多源数据增强的地铁客流时空预测大模型方法

张 媛[1]，张 婴[1]，李 皓[2*]

（1. 长沙理工大学人工智能学院，湖南 长沙 410114；2. 长沙理工大学交通学院，湖南 长沙 410114）

摘要：地铁客流同时受到历史演化、线网拓扑、站点功能和天气等因素影响，现有时空图模型多依赖物理连接或数值相关性，难以自动获取并融合非结构化环境知识。为此，提出一种数据智能体、语义增强图神经网络与大语言模型（LLM）校准相结合的地铁客流预测方法。首先，构建基于模型上下文协议（MCP）的数据智能体，自动调用地图、气象和文本嵌入服务，采集并筛选站点周边兴趣点，生成站点文本描述与空间上下文嵌入；其次，以站点语义嵌入学习稀疏邻接矩阵，并通过扩散图卷积编码器—解码器获得初步预测；最后，将站点描述、客流规律、天气和图网络预测共同组织为提示词，采用低秩自适应微调对 LLM 进行领域校准。成都地铁 1、2 号线实验表明，所提方法的 MAE、RMSE 和 MAPE 分别为 26.019、61.076 和 30.67%，在 6 项评价指标上均取得最优结果；与最优对比方法相比，三项总体误差分别降低 1.70%、3.73%和 3.43%。该方法能够识别物理距离之外的站点功能关联，并提高常规时段与多数峰值时段的预测稳定性。

关键词：地铁客流预测；数据智能体；多源信息融合；图神经网络；大语言模型；模型上下文协议

中图分类号：U293.5　　**文献标志码：**A

Large Model–Enabled Multi-Source Information Fusion for Spatiotemporal Metro Passenger Flow Forecasting

Zhang Yuan[1], Zhang Zhao[1], Li Hao[2]

（1. School of Artificial Intelligence，Changsha University of Science & Technology, Changsha 410114, China；2. School of Transportation，Changsha University of Science & Technology, Changsha 410114, China）

Abstract: Metro passenger flow is jointly affected by historical evolution, network topology, station functions, and weather. Existing spatiotemporal graph models mainly rely on physical connections or numerical correlations and therefore have difficulty automatically acquiring and integrating unstructured contextual knowledge. This paper proposes a forecasting method that combines a data agent, a semantic-enhanced graph neural network, and large language model (LLM) calibration. First, a Model Context Protocol (MCP)-based data agent automatically invokes map, meteorological, and text-embedding services to collect and screen points of interest around stations and to generate standardized station descriptions and spatial context embeddings. Second, a sparse semantic adjacency matrix is learned from these embeddings, and a diffusion graph-convolutional encoder–decoder produces preliminary forecasts. Finally, station descriptions, passenger-flow patterns, weather, and graph-network forecasts are organized into prompts, and low-rank adaptation is used to fine-tune the LLM for domain calibration. Experiments on Chengdu Metro Line 1 and 2 show that the proposed method achieves MAE, RMSE, and MAPE values of 26.019, 61.076, and 30.67%, respectively, achieving the best performance on all six evaluation metrics. Compared with the strongest baseline on each overall metric, MAE, RMSE, and MAPE are reduced by 1.70%, 3.73%, and 3.43%, respectively. The method captures functional associations beyond physical distance and improves forecasting stability in normal and most peak periods.

Key words: metro passenger flow forecasting; data agent; multi-source information fusion; graph neural network; large language model; model context protocol

地铁短时客流预测是行车计划调整、站内客流组织、设备与人员配置以及突发事件响应的重要依据。自动售检票数据具有明显的周期性，但高峰形成及异常波动还与站点功能、接驳条件、天气和日期属性密切相关。若仅根据历史序列外推，模型容易在功能差异显著的站点之间产生相似预测，也难以解释降雨、

收稿日期：2026−08−05

基金项目：国家自然科学基金项目(52502377)

节假日或大型活动引起的结构性变化。

国内针对城市轨道交通短时客流预测已开展了多种深度学习研究。张惠臻等[1]构建基于循环神经网络的预测方法，并比较不同时间粒度以及 LSTM、GRU 模型的适用性；张金雷等[2]融合 Transformer 式时序编码、深度注意力与卷积结构，开展网络级多步短时客流预测。此类方法能够刻画单站或有限站点的时间依赖，但对轨道交通网络中的跨站传播关系利用仍有不足。随着图神经网络的发展，研究重点逐渐转向时空联合建模。Guo 等[6]提出 ASTGCN，以时空注意力和图卷积同时提取周期性及空间依赖；尹宝才等[4]利用动态超边和谱域超图卷积刻画节点间的高阶隐含关系；Li 等[5]构建连接、相似性和时间相关性等多类交互图。这些方法显著增强了非欧空间建模能力，但图结构仍主要由线路连接、物理距离或客流相关系数给定，易受数据噪声影响，也难以解释物理距离较远但功能属性相似的站点关联。

外部因素融合是提升实际预测性能的另一条研究路线。冯树民等[3]针对雨雪天气构建 WI-LSTM 客流预测模型，验证了恶劣天气信息能够增强对客流波动规律的刻画并提高预测可靠性。现有研究通常将天气、节假日或兴趣点（Point of Interest，POI）编码为有限维度的人工特征后与客流序列直接拼接。该方式依赖人工查询、清洗和类别统计，难以处理持续变化的地图信息与非结构化描述；同时，简单拼接无法充分表达“站点功能相似但空间不相邻”等语义关系。

