



交通运输工程与信息学报

Journal of Transportation Engineering and Information

ISSN 1672-4747, CN 51-1652/U

《交通运输工程与信息学报》网络首发论文

题目：考虑图特征的时空交通需求集成预测模型
作者：游欣，甘蜜，张义萌，蔡信豪，程靖怡
DOI：10.19961/j.cnki.1672-4747.2025.12.010
收稿日期：2025-12-09
网络首发日期：2026-03-09
引用格式：游欣，甘蜜，张义萌，蔡信豪，程靖怡. 考虑图特征的时空交通需求集成预测模型[J/OL]. 交通运输工程与信息学报.
<https://doi.org/10.19961/j.cnki.1672-4747.2025.12.010>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

考虑图特征的时空交通需求集成预测模型

游欣¹ 甘蜜^{1,2} 张义萌^{*1,2,3} 蔡信豪⁴ 程靖怡⁵

(1.西南交通大学, 交通运输与物流学院, 成都 611756; 2.西南交通大学, 综合运输智能化四川省重点实验室, 成都 611756; 3.西南交通大学, 唐山研究院, 唐山 063000; 4.西南交通大学, 机械工程学院, 成都 611756; 5.代尔夫特理工大学, 土木工程与地球科学学院, 代尔夫特 2628 CD)

摘要:【背景】交通需求预测对于交通基础设施规划和资源分配至关重要。在时空需求预测中, 模型性能在不同站点和时间段之间差异较大, 因此在时间和空间维度上实现稳定精准预测构成了挑战。【目的】实现模型跨场景直接部署, 提升预测模型在不同时空场景的实用性, 以减少预测不确定对未来决策的影响。【方法】首先, 提出了考虑图特征的时空交通需求集成预测模型 (ST-SGAT), 采用堆叠的集成学习策略融合不同的基模型和图注意力网络 (GAT), ST-SGAT 将长短时记忆网络、梯度提升决策树、多层感知机作为基模型捕获时间特征, 并利用 GAT 对空间相关性进行编码, 自适应分配关联权重, 从而实现时空交通需求的稳健建模。随后, 通过删除 ST-SGAT 的特定组件实现消融实验, 系统分析各组件的贡献度与协同效应, 并引入降水量作为外部特征, 进一步探究模型在常规与突发气象场景下的鲁棒性与特征利用能力。【数据】研究数据来源于纽约市 424 个地铁站点 688 天的每小时客流量数据, 预处理后生成 2 h、3 h、6 h、12 h、24 h 等多时间粒度数据集。【结论】ST-SGAT 模型在交通时空需求预测中综合性能最优, 时空适配性突出。时间维度上, 于小时级至日级多粒度数据及工作日、周末数据中均展现最优综合预测精度; 空间维度上, 可自适应适配不同站点类型, 泛化能力强。集成框架能够实现多组件协同互补, 使模型在跨粒度、跨站点、跨时间类别的多场景中持续保持优异性能。同时, 模型在常规与突发降水场景下均表现出更强的鲁棒性与特征利用稳定性, 满足实际交通的多样化预测需求。

关键词: 城市交通; 需求预测; 集成学习; 多时间粒度数据; 图注意力网络

中图分类号: U292.16

文献标志码: A

DOI: 10.19961/j.cnki.1672-4747.2025.12.010

Spatiotemporal traffic demand ensemble forecasting model considering graph features

YOU Xin¹, GAN Mi^{1,2}, ZHANG Yimeng^{*1,2,3}, CAI Xinhao⁴, CHENG Jingyi⁵

(1. School of Traffic and Logistics, Southwest Jiaotong University, Chengdu 611756, China; 2. Intelligent Comprehensive Transportation Key Laboratory of Sichuan Province, Southwest Jiaotong University, Chengdu 611756, China; 3. Tangshan Institute, Southwest Jiaotong University, Tangshan 063000, China; 4. School of Mechanical Engineering, Southwest Jiaotong University, Chengdu 611756, China; 5. Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft 2628 CD, The Netherlands)

Abstract: 【Background】Traffic demand forecasting is crucial for transportation infrastructure planning and resource allocation. In spatiotemporal demand forecasting, model performance varies significantly across different stations and time periods, posing a challenge to achieving stable and accurate predictions in both temporal and spatial dimensions. 【Objective】To realize direct cross-scenario deployment of the model and enhance the practicality of forecasting models across diverse spatiotemporal scenarios, thereby reducing the impact of prediction uncertainty on future decision-making. 【Method】A spatiotemporal traffic demand ensemble forecasting model (ST-SGAT) considering graph features is proposed, which adopts a

收稿日期: 2025-12-09

录用日期: 2026-02-26

审稿日期: 2025-12-09~2025-12-17; 2026-02-14~2026-02-26

基金项目: 国家自然科学基金 (52402521, U2568219, 52372306, 52232011); 四川省自然科学基金 (2025NSFSC1969); 河北省自然科学基金 (G2024105007); 中央高校基本科研业务费 (2682025CX056)

作者简介: 游欣 (2001—), 女, 硕士, 研究方向为智慧交通与物流, E-mail: yne@my.swjtu.edu.cn

通讯作者: 张义萌 (1994—), 男, 助理教授, 研究方向为多式联运与多模式交通优化, Email: yimengzhang@swjtu.edu.cn

stacking ensemble learning strategy to integrate multiple base models with Graph Attention Network (GAT). ST-SGAT employs Long Short-Term Memory, eXtreme Gradient Boosting, and Multilayer Perceptron as base models for capturing temporal features, while leveraging GAT to encode spatial correlations and adaptively assign association weights, thereby achieving robust modeling of spatiotemporal traffic demand. Subsequently, ablation experiments are conducted by removing specific components of ST-SGAT to systematically analyze the contribution degree and synergistic effects of each component, and precipitation is introduced as an external feature to further explore the model's robustness and feature utilization capability under normal and sudden meteorological scenarios. **[Data]** The research data are derived from hourly passenger flow records of 424 subway stations in New York City over 688 days, which are preprocessed to generate multi-time granularity datasets (e.g., 2 h, 3 h, 6 h, 12 h, 24 h). **[Conclusion]** Experimental results demonstrate that the ST-SGAT model achieves the optimal comprehensive performance in spatiotemporal traffic demand forecasting with outstanding spatiotemporal adaptability. In the temporal dimension, it exhibits the best overall prediction accuracy across multi-time granularity data (from hourly to daily) as well as data for weekdays and weekends. In the spatial dimension, it can adaptively fit different station types, showing strong generalization ability. The ensemble framework enables the synergy and complementarity of multiple components, allowing the model to consistently maintain excellent performance in multiple scenarios spanning different granularities, stations, and time categories. Meanwhile, the model exhibits stronger robustness and stability in feature utilization under both normal and sudden precipitation scenarios, thus meeting the diverse forecasting needs of practical transportation systems.

Key words: urban traffic; demand forecasting; ensemble learning; multi-time granularity data; graph attention network

0 引言

交通需求预测是交通管理的关键组成部分，更是智能交通系统实现高效运转、风险防控与可持续发展的核心技术支撑，其精准性直接影响交通系统的资源配置效率与决策科学性^[1]。通过机器学习和深度学习技术预测交通需求，有助于交通管理、协调客货路线规划，在保障运输服务质量与需求覆盖的基础上，为交通运营商的服务规划、运营资源的优化配置和运输成本有效控制提供可操作的见解^[2-3]。此外，交通需求预测还能为城市交通基础设施规划、公共交通网络优化等长期发展决策提供数据支撑，推动城市交通向智能化、高效化、低碳化转型^[4]。

解决交通网络中的不确定性和路线优化问题，仍是交通运输领域的核心研究方向。以往的大多数研究通过基于场景或概率分布假设的方式处理不确定性，但这无法反映现实世界的复杂性^[5]。近年来，学者们在各个领域越来越倾向于采用“预测—优化”框架，即利用预测模型和经验拟合来描述复杂的不确定性，并将预测结果输入后续的优化模型，以实现数据驱动的、可解释的决策^[6-7]。然而，预测的不确定性会对优化结果和整体网络效率产生显著影响^[8]。较大的预测误差可能导致交通资源分配不均和网络成本增加，例如高峰时段运力供给不足、平峰时段资源闲置，或突发场景下预测失效导致调度滞后，因此，构建稳健且稳定的预测模型，解决预测误差过大、场景适配性差的实际痛点，对于高效优化和避免决策中的极端结果至关重要。

在交通需求预测领域，大量的研究致力于提高客运和货运的预测准确性和实际适用性，预测方法也从传统的统计预测方法发展到将时空数据特征与外部因素相结合的机器学习模型和深度学习模型。其中统计预测法适用于有明显规律的单变量时间序列数据，但处理非线性关系能力有限，如王莹等^[9]构建季节性自回归整合移动平均模型，对北京地铁进站客流量进行预测，但节假日客流预测效果不佳。机器学习模型可综合多类因素，挖掘复杂非线性关系，如 Wang 等^[10]提出一种基于贝叶斯优化的支持向量回归短时交通流量预测框架，在 4 个不同交通场景中验证模型有效性。深度学习模型则通过多层网络结构自动提取深层特征，广泛应用于交通需求预测中。循环神经网络^[11]、长短时

记忆 (Long Short-Term Memory, LSTM) 神经网络^[12]、门控循环单元 (Gated Recurrent Unit, GRU)^[13]等模型侧重于挖掘时间模式, 但这些模型在处理具有空间依赖性的数据时存在局限性。因此, 卷积神经网络 (convolutional neural network, CNN) 被引入, 通过卷积核和池化操作捕捉欧氏空间相关性^[14]。而图神经网络能处理非欧氏空间特征, 它专门用于对图结构数据中的节点与边关系进行建模, 能够更全面地捕捉空间特征。唐进君等^[15]通过特征模块捕捉交通流量与外部特征的关联性、图卷积神经网络 (Graph Convolutional Networks, GCN) 挖掘空间关联性、GRU 学习时间关联性, 解决了现有方法忽略时空相关性及外部特征影响的问题。胡晓伟等^[16]组合残差神经网络、CNN 和 LSTM 捕捉客流的时空特征, 在特征融合过程中应用注意力机制突出关键特征, 并引入遗传算法对模型的结构和超参数进行优化, 在不同类型地铁站点的客流预测中均表现出高有效性。当前交通流量预测研究正朝着更精细的时空建模方向演进, 通过高阶关联刻画与动态依赖捕捉突破传统模型局限。如 Xian 等^[17]提出的多尺度双超图融合时空网络, 通过多节点关联表征突破二元依赖限制, 从微观出行意图、中观社区通勤、宏观区域传播三个尺度构建超图, 实现交通网络高阶空间特征的系统性建模。Zhang 等^[18]设计的双向异质性动态图卷积网络模型, 通过静态异质性因子量化常规空间依赖, 动态异质性因子捕捉交通事故等突发场景下的依赖突变, 显著降低了异常状态下的预测误差。Liu 等^[19]提出多层时空关联感知图注意力网络, 结合空间图注意力模块提取局部与全局空间依赖, 时间图注意力模块基于节点短期相似度矩阵捕捉动态时序特征, 最终利用时空 Transformer 融合跨尺度时空动态依赖, 在长短时预测任务中均表现优异。

