



全局-局部协同感知的列车到达晚点预测

周靖翔¹, 代婷婷¹, 黄燕文², 刘 峤¹

(1. 电子科技大学 计算机科学与工程学院, 四川 成都 611731;

2. 电子科技大学(深圳)高等研究院, 广东 深圳 518110)

摘 要: 列车晚点预测对提升铁路运输效率与调度安全性至关重要。列车运行受多源异构扰动影响, 且晚点传播具有复杂的跨时空特性, 使得未来站点到达晚点时间的高精度预测面临巨大挑战。现有方法或侧重于全局路径上的延误传播趋势建模, 或聚焦于局部因果关系分析, 两者协同不足, 限制了预测性能。为此, 文章提出一种全局-局部协同感知的列车到达晚点预测方法 (GLCP-TADP)。该方法基于编码器-解码器架构构建: 首先, 局部编码器结合卷积神经网络与循环神经网络, 捕获相邻站点间的细粒度传播特征; 其次, 全局编码器引入以预测目标站点为引导的注意力机制, 从历史序列中动态聚合与当前预测相关的时空传播状态, 实现对关键传播路径的选择性感知; 最后, 通过多任务学习优化框架, 将局部延误差值作为辅助监督信号引入, 引导模型增强对局部时空传播关系的一致性学习。在真实铁路数据集上的试验结果表明, GLCP-TADP在平均绝对误差 (MAE)、均方根误差 (RMSE) 等指标上均优于主流基线模型。消融试验进一步验证了所提注意力机制与局部监督策略的有效性。

关键词: 列车晚点预测; 时空预测; 卷积神经网络(CNN); 长短期记忆网络(LSTM); 注意力机制; 多任务学习

中图分类号: U292.4

文献标志码: A

文章编号: 1000-128X(2025)05-0104-09

0 引言

在现代铁路运输系统中, 列车运行准点率不仅直接关系到乘客出行体验和铁路企业服务质量, 而且是铁路调度系统优化的核心目标。尤其在大规模列车交织运行、运能日趋紧张的复杂场景中, 列车一旦发生晚点, 极易产生串联传播效应, 导致线路拥塞, 甚至引发安全隐患。因此, 提升列车运行过程中的到站晚点时间预测能力, 对实现智能调度、优化运营计划和保障铁路系统稳定运行具有重要的现实意义与工程价值^[1]。

列车晚点的诱因呈现高度多样性和复杂性, 常见原因包括突发天气变化、设备故障、调度冲突及客流波动等。这些因素可能单独或联合作用, 影响列车在

运行过程中的实际到站时间。列车晚点预测任务旨在融合历史运行数据与外部环境因素, 实现对列车未来站点到达时间偏差的提前预判^[2]。该任务的核心挑战在于如何合理建模多源异构干扰下列车运行过程的动态演化特性, 尤其是如何刻画其具有非线性、非均匀、跨时空传播规律的延误影响机制, 从而实现高精度的未来到站晚点预测。

为应对上述挑战, 现有研究主要从以下两类技术路径展开: 一类聚焦于系统层面全局路径的延误传播建模, 旨在刻画多段运行轨迹中的长程依赖与整体传播趋势。早期传统方法多采用排队论、马尔可夫链与事件图等解析模型, 通过建模列车间的逻辑依赖关系推导延误传播路径, 如文献[3]利用马尔可夫链对横向连带晚点进行建模, 文献[4]则基于时刻表鲁棒性分析延误演化规律, 此类方法虽具有一定的理论可解释性,

基金项目: 国家重点研发计划专项“交通运载装备与智能交通系统”(2022YFB4300603); 国家自然科学基金企业创新发展联合基金(U22B2061)

通信作者: 刘 峤, 男, 博士, 教授, 博士生导师, 主要从事时间序列方面的研究; E-mail: qliu@uestc.edu.cn

引用格式: 周靖翔, 代婷婷, 黄燕文, 等. 全局-局部协同感知的列车到达晚点预测[J]. 机车电传动, 2025(5): 104-112.DOI:10.13890/j.issn.1000-128X.2025.05.011.

Citation: ZHOU Jingxiang, DAI Tingting, HUANG Yanwen, et al. Train arrival delay prediction based on global-local collaborative perception[J]. Electric drive for locomotives, 2025(5): 104-112.DOI:10.13890/j.issn.1000-128X.2025.05.011.

但通常假设网络结构静态、干扰类型单一, 难以捕捉实际铁路系统中非线性、异构扰动与时变结构的复杂性。随后, 机器学习方法(如随机森林、支持向量机等)则利用大规模历史运行数据挖掘延误行为的非线性映射, 通过集成结构提升模型稳定性, 取得了可观的预测性能^[5-7], 如文献[5]在多场景数据集上验证了随机森林的稳健性, 文献[7]则将随机森林与改进的决策树融合以提升模型性能。但此类方法依赖大量特征工程, 难以有效建模真实铁路拓扑与动态传播路径。近年来, 图神经网络与时空图卷积网络(spatial temporal graph convolutional network, STGCN)被广泛应用于建模列车运行拓扑与时序特征, 有效提升了对全局结构依赖的建模能力, 如文献[8]构建了融合图卷积与时间建模的时空预测结构, 文献[9]结合贝叶斯推理机制提升模型稳定性和迁移能力, 文献[10]进一步引入多源点过程, 增强模型对异步动态扰动的建模能力。这类方法虽具有强大的空间建模能力, 但多依赖静态邻接关系, 难以动态聚焦主导传播路径^[5,8-9]。

