

基于组合深度学习方法的地铁进站客流预测

李昌隆¹, 杨晓霞¹, 张国庆¹, 崔春月², 万佳慧¹

(1. 青岛理工大学信息与工程学院, 山东青岛 266520;

2. 青岛理工大学土木工程学院, 山东青岛 266520)

摘要: 针对地铁车站进站客流量时序预测低的问题, 提出一种新型组合深度学习方法 AFT-K-GRU⁺。基于地铁进站客流多尺度时间依赖性特征, 构建融合无注意力变压器模型(attention-free transformer, AFT)、门控循环单元(gated recurrent unit, GRU)和科尔莫戈洛夫-阿诺德网络(Kdmogorov-Arnold networks, KAN)网络 3 种核心组件的模型。AFT 模块专门用于捕捉客流数据的长期周期性规律, GRU 重点提取客流短期波动特征, KAN 网络用于增强模型表达能力。为增强模型预测结果的鲁棒性, 设计基于残差反馈的知识驱动修正控制器, 能够对输出结果进行在线自校正。通过北京地铁站实际刷卡数据验证, AFT-K-GRU⁺模型在进站客流预测方面表现较好, 相较于 Transformer 与长短期记忆网络(long short-term memory, LSTM)方法, 平均误差分别减少 25.3%和 18.7%。该模型可为地铁运营方提供有效的方法支撑, 帮助管理部门进行客流管理与日常规划, 提升地铁线路的运行效率与服务水平。

关键词: 城市轨道交通; 客流预测; 无注意力变压器模型 AFT; 科尔莫戈洛夫-阿诺德网络 KAN; 门控循环单元; 知识驱动修正

中图分类号: U231

文献标志码: A

文章编号: 1672-6073(2026)02-0046-11

Subway Station Entry Passenger Flow Prediction Based on a Hybrid Deep Learning Method

LI Changlong¹, YANG Xiaoxia¹, ZHANG Guoqing¹, CUI Chunyue², WAN Jiahui¹

(1. School of Information and Control Engineering, Qingdao University of Technology, Qingdao, Shandong 266520;

2. School of Civil Engineering, Qingdao University of Technology, Qingdao, Shandong 266520)

Abstract: To address the critical issue of predicting the time-series passenger flow at subway stations, a novel hybrid deep learning method, AFT-K-GRU⁺, is proposed to provide more accurate decision support for subway operation and management. Leveraging the multi-scale temporal dependencies inherent in subway passenger flow data, the model innovatively integrates three core components: the attention-free transformer (AFT), the gated recurrent unit (GRU), and the KAN network. Specifically, the AFT module captures long-term periodic patterns in passenger flow data, the GRU module focuses on extracting short-term fluctuation characteristics, and the KAN network enhances the model's expressive capacity. To enhance the robustness of the model's predictions, this study innovatively designs a knowledge-driven correction controller based on residual feedback, which enables online self-correction of the output. Experimental validation using real-world card swipe data from Beijing subway

收稿日期: 2025-06-06 修回日期: 2025-12-25

第一作者: 李昌隆, 男, 硕士研究生, 从事智能交通研究, lichanglong@stu.qut.edu.cn

通信作者: 杨晓霞, 女, 博士, 教授, 从事智能感知、协同控制、深度学习、强化学习在智慧轨道交通中的应用, yangxiaoxiaaza@163.com

基金项目: 国家自然科学基金(62373209); 山东省泰山学者青年专家项目(tsqn202507218); 山东省高等学校青创团队项目(2023KJ119)

引用格式: 李昌隆, 杨晓霞, 张国庆, 等. 基于组合深度学习方法的地铁进站客流预测[J]. 都市轨道交通, 2026, 39(2): 46-56.

Li Changlong, Yang Xiaoxia, Zhang Guoqing, et al. Subway station entry passenger flow prediction based on a hybrid deep learning method[J]. Urban rapid rail transit, 2026, 39(2): 46-56.

stations demonstrates that the AFT-K-GRU⁺ model effectively predicts station passenger flow, achieving mean error reductions of 25.3% and 18.7% compared to Transformer and LSTM methods, respectively. This model offers effective methodological support for subway operators, assisting management departments in passenger flow control and daily planning, thereby enhancing the operational efficiency and service quality of subway lines.