Transformer 及其时序变体为长依赖建模提供了新工具。Vaswani 等[7]提出自注意力架构后，Autoformer[8]、PatchTST[9]以及 Linear、NLinear 和 DLinear[10]等模型分别从序列分解、分块表示和线性映射角度提升了时序预测能力。近年来，Time-LLM[11]尝试利用大语言模型重编程时间序列，TimesFM[12]、MOMENT[13]、Lag-Llama[14]和 Time-MoE[15]则通过大规模预训练增强跨数据集泛化能力。然而，上述基础模型多数把数值序列视为独立或弱关联的令牌，缺少面向地铁线网的显式图结构；将 LLM 直接作为数值预测器时，还可能出现格式不稳定、局部峰值偏离或领域约束不足等问题。

综上，现有研究仍有 3 方面不足：1）多源外部信息的获取、筛选和向量化依赖人工流程，难以形成可复用的自动化数据链；2）地铁图结构过度依赖物理拓扑或历史数值相关性，难以表达由土地利用和站点功能产生的非局部联系；3）图网络的数值建模优势与 LLM 的常识推理能力尚未形成稳定分工，直接预测或简单特征拼接均难以兼顾精度、约束和解释性。

针对上述不足，创新点如下：1）构建基于 MCP 的数据智能体，将地图 POI、天气和文本嵌入服务封装为标准工具，实现站点环境信息的主动采集、语义筛选和结构化表达；2）提出语义增强的稀疏图构建方法，利用站点上下文嵌入和 α-Entmax 学习非局部功能关联，并与扩散图卷积编码器—解码器结合完成时空预测；3）采用“图网络负责数值预测、LLM 负责领域校准”的两阶段机制，以站点描述、客流规律、天气和初始预测构造监督微调样本，在保留数值模型稳定性的同时引入外部知识约束。

1 问题定义与总体框架

1.1 问题定义

设地铁线路包含N个站点，时间粒度为 15 min。在时刻t，所有站点的客流观测记为 $\boldsymbol{X}_t \in \mathbb{R}^{N\times F}$，其中 F 表示客流特征维数。给定过去 H 个时间步的历史序列、对应天气信息以及站点环境描述，预测未来 P 个时间步的客流

$$\boldsymbol{Y}_{t+1:t+P} = f\left(\boldsymbol{X}_{t-H+1:t}, \boldsymbol{W}_{t-H+1:t+P}, \boldsymbol{T}, \boldsymbol{E}\right) \tag{1}$$

式中：$\boldsymbol{W}$ 为天气及日期属性；$\boldsymbol{T} = \{T_i\}_{i=1}^{N}$ 为站点文本描述集合；$\boldsymbol{E} \in \mathbb{R}^{N\times d}$ 为文本描述对应的空间上下文嵌入，d 为嵌入维数。设置 H=P=8，即利用过去 2 h 客流预测未来 2 h 客流。

1.2 总体框架

所提方法由多源数据智能体、语义增强图神经网络和 LLM 校准 3 部分构成，如图 1 所示。数据智能体负责把外部非结构化信息转化为可复用的文本和向量；图神经网络负责稳定地学习站点间时空传播并输出初步数值预测；LLM 不替代图网络，而是依据站点属性、天气、客流规律及初步预测进行二次校准。该分工避免了 LLM 直接拟合高频数值序列的不稳定性，也避免了传统图网络缺少现实知识的问题。

