文章编号：1003-1421(2026)03-0059-09
引用格式：蔡梦影，张淼，丁怡，等. 基于时空注意力机制的GCN-LSTM地铁短时OD客流预测方法[J]. 铁道运输与经济，2026，48(3)：59-67.
CAI Mengying，ZHANG Miao，DING Yi，et al. Short-Term OD Passenger Flow Prediction Method for Subway Systems Based on Spatio-Temporal Attention Mechanism and GCN-LSTM[J]. Railway Transport and Economy，2026，48(3)：59-67.

基于时空注意力机制的GCN-LSTM地铁短时OD客流预测方法

蔡梦影[1]，张　淼[1]，丁　怡[1]，王　兵[1]，陈钉均[1]，卢广志[2]

(1.西南交通大学 交通运输与物流学院，四川 成都　611756；2.广州擎云计算机科技有限公司总经理办公室，广东 广州　510663)

摘　要：网络化地铁运营条件下，客流呈现高度复杂的时空动态特性，如何实现精准的短时OD客流预测是优化运输组织、缓解拥堵的关键基础。现有预测方法对客流时空依赖性的联合建模能力有待提升。为此，提出一种新颖的短时OD客流预测模型。该模型引入时空注意力机制，并深度融合GCN和LSTM在建模上的优势，使模型能够自适应地关注不同历史时刻对当前预测的重要性，并动态识别路网中影响目标OD对的关键站点或区域，从而更精准地捕获客流在复杂地铁网络中的非线性时空传播特征。通过实际地铁网络数据验证，结果表明，所提出的时空注意力GCN-LSTM模型相较于基准模型，显著提升了短时OD客流预测的精度，能够更有效地捕捉客流的时空波动特性，为后续网络化地铁的精细化客流管控、运力调配及协同优化策略提供了可靠的数据支撑与决策依据。

关键词：网络化；地铁；短时OD客流预测；时空注意力机制；GCN-LSTM模型
中图分类号：U293.13　　**文献标识码**：A　　**DOI**：10.16668/j.cnki.issn.1003-1421.20251021002

Short-Term OD Passenger Flow Prediction Method for Subway Systems Based on Spatio-Temporal Attention Mechanism and GCN-LSTM

CAI Mengying[1]，ZHANG Miao[1]，DING Yi[1]，WANG Bing[1]，CHEN Dingjun[1]，LU Guangzhi[2]

(1. School of Transportation and Logistics, Southwest Jiaotong University, Chengdu 611756, Sichuan, China; 2. General Manager Office, Guangzhou Qingyun Computer Technology Co., Ltd., Guangzhou 510663, Guangdong, China)

Abstract: Under subway network operation conditions, passenger flow exhibits highly complex spatio-temporal dynamics. Accurate short-term origin-destination (OD) passenger flow prediction is a fundamental requirement for transportation organization optimization and congestion mitigation. However, existing prediction methods still show limited capability in joint spatio-temporal dependency modeling of passenger flow. To address this issue, this paper proposed a novel short-term OD passenger flow prediction model. It introduced a spatio-temporal attention mechanism, and deeply integrated graph convolutional networks (GCN)-long short-term memory networks (LSTM)

收稿日期：2025-10-21　　**修订日期**：2026-01-20
通信作者：卢广志(1976—)，男，广东佛山人，广州擎云计算机科技有限公司总经理办公室高级工程师。
基金项目：国家重点研发计划课题(2022YFB4300502)；四川省科技创新人才项目(2024JDRC0020)；四川省科技计划项目(2025YFHZ0328)；广州市重点研发计划项目(202206030007)；中国铁路上海局集团有限公司科研计划课题(20.25037)

modeling advantages. The model adaptively assigned historical time-step importance and dynamically identified key stations or regions influencing target OD pairs. This allowed for a more precise capture of the nonlinear spatio-temporal passenger flow propagation in complex subway networks. Validation was performed using real subway network data. The results show that the proposed spatio-temporal attention GCN-LSTM model significantly improves prediction accuracy compared with baseline models. Spatio-temporal passenger flow fluctuations are captured more accurately. This provides reliable data support and a decision-making basis for subsequent fine-grained passenger flow control, capacity allocation, and collaborative optimization in subway network systems.

Keywords: Network; Subway; Short-Term OD Passenger Flow Prediction; Spatio-Temporal Attention Mechanism; GCN-LSTM Model

0　引言

城市轨道交通客流量持续攀升并逐步进入网络化运营阶段，出现客流拥挤现象。面对此现象，运营方为提前制定限流策略与运行计划需要掌握旅客的出行需求，因此需要进行客流预测。作为智能交通系统的重要组成，地铁短时客流预测基于历史客流数据、天气信息、假日因素等数据建模，对未来几分钟到几小时内地铁站点或线路客流的状况与变化趋势进行推算。相较于中长期预测，短时预测需应对天气、突发事件等带来的波动与更高不确定性[1]，因此模型应当具备捕捉复杂时空依赖关系的能力。