另一类方法聚焦于局部传播关系建模, 强调相邻站点间的微观延误传递。逻辑斯蒂回归、泊松分布与Box-Cox正态化等传统统计模型被用于刻画邻近站点间的因果关系与误差分布特性^[11-12]。典型方法还包括延误差值回归、小波神经网络与聚类分析等^[13], 如文献[14]基于小波神经网络与结构向量, 识别了不同类型的连带延误特征; 文献[15]结合聚类方法与贝叶斯网络(Bayesian network, BN)实现延误模式归类与演化预测。深度学习方法也被引入局部传播建模任务, 如文献[16]提出结合迭代学习控制(iterative learning control, ILC)与LSTM的多时滞建模结构, 文献[17]提出序列到序列的注意力结构, 对多步预测进行精细建模。但此类方法多忽视远程传播因子的影响, 缺乏复杂全局扰动条件下传播机制的泛化能力与结构解释性。此外, 文献[18]基于狄利克雷增量过程构建了用于延误类型聚类与异常检测的流式建模机制, 文献[19]提出基于元胞自动机的高负载铁路仿真模型, 考虑了重载条件下的多车延误传导过程。这些方法虽在建模灵活性与实景仿真方面具有优势, 但对连续预测问题的支持有限。

总体来看, 当前研究在建模机制与表示结构方面仍存在明显局限: 全局建模方法多依赖静态邻接结构, 难以有效刻画非均匀、跨尺度的传播路径; 局部建模方法虽具有可解释性, 但缺乏对远程主导因素的捕捉能力; 融合方法多为简单特征拼接或串联结构, 缺乏统一框架下的协同学习机制。

针对上述问题, 本文提出一种全局-局部协同感知的列车晚点预测(global-local collaborative perception for train arrival delay prediction, GLCP-TADP)方法。该模型由4个模块构成: 局部编码器融合卷积神经网络

和长短期记忆网络(long short-term memory, LSTM), 用于提取相邻站点间的时空传播特征; 全局编码器模块构建目标站点驱动的注意力机制, 以目标站点(第 $T+1$ 站)的天气、温度等多变量为查询向量, 以局部编码器的输出结果为键值向量, 显式建模历史传播状态对当前站点的动态非均匀影响; 解码器通过全连接网络输出全局视角下的晚点估计与局部传播估计; 多任务损失优化模块提出局部传播建模机制与差值监督策略, 构建局部传播估计与真实延误差值的监督信号, 引导模型学习局部视角下合理的延误信息传播关系, 进而提升全局视角下列车到达晚点估计的性能表现。

1 研究方法

1.1 任务定义

列车晚点预测任务旨在基于列车途经站点的多维数据, 估计下一站点的到达晚点时长。令 T 为列车途经的站点序列, $T=\{s_1, \dots, s_{|T|}\}$; s_i 为列车途经的第 i 个站点, $|T|$ 为列车已途经的站点个数。对于任意途经的站点 s_i , 均对应经纬度($s_{i, \text{lat}}$ 、 $s_{i, \text{lng}}$)、天气状况(包括天气类型 $s_{i, \text{weather}}$ 、气温 $s_{i, \text{temperature}}$ 、风速等级 $s_{i, \text{wind}}$), 其中天气类型和风速等级均为离散分类变量, 各自的类别总数本文分别记为 V_{weather} 和 V_{wind} 、列车到达该站点的延误时间($s_{i, \text{delay}}$)。 $D_{a \rightarrow b}$ 表示沿行程从站点 s_a 到站点 s_b 的总行驶距离, 其中 $a, b \in [1, |T|+1]$, 且 $a \leq b$ 。

本文的研究目标是设计一种能融合空间距离、环境状态、历史延误等多源异构信息的深度神经网络模型, 实现对下一站点到达延误时长 $\hat{y}_{|T|+1}$ 的精确预测。

1.2 模型概览

本文提出的晚点预测模型如图1所示, 该模型以编码器-解码器为基础架构, 包含局部编码器、全局编码器、解码器、多任务学习优化4个部分。局部编码器聚焦相邻站点的时空特征提取, 通过卷积神经网络和LSTM捕获局部行程的列车延误相关性; 全局编码器整合天气、距离等多源异构数据, 捕获历史序列中与目标站点相关的传播信息; 解码器利用多层感知机(multilayer perceptron, MLP)分别将全局和局部视角下的传播信息映射为晚点时长估计和晚点时长差值估计, 多任务学习优化在训练阶段联合局部和全局损失优化模型, 测试阶段输出未来站点的晚点时长估计。

1.3 局部编码器

局部编码器用于捕获列车历史运行数据中晚点信息传播的时空关联, 主要由两部分组成: 第一部分是基于卷积神经网络的空间关联, 将列车在历史站点的信息转换为特征图, 以捕获连续站点间的局部空间相