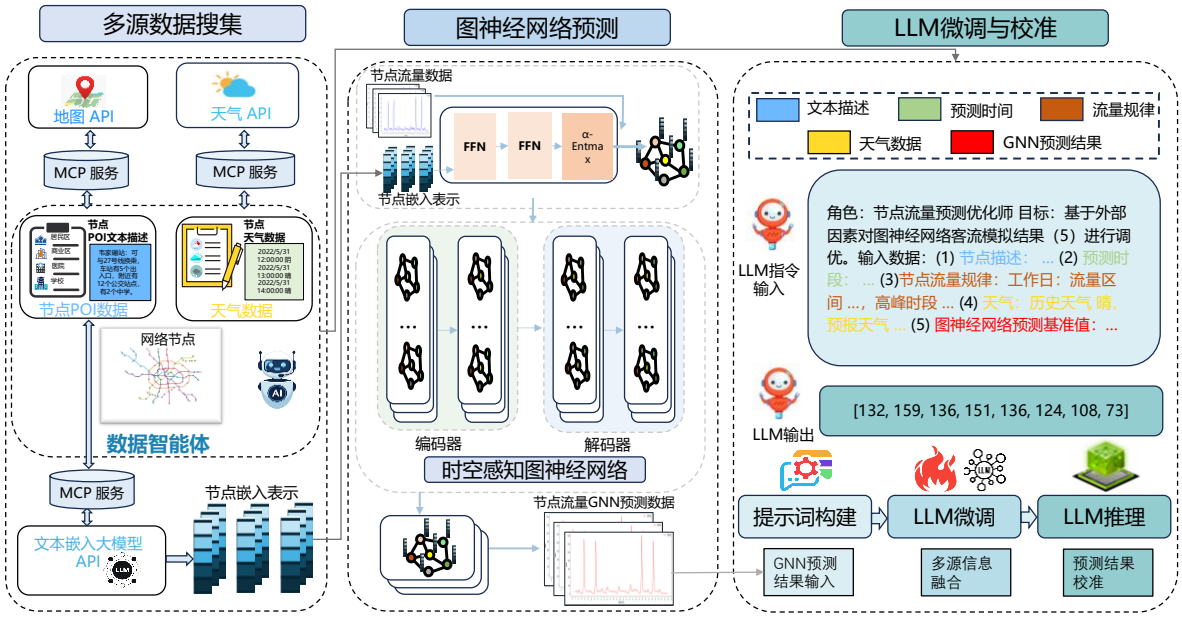

图 1 基于大模型的多源信息融合地铁客流预测框架

Fig.1 Large model-enabled multi-source information fusion framework for metro passenger flow forecasting

2 多源语义增强客流预测方法

2.1 基于 MCP 的数据智能体

数据智能体由 LLM Agent 核心和地图、气象、文本嵌入 3 类 MCP 服务组成。MCP 将外部接口封装为统一的工具描述、输入参数和返回格式，使智能体能够依据任务状态决定调用顺序并检查结果。具体流程如下。

1）站点标准化与定位。智能体首先通过地图文本检索获得线路站点列表，对返回结果执行类别筛选，仅保留“地铁站”实体；随后逐站调用地理编码工具，建立“标准站名—经纬度”映射，减少同名地点、出入口和商业设施造成的误匹配。

2）POI 采集与语义筛选。以每个站点坐标为中心，在 500 m 范围内检索 POI。智能体依据不同设施对地铁客流的产生和吸引作用，将 POI 划分为交通接驳、商业、办公、公共服务、居住和景点等类别，优先保留能够反映站点主要功能或具有明显客流吸引作用的对象，过滤小型零售餐饮、重复 POI 及附属设施；

对于类别不明确的对象，结合地图类别标签和文本描述按照统一规则进行语义判定。

3）文本描述生成。智能体将筛选后的 POI、换乘关系、出入口信息和站点功能归纳为标准化描述T_i。描述采用一致的字段顺序和表达粒度，既供 LLM 校准读取，也便于人工核验数据来源和站点功能。

4）上下文嵌入生成。通过文本嵌入 MCP 服务将T_i映射为d_i维向量

$$\boldsymbol{e}_i = \mathrm{Embed}(T_i), \qquad \boldsymbol{E} = [\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_N]^{\mathrm{T}} \tag{2}$$

静态 POI 描述可离线更新并缓存，天气等动态信息则按预测时段调用，从而将高成本的文本处理与高频客流推理解耦。

2.2 语义增强图结构构建

传统线路邻接矩阵只能表示相邻站点，基于客流相关系数构图又容易把偶然同步波动误认为稳定关系。用上下文嵌入学习站点功能关联。首先通过前馈网络完成维度压缩

$$\boldsymbol{z}_i = \mathrm{FFN}_1(\boldsymbol{e}_i) \tag{3}$$

对任意站点 *i* 和 *j*，将二者表示拼接并计算关联分数

$$\boldsymbol{h}_{ij} = \boldsymbol{z}_i \oplus \boldsymbol{z}_j, \qquad s_{ij} = \mathrm{FFN}_2(\boldsymbol{h}_{ij}) \tag{4}$$

Softmax 会为所有站点分配非零权重，容易引入大量弱连接。采用α-Entmax[16]对每行分数进行稀疏归一化

$$\boldsymbol{a}_i = \mathrm{Entmax}_\alpha([s_{i1}, s_{i2}, \ldots, s_{iN}]), \qquad \boldsymbol{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N]^{\mathrm{T}} \tag{5}$$

当$\alpha > 1$时，低相关站点权重可被压缩为 0，保留少量可解释的功能联系。这样，居住型站点、就业型站点或大型交通枢纽即使物理距离较远，也可通过语义相似性建立边；功能差异显著的相邻站点则不会被强制赋予相同传播强度。

2.3 时空图神经网络预测

在语义图 *A* 上加入自环并构造归一化传播矩阵

$$\tilde{\boldsymbol{A}} = \boldsymbol{A} + \boldsymbol{I}, \qquad \tilde{\boldsymbol{D}}_{ii} = \sum_j \tilde{A}_{ij} \tag{6}$$

第 *l* 层扩散图卷积定义为

$$\mathrm{GConv}_{\boldsymbol{A}}(\boldsymbol{X}) = \sum_{k=0}^{K-1} \boldsymbol{\Theta}_k^{(l)} \left(\boldsymbol{D}^{-1} \tilde{\boldsymbol{A}} \right)^k \boldsymbol{X} \tag{7}$$

式中：*K* 为扩散阶数，$\boldsymbol{\Theta}_k^{(l)}$为可学习参数。该为进一步刻画时间依赖，将图卷积嵌入门控循环单元。以更新门为例：操作使每个站点在一个时间步内融合多阶语义邻居信息。

$$\boldsymbol{u}_t = \sigma\left(\mathrm{GConv}_{\boldsymbol{A}}([\boldsymbol{X}_t, \boldsymbol{H}_{t-1}]) + \boldsymbol{b}_u\right) \tag{8}$$