早期的客流预测研究主要聚焦于时间序列的统计分析，采用自回归移动平均(ARIMA)[2]、支持向量机回归(SVR)[3]等方法。这些模型能有效捕捉客流的趋势和季节周期，但对高维、非线性的数据拟合能力有限，且本质上将各个站点视为独立个体，忽略了地铁网络固有的空间关联。为提升预测精度，学者们开始融合外部多源数据。柳雪丽等[4]利用改进粒子群算法优化SVR，构建了融合时段、工作日类型和天气因素的IPSO-SVR组合预测模型；黄孙等[5]则在轨道交通新线规划中，融合人口、商业、公交等多维数据构建回归模型，证实了多因素融合的有效性，但未考虑地铁客流的空间依赖性。

随着深度学习技术的发展，学者们尝试将深度学习用于客流预测。以卷积神经网络(CNN)和循环神经网络(RNN)为代表的模型凭借强大的非线性拟合与序列建模能力成为研究主流[6]。为同时捕捉时空特征，学者们提出了如ConvLSTM等混合架构[7]，通过卷积操作提取局部空间特征，再利用循环单元学习时间演化。然而，标准CNN基于规则的欧几里得网格设计，难以精准表达地铁网络中站点间通过线路连接和换乘关系形成的复杂、非欧几里得图结构，限制了其对网络化空间依赖关系的建模能力。

近年来，由于图神经网络(GNN)对图结构数据的捕捉性能好，迅速成为交通客流预测领域的前沿与核心技术[8]。图卷积网络(GCN)通过在图的谱域或空域进行信息聚合，能够学习基于网络拓扑的空间依赖性。大量研究将GCN与时间序列模型结合，构建了GCN-GRU[9]等预测模型，实现了时间特征、空间特征的提取与融合。为进一步提升性能，研究向多维度深化。一是构建更丰富、更贴合实际的空间关联图，如同时考虑物理邻接、OD流量相似性和功能相似性来定义图结构[10]；戴智丞等[11]利用VR行为试验重构旅客活动链，构建反映非邻接区域复杂交互的空间关联图，显著提升了枢纽内部的预测精度。二是引入注意力机制以动态捕捉关联重要性，如张阳等采用图注意力网络(GAT)[12]让节点自适应地关注不同邻居，设计时空注意力模块来动态加权不同历史时刻和空间位置的影响。三是设计更复杂的层级或并行架构以深度融合特征，例如陈鹤林等[13]提出的GRU-双层Transformer模型，通过分层注意力机制分别挖掘OD对自身规律和全局协同效应；王小敏等[14]则采用TCN-Transformer并行提取局部与全局时序特征。四是考虑客流的多尺度周期特性，尹浩东等[15]提出的多周期自适应图卷积网络(MPAGNet)，能识别并融合不同周期下的时空模式。此外，李京娜等[16]构建的GAT-Attention-Mamba组合模型以及牟亮等[17]考虑数据与模型不确定性的概率预测框架，均从不同角度推动着预测精度与鲁棒

性的边界。

综上所述，既有研究在利用图结构刻画空间依赖、结合深度学习模型学习时间演化方面已达成共识，并取得了显著进展。然而，在面向网络化地铁短时OD客流预测这一研究领域，仍有以下关键挑战亟待解决。①在时空特征的动态交互与协同融合方面，多数模型未能建立一种能时空维度间动态交互、相互增强与选择性融合的内在机制。②在关键信息的高效筛选与聚焦方面，面对海量、高维且极度稀疏的OD矩阵，如何使模型具备自适应能力，在众多历史时间步和全网节点中，精准识别出对当前预测目标最具影响力的少数关键时空上下文，而抑制冗余或噪声信息的干扰，是提升模型效率、可解释性与泛化能力的核心。因此，提出一种融合时空注意力机制的GCN-LSTM短时OD客流预测模型，通过图卷积网络显式建模地铁网络的非欧几里得拓扑结构，以克服传统卷积在空间特征提取上的局限；利用长短期记忆网络捕获客流在时间维度上的长期与短期依赖关系；引入时空注意力机制，使之能够自适应地为不同历史时刻和网络中的关键节点分配合适的权重，从而实现从海量时空数据中动态、精准地提取对预测目标最有效的特征信息。

1　数据预处理

客流数据来源于H市地铁AFC系统刷卡记录，这些数据完整记录了旅客的出行信息。筛选出其中5个与短时OD客流预测相关的字段，AFC系统主要数据字段如表1所示。

表1　AFC系统主要数据字段

Tab.1　Main data fields of AFC system

字段含义	数据示例
卡号	5652894
进站车站编码	323
出站车站编码	332
进站日期时间	2018-04-08 17:04:45
出站日期时间	2018-04-08 17:28:55

地铁站每天有百万余条AFC系统刷卡数据，其中存在重复、异常和缺失情况，数据清洗后才能更好地用于短时OD客流预测，数据清洗规则如表2所示。

表2　数据清洗规则

Tab.2　Data cleansing rules

筛选原因	筛选规则	筛选方式
数据缺失	数据中存在空值	删除
数据冗余	重复数据	删除
异常数据	进出站相同	删除
	出行时间大于2 h	删除
	进站时间早于出站时间	删除
	进站时间早于运营时间	删除
	出站时间晚于运营时间	删除