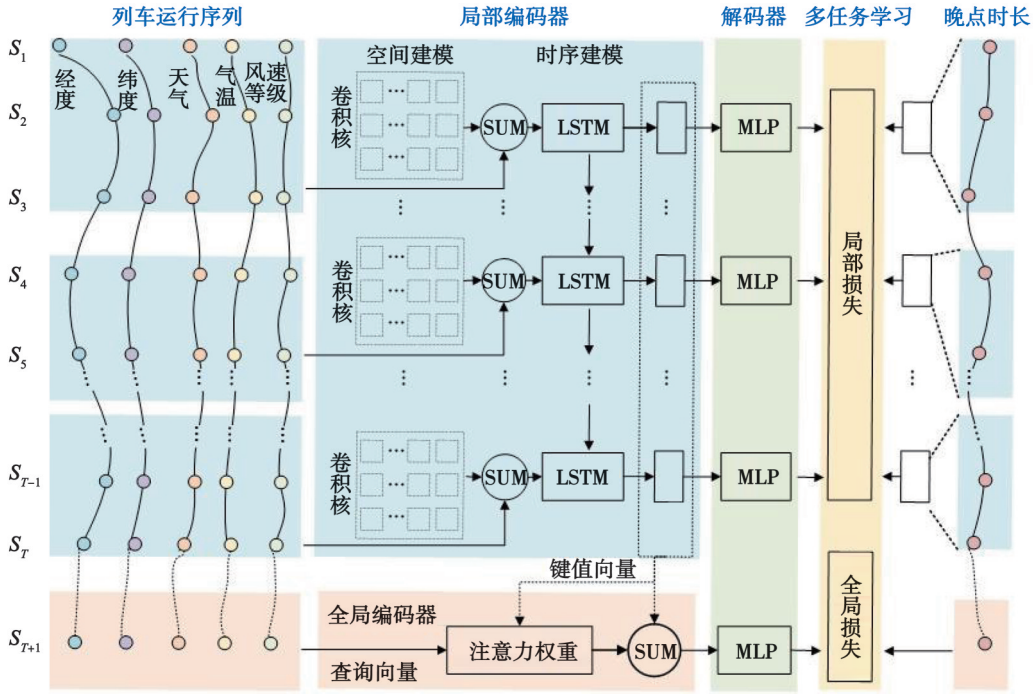


图1 模型框架

Fig. 1 Model framework

关性；第二部分是基于LSTM的时序关联，用于学习列车在站点间关于晚点的时间相关性。

1.3.1 基于卷积神经网络的空间关联

在晚点预测任务中，仅依赖经纬度坐标难以全面捕捉影响列车晚点的复杂因素，因此本文设计了基于多特征融合的卷积模块，旨在从原始轨迹数据中提取丰富的局部空间特征。该模块的计算思路如下：

①将每个站点的地理信息与环境信息（如天气、风力、气温）进行融合，形成高维特征表示。需要注意，由于天气类型和风力等级为分类的离散值，本文先利用嵌入层将离散类别特征映射为低维稠密向量，表达式为：

$$e_{s_i, \text{weather}} = E_{\text{weather}}(s_i, \text{weather}) \quad (1)$$

$$e_{s_i, \text{wind}} = E_{\text{wind}}(s_i, \text{wind}) \quad (2)$$

$$x_i = [s_i, \text{lng}, s_i, \text{lat}, s_i, \text{temperature}] \oplus e_{s_i, \text{weather}} \oplus e_{s_i, \text{wind}} \quad (3)$$

式中： E_{weather} 和 E_{wind} 为可学习的嵌入矩阵； $e_{s_i, \text{weather}} \in R^{V_{\text{weather}} \times d}$ ， $e_{s_i, \text{wind}} \in R^{V_{\text{wind}} \times d}$ ； d 为嵌入维度； $\oplus$ 为向量拼接操作； x_i 为包含地理位置和环境状况的向量。

为增强列车在每个站点处信息表达的非线性能力，本文将 x_i 通过一个全连接层进行非线性转换，即 $\text{loc}_i = \tanh(W_{\text{loc}} \cdot x_i + b_{\text{loc}})$ ， $\text{loc} \in R^{16 \times |T|}$ ，其中 W_{loc} 和 b_{loc} 为可学习的权重矩阵和偏置项。

②本文采用一维卷积网络提取连续站点间的空间关联，将特征序列 loc 视为16通道的一维信号，并使用大小为 k 的卷积核对其进行卷积操作，表达式为

$$\text{loc}_i^{\text{conv}} = \sigma_{\text{cnn}}(W_{\text{conv}} * \text{loc}_{i:i+k-1} + b) \quad (4)$$

式中： $*$ 为卷积操作； W_{conv} 为该卷积层中一组可学习

的卷积核，大小为 $R^{c \times 16 \times k}$ ； c 为滤波器数量； b 为偏置项； $\text{loc}_{i:i+k-1}$ 为 loc 抽取列车特征在第 i 个站点到第 $i+k-1$ 个站点的子序列（又称局部行程）； σ_{cnn} 为激活函数。

从上述分析可知， $\text{loc}_i^{\text{conv}}$ 本质上捕获了第 i 个局部行程的空间特征，通过连接 c 个滤波器的输出，得到大小为 $R^{c \times (|T| - k + 1)}$ 的局部行程表示。

③考虑到列车晚点时间与行程距离的强相关性，本文将归一化后的局部行程实际行驶距离 $D_{i \rightarrow i+k}$ 作为额外特征，与卷积后得到的特征图 $\text{loc}_i^{\text{conv}}$ 进行拼接，即

$$\text{loc}_i^f = \text{loc}_i^{\text{conv}} \oplus D_{i \rightarrow i+k} \quad (5)$$

1.3.2 基于LSTM的时序关联

空间表示学习模块成功提取了列车运行数据中的局部空间特征，但卷积神经网络本质上为前馈网络，缺乏对序列中长期依赖关系的建模能力。因此，本文引入基于LSTM的时序学习模块，以捕捉列车运行数据随时间演变的动态模式，学习空间表示模块输出特征序列 loc^f 中的时间关联性。

该模块的输入为上一模块输出的局部特征图 loc^f ，在每个时间步 t ，LSTM通过遗忘门、输入门、候选状态、单元状态更新及输出门等核心组件的协同计算，实现信息在序列中的选择性传递。本文采用双层堆叠的LSTM结构，第一层的输出隐藏状态序列 $H^{(0)} = \{h_1^{(0)}, \dots, h_{|T|-k+1}^{(0)}\}$ 作为第二层LSTM的输入，第二层在此基础上进行更高层次的时序抽象，捕捉更复杂的长期依赖关系，从而得到第二层LSTM输出的最终隐藏状态序列 $\{h_1, h_2, \dots, h_{|T|-k+1}\}$ ，其中 h_i 为处理完第 i 个局部信息的隐状态。

1.4 全局编码器

在局部编码器捕捉列车历史运行数据的时空传播模式后, 本文引入以注意力机制为核心的全局编码器。该模块的核心任务是根据待测站点信息, 差异化聚合历史信息, 生成与未来站点晚点时长相关的延误传播信息, 该模块由查询向量构建和注意力机制两部分组成。