重置门和候选状态采用相同的图卷积形式，最终得到

$$\boldsymbol{H}_t = \boldsymbol{u}_t \odot \boldsymbol{H}_{t-1} + (1-\boldsymbol{u}_t) \odot \tilde{\boldsymbol{H}}_t \tag{9}$$

编码器依次处理 H 个历史时间步，解码器以编码状态为初值滚动输出 P 个时间步，得到初步预测 $\boldsymbol{Y}^G \in \mathbb{R}^{N\times P}$。图网络采用平均绝对误差训练：

$$\mathcal{L}_G = \frac{1}{MNP}\sum_{m=1}^{M}\sum_{i=1}^{N}\sum_{p=1}^{P}\left|Y_{m,i,p} - \hat{Y}^G_{m,i,p}\right| \tag{10}$$

将所有站点和预测步的输出按固定顺序重塑为数值预测令牌 C，作为 LLM 校准模块的核心数值输入。

2.4 LLM 领域知识校准

LLM 提示词由 5 类已知信息和 1 类训练标签组成：任务指令、站点描述 T_i、站点客流规律、预测天气、图网络预测 C 以及训练阶段的真实未来客流。推理阶段删除真实未来客流。采用 Llama-3.1-8B-Instruct 作为基础语言模型。对于每个站点分别构造校准样本，将该站点未来 8 个时间步的图网络预测及对应上下文输入 LLM，并要求模型仅输出由方括号包围的 8 个非负数值，因此多站点预测通过逐站校准完成。该结构将文本知识置于数值预测周围，使 LLM 执行“有依据的修正”，而不是从头生成客流。

为降低微调成本，采用 LoRA[17]冻结预训练权重 $\boldsymbol{W}_0$，仅学习低秩增量

$$\boldsymbol{W} = \boldsymbol{W}_0 + \frac{\gamma}{r}\boldsymbol{B}\boldsymbol{A}_L, \qquad \boldsymbol{B}\in\mathbb{R}^{d_o\times r},\quad \boldsymbol{A}_L\in\mathbb{R}^{r\times d_i} \tag{11}$$

式中： r 为低秩维数； γ 为缩放系数。设置 r 为 32，缩放系数为 16，LoRA 作用于注意力投影层和前馈网络投影层。微调目标为输出数值序列的自回归负对数似然：

$$\mathcal{L}_{LLM} = -\sum_{\ell=1}^{L} p_\theta\left(o_\ell \mid o_{<\ell}, \text{Prompt}\right) \tag{12}$$

输出经格式解析后恢复为 $\boldsymbol{Y}$。若 LLM 输出数量不等于 8 或包含非数值内容，则回退至图网络结果；对少量负数预测执行非负截断。该安全机制保证了工程部署时的输出确定性。

3 实验与结果分析

3.1 数据与实验设置

实验采用成都地铁 1 号线、2 号线 2022 年 2 月-7 月真实客流数据，包含韦家碾至科学城、犀浦至成都行政学院共 60 个站点。客流按 15 min 聚合，每个样本输入连续 8 个时间步并预测未来 8 个时间步。站点静态信息来自地图服务检索得到的 500 m 范围 POI 及换乘、出入口等属性，动态外部信息包括预测时段的天气和日期属性。训练、验证和测试样本按时间顺序划分，所有归一化参数、站点客流规律和峰值阈值均仅由训练集确定，并在验证和测试阶段保持固定。验证和测试阶段按照与训练阶段一致的模板构造提示词，但仅包含预测时刻可获得的站点属性、预测时段、日期类型、客流规律、天气和图网络初步预测，不包含真实未来客流，以避免未来信息泄漏。

对比方法包括 4 类：1）Linear、NLinear 和 DLinear 等线性时序基线[10]；2）Transformer[7]、Autoformer[8]和 PatchTST[9]等注意力模型；3）ASTGCN[6]、DCRNN、DGCRN、STGCN、Graph WaveNet、AGCRN、MTGNN、STID 等时空图模型；4）Time-MoE[15]、TimesFM 2.5、Lag-Llama[14]和 MOMENT[13]等时序基础模型。此外纳入 MONSTER 多尺度时序基线。

所有方法采用相同的历史窗口、预测长度和数据划分，各基线模型均采用原论文或其官方公开实现中推荐的超参数设置，仅统一设置历史输入长度和预测长度，以保证各方法采用相同的预测任务和数据划分。LLM 校准模块采用 paged AdamW 8-bit 优化器进行 LoRA 微调，学习率设为 5×10^{-5}，训练 3 轮，批量大小为 32。