2　预测模型

城市轨道交通客流不仅在时间分布上变化，在空间分布上也会受到线网结构的影响。城市轨道交通客流数据量较大、维度较高，其中存在较为复杂的非线性规律且容易受到随机因素的影响，导致传统的客流预测方法难以进行准确预测。而神经网络能够有效地拟合客流数据，在捕捉客流变化的非线性特征方面具有更强的能力逐渐受到研究者的青睐。长短期记忆神经网络(LSTM)擅长于处理和预测时间序列数据，其克服了传统RNN在处理长时间序列时出现梯度下降过快乃至消失问题，具有较强的记忆能力，能够捕捉长时间跨度的依赖关系，这使得其在许多复杂的时序预测任务中表现出色[15]。图卷积神经网络(GCN)利用图的结构特性以及节点之间的关系，通过卷积操作在图结构上进行传播和聚合，从而能够从非欧几里得网络中提取空间特征信息。因此，构建基于注意力机制的GCN-LSTM模型，能够同时捕捉城市轨道交通客流数据中的时间特征以及城市轨道交通网络的空间特征，提高短时OD客流预测精度。

2.1　融合注意力机制的地铁客流预测模型构建

基于时空注意力机制的GCN-LSTM模型，其核心思想在于构建一个端到端的深度学习框架，以实现对地铁OD客流时空依赖性的分层与协同捕捉，从而提高短时OD客流预测精度。模型遵循“空间特征提取→时间特征提取→时空特征动态融合”的递进式设计。首先，空间特征提取模块将地铁网络抽象为图结构，将站点作为节点，站点间的物理连接或功能关联作为边。该模块利用图卷积操作在图的频域或空域上进行信息传播，使得每个站点的特征表征能够聚合其邻域站点的信息，从而学习到源

于网络拓扑的、非欧几里得空间意义上的深层空间依赖性。此过程将原本孤立的站点客流数据，转化为蕴含网络化空间上下文信息的特征向量，为后续时序建模奠定了坚实的基础。基于注意力机制的GCN-LSTM模型结构图如图1所示。

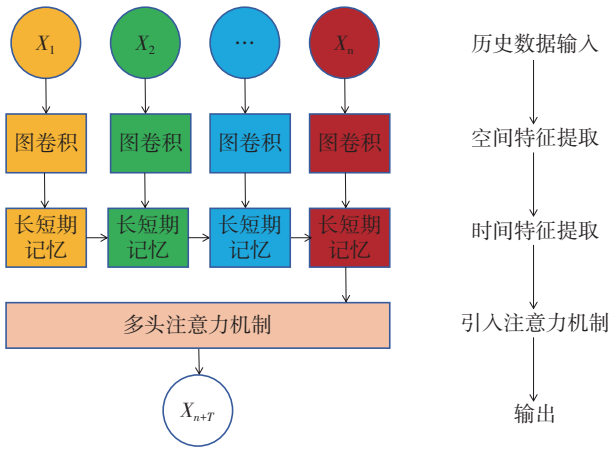


图1　基于注意力机制的GCN-LSTM模型结构图
Fig.1　Architecture of attention-based GCN-LSTM model

图卷积神经网络(GCN)模型属于空域图卷积神经网络，不依赖图谱卷积理论，而是直接在图的空间结构上定义卷积操作。与传统空域卷积需对中心节点的邻域进行有序排列不同，由于图结构数据具有不固定的邻域分布且节点间缺乏自然顺序，常规卷积难以直接应用。为此，采用基于局部一阶近似的图卷积架构，通过聚合中心节点与其邻居的特征信息，实现对局部图结构和节点属性的编码。该模型作为一种适用于图结构数据的半监督学习方法，具有良好的可扩展性，计算复杂度与图中边的数量呈线性关系，图卷积神经网络如图2所示。

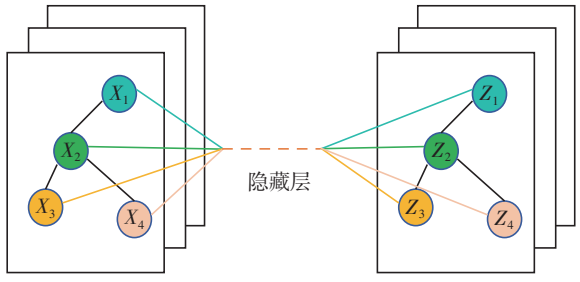


图2　图卷积神经网络
Fig.2　Graph convolutional network

对于有N个节点的无向图，根据其节点之间的连接关系能构成$N\times N$阶的邻接矩阵$\boldsymbol{A}$、度矩阵$\boldsymbol{D}$和$N\times F$特征矩阵$\boldsymbol{X}$，F为特征的维度，其中度矩阵是一个对角矩阵，对角线上的值表示图中各个节点与其相连的边的数量。K阶邻接矩阵$\boldsymbol{A}^K_{i,j}$、度矩阵$\boldsymbol{D}_{ii}$计算公式如下。

$$\boldsymbol{A}^K_{i,j}=\begin{cases}1 & i，j为K阶相邻\\ 0 & 其他情况\end{cases} \tag{1}$$

$$\boldsymbol{D}_{ii}=\sum_{j=1}^{n}\boldsymbol{A}_{ij} \tag{2}$$

为降低度值较大的节点的影响，同时避免信息流的偏向性，GCN采取对称归一化来进行聚合操作，在聚合过程中不仅考虑节点本身的度也考虑相邻节点的度从而使得信息传播更加公平稳定，同时这个过程使得GCN能够更好地从图结构中提取有效的空间特征。