1.4.1 查询向量构建

注意力机制的有效性在很大程度上取决于查询向量的质量, 一个优质的查询向量应能精确概括与预测站点最相关的上下文信息。在本文模型中, 查询向量 q 由与预测任务直接相关的两类信息构成, 即预测站点的固有信息和历史累积信息, 计算公式为

$$q = e_{(|T|+1)_{\text{weather}}} \oplus e_{(|T|+1)_{\text{wind}}} \oplus \text{norm}((|T|+1)_{\text{ing}}, (|T|+1)_{\text{lat}}, (|T|+1)_{\text{temperature}}, D_{|T| \rightarrow |T|+1}, D_{1 \rightarrow |T|+1}) \quad (6)$$

式中: $\text{norm}()$ 为特征归一化处理; $D_{|T| \rightarrow |T|+1}$ 为预测站点 $s_{|T|+1}$ 与上一站点 $s_{|T|}$ 间的距离; $D_{1 \rightarrow |T|+1}$ 为与 $s_{|T|+1}$ 首个站点间的距离; $s_{|T|+1}$ 的固有状态包括该站点的天气类型、经度、纬度; 历史累积状态包括 $D_{|T| \rightarrow |T|+1}$ 与 $D_{1 \rightarrow |T|+1}$ 。

1.4.2 注意力机制

生成查询向量 q 后, 本文首先将局部编码器输出的局部状态序列 $\{h_1, h_2, \dots, h_{|T|-k+1}\}$ 定义为键向量和值向量, 因此以预测目标站点为引导的注意力机制计算包含相关性计算与信息聚合两部分, 具体公式为

$$e_i = \frac{(W_q q) \cdot (W_k h_i)^T}{\sqrt{d_k}} \quad (7)$$

式中: e_i 为每一时刻 h_i 对于待预测站点的时空传播信息相似度得分, 该值越高代表历史位置 i 的信息与预测目标的相关性越强; W_q 和 W_k 为可学习的线性变换矩阵, 用于将 q 和 h_i 投影至相同的 d_k 维空间; $\sqrt{d_k}$ 为缩放因子。

对 e_i 进行归一化操作, 即通过一个softmax函数, 将相关性向量转换为总和为1的概率分布, 注意力权重 α_i 表达式为

$$\alpha_i = \frac{\exp(e_i)}{\sum_{j=1}^{|T|-k+1} \exp(e_j)} \quad (8)$$

最后, 将得到的注意力权重 α 应用于历史状态序列(即 $\{h_1, h_2, \dots, h_{|T|-k+1}\}$), 经加权求和得到一个与待预测站点相关的上下文向量 c_{att} , 计算公式如下:

$$c_{\text{att}} = \sum_{i=1}^{|T|-k+1} (\alpha_i h_i) \quad (9)$$

1.5 解码器

为实现对下一站点晚点时长的精准估计, 本文采

用MLP作为解码器, 分别对编码器学习到的全局表示和局部表示进行解码, 得到目标站点的晚点时长预测值和局部片段对目标站点晚点的影响程度。

1.5.1 全局解码器

利用注意力机制获得的全局上下文向量 c_{att} , 通过一个两层的MLP进行解码, 得到对下一站点晚点时长的估计值 $\hat{y}_{|T|+1}$, 如式(10)。

$$\hat{y}_{|T|+1} = \sigma(W_2 \cdot \text{ReLU}(W_1 c_{\text{att}} + b_1) + b_2) \quad (10)$$

式中: W_1 和 W_2 为全连接层的权重矩阵; b_1 和 b_2 为偏置项; σ 为输出激活函数; ReLU 为激活函数。

该结构通过残差连接避免信息遗忘, 并提升了模型的非线性表达能力。

1.5.2 局部解码器

对于每个局部行程, 本文同样采用MLP对其隐藏状态 h_i 进行解码, 用于估计局部片段中窗口右端点与左端点的时间差, 衡量局部延误传播影响。具体计算公式为

$$\Delta \hat{y}_i = W_{\text{loc}, 2} \cdot \text{ReLU}(W_{\text{loc}, 1} h_i + b_{\text{loc}, 1}) + b_{\text{loc}, 2} \quad (11)$$

式中: $\Delta \hat{y}_i$ 为第 i 个局部站点间的晚点传播估计值; $W_{\text{loc}, 1}$ 、 $W_{\text{loc}, 2}$ 为局部解码器的权重矩阵; $b_{\text{loc}, 1}$ 和 $b_{\text{loc}, 2}$ 为偏置项。

1.6 多任务优化

为提升全局视角中局部晚点传播影响的一致性, 本文采用多任务学习优化框架, 将局部延误差值引入作为辅助监督信号, 引导模型增强对局部时空传播关系的对齐。具体而言, 全局损失 L_{en} 和局部损失 L_{local} 分别定义如下:

$$L_{\text{en}} = \frac{|\hat{y}_{|T|+1} - y_{|T|+1}|}{y_{|T|+1}} \quad (12)$$

$$L_{\text{local}} = \frac{1}{|T|-l+1} \sum_{i=1}^{|T|-l+1} \left| \frac{\Delta \hat{y}_i - \Delta y_i}{\Delta y_i} \right| \quad (13)$$

$$\Delta y_i = y_{i+l-1} - y_i \quad (14)$$

式中: y_{i+l-1} 和 y_i 分别为窗口右端点和左端点的实际延误时长, 二者差值为局部损失的真实标签; l 为局部视角中滑动窗口的长度。

最后, 利用局部损失项和全局损失项的加权和 L 来优化模型, 如式(15)。

$$L = \beta \cdot L_{\text{local}} + (1 - \beta) \cdot L_{\text{en}} \quad (15)$$