为了构建稳健的预测模型, 基于集成学习预测方法应运而生, 它通过结合多种数据源和策略采取多种预测技术来提高准确性和稳定性^[20]。在集成学习中, 装袋法 (Bagging)、提升法 (Boosting) 和堆叠法 (Stacking) 是三种经典策略, Bagging 通过并行训练降低模型方差, 但由于基模型多为同类型模型, 难以有效挖掘不同基模型间的互补性; Boosting 通过串行迭代降低模型偏差, 易受噪声样本影响, 基模型无法并行训练; 而 Stacking 能够在统一框架内自适应融合多类基学习器的输出, 充分利用不同模型的优势互补, 同时支持基模型并行训练, 显著提升预测的泛化能力^[21]。Moretti 等^[22]利用 Bagging 对神经网络与简单统计模型的集成模型进行改进, 根据误差自适应选择预测机制, 有效降低了城市交通流的短期预测误差。Boosting 采用顺序训练模式, 根据前面模型的误差调整样本权重, 促使后续模型聚焦于分类错误的样本, 进而降低偏差, 如 Lee 和 Kim^[23]提出了融合自适应提升算法与残差决策树的模型, 显著提升了共享单车系统在极端需求下的预测精度。Stacking 则并行训练模型, 并使用元学习器整合模型结果, 如 Lei 等^[24]提出一种基于树模型的 Stacking 集成学习预测方法, 以轻量级梯度提升机、随机森林、极限梯度提升机 (eXtreme Gradient Boosting, XGBoost) 为基学习器并行训练, 最后以线性回归作为元学习器, 有效提升了内河船舶到港时间预测的准确性与稳健性; 李浩等^[25]为解决单一模型预测航空器滑出时间精度提升受限的问题, 提出结合最大互信息系数、迭代注意力特征融合模块与堆叠集成学习框架的预测模型, 将支持向量回归、随机森林、多层感知机 (Multilayer Perceptron, MLP) 和 XGBoost 作为基学习器构造特征, MLP 作为元学习器。

现有的交通需求预测研究在时空建模和模型融合方面取得了进展, 但结合实际应用场景来看, 在基于集成的稳健建模来提高预测稳定性方面仍存在三大挑战。首先, 多数集成方法依赖同质基模型的简单组合, 缺乏对异构模型优势互补特性的深度挖掘, 难以全面捕捉交通需求复杂时空特征,

进而导致预测精度不足；其次，现有使用集成模型的研究中，元学习器多采用线性回归等简单结构，未能有效建模交通网络中不同站点间的复杂空间依赖关系，导致空间关联性捕捉不足；最后，单一模型或传统集成模型的场景适配性差，在不同时间粒度、不同站点类型及突发场景下，预测稳定性不足，难以满足交通管理中全场景稳健预测的实际需求。在此背景下，本研究提出基于 Stacking 集成策略和图注意力网络（Graph Attention Network, GAT）的时空交通需求预测模型（Spatiotemporal Stacking-Based Graph Attention Network, ST-SGAT），该模型以 Stacking 集成为整体框架，基模型层采用多种时序预测模型完成初步预测，元模型层引入 GAT 实现空间建模。与传统的单一模型不同，ST-SGAT 整合了不同的模型，以利用它们的互补优势，这种方法克服了过拟合和空间建模不完整等局限性。依托 Stacking 框架的自适应融合能力与 GAT 的空间建模优势，提升模型在不同时间粒度、不同站点类型下的预测稳定性，为交通资源调度、运力优化提供更可靠的预测支撑，解决传统模型场景适配性差的问题。

本研究的核心贡献如下：第一，提出一种基于 Stacking 策略的集成学习方法，通过融合异构单一学习器实现交通需求预测，模型框架具备高度灵活性，支持各类型的预测模型作为基模型嵌入；第二，创新性地引入 GAT 作为元学习器，有效建模不同站点间的空间依赖关系，弥补传统模型对空间相关性捕捉不足的缺陷；第三，针对交通时空需求，所选取的异构基模型在时间特征提取中各有侧重、优势互补，具有较强的特征表征能力；第四，基于真实世界数据集开展充分验证，不仅证实了堆叠集成方法在交通需求预测中的稳健性与有效性，还量化分析了不同模块及特征对预测性能的贡献度，为相关研究提供实践参考。

1 时空交通需求集成预测模型

时空需求预测模型的性能受数据集特性与应用场景属性的双重约束，面对场景动态变化或数据集异质性问题时，集成学习模型展现出显著适配优势。集成学习的核心思想在于融合多个基模型的互补优势以提升整体预测性能，该方法可有效规避单一模型受随机因素影响导致的预测偏差，进而强化模型的泛化能力与预测稳定性^[20]。

本文采用 Stacking 集成策略并结合 GAT 构建时空交通需求预测框架，该框架由基模型模块与元模型模块构成。基模型模块以时间特征捕捉为核心目标，同时需满足复杂度适中、能处理非平稳交通数据、训练效率较高且成本可控的要求，因此选取 LSTM、XGBoost 及 MLP 作为基模型。LSTM 擅长捕捉时序数据中的长期依赖关系与动态波动规律，能够有效建模交通客流的时序演变趋势，尤其适合处理具有周期性、延迟性的交通时间序列；XGBoost 具备较强的非线性拟合能力与抗噪声能力，能够高效挖掘特征与预测目标间的复杂统计关系，对异常值与数据异质性具有良好鲁棒性；MLP 的结构简洁，能够以端到端方式学习特征映射关系，提供稳定的基准预测结果。其中，XGBoost 与 MLP 侧重挖掘时间序列特征和预测目标的统计规律，LSTM 则专注于学习时序数据的长期依赖关系，三类模型在建模机理上形成互补，协同实现对时间特征的高效提取。元模型模块使用 GAT 模型，专注于挖掘数据中多个地点的时间序列之间的空间依赖关系。通过整合两类模型的特性，能够捕捉复杂的时空关系，同时具备良好的灵活性，在不同交通场景下，可替换基模型以满足多样化预测需求，最终实现更精准、稳定的交通需求预测。该框架的核心逻辑是提取时空数据的周期性、关联性特征，

本质上不局限于特定城市或交通方式，只要数据为等时间间隔收集的时空数据，满足时序依赖与空间关联，即可适配该模型。模型整体架构如图 1 所示。

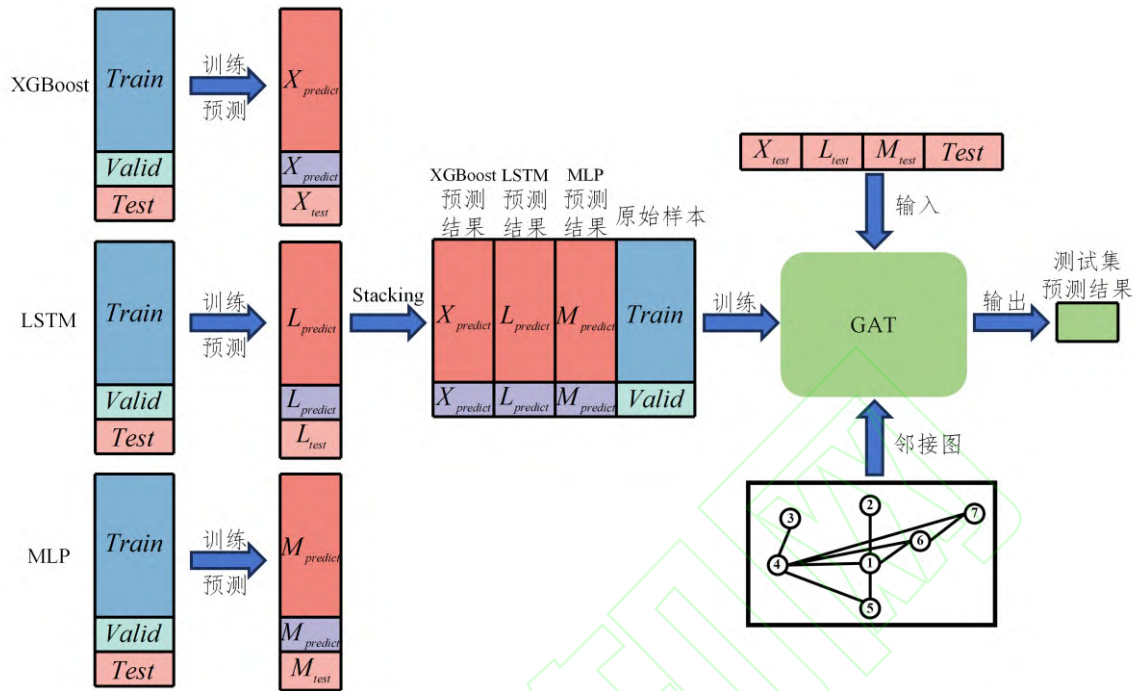


图 1 时空交通集成预测模型训练框架

Fig. 1 Training framework for spatiotemporal traffic ensemble forecasting model

1.1 时空特征提取

交通客货运需求数据具有时间周期性，过去时刻的需求与未来的需求具有一定的相关性，对历史数据进行采样可以提取出对预测有用的时间特征^[26]。考虑到时间序列的不同周期性特征，采样时按照不同的时间间隔进行，除了邻近相似的历史数据外，还提取历史周期中同一时刻的数据，特征提取示意图如图 2 所示，周期性特征包括日、周、月和年四项周期性等，可根据时间序列的时间粒度，选择合适的周期性特征。

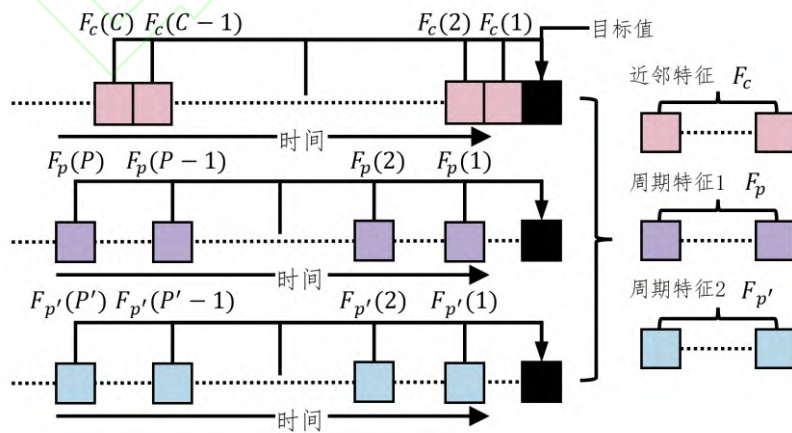


图 2 时间特征提取

Fig. 2 Temporal feature extraction

对于不同的观测实例，令 N 为样本数量， M 为地点数量， F 为基模型的输入特征数量。将全局输入的样本定义为 $X \in \mathbb{R}^{N \times M \times F}$ ，其中 $x_{n,m} \in \mathbb{R}^F$ 表示第 n 个样本、第 m 个地点的特征向量，由提取的时间特征拼接而成。

空间特征刻画了站点间的需求关联特性，其通常以站点邻接图为表征形式。邻接图的构建多依赖先验知识，例如基于站点实际连接关系、需求相似度等信息构建相应图谱，进而通过 GAT 对潜在的空间依赖关系进行挖掘与提取，不同构建方式的邻接图能够映射出具有差异性的空间特征^[26]。空间邻接矩阵为 $A \in \mathbb{R}^{M \times M}$ ，其中 $A_{i,j} = 1$ 表示地点 i 和地点 j 关联。