采用 MAE、RMSE 和 MAPE 评价总体预测。考虑运营管理更加关注高客流时段，进一步在峰值样本集合 Ω_{peak} 上计算 Peak-WAPE、Peak-MAE 和 Peak-RMSE。峰值集合由训练集客流分布的高分位阈值确定，避免使用测试集信息。各指标定义为

$$\text{MAE}=\frac{1}{Q}\sum_{q=1}^{Q}\left|y_q-\hat{y}_q\right|,\quad \text{RMSE}=\sqrt{\frac{1}{Q}\sum_{q=1}^{Q}\left(y_q-\hat{y}_q\right)^2} \tag{13}$$

$$\text{MAPE}=\frac{100\%}{Q}\sum_{q=1}^{Q}\frac{\left|y_q-\hat{y}_q\right|}{\left|y_q\right|+\varepsilon},\quad \text{WAPE}=\frac{\sum_{q}\left|y_q-\hat{y}_q\right|}{\sum_{q}\left|y_q\right|+\varepsilon}\times100\% \tag{14}$$

式中：Q 为样本点数；为防止分母为 0 的极小常数。Peak 类指标仅将求和范围替换为 Ω_{peak}。

3.2 总体对比结果

表 1 给出了成都地铁 1 号线和 2 号线测试样本联合统计的预测结果。所提方法在 6 项指标上均取得最优结果，表明其在两条线路联合测试样本上具有较好的总体预测精度和高客流时段预测性能。

表 1 所提方法与对比方法的地铁客流预测结果

Tab.1 Metro passenger flow forecasting results of the proposed and baseline methods

模型	MAE	RMSE	MAPE	Peak-WAPE	Peak-MAE	Peak-RMSE
PatchTST	73.584	190.579	118.74	43.51	399.403	640.358
Autoformer	55.940	126.421	120.79	24.62	226.024	393.377
Transformer	35.685	81.707	56.61	13.21	121.303	219.329
MONSTER	59.708	147.169	116.17	26.81	246.078	452.435
DLinear	96.705	215.521	243.88	59.39	545.123	774.485
NLinear	104.475	243.070	168.21	56.99	523.154	754.291
Linear	100.760	221.577	260.18	62.29	571.775	801.462
ASTGCN	35.696	73.541	60.17	13.19	121.052	188.605
DCRNN	38.281	95.360	36.56	18.97	174.151	288.396
DGCRN	27.782	66.312	33.97	9.34	85.736	150.020
STGCN	41.742	86.060	68.18	14.55	133.594	220.557
Graph WaveNet	27.585	65.476	31.76	9.80	89.990	153.859
AGCRN	26.470	63.440	34.86	8.90	81.725	150.366
MTGNN	46.406	132.154	125.20	21.10	193.712	459.026

模型	MAE	RMSE	MAPE	Peak-WAPE	Peak-MAE	Peak-RMSE
STID	29.539	73.983	31.97	11.27	103.424	198.203
Time-MoE	71.320	177.041	144.25	42.22	387.582	616.543
TimesFM 2.5 LoRA	99.505	258.371	145.67	42.42	389.353	625.930
Lag-Llama	87.516	231.718	129.82	60.00	550.768	824.345
MOMENT	75.377	184.585	120.78	43.08	395.481	604.703
所提出的方法	26.019	61.076	30.67	8.81	80.830	148.003

在 3 项总体指标中，AGCRN 和 Graph WaveNet 是表现较强的对比模型。相比 AGCRN，所提出的方法的 MAE 和 RMSE 分别降低 1.70%和 3.73%；相比 Graph WaveNet，MAPE 降低 3.43%。同时，所提出的方法明显优于 ASTGCN、DCRNN 和 STGCN 等主要依赖物理拓扑关系的图模型，说明基于站点功能和环境信息构建的语义图能够补充传统线路连接难以刻画的非局部站点关联。

在峰值指标方面，所提出的方法的 Peak-WAPE 和 Peak-MAE 较 AGCRN 分别降低 1.01%和 1.10%，Peak-RMSE 较 DGCRN 降低 1.34%。说明所提方法不仅能够降低整体平均误差，在高客流时段也能保持较稳定的预测性能。

时序基础模型在本数据上的表现未显示出预训练规模优势。TimesFM 2.5、Lag-Llama 和 MOMENT 的总体误差均高于专门训练的 Transformer 与 ASTGCN，原因可能在于：地铁站点客流具有强烈的本地周期和站间耦合，而通用基础模型对成都地铁的时间粒度、峰谷结构和线网关系缺少先验。把 LLM 置于图网络之后作为校准器，避免其直接承担全部数值建模任务，因此能更有效地发挥语义推理能力。

3.3 消融实验

为验证“图网络初步预测—LLM 领域校准”联合机制的有效性，选取桐梓林站作为代表性站点，在相同测试时段和数据划分下设置 3 种预测方式进行比较。GNN only 仅采用语义增强图网络，根据历史客流进行预测；LLM Direct 不使用图网络初步结果，而是将历史客流和额外信息输入 LLM 直接预测未来客流；Ours 首先利用图网络获得初步预测，再由 LLM 结合额外信息进行校准。结果见表 2。

表 2 不同预测机制的消融实验结果

Tab.2 Ablation results of different forecasting mechanisms

方法	历史客流	外部信息	GNN 预测	MAE	Peak-WAPE	Peak-MAE	Peak-RMSE
仅 GNN	√	×	×	31.448	7.179	106.543	132.798
LLM 直接预测	√	√	×	63.331	24.64	365.709	497.263
所提出的方法	×*	√	√	29.813	6.678	99.104	125.299