式中: β 为组合系数, 用于表征模型优化过程中 L_{local} 、 L_{en} 的占比。

2 试验

2.1 数据集描述

本文使用中国铁路客票系统的高速列车实际运行数据, 该数据涵盖727个火车站2019年10月8日至

2020 年 1 月 27 日的列车运行记录^[20]，包括列车运行日期、车次、车站名称、车站编号、预计到达时间和出发时间、实际到达时间和出发时间、停靠时间、是否到达/出发晚点等信息，同时将天气、温度、风速等级作为影响列车运行的因素。

为从空间、时间、环境维度探索它们对列车到达晚点的影响，本文对原始数据进行多次预处理，其中空间特征通过站点名称与经纬度映射表为每条记录补充经度和纬度信息，进而获得相邻站点间的地理距离及累计行驶距离；时间特征提取计划到达与实际到达时间，统一转换为分钟数后得到每一站的到达延误时长，并剔除晚点时长超出 $[-200, 200]\text{min}$ 的运行数据；环境特征包括天气、温度、风速等级等，所有类别型特征均通过编码映射为数值型输入。此外，按照时间顺序将列车运行数据集划分为训练集、验证集和测试集，其中 70% 用于训练，15% 用于验证，剩余 15% 用于测试。数据集的统计特征、列车到达站点晚点分布如表 1 和表 2 所示。

表 1 数据集的基本统计描述
Table 1 Basic statistical description of the dataset

特征	列车运营 线路数	列车 数	天气 类别	风速 等级	列车运营 平均站点数	列车平均 运营公里数
训练集 数量	110 687	1 964	24	42	9.24	865.47
验证集 数量	27 653	1 849	24	42	9.28	865.92
测试集 数量	27 688	1 862	24	42	9.23	861.81

表 2 数据集中关于列车到达晚点的时间分布统计
Table 2 Statistics on temporal distribution of train arrival
delays in the dataset

晚点时长/ min	数量	占比/%	平均晚点时 长/min	平均早到 时长/min
[0,4]	264 649	17.24	3.23	1.94
[5,9]	33 713	2.20		
[10,29]	13 526	0.88		
[30,59]	1 806	0.12		
≥ 60	293	0.02		
不晚点	1 220 930	79.54		

2.2 评估指标

与列车晚点预测任务的常用评价指标一致，本文采用平均绝对误差（mean absolute error, MAE）、均方根误差（root mean squared error, RMSE）、均方误差（mean squared error, MSE）衡量列车晚点预测值与实际值之间的差异。相关公式为：

$$A_{MAE} = \frac{1}{n} \sum_{i=1}^n |x_i - \hat{x}_i| \tag{16}$$

$$A_{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \hat{x}_i)^2} \tag{17}$$

$$A_{MSE} = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{x}_i)^2 \tag{18}$$

上述公式中， A_{MAE} 衡量晚点时长估计值与真实值间平均绝对偏差，反映模型偏差的平均水平； A_{RMSE} 以标准差形式体现估计值与真实值的波动性与稳定性； A_{MSE} 代表各预测点波动强度的平均值。三者数值越小，表明模型预测性能越优。

2.3 试验环境设置

本文试验环境为 NVIDIA GeForce RTX 1650 GPU、PyTorch 1.8.0。模型训练阶段，优化器选用 Adam，初始学习率设为 0.001，批量大小为 1 600，最大训练轮数为 100；训练过程中采用早停策略以防止过拟合，选取验证集上性能最优的模型参数用于最终测试评估。为保障对比公平性，所有试验均在相同数据划分和参数设置下开展。

2.4 基准模型描述

为验证所提 GLCP-TADP 模型的有效性，采用以下基线模型进行对比：

- ①BN：刻画变量间条件依赖关系的概率图模型，仅利用轨迹统计特征进行晚点预测，未对时序或空间依赖进行建模。
- ②深度置信网络（deep belief network, DBN）：由多层受限玻尔兹曼机（restricted Boltzmann machine, RBM）堆叠而成的深度生成模型。
- ③MLP：由多层全连接神经元组成的前馈神经网络。
- ④LSTM：用于序列建模的循环神经网络变体，可学习轨迹数据中的长期时序依赖。
- ⑤CLF-Net^[21]：联合 LSTM 和卷积神经网络，旨在同时捕获列车运行数据的全局与局部依赖特征。
- ⑥BN-STGCN^[9]：融合动态 BN 和 STGCN 的深度学习模型，通过 BN 学习站点间的时变因果依赖关系，利用图卷积网络（graph convolutional network, GCN）和门控循环单元（gated recurrent unit, GRU）分别建模空间相关性和时序依赖性，并采用通道注意力机制进行特征加权，实现列车晚点预测。

2.5 总体性能对比与分析

为评估 GLCP-TADP 的有效性，本文开展了多个基线模型在实际列车运行数据集上的对比试验，并采用 MAE、RMSE、MSE 指标进行衡量。GLCP-TADP 与基线模型在数据集上的性能对比如表 3 所示。试验结果表明，所提 GLCP-TADP 模型在 3 个指标上均取得最优表现。与 DBN、MLP 等神经网络方法相比，BN 在所

有指标的表现均显著落后,这主要是由于传统方法仅能利用列车历史运行序列的统计特征,无法有效捕获时空依赖与复杂的非线性关系。具体而言,LSTM通过门控机制捕捉历史晚点的长期依赖,相较于BN模型,其MAE降至1.25, RMSE为2.90,验证了门控机制对长期依赖关系的捕捉能力。此外,MLP相较于LSTM性能表现更优,推测原因是LSTM存在长距离遗忘问题,而MLP的全连接结构可改善这一不足。CLF-Net的MAE优于LSTM、MLP,但RMSE和MSE表现出不同性能,这表明时空关联与非线性关联均有助于提升晚点时长估计性能,验证了本文所提方法的必要性。