1.2 基模型模块

对于每个基模型，在各个地点上分别独立进行训练和预测，并在地点维度将结果进行拼接， H 为基模型数量，则第 h 个基模型的输出为 $X_h \in \mathbb{R}^{N \times M \times 1}$ 。

1.2.1 长短时记忆网络

LSTM 适用于处理序列建模任务，通过引入“门控”机制来控制信息的流动，包括遗忘门、输入门和输出门，从而能够更有效地学习序列数据中的长期依赖关系^[27]，模型的数学公式见 (1) ~ (6)。

$$f_{t,m} = \sigma(W_f \cdot [h_{t-1,m}, x_{t,m}] + b_f) \quad (1)$$

$$i_{t,m} = \sigma(W_i \cdot [h_{t-1,m}, x_{t,m}] + b_i) \quad (2)$$

$$C_{t,m} = \tanh(W_C \cdot [h_{t-1,m}, x_{t,m}] + b_C) \quad (3)$$

$$C_{t,m} = f_{t,m} \odot C_{t-1,m} + i_{t,m} \odot C_{t,m} \quad (4)$$

$$o_{t,m} = \sigma(W_o \cdot [h_{t-1,m}, x_{t,m}] + b_o) \quad (5)$$

$$h_{t,m} = o_{t,m} \odot \tanh(C_{t,m}) \quad (6)$$

式中：下标 f 、 i 、 o 分别表示遗忘门、输入门和输出门； $x_{t,m}$ 为模型的输入特征； $h_{t-1,m}$ 、 $C_{t-1,m}$ 分别为上一时间步的隐藏状态和细胞状态； σ 为 sigmoid 激活函数，表示遗忘门和输出门的信息通过量； w 和 b 是可训练的权重和偏置； \odot 表示逐元素相乘。

最终的单步单地点的预测值为 $X_{lsm}[t, m, 1] = h_{t,m}$ 。

1.2.2 多层感知机

MLP 利用非线性激活函数自动学习复杂非线性关系，通过反向传播算法调整权重，可处理高维输入，鲁棒性较强，适合拟合复杂函数，在图像、时序等数据的回归任务中表现出色^[28]。对于一个具有 L 层的 MLP，逐层传递规则计算见公式 (7) ~ (8)。

$$z^{(l)} = W^{(l)}a^{(l-1)} + b^{(l)} \quad (7)$$

$$a^{(l)} = f^{(l)}(z^{(l)}) \quad (8)$$

式中： $a^{(l-1)}$ 是第 $l-1$ 层的输出； $W^{(l)}$ 、 $b^{(l)}$ 、 $f^{(l)}$ 分别是第 l 层的权重矩阵、偏置向量和激活函数，常用的激活函数包括 ReLU、Sigmoid 和 Tanh； $a^{(0)}$ 是输入特征 $x_{t,m}$ ；输出层 $a^{(L)}$ 为模型单步单地点的预测值 $X_{\text{mp}}[t, m, 1]$ 。

1.2.3 梯度提升决策树

XGBoost 是一种基于决策树的集成机器学习算法，通过迭代地训练一系列决策树，每一棵树都学习并修正前一棵树的残差^[29]。该模型适用于结构化或表格数据，具有对稀疏数据的高效处理、大规模并行计算能力以及工业级的鲁棒性，模型的最终预测结果是 K 棵树的预测结果之和， f_k 是第 k 棵决策树，计算公式见式 (9)。

$$X_{\text{xgb}}[t, m, 1] = \sum_{k=1}^K f_k(x_{t,m}) \quad (9)$$

1.3 元模型模块

空间特征的学习通常采用图表示学习方法，通过将交通网络节点向量化表示，并结合节点特征矩阵完成空间依赖关系的建模，常用的模型有 GCN 和 GAT。与传统 GCN 依赖固定邻接矩阵的方式不同，GAT 通过自适应注意力机制为不同邻接节点分配动态权重，能够更灵活地捕捉站点间非均匀、异质的空间关联，适配交通流在空间维度的非均匀分布特性，尤其适用于复杂的城市轨道交通网络。但 GAT 的注意力权重的计算依赖节点特征的相似度，若交通数据中节点特征的区分度不足，可能出现注意力分配同质化的问题，而 GCN 基于拓扑结构的聚合方式在特征信息有限时更易保持稳定，并且 GAT 的注意力机制会引入额外计算开销，在大规模交通网络中，其训练效率略低于 GCN。在本研究的实验场景中，地铁站点的特征具备较强区分度，同时训练效率的影响在可接受范围内，因此最终选择 GAT 作为空间特征提取模块。

传统的机器学习将不同地点的样本看作是独立同分布的，而 GAT 对整个数据集进行预测，考虑节点之间的相互联系，其节点特征可以相互影响。基模型的预测结果和全局输入沿着最后一个维度拼接，构成元模型的输入特征矩阵。

$$X' = [X_1, X_2, \dots, X_h, X] \in \mathbb{R}^{N \times M \times (F+H)} \quad (10)$$

GAT 通过注意力分数的计算，使节点的邻居节点的权重不同，从而无需依赖固定的路网拓扑归一化操作，可灵活适配早晚高峰时段路网拓扑的动态变化^[30]。多头注意力机制借鉴了 transformer 的架构，多个头可以提供不同的视角来平衡节点对某个邻居的过度偏好，在 GAT 的中间层，不同头的输出被拼接起来，而在最终层则设置为单头。

对每个时间步 t ，首先进行特征投影，让模型学到更好的特征表示，转换后的特征向量见公式 (11)。

$$h_{t,m}^{(k)} = W^{(k)} \cdot x_{t,m}' \quad (11)$$

式中： $W^{(k)} \in \mathbb{R}^{d_h \times (F+H)}$ 是第 k 头的权重矩阵； d_h 为隐藏层的维度。

使用转换后的特征向量计算节点 m 对节点 j 的注意力得分 $e_{m,j}^{(k)}$ ，再将其与所有邻居计算出的注意力得分进行归一化处理，得到注意力权重系数 $\alpha_{m,j}^{(k)}$ ，计算公式为

$$e_{m,j}^{(k)} = \text{LeakyReLU}(a^{(k)T} [h_{t,m}'^{(k)} \parallel h_{t,j}'^{(k)}]) \quad (12)$$

$$\alpha_{m,j}^{(k)} = \text{softmax}_j(e_{m,j}^{(k)}) = \frac{\exp(e_{m,j}^{(k)})}{\sum_{j' \in \mathcal{N}(m)} \exp(e_{m,j'}^{(k)})} \quad (13)$$

式中： \parallel 表示向量拼接； $a^{(k)} \in \mathbb{R}^{2d_h}$ 为第 k 头的注意力参数向量； LeakyReLU 为激活函数； $\mathcal{N}(m)$ 为节点 m 的所有邻接节点集合。

对于 K 个注意力头，按照注意力机制加权求和得到节点 m 新的特征向量为 $h_{t,m}''$ ，激活函数 σ 通常使用 ELU，计算公式为

$$h_{t,m}'' = \sum_{k=1}^K \sigma(\sum_{j \in \mathcal{N}(i)} \alpha_{m,j}^{(k)} \cdot h_{t,m}'^{(k)}) \quad (14)$$

在多层多头图注意力聚合时，为避免梯度在深层网络中的消失或爆炸，除最后一层外，通过残差连接将前一层特征叠加到当前层的输出， proj_l 是第 l 层的残差投影函数，则第 l 层 GAT 的输出见式 (15)。

$$h_{t,m}'''^{(l)} = h_{t,m}''^{(l)} + \text{proj}_l(h_{t,m}''^{(l-1)}) \quad (15)$$

当为最后一层 GAT 时，使用单头注意力，并且输出维度为 1，则特征聚合的值为时间步 t ，地点 m 的最终预测值 $y_{t,m}$ ，元模型最终输出为所有时间步、所有位置的预测值 $Y \in \mathbb{R}^{N \times M \times 1}$ 。

1.4 评估指标

交通时空流量需求预测属于回归类预测问题，为全面评估模型的性能，采用了两类评价指标。其中，平均绝对误差 (MAE)、均方根误差 (RMSE)、决定系数 (R^2) 是回归任务的常用指标，MAE 反映预测值与真实值的绝对偏差程度，RMSE 通过平方放大极端误差的影响以更敏感地捕捉偏离情况， R^2 则衡量模型对数据变异的解释能力，三者共同从整体预测精度维度完成对模型的评估。此外，本研究引入 MAE 标准误 (Standard Error of MAE, SEMAE)、AvgNRMSE 和 WstNRMSE^[31] 作为鲁棒性指标，聚焦模型的稳定性和泛化能力。SEMAE 通过将同一场景的数据按时间 (SEMAE-T) 或地点 (SEMAE-L) 划分子数据集，衡量模型在同一场景的不同时空下的性能稳定性，其数值越小，说明模型在时间或空间维度的误差波动越弱；AvgNRMSE 通过对不同场景的 RMSE 进行归一化后取均值，反映模型在不同场景上的平均泛化能力；WstNRMSE 则聚焦各场景中的最差表现，以评估模型在极端场景下的鲁棒性。对于 AvgNRMSE 和 WstNRMSE，其子数据集为不同场景的数据，数值越接近 1，表明模型性能越好。各指标计算公式如下。

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (16)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (17)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (18)$$

$$\text{SEMAE} = \frac{\sqrt{\frac{1}{|\mathcal{D}|-1} \sum_{d \in \mathcal{D}} (\text{MAE}_d - \overline{\text{MAE}})^2}}{\sqrt{|\mathcal{D}|}} \quad (19)$$

$$\text{AvgNRMSE}_x = \frac{1}{|\mathcal{D}|} \sum_{d \in \mathcal{D}} \frac{\text{RMSE}_{x,d}}{\min_{x' \in \mathcal{X}} (\text{RMSE}_{x',d})} \quad (20)$$

$$\text{WstNRMSE}_x = \max_{d \in \mathcal{D}} \left(\frac{\text{RMSE}_{x,d}}{\min_{x' \in \mathcal{X}} (\text{RMSE}_{x',d})} \right) \quad (21)$$

式中： y_i 和 \hat{y}_i 分别表示第 i 个样本的真实值和预测值； n 为样本数量； \bar{y} 为所有样本的真实值的平均值； \mathcal{D} 为不同数据集的集合； d 为子数据集，子数据集的数量为 $|\mathcal{D}|$ ，以衡量在时间和空间维度的误差波动情况； \mathcal{X} 是所有预测方法的集合； $\text{RMSE}_{x,d}$ 表示预测方法 x 在子数据集 d 上的 RMSE 。

2 数据获取与分析

案例分析使用 2022 年 2 月 1 日至 2023 年 12 月 20 日纽约市地铁站点的客运出行数据，数据间隔为 1 h，数据来源为时空人群流动预测工具箱 UCTB 中整理的数[24]，共 424 个站点，部分原始数据如图 3 所示。工作日客流以通勤出行需求为主导，呈现“早高峰-平峰-晚高峰”的双峰周期性特征；周末客流则以休闲、购物等非通勤需求为核心，流量规模整体低于工作日，但客流分布更分散、高峰时段不显著，时间演化规律的不确定性更强。