传统RNN通过前馈神经网络将历史信息传递到后续时间步，但在处理长序列时，梯度问题导致模型难以捕捉长期依赖关系。LSTM通过引入门控机制有效缓解这一问题，使其能够更好地记忆和利用长序列数据。LSTM记忆单元示意图如图3所示。

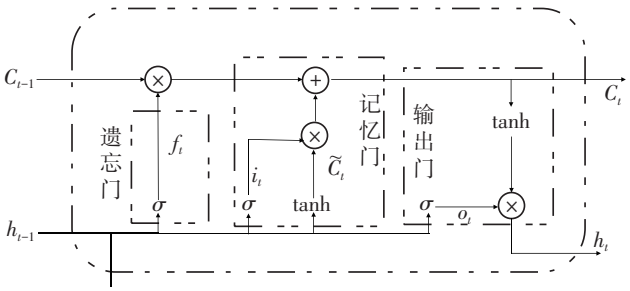


图3　LSTM记忆单元示意图
Fig.3　LSTM memory cell

遗忘门计算公式如下。

$$f_t=sigmoid(W_f\cdot[h_{t-1},\ x_t]+b_f) \tag{3}$$

式中：$sigmoid$将任意实数映射到(0，1)区间内；W_f为权重；h_{t-1}为前一时间步的隐藏状态；x_t为当前时间步的输入；b_f为偏置；f_t为[0，1]之间的向量，接近1表示保留，接近0表示遗忘。

记忆门决定要存储的细胞状态信息，其计算公式如下。

$$i_t=sigmoid(W_i\cdot[h_{t-1},\ x_t]+b_i) \tag{4}$$

$$\widetilde{C}_t=\tanh(W_c\cdot[h_{t-1},\ x_t]+b_c) \tag{5}$$

$$C_t=i_t\cdot\widetilde{C}_t+f_t\cdot C_{t-1} \tag{6}$$

式中：W_i，W_c为权重；b_i，b_c为偏置；i_t为[0，1]之间的向量，接近1表示记忆，接近0表示不保留；tanh将实数映射到(-1，1)区间；$\widetilde{C}_t$为激活后的细胞状态；C_{t-1}为上一时刻的细胞状态；C_t为当前时刻的细胞状态。

输出门主要用于控制从记忆单元中输出的信息量，其计算公式如下。

$$O_t = sigmoid(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{7}$$

$$h_t = O_t \cdot \tanh(C_t) \tag{8}$$

式中：W_o为权重；b_o为偏置；O_t为[0，1]之间的门控值；h_t为当前时间步的隐藏状态。

LSTM通过正向传播过程中的门控机制来处理数据，生成隐藏状态和细胞状态。在反向传播过程中，损失函数会随着序列长度的增加而变化，优化器通过计算梯度并更新网络参数来最小化损失函数。与RNN不同，LSTM通过其独特的结构能够在反向传播时有效截断误差的传播路径，从而缓解梯度消失与梯度爆炸的问题，模型可以更好地捕捉长序列中的依赖关系。

注意力机制是模拟人类大脑的思维模式，旨在通过削弱无关信息的影响，增强对问题求解有用信息的关注，在计算过程中通过对输入序列中的每个时间步分配不同的权重，从而决定它们对最终输出的影响程度。具体来说计算某个查询Query、键K和值V之间的关系时，使用以下的公式。

$$Attenion(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = softmax(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}})\boldsymbol{V} \tag{9}$$

式中：$\boldsymbol{Q}$为查询矩阵；$\boldsymbol{K}$为键矩阵；$\boldsymbol{V}$为值矩阵；$\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}$是查询和键之间的点积相似度；$softmax$是归一化操作，确保权重和为1。

多头注意力机制通过线性变换将输入数据分解为多个独立的查询Query、键K和值V向量组，每个组对应一个头。每个头通过注意力机制计算其自身的注意力权重，并生成对应的输出。所有头的输出经过拼接后，再通过一个线性变换层生成最终的输出。这种设计不仅提高了计算效率，还使得模型能够从多个角度关注输入序列的不同部分，从而捕捉到更丰富和深层次的特征信息。其计算公式如下。

$$MultiHead(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = Concat(Attenion_1(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}), Attenion_2(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}), \cdots, Attention_h(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}))W^O \tag{10}$$

其中每个注意力头h的计算为

$$Attenion_h(\boldsymbol{Q}_h, \boldsymbol{K}_h, \boldsymbol{V}_h) = soft\max(\frac{\boldsymbol{Q}_h\boldsymbol{K}_h^{\mathrm{T}}}{\sqrt{d_k}})\boldsymbol{V}_h \tag{11}$$