与 LSTM、MLP、CLF-Net、BN-STGCN 相比, GLCP-TADP在3项指标上均表现最佳。例如,相较于LSTM, GLCP-TADP的MAE、RMSE、MSE分别降低28.8%、19.3%、33.1%;相较于时空图方法BN-STGCN,3项指标分别降低约6.3%、8.6%、10.9%。这主要得益于两方面设计:一方面, GLCP-TADP构建了全局引导的注意力机制,用于建模列车运行路径中历史传播状态对目标站点延误的动态影响,可捕捉非线性、跨时空的传播结构;另一方面, GLCP-TADP引入了局部差值监督机制,引导模型显式学习相邻站点间的晚点传播规律,增强局部建模的一致性。

表3 GLCP-TADP与基线模型在数据集上的性能对比
Table 3 Performance comparison of GLCP-TADP and baseline models on the dataset

模型	A_{MAE}/min	A_{RMSE}/min	A_{MSE}/min^2
BN	11.56	12.34	241.11
DBN	4.84	5.87	45.09
MLP	1.21	2.75	8.92
LSTM	1.25	2.90	10.48
CLF-Net	1.19	2.82	9.50
BN-STGCN	0.95	2.56	7.87
GLCP-TADP	0.89	2.34	7.01

2.6 消融试验与分析

2.6.1 注意力机制的有效性验证

为验证GLCP-TADP模型中注意力机制的有效性,本文开展了消融试验,移除原模型中的全局注意力模块(记作GLCP-TADP-NoAttention)。该模型在全局编码阶段不采用自注意力机制加权融合历史轨迹状态,直接以局部编码器最后一个时刻的隐藏状态作为全局表示,其他部分与原始模型保持一致。

GLCP-TADP中注意力机制的有效性对比如表4所示。移除注意力机制后,完整模型(GLCP-TADP)的预测误差显著降低, A_{MAE} 、 A_{RMSE} 、 A_{MSE} 分别下降22.41%、16.07%、26.21%。性能差异的原因主要有两点:一是显式建模目标站点的信息表示,为识别不同站点的晚点信息相关性奠定基础;二是注意力机制能从历史运

行数据中捕获与目标站点延误传播相关的信息,两者综合实现对包含时空关联的历史信息的差异性聚合。该试验表明,所设计的注意力机制通过建模列车历史晚点数据对未来站点晚点的时空传播影响,有效提升预测性能,验证了其引入的必要性。

表4 GLCP-TADP中注意力机制的有效性对比
Table 4 Effectiveness comparison of attention mechanisms in GLCP-TADP

模型	A_{MAE}/min	A_{RMSE}/min	A_{MSE}/min^2
GLCP-TADP-NoAttention	1.16	2.80	9.50
GLCP-TADP	0.90	2.35	7.01
误差降低率	22.41%	16.07%	26.21%

此外,为进一步验证注意力机制的有效性,本文还对设计试验探究查询向量(q)的不同表征方式对模型性能的影响。其中, GLCP-TADP-Distance-Free-Query-Vector模型从GLCP-TADP原始查询向量中移除历史运行数据与目标站点间的空间信息(目标站点与最近站点、首个站点的区间距离),仅保留目标站点固有信息,不考虑历史站点对目标站点的影响。

不同查询向量对模型性能的影响如图2所示,可见去除距离特征后,3项指标均出现不同程度退化, A_{MAE} 由0.90上升至0.94, A_{RMSE} 由2.35上升至2.59, A_{MSE} 由7.01上升至8.36。这表明相对距离特征对提升模型预测精度具有重要作用,原因可能是距离特征能有效反映预测站点在历史轨迹中的空间关系。将距离信息融入 q ,有助于模型理解列车所处的物理运行阶段,为延误传播趋势和空间级联效应的建模能力提供基础。当 q 中缺失距离特征时,模型难以准确捕捉全局行程进度和历史累计状态,因此在复杂动态场景下的预测误差显著增加。因此,试验结果进一步验证了 q 中融合距离特征的必要性与有效性。

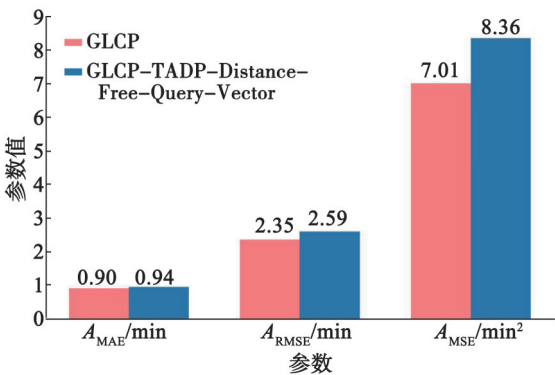
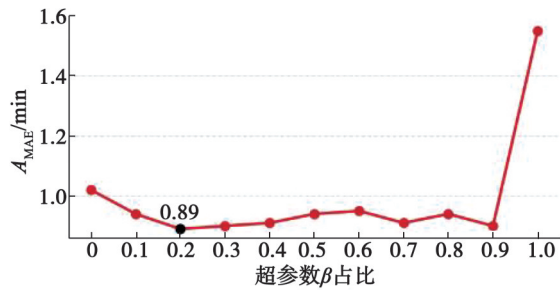