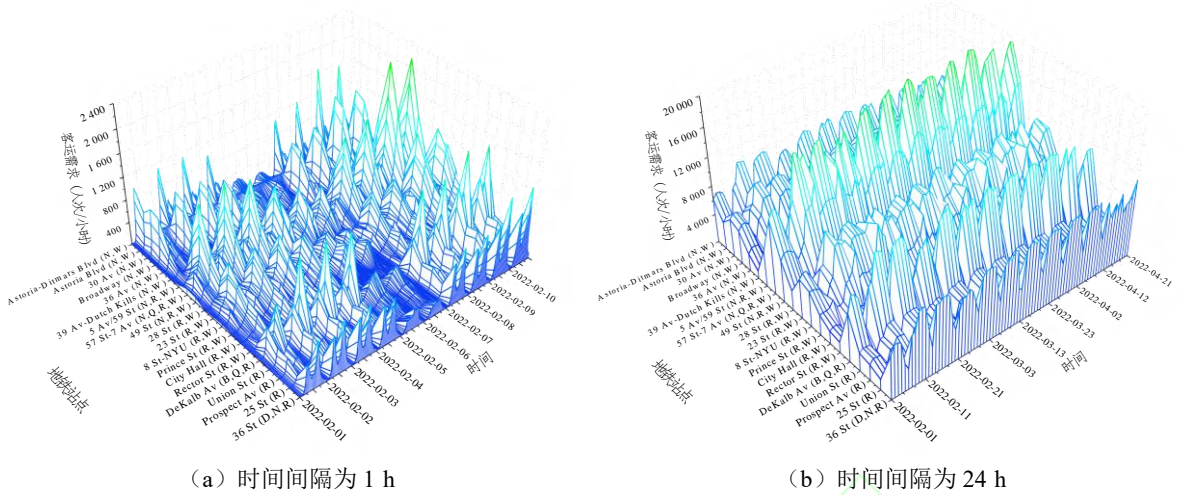


图 3 纽约市地铁客流需求部分数据

Fig. 3 Partial data of subway passenger flow demand in New York City

2.1 数据预处理

客运需求时序数据具备周期性波动与趋势性特征，而数据的完整性、一致性是保障预测模型输入质量与预测精度的核心前提，因此需对原始数据开展缺失值处理。原始数据中，负数值标记为缺失样本。地铁客流存在明显的周期性，如工作日早高峰、周末低峰等规律，因此本文采用 7 天周期均值插补法。该方法能够最大程度保留数据的固有周期特征，避免插补值引入虚假波动，且基于真实样本计算，过程简洁、效率较高，插补值的合理性与可解释性更强。具体处理逻辑为：首先定位缺失值在以 7 天为一周期的对应时刻位置（如每周一 7 点至 8 点）；随后提取历史数据集中所有完整周期内该时刻位置的非缺失样本；最终以这些非缺失样本的算术平均值作为插补值，替换原始缺失值，确保数据时序连续性与周期性特征的保留。针对同一时刻历史数据均缺失的极端情况，本文采用该站点同日期相邻时刻的客流均值进行补充；若相邻时刻数据仍缺失，则采用该站点全时段客流均值完成插补。本研究中纽约地铁原始数据的整体缺失值占比仅为 2.15%，且未出现同一时刻各周期数据均缺失的极端情况。上述插补策略可在极低缺失率的前提下有效保障数据完整性与时序特征，且不会对整体数据分布与周期性特征造成显著影响。

本研究采用的数据集涵盖 688 天的每小时客流需求记录，为适配预测模型的输入格式，需对时序数据进行样本重构。结合地铁客流的周期性特征，研究采用滑动窗口法构建输入样本：以历史 28 天（672 小时）的客流数据作为输入特征，对应预测未来下一时段的客流需求。基于不同时间间隔的预测场景，样本生成与数据集划分规则如下：当预测时间间隔为每小时，共生成 15 840 个有效样本，按“训练集—验证集—测试集”的比例划分为 12 960 个、1 440 个、1 440 个；当预测时间间隔为 1 天时，共生成 660 个有效样本，对应划分为 540 个训练样本、60 个验证样本、60 个测试样本；其余时间间隔的样本生成与划分方式，均遵循上述一致规则执行。

2.2 时空特征选择

在时序预测任务中，历史数据的输入规模与模型性能、计算效率存在密切关联。若输入的历史数据量过大，不仅会显著增加模型的参数规模与计算复杂度，导致训练与推理耗时延长、硬件资源开销上升，还可能引入冗余信息与噪声干扰，削弱关键特征对预测目标的贡献权重，进而影响模型

的泛化能力与预测精度。本模型考虑 τ 为历史 28 天来表示近四周历史客流对预测目标的影响，为确定哪些历史数据构成的特征对预测的作用最大，需要计算出时间步 t 所处的当前客流与前 28 天每个时间步的客流之间的相关系数。但单一时间步的相关系数可能受短期波动影响如某一天突发降雨导致该时刻客流异常，进而使该时间步的相关系数偏离常规，因此对不同时间步计算出来的相同滞后长度的相关系数进行平均得到相关系数计算结果，以削弱随机噪声的干扰。该过程通过量化当前与历史客流的线性关联强度，为预测模型筛选出最具价值的历史数据范围，避免盲目使用冗余历史信息导致模型复杂度过高，或因历史数据不足丢失关键规律。

当时间间隔为每小时 (1 h) 时的部分滞后的相关系数计算结果如图 4 (a) 所示，从近邻特征看，相关系数随着时间间隔越来越接近当前周期 (t) 而增加，从日周期和周周期特征看，客流相关性呈现出显著的周期性波动特征。相关系数在滞后长度为 24 的整数倍时反复出现波峰，在每相隔 7 天的相关系数又比前后几天的波峰更高，且峰值稳定在 0.95 左右。同时，曲线中波谷的相关系数维持在 0.65 左右，说明非周期节点的客流相关性较弱。因此在特征选择时，优先纳入前 1 天、前 7 天等周期整数倍滞后的历史客流数据，而对于非周期滞后的历史数据，仅考虑最近的时间步，根据数据情况选择相关系数大于 0.75 的时间步。

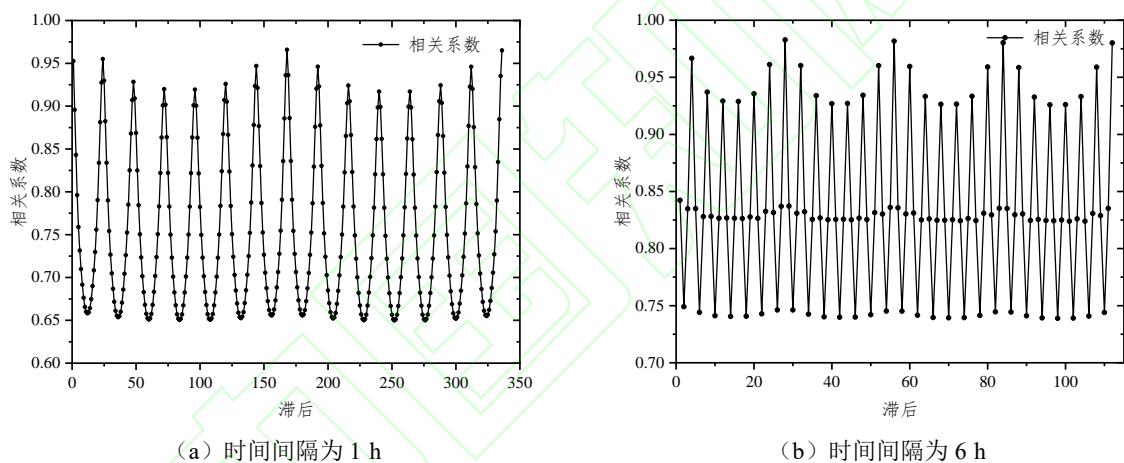


图 4 当前时刻与不同滞后期的皮尔逊相关系数

Fig. 4 Pearson correlation coefficients between the current moment and different lag periods

为进一步保障特征筛选的合理性，采用方差膨胀因子(VIF)评估多重共线性，并通过互信息(MI)量化客流特征与预测目标的非线性关联强度。以 24 h 时间间隔为例，根据皮尔逊系数所选的滞后特征的平均 VIF 值如表 1 所示，特征存在中等程度共线性，但结合交通客流的强周期性特征，滞后 7 (1 周)、滞后 14 (2 周) 等周周期特征对预测目标具有不可替代的物理意义，且时序数据存在时间自相关性，因此并未剔除高 VIF 特征。此外，周周期特征的 MI 值显著高于其他特征，与皮尔逊相关系数的周期性波峰特征高度一致，验证了所选周期特征的有效性，同时也说明皮尔逊相关系数筛选的线性特征与非线性关联特征具有高度重合性，进一步支撑了特征选择的合理性。

表 1 24 h 时间间隔数据选取的滞后特征对应的 VIF 和 MI

Tab. 1 VIF and MI for lag features corresponding to data sampled at 24-hour intervals

滞后	VIF	MI	滞后	VIF	MI
1	35.9	0.77	6	40.6	0.49
2	39.5	0.34	7	65.9	1
3	40.7	0.33	14	59.5	0.88
4	41.2	0.33	21	58.3	0.83
5	40.9	0.29	28	55.4	0.78

因此对于不同时间间隔的数据，结合 VIF 与 MI 的分析结果，选取的时间特征如表 2 所示。

表 2 不同时间间隔数据选取的时间特征

Tab. 2 Temporal features selected from data with different time intervals

时间间隔	近邻特征	日周期	周周期	时间间隔	近邻特征	日周期	周周期
1 h	5	6	4	4 h	2	6	4
2 h	3	6	4	6 h	1	6	4
3 h	2	6	4	24 h	6	0	4

空间特征反映了站点与站点流量之间的关系，通过实际物理连接关系构建空间邻接图和邻接矩阵，构建方式如图 5 所示。

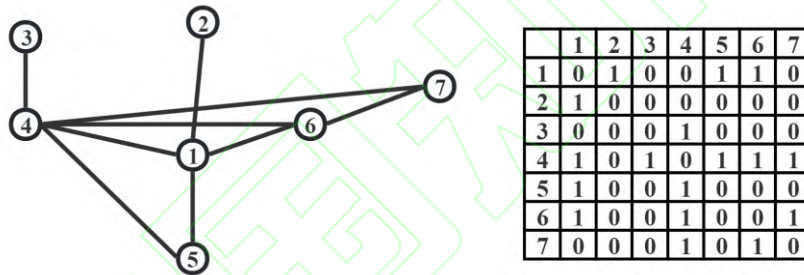


图 5 邻接矩阵构建

Fig. 5 Adjacency matrix construction

3 实验分析

3.1 实验环境及对比模型

该模型使用 pytorch 框架，并在 Intel Core i7-13620H 处理器、NVIDIA GeForce RTX 4050 显卡的环境下实现模型训练和验证。所有对比模型使用 Python 实现或从现有包中导入。

为了将所提出的 ST-SGAT 模型与其他经典模型在交通需求预测中的预测性能进行比较，本研究选取了 LSTM、MLP、XGBoost 作为基模型，GAT 作为元模型，并将这些模型和 GCN、ASTGCN (Attention-based spatial-temporal graph convolutional networks)、GraphWaveNet 等预测模型作为对比模型，ASTGCN 和 GraphWaveNet 的参数与提出该模型的论文保持一致，其他模型的参数基于 Optuna 库进行优化，根据经验，参数的搜索范围设置如表 3 所示，训练均使用早停策略，忍耐值均为 30 次。