式中：$\boldsymbol{Q}_h$，$\boldsymbol{K}_h$，$\boldsymbol{V}_h$是通过不同线性变换得到的多个查询、键和值；$Concat$操作将多个头的输出拼接起来；W^O是输出的线性变换矩阵。

这种方式使得模型能够同时关注到输入数据的不同部分，从多个子空间中捕获数据的多样特征，每个注意力头专注于特定的特征维度，从而增强模型对复杂信息的处理能力。多个头协同作用，能够在处理序列数据时聚焦于不同的时间步或特征，提升模型的表达能力以及对关键信息的捕捉精度。同时，将注意力分配扩展到多个相关节点，减少单次注意力运算的随机误差，进一步提高模型预测的准确性和学习效率。

2.2　融合注意力机制的地铁客流预测模型实现步骤

（1）空间特征提取模块。在空间特征提取时，一般采用一阶邻接矩阵作为模型的邻接矩阵输入，仅考虑一阶邻接矩阵时，节点的信息传递仅限于其直接相邻的站点。而城市轨道交通线网结构复杂，节点的关系往往不仅限于直接相邻的节点，而是通过一系列中间节点构建的间接关系，线路之间可以通过换乘站建立联系。因此，采取3阶邻接矩阵作为模型的输入，希望信息通过更多的节点进行传播，从而捕捉到更加复杂和远距离的空间依赖关系，提高模型的预测精度。除此之外，空间距离也是影响旅客出行的一个重要因素，如果一个车站与其相近的车站客流波动情况相似，那么在拥挤传播时影响也更大，因此将站点之间的距离作为特征之一。同时，由于距离越近影响越大，距离越远影响越小，所以在处理时，将站点之间距离的倒数来生成权重矩阵。最后将客流数据按照OD进行统计，分时段提取OD矩阵，并且利用掩码层将值为0的OD处理不进入计算，降低低值客流对高值客流的影响。

将GCN模型属性设置为2层图卷积，第1层用来接收输入的邻接矩阵与特征矩阵，初步提取的空间特征，生成新的节点特征；第2层则在第1层输出的基础之上，进一步捕捉更深层次的站点之间的

空间依赖性。其具体计算公式如下。

$$H^1 = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}XW^{(0)}\right) \tag{12}$$

$$H^2 = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(1)}W^{(1)}\right) \tag{13}$$

（2）时间特征提取模块。在GCN层输出的节点特征信息已经包含邻接节点的关系和空间信息的基础之上，为进一步提取时序信息，构建了一个双层的LSTM模型。LSTM通过学习历史时间步的客流数据，捕捉每个站点在时间维度上的动态特征。同时由于每个节点在经过GCN层提取后已经融合了其他邻接节点的空间特征，LSTM层不仅仅关注单一节点的时序特征，还能够在时间维度上融合其他邻接节点的动态特征，从而进一步提升了时空特征提取能力，并且双层LSTM能够使得模型捕捉长时间序列中复杂的时空关联。

（3）多头注意力机制模块。为进一步增强时序数据中不同时间步之间的相关性，在LSTM之后引入了多头注意力机制，将注意力头数设为8个。模型通过8个独立的注意力头分别关注输入序列中不同部分的信息，并且每个注意力头都会根据不同的权重分配，动态调整对序列中其他时间步的关注程度，使得模型在每个时间步上为未来的预测任务分配不同的关注焦点，从而提升对关键信息的捕捉能力，强化时间序列的关联性，使得预测结果更加精准。

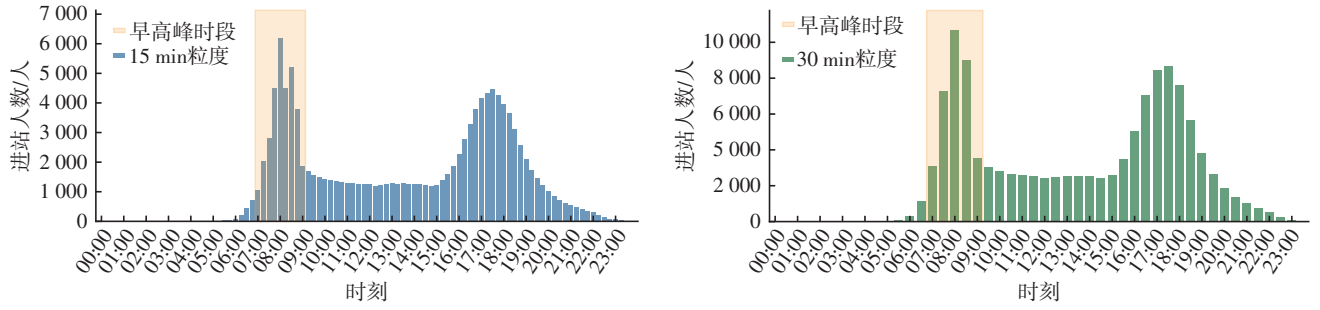
3　案例分析

3.1　数据分析与处理

为验证所构建的基于注意力机制的GCN-LSTM短时OD客流预测模型在常态运营场景下的有效性，利用H市地铁AFC系统刷卡数据作为依据，H市地铁网络作为研究对象，路网共包括136个站点。