图2 不同查询向量对模型性能的影响
Fig. 2 Effects of different query vectors on model performance

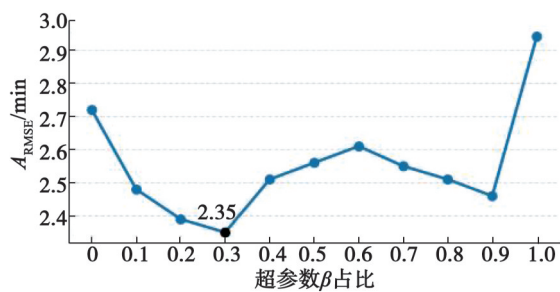
2.6.2 局部监督任务的有效性分析

为探究局部视角辅助监督任务的作用,本文开展了GLCP-TADP模型超参数 β 的对比试验。通过调节 β ,可平衡模型对局部与全局预测任务的关注程度,其中 β

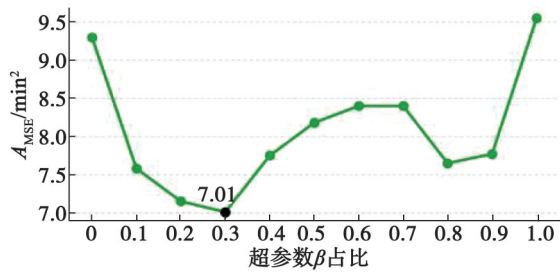
$\beta=1$ 代表仅包含局部晚点影响程度预测损失, $\beta=0$ 代表仅包含全局延误预测损失。图3展示了 β 从0到1变化时, 模型在 A_{MAE} 、 A_{RMSE} 和 A_{MSE} 3项指标上的性能表现。



(a) 不同局部损失占比对模型在 A_{MAE} 指标上的影响



(b) 不同局部损失占比对模型在 A_{RMSE} 指标上的影响



(c) 不同局部损失占比对模型在 A_{MSE} 指标上的影响

图3 不同局部损失占比对模型性能的影响

Fig. 3 Effects of different local loss ratios on model performance

由图3可知, 仅使用全局损失 ($\beta=0$) 时, 模型倾向于优化全行程的平均延误, 忽略站点间局部延误传播影响, 导致局部晚点时长估计能力不足; 仅使用局部损失 ($\beta=1$) 时, 模型虽能捕捉站点级局部细节, 但缺乏对全局时空依赖的整体把控。

当 β 取0或1时, 模型性能均显著下降, A_{MAE} 分别为1.022 4和1.530 4, A_{MSE} 分别为9.30和9.55, 均远高于联合优化的最优结果; 随着 β 从0增至0.3, 模型性能持续提升, 表明适当增加局部损失权重有助于提升站点级延误的刻画能力; 但当 $\beta>0.3$ 时, 模型性能反而下降, 说明过度关注局部损失会削弱全局行程延误的建模能力, 导致整体预测精度下降。上述结果表明, 适当权衡两者权重 (如 $\beta=0.3$), 可使模型既精细刻画局部关联, 又保持全局关联, 进而实现最优的列车到

达晚点时长估计。

3 结束语

本文针对列车到达晚点预测任务, 提出融合全局延误趋势与局部传播细节的协同感知建模框架 GLCP-TADP, 解决了现有方法在模型结构与延误建模机制上的协同不足问题, 提升了预测性能。该方法构建了由卷积神经网络与 LSTM 联合组成的局部编码器, 用于刻画邻近站点间的细粒度传播模式; 设计了目标站点特征引导的注意力全局编码器, 实现历史延误路径中关键传播信息的动态聚合; 同时, 在解码阶段引入多任务优化机制, 通过延误差值监督策略强化局部传播关系的一致性建模。真实列车运行数据集上的实证结果表明, GLCP-TADP 的预测精度优于现有基准方法, 且有效验证了全局注意力机制与局部差值监督的建模作用。然而, 当前框架尚未充分建模列车间的动态耦合关系, 也未显式引入站点停靠时长、出发延误等影响调度的关键因素。未来研究将进一步探索基于多列车图结构的传播建模方法, 融合更丰富的调度上下文与动态属性信息, 提升模型在复杂运行环境下的适应能力与推理能力, 为智慧铁路系统中的实时调度优化与风险预警等关键场景提供技术支持。

参考文献:

- [1] 余祖俊, 唐涛, 李开成, 等. 智慧高速铁路运行控制系统发展趋势综述[J]. 铁道学报, 2024, 46(1): 1-12.
YU Zujun, TANG Tao, LI Kaicheng, et al. Overview of development trends of intelligent high-speed rail train control systems[J]. Journal of the China railway society, 2024, 46(1): 1-12.
- [2] KREUZER T, PAPAPETROU P, ZDRAVKOVIC J. Artificial intelligence in digital twins: a systematic literature review[J]. Data & knowledge engineering, 2024, 151: 102304.
- [3] 文超, 李忠灿, 黄平, 等. 数据驱动的列车晚点传播研究[J]. 中国安全科学学报, 2019, 29(增刊 2): 1-9.
WEN Chao, LI Zhongcan, HUANG Ping, et al. Progress and perspective of data-driven train delay propagation[J]. China safety science journal, 2019, 29(Suppl 2): 1-9.
- [4] ARTAN M Ş, ŞAHİN İ. Exploring patterns of train delay evolution and timetable robustness[J]. IEEE transactions on intelligent transportation systems, 2022, 23(8): 11205-11214.
- [5] WU J Q, WANG Y H, DU B, et al. The bounds of improvements toward real-time forecast of multi-scenario train delays[J]. IEEE transactions on intelligent transportation systems, 2022, 23(3): 2445-2456.
- [6] 廖璐, 张亚东, 葛晓程, 等. 基于GBDT的列车晚点时长预测模型研究[J]. 铁道标准设计, 2021, 65(8): 149-154.