表 3 参数搜索范围

Tab. 3 Parameter search range

参数	搜索范围	参数	搜索范围
隐层神经元	32, 64, 128, 256	学习率	0.1, 0.01, 0.001, 0.000 1

批处理大小	32, 64, 128, 256	丢弃率	0.1, 0.2, 0.3, 0.4
隐层数量	1 至 8 的整数	注意力头	1 至 4 的整数

以 1 h 和 24 h 时间间隔的数据为例，对比了 ST-SGAT 与性能较优的 ASTGCN 和 GraphWaveNet 模型在参数规模与训练时间上的差异，结果如表 4、表 5 所示。其中，各基模型采用单站点独立训练方式，参数规模保持一致，训练时间取 424 个站点中最长训练时长；ST-SGAT 模型的总参数量为基模型参数与元模型参数之和，训练时间因基模型可并行执行，故为基模型最长训练时间与元模型训练时间的总和。结果表明，在满足算力并行条件与站点数量适配的前提下，尽管 ST-SGAT 的总参数量相对较多，但模型收敛速度更快，整体训练效率显著优于 ASTGCN、GraphWaveNet 等深度时空预测模型。

表 4 不同模型的参数规模

Tab. 4 Parameter Sizes of Different Models

时间间隔	ASTGCN	GraphWaveNet	LSTM (单地点)	XGBoost (单地点)	MLP (单地点)	元模型
1 h	792 236	292 129	199 297	300 000	13 569	22 209
24 h	791 081	291 969	199 297	300 000	13 249	20 929

表 5 不同模型的训练时间 (min)

Tab. 5 Training Time of Different Models (minutes)

时间间隔	ASTGCN	GraphWaveNet	LSTM	XGBoost	MLP	ST-SGAT
1 h	399	382	2.6	0.018	4.5	50.5
24 h	7	32	0.14	0.01	0.13	12.14

3.2 结果分析

为全面且分层验证所提集成模型的预测能力与泛化特性，采用“整体—细分”的递进式分析框架，先基于全量数据评估模型的整体性能与不同时间间隔数据集下的效果对比，再通过时空维度的精细化数据筛选，评估模型在不同时空场景下的表现，从站点类型、时间类别两个角度展开针对性评估。

3.2.1 整体评估分析

为系统评估模型在不同数据粒度下的适配能力，针对 1 h、2 h、3 h、4 h、6 h、24 h 等不同时间间隔的数据集，开展全量数据整体性能评估。考虑到不同数据集的 MAE 与 RMSE 存在数量级差异，为直观呈现跨粒度模型性能对比，在绘图时对指标进行最大最小归一化处理，评估结果见表 6、图 6。

表 6 不同模型在不同时间间隔的 RMSE 指标对比

Tab. 6 Comparison of RMSE metrics of different models under different time intervals

模型名称	RMSE						整体评估	
	1 h	2 h	3 h	4 h	6 h	24 h	AvgNRMSE	WstNRMSE
LSTM	117.64	230.88	371.82	493.72	716.77	2 118.56	1.43	1.50
XGBoost	85.09	175.75	266.57	341.07	507.67	1 838.89	1.06	1.13
MLP	78.95	191.91	281.04	369.43	530.90	1 664.34	1.08	1.15
GAT	111.20	270.42	378.32	540.79	695.05	1 951.91	1.46	1.64
GCN	246.08	472.14	691.14	948.78	1 386.95	4 987.12	2.91	3.12
ASTGCN	102.03	252.94	375.18	463.02	655.70	1 891.52	1.37	1.52
GraphWaveNet	103.24	264.10	374.13	483.79	626.26	1 631.93	1.35	1.59

ST-SGAT	81.62	166.23	255.94	329.08	485.25	1 682.54	1.01	1.03
---------	-------	---------------	---------------	---------------	---------------	----------	-------------	-------------

注：加粗数据表示最优的指标值。

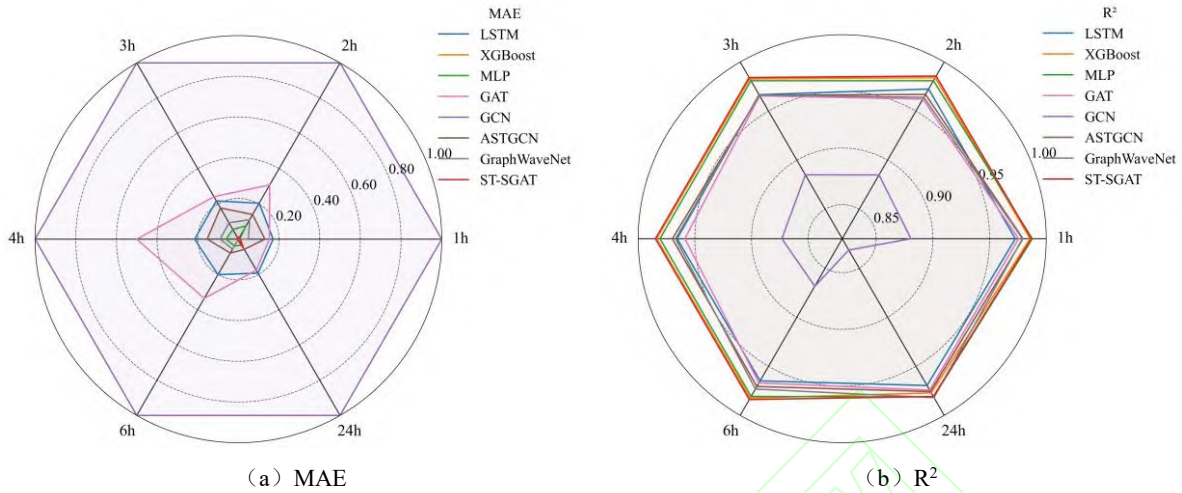


图 6 不同模型在不同时间间隔的 MAE 和 R^2 指标对比

Fig. 6 Comparison of MAE and R^2 metrics of different models under different time intervals

ST-SGAT 模型在多维度指标中表现突出，在核心预测精度层面，其在 1 h 粒度下的 RMSE、 R^2 指标仅次于 MLP 模型，24 h 粒度下仅次于 MLP 和 GraphWaveNet，而在 2 h 至 6 h 粒度中均实现最优性能，RMSE 指标保持最低水平， R^2 值稳定在 0.98 以上，且该区间内 XGBoost 模型性能亦优于 MLP，证实不同时间细粒度下模型的预测优势具有场景差异性，而 ST-SGAT 能够适配多数场景并达成最优预测效果。在鲁棒性层面，ST-SGAT 的 AvgNRMSE 达 1.01、WstNRMSE 为 1.03，是所有对比模型中唯一两个鲁棒性指标均逼近理想值 1 的模型，显著优于其他模型，显示了其在不同类型数据集上的稳定适配能力。

在 1 h 和 24 h 的场景下，客流波动受局部出行需求、宏观时间趋势影响，特征相关性高，MLP 的简单全连接结构即可有效拟合数据分布，无需复杂模块捕捉高阶时空关联，因此 MLP 的表现略优于 ST-SGAT。从理论角度而言，简单模型的优势在于低复杂度带来的低过拟合风险，当数据噪声小、特征维度低时，其泛化性能未必弱于复杂模型。但简单模型的优势具有显著场景局限性，当数据呈现高维时空耦合特征时，其无法捕捉时间依赖与空间关联的高阶交互，预测精度与鲁棒性显著下降。

从性能与计算成本来看，结合表 4、表 5，时间粒度的选择直接影响预测精度与计算开销。时间粒度越细，如 1 h 粒度，模型能够捕捉更精细的客流波动，预测精度相对更高，各模型的 RMSE 整体低于粗粒度对应指标，但此时数据量更大，模型训练与推理时间更长，计算成本更高；时间粒度越粗，如 24 h 粒度，数据量显著减少，训练速度更快、计算成本更低，但其会损失部分短期客流波动信息，可能导致预测精度出现下降。而 ST-SGAT 在不同粒度下均实现了较好的平衡，在保障高预测精度与强鲁棒性的同时，有效控制了计算成本。

进一步从时空细分维度解析全量数据预测精度，结果如表 7。时间维度的 SEMAE-T 指标中，ST-SGAT 在所有时间粒度中均保持最优水平；空间维度的 SEMAE-L 指标中，ST-SGAT 的数值均显著低于 LSTM、GAT、ASTGCN 等对比模型，在 1 h、6 h 和 24 h 场景下略高于 MLP，保持最优或次优水平，体现其在不同站点流量场景中的强泛化性能。综上，ST-SGAT 模型在时空交通需求预测的整体精度、时间维度与空间维度的稳定性和泛化性上均展现出突出优势，综合性能优于其他对比模型。

表 7 不同模型在不同时间间隔的时空维度上的 SEMAE 指标对比

Tab. 7 Comparison of SEMAE metrics of different models in the spatiotemporal dimension with different time intervals

模型名称	1 h		2 h		3 h		4 h		6 h		24 h	
	T	L	T	L	T	L	T	L	T	L	T	L
LSTM	0.81	2.58	2.65	4.73	5.56	7.95	7.54	11.23	14.63	17.79	84.41	56.22
XGBoost	0.52	1.83	1.59	3.93	3.18	6.00	4.40	7.82	8.67	12.00	66.72	51.26
MLP	0.56	1.61	2.51	3.92	4.77	5.80	6.27	8.06	12.66	11.44	89.80	44.60
GAT	0.78	2.54	3.32	5.16	5.96	7.42	8.52	9.49	14.05	13.49	80.21	51.37
GCN	1.93	6.34	5.06	12.42	9.02	18.45	12.76	26.86	23.73	39.33	72.98	194.17
ASTGCN	0.72	2.19	3.32	5.15	6.05	7.75	8.20	9.47	16.06	14.28	98.04	52.78
GraphWaveNet	0.72	2.19	3.65	5.13	6.52	7.49	9.12	9.81	15.44	13.65	86.39	43.60
ST-SGAT	0.48	1.72	1.41	3.62	3.09	5.71	4.14	7.51	8.04	11.45	66.07	47.03

注：加粗数据表示最优的指标值。

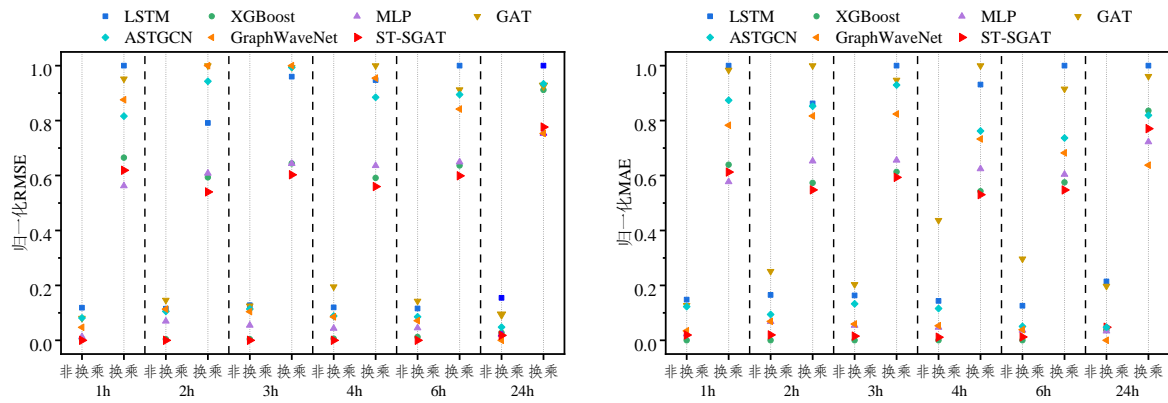
3.2.2 站点类型分析

在空间维度，基于地铁网络邻接矩阵量化各站点的拓扑关联特征，通过统计节点邻居数，将站点划分为换乘点（邻居数 >2 ）与非换乘点（邻居数 ≤ 2 ）两类异质性空间单元。其中，换乘点作为多线路交汇的枢纽节点，其客流需求受不同线路列车接驳、跨线换乘行为等多重因素耦合影响，呈现出流量规模大、时空分布复杂、波动不确定性强的特征；而非换乘点的客流主要依赖沿线局部出行需求，流量趋势相对平稳，空间特征复杂度显著低于换乘点，两类站点的预测难度存在本质差异。