早、晚高峰期间城市轨道交通客流具有显著的短时集聚性与动态变化性。路网进站客流图如图4所示，H市地铁网络工作日的进站客流呈现明显的双峰形态。为进行短时OD客流预测，仍需对数据进一步处理，需要合理划分历史客流数据并生成OD分时矩阵，时间间隔过大会导致客流特征捕捉不够精细，难以反映客流实际波动规律；而时间间隔过小则会显著增加数据处理复杂度和计算难度。如图4中30 min间隔可能掩盖其内部细微的波动，在8:00—9:00的早高峰核心时段，客流经历了先下降再上升而后下降后骤减的过程，15 min的时间粒度能够更好地捕捉这类高峰内的客流上升、峰值和回落阶段，从而更精细地反映客流的真实动态，为限流与运行计划优化提供更灵敏的数据输入。而如果时间粒度过小将导致数据维度急剧膨胀，会显著增加图卷积神经网络与长短期记忆网络的训练复杂度和时间成本，同时可能因数据过于稀疏而影响模型稳定性。因此，选择15 min时间粒度来刻画客流短时波动特征与控制模型计算复杂度，满足运营调整对数据时效性与准确性的要求，又避免了因粒度过细导致的数据处理负担过重和模型训练效率低下问题。

H市地铁运营时间为6:00—24:00，选取对客流特征捕捉更精细的15 min时间粒度作为时间间隔，一天共计72个时间间隔，用于结构化历史客流数据。短时OD预测模型输出的是未来时段的OD矩阵，其时间粒度与历史数据一致。同时GCN的输入需要邻接矩阵与特征矩阵，根据拓扑图中车站之间的连接关系可以获取3阶邻接矩阵，而OD对之间的距离利用站点经纬度信息进行求取，而后利用其计算结果的倒数来生成权重矩阵；将每个时间



a　15 min时间粒度客流　　b　30 min时间粒度客流

a　Passenger flow fluctuations at 15-munute intervals　　b　Passenger flow fluctuations at 30-munute intervals

图4　路网进站客流图

Fig.4　Passenger entry flow map of subway network

步的客流量信息作为特征矩阵，同时将特征矩阵时间步与批次大小整合为一个维度，以便每个时间步的特征矩阵可以单独进行处理。

最后为提高模型的稳定性与泛化能力，利用min-max公式将OD客流数据进行归一化，计算公式如下所示。

$$y_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{14}$$

式中：x_i为序列原始值；$x_{\min}$，$x_{\max}$为序列中的最小值与最大值；y_i序列经过归一化后的值。

3.2　评价指标选取

OD客流预测是对未知值的拟合，其结果与真实数据之间存在一定的差值，这部分差值是客观存在的预测误差，而预测误差能够反映模型的预测效果。为评估预测效果，选取均方误差(MSE)作为损失函数，平均绝对误差(MAE)和均方根误差(RMSE)作为评价指标，这3个指标分别用r_{MSE}，r_{MAE}，r_{RMSE}表示，值越大说明误差越大，值越小则误差越小。计算公式如下所示。

$$r_{\mathrm{MSE}} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(od_{i,j}^{d,t} - \overline{od_{i,j}^{d,t}})^2 \tag{15}$$

$$r_{\mathrm{MAE}} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left| od_{i,j}^{d,t} - \overline{od_{i,j}^{d,t}} \right| \tag{16}$$

$$r_{\mathrm{RMSE}} = \frac{\sum_{t=1}^{k}\sum_{i=1}^{n}\sum_{j=1}^{n} od_{i,j}^{d,t}}{\sum_{t=1}^{k}\sum_{i=1}^{n}\sum_{j=1}^{n}\overline{od_{i,j}^{d,t}}}\sqrt{r_{\mathrm{MSE}}} \tag{17}$$

式中：k表示一天的时间间隔；n表示车站总数；$od_{i,j}^{d,t}$表示d日t时间间隔由i站到j站的od量预测值；$\overline{od_{i,j}^{d,t}}$表示$d$日$t$时间间隔由$i$站到$j$站的$od$量真实值。

3.3　模型参数设置

利用深度学习框架和接口来构建基于注意力机制的GCN-LSTM模型，GCN和LSTM均为2层。设置训练轮数为60，并且节假日导致部分时间段客流发生突变与日常客流波动情况不同，将其剔除后，数据量较小，为更好学习客流变化情况，将批次大小设置为6，同时利用MSE衡量平均误差，MAE衡量误差的绝对大小而RMSE衡量异常值，LSTM模型每层的神经元设置为128个，选择Adam优化器，学习率设置为0.001，损失函数图如图5所示。