Keywords: urban rail transit; passenger flow forecasting; attention-free transformer; KAN network; gated recurrent unit; knowledge-driven correction

随着城市轨道交通的迅速发展以及城市人口的持续增长,传统地面交通已难以满足居民日益增长的出行需求。地铁作为一种高效便捷的公共交通方式,已逐渐成为城市居民出行的主要选择。然而,地铁在承载大规模乘客的同时,也伴随着诸多安全隐患。特别是在空间有限的地下站内环境中,乘客数量的快速增加极易导致拥堵甚至踩踏等安全事故^[1]。因此,开展地铁进站客流预测研究对于优化运营调度、提高运力配置效率、预警潜在风险事件具有重要现实意义^[2-4]。

目前,地铁客流预测方法主要分为基于传统统计学方法与基于机器学习和深度学习方法 2 大类。传统方法如自回归滑动平均模型(auto-regression and moving average, ARMA)^[5]、向量自回归模型(vector autoregressive, VAR)^[6]等在早期应用中取得了一定成效。然而,在面对地铁客流数据所呈现出的非线性、复杂性与高维特征时,这些模型常常面临建模能力有限、预测精度不高以及适应性较差等问题。

近年来,机器学习与深度学习以其特征提取与时序建模能力,逐渐成为客流预测研究的主流技术。其中,支持向量回归(support vector regression, SVR)^[7-8]、XGBoost 模型^[9]、k 近邻回归(k-nearest neighbors, KNN)^[10]等机器学习方法被广泛应用于小规模数据的短期预测任务,而循环神经网络(recurrent neural network, RNN)^[11]、长短期记忆网络(long short-term memory network, LSTM)^[12]、时序卷积网络(temporal convolutional network, TCN)^[13]、Transformer^[14-15]以及组合深度学习模型^[16]等更多应用于需要处理大规模数据的长序列预测任务中。值得注意的是,Transformer 结构凭借强大的全局建模能力受到广泛关注,但其计算复杂度高、对超长序列处理效率不佳的问题也日益显现。

根据目前研究存在的问题,本研究提出一种新型地铁进站客流时序组合预测模型 AFT-K-GRU⁺。将 AFT^[17]、GRU^[18]和 KAN^[19]机制进行深度融合,并设计了知识驱动修正控制器(knowledge-driven controller, KDC)模块。首先,在地铁客流时序特征捕捉方面,AFT 模块替代传统 Transformer 中复杂的多头自注意力机

制,降低了训练拟合过程中时间与空间复杂度,从而大幅降低了序列交互开销;其次,在时序建模方面,选择采用门控机制更为简洁、参数更少、训练速度更快的 GRU 方法,而非传统的 LSTM 方法,从而减小了运算延迟;然后,在输出映射部分,传统客流预测模型一般采用全连接层结构,本文通过引入 KAN 网络进行替代,以更少参数实现了更优秀的拟合能力;最后,基于真实客流数据与预测模型结果之间的残差分布,设计了基于残差反馈的知识驱动修正器 KDC 机制,实现了对模型输出结果的在线自校正,从而有效抑制滚动误差累积,改善预测结果的稳定性与一致性。

1 AFT-K-GRU⁺模型框架

通过多模块协同作用,基于 AFT-K-GRU⁺的组合预测模型能够有效提升模型在地铁客流预测任务中的预测精度与趋势拟合度,为地铁站内的客流调度与风险预警提供可靠的数据支持。模型的整体框架结构如图 1 所示。

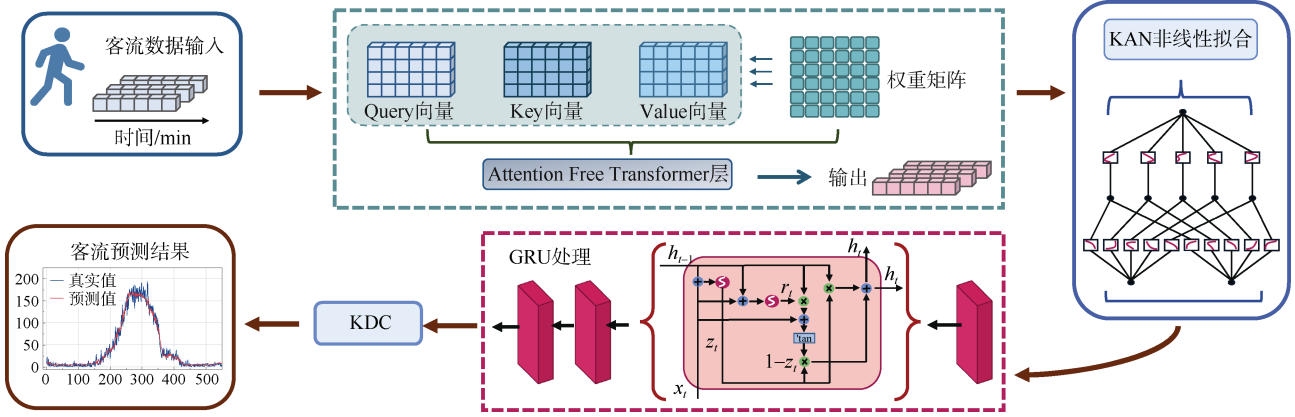
1.1 AFT 模型

设进站客流量输入数据为 $X = \{x_1, x_2, \dots, x_t\}$, 模型通过嵌入层与位置编码将原始数据序列转变为可学习表示。

$$e_t = W_{\text{input}} x_t + b_{\text{input}} \quad (1)$$

$$E_t^{\text{PE}}(i) = \begin{cases} \sin\left(\frac{t}{10\,000 \frac{i}{d_m}}\right), i \in (0, 2, 4, \dots) \\ \cos\left(\frac{t}{10\,000 \frac{i-1}{d_m}}\right), i \in (1, 3, 5, \dots) \end{cases} \quad (2)$$