注：√表示相应信息直接输入当前预测或校准模块；所提出的方法中的历史客流由图网络进行时空建模，不直接输入 LLM。

由表 2 可知，与仅 GNN 方法相比，所提出的方法 MAE 降低 5.20%，Peak-WAPE、Peak-MAE 和 Peak-RMSE 分别降低 6.98%、6.98%和 5.65%，说明 LLM 能够利用外部上下文对图网络初步预测进行有效校准。相比 LLM Direct，Ours 在总体误差和峰值指标上表现更稳定，表明直接由 LLM 完成数值预测难以

充分刻画地铁客流的时空变化，而“图网络负责数值预测、LLM 负责领域校准”的两阶段机制能够更好地结合数值建模与外部知识。

3.4 语义图结构分析

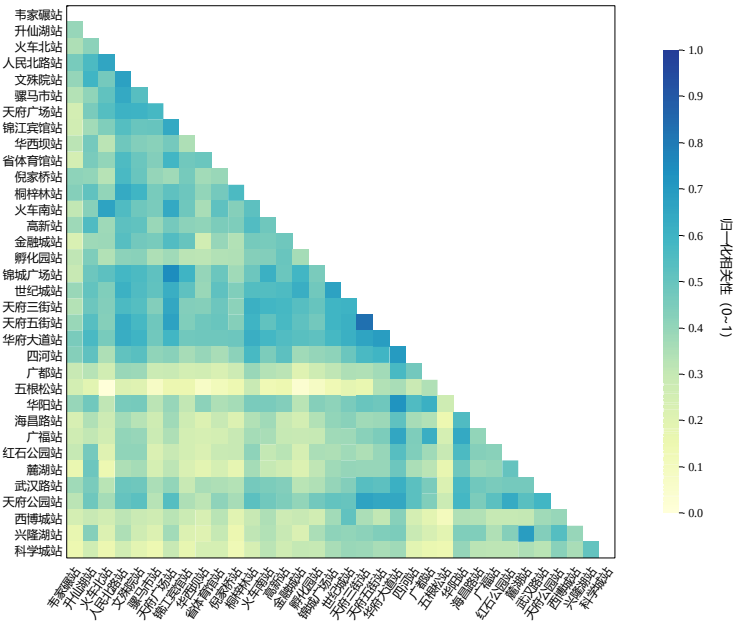

图 2 基于站点上下文嵌入的语义相关性热力图

Fig.2 Semantic correlation heatmap based on station context embeddings

图 2 给出了由站点文本嵌入学习得到的相关性热力图。图中除主对角线外还存在多条非局部高值带，说明语义图并未简单复制线路顺序。例如，具有居住和通勤功能的外围站点、承担商业和公共服务功能的中心站点分别形成相似关系；该关系可跨越多个物理站间距。α-Entmax 将大量弱关联压缩为 0，使保留连接更集中，减少全连接图中噪声传播。

与仅依据线路相邻关系构建的物理图相比，语义图从站点功能和土地利用角度刻画站点间关系，可建立物理距离之外的非局部联系。因此，采用语义邻接矩阵作为图网络的空间传播结构，使站点间的信息传播不再受线路相邻关系限制。实验结果表明，基于语义关系的信息聚合能够有效刻画功能相似站点之间的关联。

3.5 不同预测步长与典型站点预测分析

为分析模型在不同预测提前量下的性能，分别统计 15、30、45、60、90 和 120 min 对应预测步的 MAE、RMSE 和 MAPE，结果如图 3 所示。其中，6 种预测步长分别对应未来第 1、2、3、4、6 和 8 个时间步。通过比较不同预测步长下的误差变化，分析模型在短期和较长预测范围内的预测稳定性。

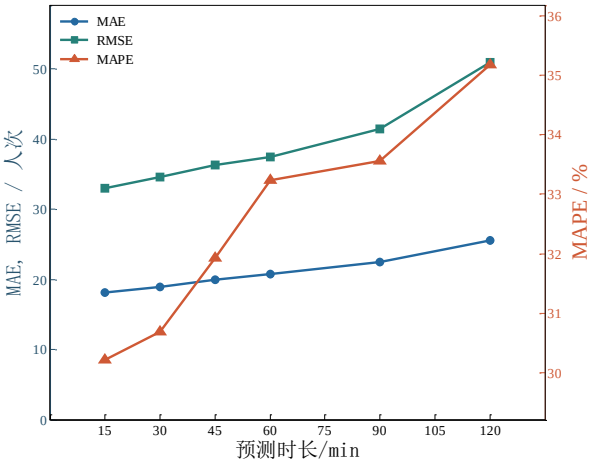

图 3 不同预测步长下的预测性能

Fig.3 Forecasting performance at different prediction horizons

由图 3 可知，随着预测步长增加，各项预测误差指标整体呈现一定的上升趋势，但不同预测步之间存在局部波动。所提出的方法在不同预测步长下均保持较低的 MAE、RMSE 和 MAPE，且在较长预测步长下仍具有较稳定的预测性能，表明所提出的语义增强图网络与 LLM 校准机制能够有效缓解多步预测过程中误差增加的问题。

为进一步直观分析模型在不同功能属性站点上的预测表现，从居住型、商业型、枢纽型和景点型站点中分别选取一个具有典型客流特征且数据完整的代表性站点，展示连续 500 个测试时间步的真实客流、AGCRN 预测结果和所提出的方法预测结果，其中每个时间步对应 15 min。如图 4 所示。