为深入验证模型在异质性空间场景下的适配能力与鲁棒性，针对不同时间间隔的数据集，分别筛选换乘点与非换乘点对应的客流数据，开展分场景指标计算与对比分析。结果如图 7 所示，单一基模型的预测性能存在空间场景偏好，以 1 h 细粒度数据为例，XGBoost 模型在空间特征简单的非换乘点场景中表现最优，而 MLP 模型则在换乘点的复杂空间关联建模中更具优势，反映出单一模型难以同时适配两类差异显著的空间场景；在 2 h、3 h、4 h、6 h 的时间间隔下，XGBoost 在非换乘点的表现明显优于 MLP，但在换乘点的表现与 MLP 相差不大；在 24 h 的时间间隔下，MLP 在两类站点的性能均优于 XGBoost。

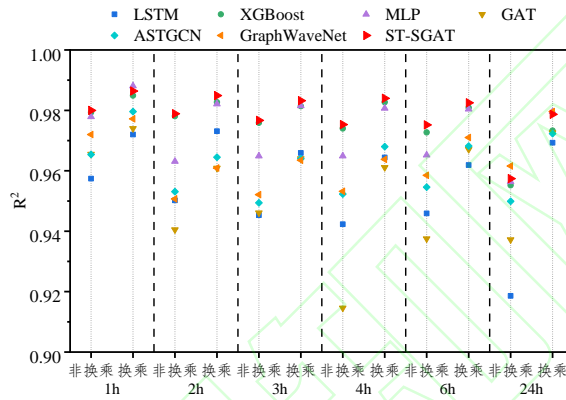
相比之下，本文提出的 ST-SGAT 模型在两类空间场景中均展现出卓越的预测精度与稳定性：在非换乘点场景中，其指标均处于最优或次优水平，精准捕捉了平稳客流的时间演化规律；在换乘点场景中，尽管面临更强的客流不确定性，该模型的各项评估指标仍保持在最优或次优梯队，有效克服了复杂空间关联带来的预测挑战。进一步分析鲁棒性指标，在换乘点与非换乘点两类场景下，ST-SGAT 模型在不同时间细粒度数据集上的指标值均最接近理想值 1，表明其能够有效适配不同空间特征场景与数据粒度的变化，在异质性空间环境中保持稳定可靠的预测性能。

ST-SGAT 在训练与预测过程中通过多基模型的差异化优势与元模型的自适应融合，实现对空间异质性的动态适配。不同基模型对空间复杂度的建模能力存在差异，XGBoost 更适合空间结构简单、客流规律稳定的非换乘点场景，MLP 则对换乘点复杂关联与强波动特征具有更强的捕捉能力。元模型 GAT 在融合阶段能够根据站点拓扑结构动态调整各基模型的贡献权重，使模型在不同空间特征场景下均能发挥最优组合效应，从而有效应对空间异质性带来的预测挑战。



(a) 预测结果归一化 RMSE 对比

(b) 预测结果归一化 MAE 对比



(c) 预测结果归一化 R^2 对比

图 7 不同站点类型的预测评价指标对比

Fig. 7 Comparison of prediction evaluation metrics for different station types

3.2.3 时间类别分析

在时间维度，通过反向时间推导法精准匹配各时间步的星期属性，依据客流生成的核心驱动因素，将数据划分为通勤客流主导的工作日（周一至周五）与休闲客流主导的周末（周六至周日）两类典型时间场景，分析模型在异质性时间场景下的适配能力与泛化性能。

评估结果如图 8 所示，多数模型在周末的预测误差低于工作日，这一现象主要源于周末客流基数更小，但周末的非周期性波动更强，对模型的泛化能力要求更高。单一基模型在两类时间场景中的预测性能呈现明显的一致性，除 ST-SGAT 模型外，其余单一模型在工作日与周末场景中的预测能力保持相似水平，即在工作日表现最优的模型，在周末场景中也维持相对最优水平。

值得注意的是，在 24 h 粗粒度数据场景中，模型性能呈现特殊分化。除 GraphWaveNet 模型外，ST-SGAT 模型在工作日场景下的预测表现最优，而在周末场景中，其预测性能略低于 MLP 模型。MLP 模型对该场景下的客流演化特征捕捉更为精准，而其他单一基模型的适配性相对较差，导致集成模型未能完全超越最优单一基模型的专项性能，但仍处于次优水平。从整体泛化能力来看，ST-SGAT 模型的鲁棒性指标在所有时间场景与数据细粒度下均最接近理想值 1，充分证明其能够有效应对时间规律的动态变化与数据粒度的差异，在复杂时间场景中保持稳定可靠的预测性能。

ST-SGAT 在训练与预测过程中通过基模型在不同时间场景与粒度下的差异化表现，实现对时间

异质性的自适应调节。从预测结果来看，MLP 在 1 h 与 24 h 粒度下表现更优，XGBoost 则在 2 h、3 h、4 h、6 h 粒度下优势更为明显，LSTM 在周末场景中的整体性能相对工作日更突出。这些差异反映出不同基模型对不同时间规律的捕捉能力存在区别，MLP 更适合近邻特征占比高的场景，XGBoost 对周期性特征占主导的场景具有更强拟合能力，LSTM 则对周末非规律波动具有一定适应性。元模型 GAT 能够在训练阶段自动学习各基模型在不同时间场景与时间粒度下的优势权重，并在预测阶段动态分配各基模型的贡献比例，使模型在工作日与周末、细粒度与粗粒度之间均能发挥最优组合效应，从而有效适配时间规律的异质性变化。相较于单一基模型的场景局限性，ST-SGAT 模型展现出更强的环境适应性与泛化能力，更能满足实际应用中客流规律随时间动态变化的预测需求。

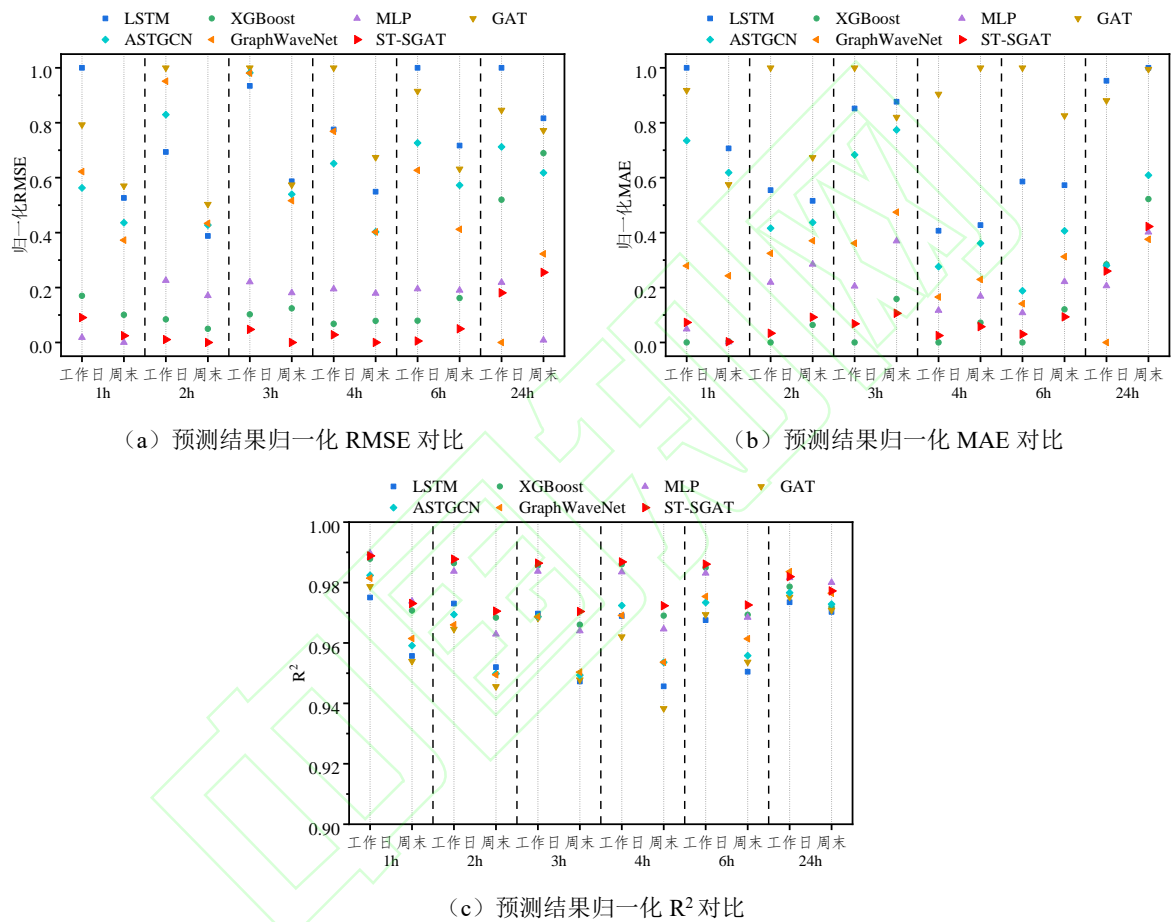


图 8 不同时间类别的预测评价指标对比

Fig. 8 Comparison of prediction evaluation metrics for different time categories

3.3 消融实验

为量化各基模型组件对集成模型整体性能的贡献度，揭示不同模块间的协同与制约机制，本研究设计了系列消融实验，通过依次移除集成框架中的 MLP、XGBoost、LSTM、GAT 组件，构建 No-MLP、No-XGBoost、No-LSTM、No-GAT 四种消融模型，与原始 ST-SGAT 模型进行跨场景对比分析，结果见表 8。

实验结果表明，各组件的性能贡献具有显著的时间粒度依赖性。在 1 h 和 24 h 数据场景中，移除元模型 GAT 后的预测误差最大；而在其他时间粒度下，移除 XGBoost 后模型性能下降最为明显。值得注意的是，1 h、24 h 场景中，移除 LSTM 的模型预测表现最优；同时，MLP 与 XGBoost 的优

劣也随时间粒度变化，1 h 短时间尺度下 MLP 的预测效果更优，中长时粒度下则是 XGBoost 的表现更为突出。

表 8 消融实验评估结果

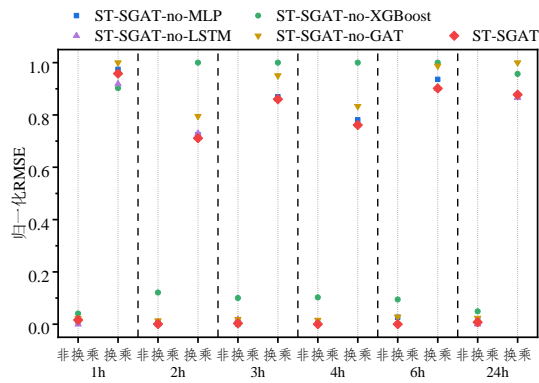
Tab. 8 Evaluation results of ablation experiments