式中, e_t 是输入 x_t 经过映射得到的结果; W_{input} 与 b_{input} 是输入的权重与偏置项; $E_t^{\text{PE}}(i)$ 是时间步 t 时刻的位置编码量; i 代表维度编号,对于奇数项与偶数项模型有不同的处理方法; d_m 为模型嵌入维度, $10\,000^{i/d_m}$ 代表控制不同维度周期的缩放因子。

图 1 基于 AFT-K-GRU⁺模型的进站客流预测框架Figure 1 Passenger inflow prediction based on the AFT-K-GRU⁺ model

将位置编码与嵌入层输出结果相加,可以得到 AFT 层输入 h_t 。

$$h_t = e_t + E_t^{\text{PE}} \quad (3)$$

地铁进站客流在长时序上会呈现出较为明显的时间依赖性,Transformer 模型基于自注意力机制,能够有效地捕捉序列中不同时间步之间的长程依赖。这对于地铁进站客流预测非常重要,传统的模型(如 LSTM)在捕捉这些长期依赖时会受到梯度消失问题的影响。

AFT 是一种替代 Transformer 中传统自注意力机制的模型,它可以用来捕捉时间序列中的全局依赖关系。AFT 的核心思想是使用轻量级的加权求和替代自注意力机制中的点积注意力操作。传统 Transformer 中的自注意力机制需要计算所有位置对之间的关系,计算量较大;AFT 层则采用了一种局部加权的机制,降低了计算复杂度。注意力机制原理^[17]如式(4)~(7)所示。

$$Q_t = h_t \cdot W_q \quad (4)$$

$$K_t = h_t \cdot W_k \quad (5)$$

$$V_t = h_t \cdot W_v \quad (6)$$

$$h_t^{\text{AFT}} = \sigma(Q_t) \odot \frac{\sum_{t'=1}^T \exp(K_{t'} + b') \odot V_{t'}}{\sum_{t'=1}^T \exp(K_{t'} + b')} \quad (7)$$

式中, W_q 、 W_k 、 W_v 为可学习参数矩阵,分别用于将输入特征映射到查询、键和值空间。其中, W_k 决定历史信息的重要性权重分布, W_v 表征被聚合的信息内容,而 W_q 通过门控机制调节聚合信息对当前时刻输出的影响程度。 T 是输入序列的总长度, t' 是求和索引,表示所有被聚合的信息来源位置; b' 为一个可学习偏置项; Q_t 为时间步 t 的查询向量, K_t 为时间步 t 的键向量, V_t 为时间步 t 的值向量。

AFT 根据当前时间步与其他时间步之间的关系计算加权系数,得到当前时间步输出 h_t^{AFT} 。

1.2 KAN 网络

KAN 网络是一种新型的神经网络架构,其设计思路来源于 Kolmogorov-Arnold^[20]表示定理。与传统的多层感知机(multi-layer perceptron, MLP)不同, KAN 在网络的连接上引入了可学习的非线性函数,从而增强了模型的表达能力和可解释性。KAN 的设计理念是将传统神经网络中的固定激活函数替换为可学习的一元函数,从而更灵活地建模复杂的非线性关系。在本文所构建的预测模型中, KAN 模块位于 AFT 模块之后、GRU 模块之前,在整个模型中相当于一个非线性特征变换层,它接收 AFT 输出的高维时间特征,对其进行一组可学习核函数的非线性变换,使后续 GRU 更容易捕捉复杂的时序关系,相当于传统深度学习架构中 MLP 的功能位置。但与普通 MLP 不同, KAN 的特点是它不使用固定的激活函数,而是用一组可学习的一维函数近似任意多维连续函数,从而增强非线性建模能力和可解释性。需要指出的是, KAN 模块的输入是来自 AFT 模块输出的全局时序特征, KAN 模块通过可学习的核函数映射实现非线性变换,并且其输出保持与输入相同维度的同时,在特征空间上实现了更丰富的非线性重构。KAN 模块的输出作为 GRU 的输入,能够帮助 GRU 在后续时间建模阶段更充分地捕获短期动态和复杂非线性关系。KAN 模型的基本原理^[19]如式(8)所示。

$$f(x_1, \dots, x_n) = \sum_{q=0}^{2n} \Phi_q \left(\sum_{p=1}^n \Phi_{q,p}(x_p) \right) \quad (8)$$