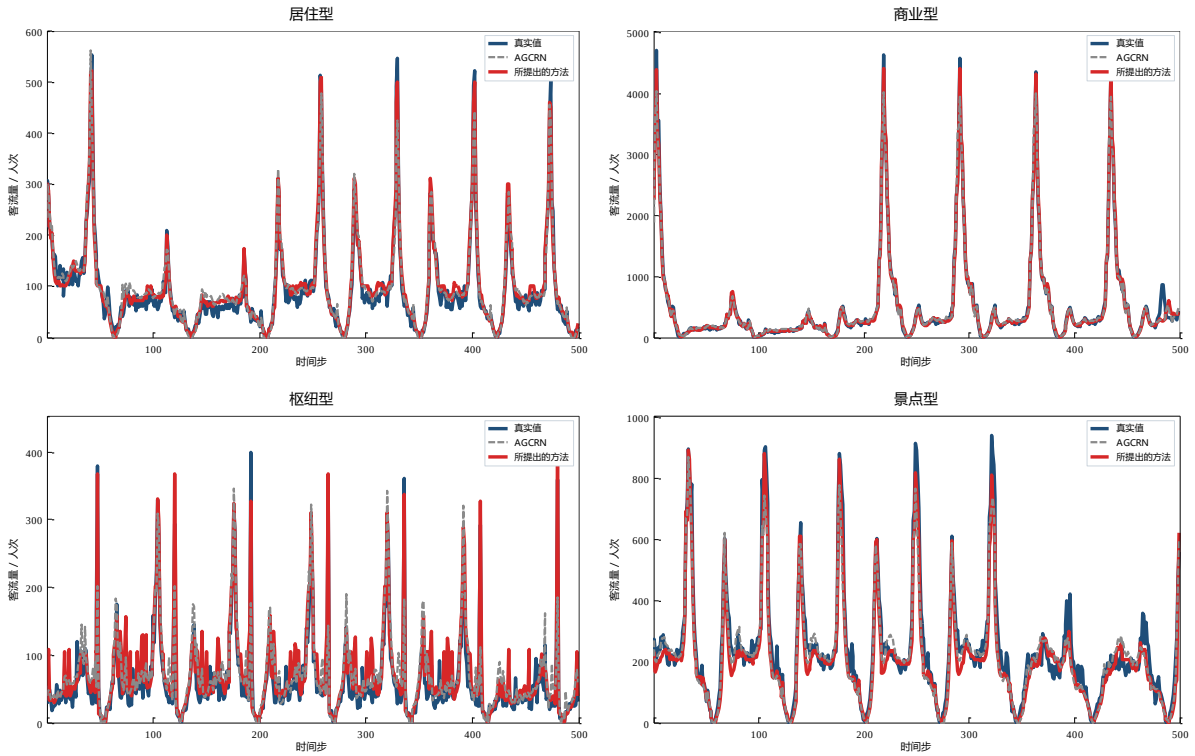

图 4 不同功能类型典型站点的客流预测结果

Fig.4 Passenger flow forecasting results at representative stations with different functional types

由图 4 可知，不同功能类型站点呈现出明显不同的客流变化特征。居住型站点具有较为规律的周期性峰谷变化，所提出的方法能够较好地跟随主要峰值及低谷变化；商业型站点客流规模较大，并存在多个尖锐高峰，所提方法对主要峰值出现时刻及幅值变化具有较好的跟踪能力。枢纽型站点客流波动更为频繁，局部随机性较强，所提出的方法能够保持总体变化趋势，但个别突增峰值仍存在一定偏差。景点型站点具有较明显的周期性高峰，所提出的方法能够较好地刻画峰前上升、峰值位置及峰后回落过程。与 AGCRN 相比，所提出的方法在多数主要峰谷位置与真实曲线具有较高的一致性，表明引入站点语义信息与 LLM 校准后能够适应不同客流形态下的预测需求。

从机制上看，图网络首先保证了预测曲线的连续性和时空一致性；LLM 再依据站点功能、时段和天气对整体幅值及峰谷趋势进行有限修正。二者的串联分工使模型既能保持数值预测的稳定性，又能在外部条件变化时获得语义修正依据。

3.6 方法适用性与局限

所提方法适用于具备连续客流数据、可访问地图与天气服务的城市轨道交通线路。站点 POI 文本可定期离线更新，在线阶段仅需输入近期客流和短时天气，因此数据智能体不会成为每个 15 min 预测周期的主要计算瓶颈。站点描述和保留的语义边还可供运营人员检查，从而为模型输出提供一定解释依据。

该方法仍存在 3 点局限：1）POI 搜索结果和文本嵌入会受到外部服务更新频率及类别偏差影响，应建立缓存、版本记录和人工抽查机制；2）LLM 校准增加了推理资源消耗，且少数极端峰值仍可能被过度修正；3）当前实验主要基于成都地铁两条线路，跨城市、换乘网络和突发事件场景下的泛化能力仍需验证。后续可引入大型活动、故障和社交感知数据，采用蒸馏或小模型替代部分 LLM 推理，并将峰值加权、概率预测和不确定性区间纳入训练目标。