模型名称	RMSE						整体评估	
	1 h	2 h	3 h	4 h	6 h	24 h	AvgNRMSE	WstNRMSE
No-MLP	82.17	168.54	256.87	334.66	504.56	1 693.15	1.02	1.04
No-XGBoost	80.67	220.47	297.39	412.39	547.51	1 809.21	1.16	1.33
No-LSTM	79.01	168.29	258.82	330.13	486.39	1 665.17	1.00	1.01
No-GAT	83.90	178.64	272.48	347.44	521.03	1 856.72	1.07	1.09
ST-SGAT	81.62	166.23	255.94	329.08	485.25	1 682.54	1.00	1.03

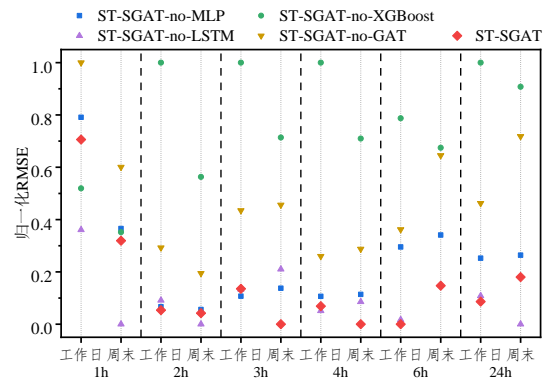
注：加粗数据表示最优的指标值。

进一步分析各组件间的相互作用机制可见，ST-SGAT 的 Stacking 结构通过基模型互补、元模型融合形成了协同体系。基模型在不同时间粒度下呈现差异化的预测优势，例如在 1 h 短时间尺度中，MLP 的预测表现优于 XGBoost，这是由于短时间尺度下近邻历史特征相关性更强，MLP 的线性拟合能力能够更高效地捕捉局部时序波动；而在中长时尺度下，XGBoost 对非线性周期特征的挖掘能力更为突出，因此成为整体性能的核心支撑。LSTM 的核心优势是捕捉长时序依赖，但其适配的场景是时序规律复杂、隐性长期关联显著的数据，本研究中实验数据的周周期与日周期等显性规律已通过预处理环节充分提取，导致 LSTM 的核心优势难以充分发挥，且 LSTM 的门控机制引入了较多模型参数，在有效时序信息有限的情况下，容易产生轻微过拟合，进而对集成模型的整体预测精度造成微弱负面影响。因此，LSTM 虽能提供时序补充信息，但其贡献相对有限，甚至在预测偏差较大时，移除后模型整体性能反而出现小幅提升。GAT 作为元模型，通过学习基模型输出的自适应权重并结合站点空间拓扑结构，实现了时序预测与空间关联的深度融合，移除 GAT 后各时间粒度性能均显著下降，表明空间融合对最终预测精度具有关键作用。同时，GAT 的动态权重分配机制能够有效抑制 LSTM 带来的微弱负面影响，强化 XGBoost 与 MLP 的稳定预测贡献，从而保障整体模型在不同场景下的稳健性与泛化能力。

从站点类型与时间类别的细分场景来看，结果见图 9、表 9，在换乘点场景中，No-LSTM 模型的预测精度略高于 ST-SGAT，表明 LSTM 组件在复杂空间枢纽的客流预测中存在微弱反贡献，但该反作用程度有限，未对整体性能造成显著影响；而 No-XGBoost 模型在全场景中表现最差，凸显 XGBoost 组件对集成预测的关键支撑作用。在工作日与周末两类时间场景中，ST-SGAT 模型的预测性能均优于或近似于所有消融模型，展现出稳定的场景适配性。



(a) 不同站点类型预测结果评估对比



(b) 不同时间类别预测结果评估对比

图 9 细分场景消融实验 RMSE 评估对比

Fig. 9 Comparison of RMSE evaluation results for ablation experiments in subdivided scenarios

表 9 细分场景消融实验鲁棒性指标评估对比

Tab. 9 Comparison of robustness metric evaluation results for ablation experiments in subdivided scenarios

模型名称	AvgNRMSE				WstNRMSE			
	换乘点	非换乘点	工作日	周末	换乘点	非换乘点	工作日	周末
No-MLP	1.02	1.03	1.03	1.02	1.06	1.07	1.04	1.04
No-XGBoost	1.13	1.26	1.19	1.11	1.29	1.43	1.38	1.20
No-LSTM	1.01	1.01	1.01	1.02	1.02	1.05	1.02	1.04
No-GAT	1.08	1.06	1.07	1.09	1.10	1.08	1.08	1.16
ST-SGAT	1.01	1.01	1.01	1.02	1.05	1.04	1.03	1.05

注：加粗数据表示最优的指标值。

综合来看，ST-SGAT 集成模型的各组件通过协同融合形成互补优势：即便部分组件在特定场景下存在微弱反贡献，其影响也可通过框架的自适应调节机制被有效削弱，且幅度可忽略不计。原始模型在所有时间粒度数据集下均保持最优或次优的 RMSE、AvgNRMSE 和 WstNRMSE，充分印证了集成框架在跨时间粒度、跨站点类型场景中的泛化优势，以及各组件协同作用对提升模型稳定性的核心价值。

3.4 突发场景模型性能分析

2023 年 9 月 29 日纽约市极端暴雨事件导致地铁近半数线路完全停运、其余线路运行严重延误，当日地铁客流规模出现大幅锐减，这一突发事件凸显了客流预测模型对极端气象干扰的适配能力具有重要实际应用价值。为进一步优化模型预测性能，同时探究 ST-SGAT 模型在突发场景下的预测表现，本研究将降水量作为外部特征纳入模型训练，该气象数据来源于美国国家海洋和大气管理局 (NOAA)，由纽约中央公园气象站逐日采集获取。基于数据匹配性，选取该气象站周边 100 个地铁站点的 24 h 时间间隔客流数据展开后续分析。

参考美国国家气象局通用的 24 小时累计降水量分级标准，对降水事件进行多尺度划分，分别设置 2 类与 4 类降水分级开展对比分析，具体分级标准如表 10 所示，其中将暴雨及以上等级界定为异常降水场景。针对不同降水分级下的模型预测结果计算评估指标，量化分析模型在不同降水场景下的预测性能差异，结果如表 11、12 所示。

表 10 降雨量分级标准

Tab. 10 Rainfall classification standard

降水量 (英寸)	4 类	2 类
$PRCP < 0.1$	无雨或小雨	常规降水
$0.1 \leq PRCP < 0.5$	中雨	
$0.5 \leq PRCP < 1.0$	大雨	
$1.0 \leq PRCP$	暴雨及以上	异常降水

表 11 模型在 4 类分级的降水场景下的预测性能

Tab. 11 Model prediction performance under four-level precipitation scenarios

模型名称	降水等级	MAE		RMSE		R ²	
		含降水	无降水	含降水	无降水	含降水	无降水
ST-SGAT	无雨或小雨	1 387.96	1 473.20	2 386.60	2 538.61	0.99	0.98
ST-SGAT	中雨	1 962.93	2078.09	3 065.72	3 320.78	0.97	0.96
ST-SGAT	大雨	1 605.35	1 689.84	2 504.72	2 827.49	0.99	0.98
ST-SGAT	暴雨及以上	3 792.19	3 786.52	6 879.56	6 865.90	0.85	0.85
GraphWaveNet	无雨或小雨	1 187.58	1 286.77	2 160.30	2 331.97	0.99	0.99
GraphWaveNet	中雨	1 604.82	1 535.34	2 623.82	2 553.11	0.98	0.98
GraphWaveNet	大雨	1 006.33	1 006.26	1 693.13	1 693.65	0.99	0.99
GraphWaveNet	暴雨及以上	3 834.49	4 156.29	7 757.46	8 348.53	0.81	0.78

表 12 模型在 2 类分级的降水场景下的预测性能

Tab. 12 Model prediction performance under two-level precipitation scenarios

模型名称	降水等级	MAE		RMSE		R ²	
		含降水	无降水	含降水	无降水	含降水	无降水
ST-SGAT	常规降水	1 450.94	1 538.81	2 461.18	2 633.78	0.98	0.98
ST-SGAT	异常降水	3 792.19	3 786.52	6 879.56	6 865.90	0.85	0.85
GraphWaveNet	常规降水	1 215.12	1 293.93	2 183.68	2 323.12	0.99	0.99
GraphWaveNet	异常降水	3 834.49	4 156.29	7 757.46	8 348.53	0.81	0.78

从常规降水场景预测结果来看, GraphWaveNet 模型在 24 h 客流预测任务中本就具备一定优势, 因此在 2 类分级的常规降水场景及 4 类分级的无雨或小雨、中雨、大雨场景中, 其各项预测指标均表现出数值优势。整体而言, 降水量特征对两种模型的常规降水场景预测具备正向补充价值, 但二者的特征增益表现存在显著差异。ST-SGAT 模型在 4 类分级的无雨或小雨、中雨、大雨等常规降水场景下, 纳入降水特征后均实现稳定正向提升, MAE 提升率保持在 5.0%~5.79% 区间内。GraphWaveNet 模型仅在无雨或小雨场景实现显著正向增益, 在中雨、大雨场景均出现特征增益负向波动, 二者的增益差异直观反映出 ST-SGAT 模型对降水量特征的利用适配性与稳定性更优。

异常降水的预测表现是衡量模型应对极端干扰能力与鲁棒性的核心依据。在 2 类分级的异常降水场景及 4 类分级的暴雨及以上场景中, ST-SGAT 模型展现出显著的突发场景预测优势。从预测精度角度分析, 使用降水量特征的 ST-SGAT 模型的评估指标均优于同特征下 GraphWaveNet 模型, 即使 ST-SGAT 模型未纳入降水量特征, 其在突发场景下的基础预测精度仍优于含降水特征的 GraphWaveNet 模型, 体现出 ST-SGAT 模型在降水突发场景下具有更强的预测能力。从突发事件发生时的预测性能衰减角度分析, 以 2 类降水分级为例, 从常规降水到异常降水这一突发场景, ST-SGAT 模型的 MAE、RMSE 的增幅分别为 161.3%、179.5%, 而 GraphWaveNet 模型的增幅分别达

215.6%、255.3%，ST-SGAT 模型的性能衰减幅度显著低于 GraphWaveNet 模型，说明随降水强度从常规向极端突变，ST-SGAT 模型能更好地保持预测稳定性，在气象突发干扰下表现出更优的抗衰减能力。

从降水量特征的全域增益效果来看，ST-SGAT 模型的特征利用稳定性进一步凸显。其仅在暴雨及以上的等级中出现轻微负向波动，可能由于突发场景样本量稀缺，训练集中仅占 3.9%，测试集占 6.7%，模型难以充分学习极端场景下降水量与客流波动的复杂非线性关联。此外，ST-SGAT 模型在无降水量特征时的突发场景的预测精度已较高，特征补充的提升空间较小，属于小样本场景下的正常数值波动。而 GraphWaveNet 模型虽在气象突发场景中表现出显著的特征增益效果，但其中雨、大雨等常规降水等级即出现明显的负增益，说明其对降水量特征的利用在常规降水场景中存在适配性不足的问题。