- LIAO Lu, ZHANG Yadong, GE Xiaocheng, et al. Research on prediction model of train delay time based on GBDT[J]. Railway standard design, 2021, 65(8): 149-154.
- [7] 刘鲁岳, 肖宝弟, 岳丽丽. 基于改进RF-XGBoost算法的列车运行晚点预测研究[J]. 铁道标准设计, 2023, 67(3): 38-43.
- LIU Luyue, XIAO Baodi, YUE Lili. Research on train delay prediction based on improved RF-XGBoost algorithm[J]. Railway standard design, 2023, 67(3): 38-43.
- [8] ZHANG D L, PENG Y J, ZHANG Y M, et al. Train time delay prediction for high-speed train dispatching based on spatio-temporal graph convolutional network[J]. IEEE transactions on intelligent transportation systems, 2022, 23(3): 2434-2444.
- [9] LI J M, XU X Y, DING X, et al. Bayesian spatio-temporal graph convolutional network for railway train delay prediction[J]. IEEE transactions on intelligent transportation systems, 2024, 25(7): 8193-8208.
- [10] ZHANG D L, DU C Y, PENG Y J, et al. A multi-source dynamic temporal point process model for train delay prediction[J]. IEEE transactions on intelligent transportation systems, 2024, 25(11): 17865-17877.
- [11] 石晶, 彭其渊, 文超, 等. 铁路列车晚点时间预测方法研究[J]. 铁道运输与经济, 2020, 42(7): 33-39.
- SHI Jing, PENG Qiyuan, WEN Chao, et al. Prediction method of passenger train delay propagation[J]. Railway transport and economy, 2020, 42(7): 33-39.
- [12] 唐涛, 甘靖. 基于国内外铁路运营数据的列车运行时间预测模型[J]. 中国安全科学学报, 2022, 32(6): 123-130.
- TANG Tao, GAN Jing. A train running time prediction model based on domestic and foreign railway operation data[J]. China safety science journal, 2022, 32(6): 123-130.
- [13] GHAMDI M A, PARR G, WANG W J. Heterogeneous machine learning ensembles for predicting train delays[J]. IEEE transactions on intelligent transportation systems, 2024, 25(6): 5138-5153.
- [14] 张琦, 陈峰, 张涛, 等. 高速铁路列车连带晚点的智能预测及特征识别[J]. 自动化学报, 2019, 45(12): 2251-2259.
- ZHANG Qi, CHEN Feng, ZHANG Tao, et al. Intelligent prediction and characteristic recognition for joint delay of high speed railway trains[J]. Acta automatica sinica, 2019, 45(12): 2251-2259.
- [15] HUANG P, SPANNINGER T, CORMAN F. Enhancing the understanding of train delays with delay evolution pattern discovery: a clustering and Bayesian network approach[J]. IEEE transactions on intelligent transportation systems, 2022, 23(9): 15367-15381.
- [16] CHEN Y, HUANG D Q, XU C, et al. Iterative learning tracking control of high-speed trains with nonlinearly parameterized uncertainties and multiple time-varying delays[J]. IEEE transactions on intelligent transportation systems, 2022, 23(11): 20476-20488.
- [17] XU N, CHEN C Y, ZHANG Y, et al. A sequence-to-sequence car-following model for addressing driver reaction delay and cumulative error in multi-step prediction[J]. IEEE transactions on intelligent transportation systems, 2024, 25(9): 12203-12215.
- [18] SHEN P J, SONG L Y, WANG B. Dynamic clustering and anomaly detection of train delays in stream data: an incremental dirichlet process approach[J]. IEEE transactions on intelligent transportation systems, 2025, 26(8): 12162-12182.
- [19] 彭浩宇, 何世伟, 蔡近近, 等. 考虑晚点传播的重载铁路列车运行仿真与调整研究[J]. 铁道科学与工程学报, 2024, 21(11): 4431-4441.
- PENG Haoyu, HE Shiwei, CAI Jinjin, et al. Simulation and timetable rescheduling of heavy-haul railway train operation considering delay propagation[J]. Journal of railway science and engineering, 2024, 21(11): 4431-4441.
- [20] ZHANG D L, PENG Y J, XU Y, et al. A high-speed railway network dataset from train operation records and weather data[J]. Scientific data, 2022, 9(1): 244.
- [21] HUANG P, WEN C, FU L P, et al. A deep learning approach for multi-attribute data: a study of train delay prediction in railway systems[J]. Information sciences, 2020, 516: 234-253.

Train arrival delay prediction based on global-local collaborative perception

ZHOU Jingxiang¹, DAI Tingting¹, HUANG Yanwen², LIU Qiao¹

(1. School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China; 2. Shenzhen Institute for Advanced Study, University of Electronic Science and Technology of China, Shenzhen, Guangdong 518110, China)

Abstract: Train arrival delay prediction, as a key capability of intelligent railway systems, is crucial for ensuring transport efficiency and safety. As train operation is influenced by multi-source heterogeneous disturbances, arrival delays propagate, exhibiting complex cross-temporal and spatial dependencies, which makes high-precision prediction of arrival delays at downstream stations significantly challenging. Existing methods model either global delay propagation trends along paths or local causal relationships. The lack of global-local synergy limits prediction performance of both groups. This paper proposes a global-local collaborative perception method for train arrival delay prediction (GLCP-TADP) based on an encoder-decoder architecture. The local encoder that combines a convolutional neural network with a recurrent neural network captures fine-grained propagation features between adjacent stations. The global encoder introduces an attention mechanism guided by the target prediction station to dynamically aggregate spatiotemporal propagation states related to the current prediction from historical sequences, achieving selective perception of key propagation paths. Within a multi-task learning optimization framework, local delay errors are used as auxiliary supervision signals to guide model learning and enhance consistency with local spatiotemporal propagation relationships. Experimental results on datasets from operational railways show that GLCP-TADP outperforms mainstream baseline models on indicators including mean absolute error (MAE) and root mean square error (RMSE). Ablation experiments further verifies the effectiveness of the attention mechanism and the local supervision strategy in improving the performance of train arrival delay estimation.

Keywords: train arrival delay prediction; temporal-spatial prediction; convolutional neural network (CNN); long short-term memory (LSTM); attention mechanism; multi-task learning