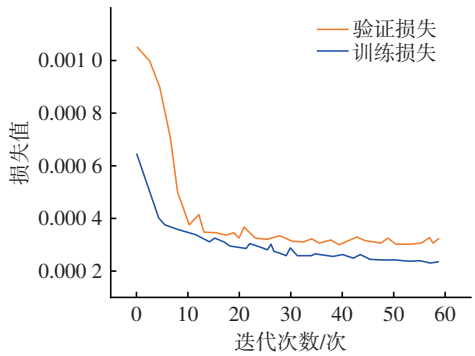
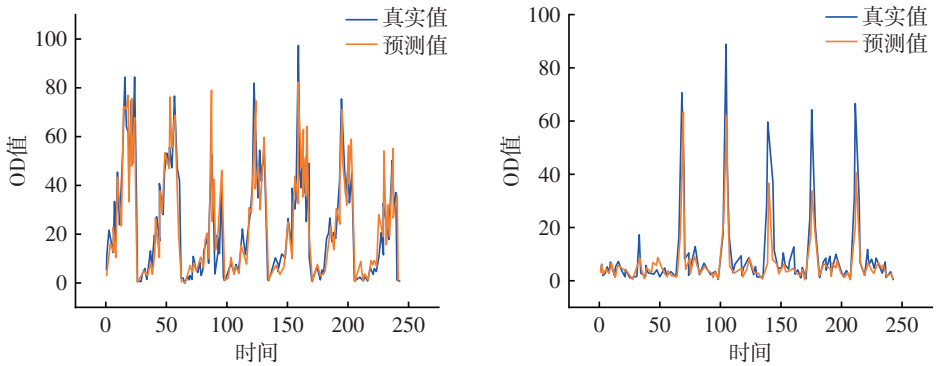


图5　损失函数图

Fig.5　Loss function plot

3.4　结果分析

由图5可知，损失函数值随着迭代次数的增加，先明显下降，然后渐渐趋于稳定。在训练次数达到30次后，构建的模型已处于稳定状态，此时的损失值为0.000 308 52。同时为更直观地观察预测结果，将预测值反归一化后选择2个OD进行可视化展示，OD预测效果图如图6所示。



a　第一组OD可视化展示

a　OD visualization for Group 1

b　第二组OD可视化展示

b　OD visualization for Group 2

图6　OD预测效果图

Fig.6　OD prediction results

从图6可以看出，基于注意力机制的GCN-LSTM模型对于短时OD客流量预测有较好的效果，其优越性源于其对客流时空动态本质更精细的刻画能力。图卷积组件确保了模型能够理解客流的空间传播受限于物理路网结构，而非均匀扩散；长短期记忆网络则有效建模了客流的时序惯性、周期性与趋势性。而注意力机制的引入，是实现性能跨越的关键。这种动态加权机制使得模型在面对客流数据中的非线性、非平稳性及突发波动时，具备更强的特征选择与适应能力，从而实现了比ConvLSTM等静态或均权模型更高的预测精度。同时利用ConvLSTM模型进行对比，不同预测模型精度指标对比如表3所示。

表3　不同预测模型精度指标对比

Tab.3　Comparison of accuracy metrics among different prediction models

模型	MSE	MAE	RMSE
ConvLSTM	7.240 5	0.678 2	2.690 8
基于注意力机制的GCN-LSTM	4.422 3	0.363 1	2.102 9

从预测精度指标可以看出，基于注意力机制的GCN-LSTM模型相较于ConvLSTM模型取得了更好的效果。

4　结束语

针对网络化地铁短时OD客流预测关键问题，构建基于自注意力机制的GCN-LSTM短时OD客流预测模型，利用H市地铁网络进行实例验证，结果表明短时OD客流预测模型具有较好的效果，误差较小，预测结果可作为基础数据进行输入。由于OD客流数据无法实时获取，采用历史AFC刷卡数据对其进行预测，使得短时OD客流预测结果准确性中存在不足，在后续的研究中可以将视频技术与进出站数据相结合从而实现动态捕捉客流波动情况。同时提出的模型针对常态运营场景下的地铁短时OD客流预测问题，展现了良好的性能。模型的优势在于能有效学习并预测由固定通勤等规律性需求驱动的客流时空演变规律。但对于节假日、大型活动或突发事件等非常态场景，客流模式会发生剧烈变化甚至偏离历史规律，此时模型的预测精度可能面临挑战，未来可针对突发状况、节假日以及大型活动的OD客流进行预测，从而为运营方制定应对方案提供数据支撑。

参考文献：

[1] 徐新颖. 基于卷积神经网络的城市轨道交通短时客流预测方法研究[D]. 福州：福州大学，2020.

[2] 马超群，李培坤，朱才华，等. 基于不同时间粒度的城市轨道交通短时客流预测[J]. 长安大学学报(自然科学版)，2020，40(3)：75-83.
MA Chaoqun，LI Peikun，ZHU Caihua，et al. Short-Term Passenger Flow Forecast of Urban Rail Transit Based on Different Time Granularities[J]. Journal of Chang'an University (Natural Science Edition)，2020，40(3)：75-83.

[3] 杨　军，侯忠生. 基于小波分析的最小二乘支持向量机轨道交通客流预测方法[J]. 中国铁道科学，2013，34(3)：122-127.
YANG Jun，HOU Zhongsheng. A Wavelet Analysis Based LS-SVM Rail Transit Passenger Flow Prediction Method[J]. China Railway Science，2013，34(3)：122-127.

[4] 柳雪丽，徐　亮，孔祥飞，等. 基于IPSO-SVR组合算法的城市轨道交通客流预测研究[EB/OL]. 甘肃科学学报，(2025-11-26)[2025-12-22]. https：//lib. swjtu. edu. cn/articlesearch/web_searchingDetail? id=2031626963698.

