Sohn K. Forecasting road traffic speeds by considering area-wide spatio–temporal dependencies based on a graph convolutional neural network (GCN). *Transp Res Part C Emerg Technol* 2020;114:189–204.
 - [39] Peng H, Du B, Liu M, Liu M, Ji S, Wang S, et al. Dynamic graph convolutional network for long-term traffic flow prediction with reinforcement learning. *Inf Sci* 2021;578:401–16.
 - [40] Wu Z, Pan S, Long G, Jiang J, Zhang C. Graph WaveNet for deep spatial–temporal graph modeling. 2019. arXiv:1906.00121.
 - [41] Wang J, Zhang Y, Wei Y, Hu Y, Piao X, Yin B. Metro passenger flow prediction via dynamic hypergraph convolution networks. *IEEE Trans Intell Transp Syst* 2021;22(12):7891–903.
 - [42] Reza S, Ferreira MC, Machado JJM, Tavares JMRS, Machado JJM, Tavares JMRS. A multi-head attention-based transformer model for traffic flow forecasting with a comparative analysis to recurrent neural networks. *Expert Syst Appl* 2022;202:117275.
 - [43] Ye X, Fang S, Sun F, Zhang C, Xiang S. Meta graph transformer: a novel framework for spatial–temporal traffic prediction. *Neurocomputing* 2022;491:544–63.
 - [44] Zhang H, Zou Y, Yang X, Yang H. A temporal fusion transformer for short-term freeway traffic speed multistep prediction. *Neurocomputing* 2022;500:329–40.
 - [45] Yan H, Ma X, Pu Z. Learning dynamic and hierarchical traffic spatiotemporal features with transformer. *IEEE Trans Intell Transp Syst* 2022;23(11):22386–99.
 - [46] Xie Y, Niu J, Zhang Y, Ren F. Multisize patched spatial–temporal transformer network for short- and long-term crowd flow prediction. *IEEE Trans Intell Transp Syst* 2022;23(11):21548–68.
 - [47] Xu M, Dai W, Liu C, Gao X, Lin W, Qi GJ, et al. Spatial–temporal transformer networks for traffic flow forecasting. 2020. arXiv:2001.02908.
 - [48] Chu KF, Lam AYS, Li VOK. Deep multi-scale convolutional LSTM network for travel demand and origin–destination predictions. *IEEE Trans Intell Transp Syst* 2020;21(8):3219–32.
 - [49] Hu J, Yang B, Guo C, Jensen CS, Xiong H. Stochastic origin–destination matrix forecasting using dual-stage graph convolutional, recurrent neural networks. In: *Proceedings of the 2020 IEEE 36th International Conference on Data Engineering (ICDE-2020)*; 2020 Apr 20–24; Dallas, TX, USA. New York City: IEEE; 2020. p. 1417–28.
 - [50] Yao X, Gao Y, Zhu D, Manley E, Wang J, Liu Y. Spatial origin–destination flow imputation using graph convolutional networks. *IEEE Trans Intell Transp Syst* 2021;22(12):7474–84.
 - [51] Ke J, Qin X, Yang H, Zheng Z, Zhu Z, Ye J. Predicting origin–destination ride-sourcing demand with a spatio–temporal encoder–decoder residual multi-graph convolutional network. *Transp Res Part C Emerg Technol* 2021;122:102858.
 - [52] Huang Z, Wang D, Yin Y, Li X. A spatiotemporal bidirectional attention-based ride-hailing demand prediction model: a case study in Beijing during COVID19. *IEEE Trans Intell Transp Syst* 2022;23(12):25115–26.
 - [53] Huang Z, Zhang W, Wang D, Yin Y. A GAN framework-based dynamic multigraph convolutional network for origin–destination-based ride-hailing demand prediction. *Inf Sci* 2022;601:129–46.
 - [54] Qurashi M, Lu QL, Cantelmo G, Antoniou C. Dynamic demand estimation on large scale networks using principal component analysis: the case of nonexistent or irrelevant historical estimates. *Transp Res Part C Emerg Technol* 2022;136:103504.
 - [55] Qurashi M, Ma T, Chaniotakis E, Antoniou C. PC–SPSA: employing dimensionality reduction to limit SPSA search noise in DTA model calibration. *IEEE Trans Intell Transp Syst* 2020;21(4):1635–45.
 - [56] Zhang J, Chen F, Guo Y, Li X. Multi-graph convolutional network for short-term passenger flow forecasting in urban rail transit. *IET Intell Transp Syst* 2020;14(10):1210–7.
 - [57] Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. 2016. arXiv:1609.02907.
 - [58] Shi X, Chen Z, Wang H, Yeung DY, Wong WK, Woo WC. Convolutional LSTM network: a machine learning approach for precipitation nowcasting. In: *Proceedings of the 28th International Conference on Neural Information Processing Systems*; 2015 Dec 7–12; Montreal, QC, Canada. New York City: Association for Computing Machinery (ACM); 2015. p. 802–10.
 - [59] Zheng H, Lin F, Feng X, Chen Y. A hybrid deep learning model with attention-based Conv-LSTM networks for short-term traffic flow prediction. *IEEE Trans Intell Transp Syst* 2021;22(11):6910–20.
 - [60] Lu X, Ma C, Qiao Y. Short-term demand forecasting for online car-hailing using Conv-LSTM networks. *Physica A* 2021;570:125838.
 - [61] Li Y, Chai S, Wang G, Zhang X, Qiu J. Quantifying the uncertainty in long-term traffic prediction based on PI-ConvLSTM network. *IEEE Trans Intell Transp Syst* 2022;23(11):20429–41.
 - [62] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Aidan N, et al. Attention is all you need. In: *Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017)*; 2017 Dec 4–9; Long Beach, CA, USA. Online: NeurIPS Proceedings; 2017.
 - [63] Chen Y, Lv Y, Wang X, Li L, Wang FY. Detecting traffic information from social media texts with deep learning approaches. *IEEE Trans Intell Transp Syst* 2019;20(8):3049–58.
 - [64] Yao W, Qian S. From twitter to traffic predictor: next-day morning traffic prediction using social media data. *Transp Res Part C Emerg Technol* 2021;124:102938.
 - [65] Zhou H, Zhang S, Peng J, Zhang S, Li J, Xiong H, et al. Informer: beyond efficient transformer for long sequence time-series forecasting. *Proc Conf AAAI Artif Intell* 2021;35(12):11106–15.