4 结束语

本文为解决现有预测模型在不同数据集、不同时间粒度下的预测能力各异的问题，基于“堆叠集成学习”思想，并结合 GAT 在时空预测中的优势，提出了一种基于 Stacking 堆叠集成策略和 GAT 的时空需求预测模型（ST-SGAT）。得出了以下结论：

（1）本文采用 Stacking 堆叠集成策略，融合多类型的时间预测基模型与 GAT，实现交通客运时空需求的精准预测。在特征层面同时提取多粒度周期性时间特征与站点拓扑空间特征，基于纽约地铁多站点、多时间粒度数据的实验验证表明，该集成框架能够有效处理时空数据中存在的时间与空间异质性，为复杂交通场景下的精准需求预测提供了新的模型构建思路与方法参考。

（2）ST-SGAT 模型在多时间粒度与时空不同场景下均表现出最优综合性能与强泛化能力。在小时级至日级的不同时间粒度下，ST-SGAT 的预测精度在多数场景中达到最优，仅在少数粒度下略逊于单一优势基模型，突破了传统单一模型的场景局限性，验证了基模型优势互补与空间特征融合设计的合理性。在时间维度，模型可稳定捕捉工作日与周末的客流规律差异；在空间维度，可有效适配换乘点与非换乘点的客流特征差异。

（3）消融实验表明，模型组件贡献存在场景依赖性，但集成框架的自适应调节机制可实现全局最优。Stacking 的集成框架与 GAT 的动态权重分配机制能够有效抑制局部性能不足带来的影响，实现多组件协同互补，使整体模型在全场景下均保持最优或次优性能，体现了集成框架的稳定性与自适应能力。

（4）引入降水量外部特征后，ST-SGAT 在常规与突发降水场景下均展现出更优的鲁棒性与特征利用能力。在暴雨等突发异常场景中，ST-SGAT 不仅基础预测性能更高，且性能衰减幅度显著低于 GraphWaveNet，表现出更强的抗干扰能力与鲁棒性，体现 ST-SGAT 在应对气象突发事件与外部扰动方面具备更高的实用价值与应用潜力。

然而，本研究仍存在局限性。首先，模型在特征构建层面仍以历史客流与静态拓扑为主，虽已

引入降水量进行拓展分析,但尚未系统融合节假日、特殊事件等多源外部影响因素,在动态环境下的适应性仍有提升空间。其次,当前空间建模仅基于物理邻接关系构建静态图结构,未考虑客流需求相似性、站点间动态客流转移、功能属性关联等动态空间依赖关系,对复杂空间交互的表征能力存在不足。第三,模型对数据质量存在一定依赖性,在传感器故障、数据缺失、异常采样等低质量数据场景下的鲁棒性尚未充分验证,难以完全覆盖真实交通系统中的复杂数据状况。第四,模型以预测精度为核心目标,可解释性分析较为有限,组件贡献机理与决策过程不够透明。第五,模型在计算效率与工程部署方面仍存在优化空间,在大规模路网实时推理时可能面临延迟与资源消耗问题,工程实用性仍需进一步提升。

未来研究可从以下方向继续深化。一是在捕捉时空特征时,若算力足够,可以融合多源外部因素与动态图结构,构建更全面的时空依赖表征体系,提升模型在复杂扰动场景下的预测能力;二是增强模型可解释性与数据鲁棒性,优化模型轻量化与推理效率,降低实际部署成本,推动模型在交通管理系统中的落地应用;三是验证模型在不同地区、不同交通模式下的泛化能力,提升模型可扩展性;四是结合概率预测与不确定性量化,构建鲁棒性更强的动态预测体系,为交通路径规划、班期优化、运力调度等上层决策提供更坚实的支撑。

参考文献

- [1] BOUKERCHE A, WANG J. Machine learning-based traffic prediction models for intelligent transportation systems[J]. *Computer Networks*, 2020, 181: 107530.
- [2] ROCCO DI TORREPADULA F, NAPOLITANO E V, DI MARTINO S, et al. Machine learning for public transportation demand prediction: a systematic literature review[J]. *Engineering Applications of Artificial Intelligence*, 2024, 137: 109166.
- [3] WANG H, YI W, TIAN X, et al. Prescriptive analytics for intelligent transportation systems with uncertain demand[J]. *Journal of Transportation Engineering, Part A: Systems*, 2023, 149(12): 04023118.
- [4] 赵鹏军, 陈霄依, 王祎勃, 等. 城市系统交通需求模拟预测技术框架构建及应用[J]. *地球信息科学学报*, 2025, 27(3): 539-552.
ZHAO Pengjun, CHEN Xiaoyi, WANG Yiqing, et al. Construction and application of an urban system-based travel demand forecasting technology framework[J]. *Journal of Geo-information Science*, 2025, 27(3): 539-552.
- [5] ZHU Y, GOVERDE R M P. Dynamic and robust timetable rescheduling for uncertain railway disruptions[J]. *Journal of Rail Transport Planning & Management*, 2020, 15: 100196.
- [6] FILOM S, RAZAVI S. A learning-based robust optimization framework for synchromodal freight transportation under uncertainty[J]. *Transportation Research Part E: Logistics and Transportation Review*, 2025, 195: 103967.
- [7] ISLAM S, AMIN S H, WARDLEY L J. A supplier selection & order allocation planning framework by integrating deep learning, principal component analysis, and optimization techniques[J]. *Expert Systems with Applications*, 2024, 235: 121121.
- [8] GUO W, ATASOY B, VAN BLOKLAND W B, et al. Anticipatory approach for dynamic and stochastic shipment matching in hinterland synchromodal transportation[J]. *Flexible Services and Manufacturing Journal*, 2022, 34(2): 483-517.
- [9] 王莹, 韩宝明, 张琦, 等. 基于 SARIMA 模型的北京地铁进站客流量预测[J]. *交通运输系统工程*

与信息, 2015, 15(6): 205-211.

WANG Ying, HAN Baoming, ZHANG Qi, et al. Forecasting of entering passenger flow volume in Beijing subway based on SARIMA model[J]. Journal of Transportation Systems Engineering and Information Technology, 2015, 15(6): 205-211.

- [10] WANG D, WANG C, XIAO J, et al. Bayesian optimization of support vector machine for regression prediction of short-term traffic flow[J]. Intelligent Data Analysis, 2019, 23(2): 481-497.
- [11] NEJADETTEHAD A, MAHINI H, BAHRAK B. Short-term demand forecasting for online car-hailing services using recurrent neural networks[J]. Applied Artificial Intelligence, 2020, 34(9): 674-689.
- [12] LIU X, GHERBIA A, LI W, et al. Multi features and multi-time steps LSTM based methodology for bike sharing availability prediction[J]. Procedia Computer Science, 2019, 155: 394-401.
- [13] 谭雪, 张小强. 基于 GRU 深度网络的铁路短期货运量预测[J]. 铁道学报, 2020, 42(12): 28-35.
TAN Xue, ZHANG Xiaoqiang. GRU deep neural network based short-term railway freight demand forecasting[J]. Journal of the China Railway Society, 2020, 42(12): 28-35.
- [14] LIU Y, WU C, WEN J, et al. A grey convolutional neural network model for traffic flow prediction under traffic accidents[J]. Neurocomputing, 2022, 500: 761-775.
- [15] 唐进君, 段一鑫, 商淑杰, 等. 多特征融合时空图卷积的高速公路交通流量预测[J]. 公路交通科技, 2025, 42(6): 22-31.
TANG Jinjun, DUAN Yixin, SHANG Shujie, et al. Expressway traffic volume prediction based on feature-fused spatio-temporal graph convolutional networks[J]. Journal of Highway and Transportation Research and Development, 2025, 42(6): 22-31.
- [16] 胡晓伟, 吴则洋, 卢泓博, 等. 基于遗传算法优化深度神经网络的站点客流预测[J]. 交通运输工程与信息学报, 2025, 23(1): 72-84.
HU Xiaowei, WU Zeyang, LU Hongbo, et al. Passenger flow prediction for stations using genetic algorithm-optimized deep neural networks[J]. Journal of Transportation Engineering and Information, 2025, 23(1): 72-84.
- [17] XIAN J, YE Y, ZHANG W, et al. MDHGFN: Multiscale dual hypergraph fusion spatiotemporal network for traffic flow prediction[J]. Chaos, Solitons & Fractals, 2025, 201: 117228.
- [18] ZHANG H, LIN Z, XIE H, et al. Two-way heterogeneity model for dynamic spatiotemporal traffic flow prediction[J]. Knowledge-Based Systems, 2025, 320: 113635.
- [19] LIU J, WANG Y, ZHU J, et al. A multilayer spatiotemporal correlation-aware graph attention network for traffic flow prediction[J]. IEEE Transactions on Neural Networks and Learning Systems, 2025: 1-15.
- [20] FENG C, CUI M, HODGE B M, et al. A data-driven multi-model methodology with deep feature selection for short-term wind forecasting[J]. Applied Energy, 2017, 190: 1245-1257.
- [21] 杜培军, 慕号伟, 郭山川, 等. 集成学习遥感应用: 进展与展望[J]. 遥感学报, 2025, 29(6): 1614-1635.
DU Peijun, MU Haowei, GUO Shanchuan, et al. Ensemble learning in remote sensing applications: Progress and prospects[J]. National Remote Sensing Bulletin, 2025, 29(6): 1614-1635.
- [22] MORETTI F, PIZZUTI S, PANZIERI S, et al. Urban traffic flow forecasting through statistical and neural network bagging ensemble hybrid modeling[J]. Neurocomputing, 2015, 167: 3-7.
- [23] LEE D, KIM K. AdaBoost.RDT: AdaBoost integrated with residual-based decision tree for demand prediction of bike sharing systems under extreme demands[J]. IEEE Access, 2024, 12: 144316-144336.
- [24] LEI J, CHU Z, WU Y, et al. Predicting vessel arrival times on inland waterways: a tree-based stacking approach[J]. Ocean Engineering, 2024, 294: 116838.

- [25] 李浩, 卢朝阳, 谈翌平, 等. 基于 MIC-iAFF-Stacking 集成学习的航空器滑出时间预测[J]. 交通运输工程与信息学报, 2024, 22(4): 142-153. LI Hao, LU Chaoyang, TAN Yiping, et al. Aircraft departure taxi time prediction based on MIC-iAFF-Stacking ensemble learning[J]. Journal of Transportation Engineering and Information, 2024, 22(4): 142-153.
- [26] 陈李越, 柴迪, 王乐业. UCTB: 时空人群流动预测工具箱[J]. 计算机科学与探索, 2022, 16(4): 835-843. CHEN Liyue, CHAI Di, WANG Leye. UCTB: spatiotemporal crowd flow prediction toolbox[J]. Journal of Frontiers of Computer Science and Technology, 2022, 16(4): 835-843.
- [27] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [28] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning representations by back-propagating errors[J]. Nature, 1986, 323(6088): 533-536.
- [29] CHEN T, GUESTRIN C. XGBoost: A scalable tree boosting system[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco California USA: ACM, 2016: 785-794.
- [30] TIAN K, GUO J, YE K, et al. ST-MGAT: Spatial-temporal multi-head graph attention networks for traffic forecasting[C]//2020 IEEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI). Baltimore: IEEE, 2020: 714-721.
- [31] WANG L, CHAI D, LIU X, et al. Exploring the generalizability of spatio-temporal traffic prediction: meta-modeling and an analytic framework[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(4): 3870-3884.