[5] 黄　孙，陈善宇，谢静静，等. 基于多维数据的轨道交通客流预测及诱增研究：以宁波轨道交通8号线为例[J]. 人民公交，2025(21)：97-102.

[6] 张锣琰. 基于图神经网络的地铁客流预测研究综述[J]. 中国储运，2025(11)：32.

[7] CHU K F，LAM A Y S，LI V O K. Deep Multi-Scale Convolutional LSTM Network for Travel Demand and Origin-Destination Predictions[J]. IEEE Transactions on Intelligent Transportation Systems，2020，21(8)：3219-3232.

[8] KE J T，QIN X R，YANG H，et al. Predicting Origin-Destination Ride-Sourcing Demand with a Spatio-Temporal Encoder-Decoder Residual Multi-Graph Convolutional Network[J]. Transportation Research Part C：Emerging Technologies，2021，122：102858.

[9] LIU L B，ZHU Y Y，LI G B，et al. Online Metro Origin-Destination Prediction *via* Heterogeneous Information Aggregation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2023，45(3)：3574-3589.

[10] 彭　颢，贺玉龙，宋太龙，等. 基于动态图神经常微分方

程的地铁短时客流预测方法[J]. 交通信息与安全，2024，42(1)：150-160.
PENG Hao，HE Yulong，SONG Tailong，et al. Forecasting for Short-Term Passenger Flow of Subway Based on Dynamic Graph Neural Ordinary Differential Equations[J]. Journal of Transport Information and Safety，2024，42(1)：150-160.

[11] 戴智丞，李得伟，李　华，等. 基于行为表征空间关联网络的综合交通枢纽客流短时预测框架[EB/OL]. 交通运输工程学报，(2025-12-18)[2025-12-22]. https：//kns. cnki. net/kcms2/article/abstract？v=VYuoLtjwl8PBzHWrDzIxCY6g-H08pJsCAiKQ7F_vJ1diT8j6aP7wpipZW0M_3ONHpsCPOEl5VMYGhXhInyf-ESyo5vKgS2Jnxq0dx-jAPZ11n2IH7eZs4YTXUittgrsa2OAzmunGQBRoyPBTk1YjdaluQEemYkRdk0Niog7nVIKVdrjViVTctw== &uniplatform=NZKPT&language=CHS.

[12] 张　阳，陈燕玲. 多头图注意机制的深度学习地铁客流预测方法[J]. 铁道运输与经济，2024，46(2)：167-175.
ZHANG Yang，CHEN Yanling. Deep Learning Subway Passenger Flow Forecasting Method Based on Multi-Head Graph Attention Mechanism[J]. Railway Transport and Economy，2024，46(2)：167-175.

[13] 陈鹤林，陈军华，高天泽. 模型融合时空特征的GRU-双层Transformer轨道交通OD客流预测[EB/OL]. 铁道科学与工程学报，(2025-10-13)[2025-12-22]. https：//lib. swjtu. edu. cn/articlesearch/web_searchingDetail？id=2031621064173.

[14] 王小敏，张悦晗. 基于时间卷积-Transformer模型的多场景地铁短时进站客流预测[J]. 同济大学学报(自然科学版)，2025，53(11)：1737-1745.
WANG Xiaomin，ZHANG Yuehan. Multi-Scenario Short Term Inbound Passenger Flow Prediction of Subway Based on TCN-Transformer Model[J]. Journal of Tongji University (Natural Science)，2025，53(11)：1737-1745.

[15] 尹浩东，范乃涯，昌锡铭，等. 基于多周期自适应图卷积网络的城市轨道交通短时客流预测[EB/OL]. 系统科学与数学，(2025-11-20)[2025-12-22]. https：//kns. cnki. net/kcms2/article/abstract？v=VYuoLtjwl8MkK8GXziDyL_963nMJGP87fpxOpIJXAS8oJPwcLkt3bYjoBYagrVLvg7rccFFPpTpN_JIg6tru9MwO_61wBl4eEgQY3K4AKmFTD1DXgRUyMmulbFiVZ6U1oYOYgK7Vbs6d6J_CUHNAdmTm4MzTOwoFlGSt1-wGTyzt0eo9YVrGeA==&uniplatform=NZKPT&language=CHS.

[16] 李京娜，秦利燕，王　颖，等. 基于GAT-Attention-Mamba组合模型的轨道交通短时客流预测[EB/OL]. 青岛大学学报(自然科学版)，(2025-10-17)[2025-12-22]. https：//lib. swjtu. edu. cn/articlesearch/web_searchingDetail？id=2031621134730.

[17] 牟　亮，康彧瑞，闫梓续，等. 考虑数据和模型不确定性的城轨线网客流短时预测方法[J]. 交通运输系统工程与信息，2025，25(6)：118-128.
MU Liang，KANG Yurui，YAN Zixu，et al. Method for Short-Term Passenger Flow Prediction in Urban Rail Transit Networks Considering Data and Model Uncertainty[J]. Journal of Transportation Systems Engineering and Information Technology，2025，25(6)：118-128.

(责任编辑　汪沁萱)