4 结论

针对地铁客流预测中外部信息难获取、物理图难表达功能关联以及数值模型缺少领域知识的问题，提出了数据智能体、语义增强图网络和 LLM 校准相结合的预测方法，主要结论如下。

1）基于 MCP 的数据智能体能够自动完成站点标准化、500 m 范围 POI 检索、语义筛选、文本描述生成和嵌入向量构建，将分散的非结构化环境信息转换为图模型和 LLM 均可使用的统一表示，减少了人工数据处理链。

2）由站点上下文嵌入学习的稀疏语义图能够捕捉物理距离之外的功能关联。与仅强调拓扑关系的 ASTGCN 相比，所提方法的 MAE、RMSE 和 MAPE 分别降低 27.11%、16.95%和 49.03%，说明语义关系可有效补充线路连接信息。

3）“图网络初步预测—LLM 领域校准”的串联机制优于让通用基础模型直接承担客流预测。所提方法的 MAE、RMSE 和 MAPE 分别达到 26.019、61.076 和 30.67%，相较最优总体对比方法分别降低 1.70%、3.73%和 3.43%；Peak-WAPE、Peak-MAE 和 Peak-RMSE 也分别降至 8.81%、80.830 和 148.003。

4）所提方法在 6 项指标上均取得最优结果，表明其在总体预测和高客流时段均具有较好的预测稳定性。但当前实验仍主要基于成都地铁 1 号线和 2 号线，模型在跨城市、复杂换乘网络及突发异常场景下的泛化能力仍需进一步验证。未来可结合峰值感知损失、概率预测和轻量化校准模型，并在更大规模、多线路及实时运营环境中检验方法的可迁移性和计算效率。

参考文献：

[1] 张惠臻,高正凯,李建强,王晨曦,潘玉彪,王成,王靖.基于循环神经网络的城市轨道交通短时客流预测[J].吉林大学学报(工学版),2023,(2):430-438.

[2] 张金雷,陈奕洁,Panchamy Krishnakumari,金广垠,王骋程,杨立兴.基于注意力机制的城市轨道交通网络级多步短时客流时空综合预测模型[J].地球信息科学学报,2023,25(4):698-713.

[3] 冯树民,刘浩,李来成.雨雪天气下轨道交通客流预测模型[J].哈尔滨工业大学学报,2022,54(9):1-6.

[4] 尹宝才,王竟成,张勇,胡永利,孙艳丰.基于谱域超图卷积网络的交通流预测模型[J].北京工业大学学报,2024,83(2):152-164.

[5] LI P, WANG S P, ZHAO H T, et al. IG-Net: IG-Net: An interaction graph network model for metro passenger flow forecasting[J]. IEEE Transactions on Intelligent Transportation Systems, 2023, 24(4): 4147-4157.

[6] GUO S N, LIN Y F, FENG N, et al. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2019, 33(1): 922-929.

[7] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Advances in Neural Information Processing Systems 30. Red Hook: Curran Associates, 2017: 5998-6008.

[8] WU H X, XU J H, WANG J M, et al. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting[C]//Advances in Neural Information Processing Systems 34. Red Hook: Curran Associates, 2021: 22419-22430.

[9] NIE Y Q, NGUYEN N H, SINTHONG P, et al. A time series is worth 64 words: Long-term forecasting with transformers[C]//The Eleventh International Conference on Learning Representations. Kigali: ICLR, 2023.

[10] ZENG A L, CHEN M X, ZHANG L, et al. Are transformers effective for time series forecasting?[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2023, 37(9): 11121-11128.

[11] JIN M, WANG S Y, MA L T, et al. Time-LLM: Time series forecasting by reprogramming large language models[C]//The Twelfth International Conference on Learning Representations. Vienna: ICLR, 2024.

[12] DAS A, KONG W H, SEN R, et al. A decoder-only foundation model for time-series forecasting[C]//Proceedings of the 41st International Conference on Machine Learning. PMLR, 2024, 235: 10148-10167.

[13] GOSWAMI M, SZAFER K, CHOUDHRY A, et al. MOMENT: A family of open time-series foundation models[C]//Proceedings of the 41st International Conference on Machine Learning. PMLR, 2024, 235: 16115-16152.

[14] RASUL K, ASHOK A, WILLIAMS A R, et al. Lag-Llama: Towards foundation models for probabilistic time series forecasting[EB/OL]. (2024-02-07)[2026-07-29]. https://arxiv.org/abs/2310.08278.

[15] SHI X M, WANG S Y, NIE Y Q, et al. Time-MoE: Billion-scale time series foundation models with mixture of experts[C]//The Thirteenth International Conference on Learning Representations. Singapore: ICLR, 2025.

[16] PETERS B, NICULAE V, MARTINS A F T. Sparse sequence-to-sequence models[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: Association for Computational Linguistics, 2019: 1504-1519.

[17] HU E J, SHEN Y L, WALLIS P, et al. LoRA: Low-rank adaptation of large language models[C]//The Tenth International Conference on Learning Representations. Virtual Event: ICLR, 2022.



第一作者：张媛（2002—），女，硕士研究生，研究方向为交通流量预测。



通信作者：李皓（1994—），男，讲师，博士，硕士生导师，研究方向为交通流量预测和交通系统优化。