式中, $f(x_1, \dots, x_n)$ 为一个连续的多元函数; $\Phi_{q,p}$ 与 Φ_q 均为一元连续函数。

Kolmogorov-Arnold 定理表明, 一个多元连续函数可以表示为有限个一元连续函数的加法运算的复合函数。此外, 设经过 ATF full 层处理的输入为 h_i^{AFT} , 则经过 KAN 层输出可以表示为式(9)。

$$H_k^{\text{KAN}} = \sum_{j=1}^n \Phi_{k,j} \left(\sum_{i=1}^T \Phi_{j,i} (h_i^{\text{AFT}}) \right), k=1, \dots, m \quad (9)$$

式中, $H^{\text{KAN}} = [H_1^{\text{KAN}}, \dots, H_k^{\text{KAN}}, \dots, H_m^{\text{KAN}}]^T$ 是输出向量, $H^{\text{AFT}} = [h_1^{\text{AFT}}, \dots, h_i^{\text{AFT}}, \dots, h_n^{\text{AFT}}]^T$ 是输入向量, $\Phi_{k,j}$ 与 $\Phi_{j,i}$ 为可学习一元函数, m 是输出维度, $\Phi_{j,i}$ 为输入 h_i^{AFT} 到隐藏节点 j 的一元函数; Φ_j 为隐藏节点 j 到输出层的一元函数; n 是隐藏节点的数量。

这种结构使 KAN 网络相比于 MLP 方法能够以更少的参数逼近更复杂的函数, 提高了模型的效率和可解释性。

在信息变换中, 可能存在对输入特征信息的压缩与变形, 从而造成部分原始信息的弱化或丢失。因此, 为了防止 KAN 在非线性映射过程中可能对输入特征信息产生压缩损失, 在 KAN 层的输入与输出之间引入了残差连接, 处理后的输出 H_R^{KAN} 如式(10)所示。

$$H_R^{\text{KAN}} = \text{KAN}(\mathbf{x}) + \mathbf{x} \quad (10)$$

其中, $\mathbf{x}$ 是输入向量; $\text{KAN}(\mathbf{x})$ 是 KAN 网络的前向映射输出; y_{res} 是包含原始信息与变换信息的输出。

这种结构机制使得网络在拟合复杂关系的同时兼具对原始特征的保持能力, 缓解了模型中可能存在的信息丢失风险。

1.3 GRU 模型

地铁进站客流除了在长周期上具有一定的依赖信息, 在短时间上也体现出一定的变化规律。GRU 是一种用于处理序列数据的神经网络结构, 属于循环神经网络的一种变体。GRU 的主要特点是引入了重置门和更新门机制, 这决定了在每个时间步长上网络会保留或遗忘多少过去时间的信息。GRU 的基本原理^[18]可定义为:

$$z_t = \sigma(\mathbf{w}_z \mathbf{x}_t + \mathbf{u}_z \mathbf{h}_{t-1} + \mathbf{b}_z) \quad (11)$$

$$r_t = \sigma(\mathbf{w}_r \mathbf{x}_t + \mathbf{u}_r \mathbf{h}_{t-1} + \mathbf{b}_r) \quad (12)$$

式中, z_t 是更新门的输出; σ 是激活函数; $\mathbf{x}_t$ 是当前时

刻的输入; $\mathbf{h}_{t-1}$ 是上一时刻的隐藏状态; $\mathbf{w}_z$ 与 $\mathbf{u}_z$ 是更新门权重; $\mathbf{b}_z$ 是更新门偏置项; $\mathbf{r}_t$ 是重置门的输出; $\mathbf{w}_r$ 与 $\mathbf{u}_r$ 是重置门权重; $\mathbf{b}_r$ 是重置门偏置项。

候选隐藏状态 $\tilde{\mathbf{h}}_t$ 由当前输入与历史信息生成, 如式(13)所示。

$$\tilde{\mathbf{h}}_t = \tanh(\mathbf{w}_h \mathbf{x}_t + \mathbf{u}_h (\mathbf{r}_t \odot \mathbf{h}_{t-1}) + \mathbf{b}_h) \quad (13)$$

式中, $\mathbf{w}_h$ 与 $\mathbf{u}_h$ 是隐藏状态权重; $\mathbf{b}_h$ 是隐藏状态偏置项; $\tanh$ 是双曲正切函数。

最终隐藏状态的输出如式(14)一式(15)所示。

$$\text{GRU}(\mathbf{x}_t) = (1 - z_t) \odot \mathbf{h}_{t-1} + z_t \odot \tilde{\mathbf{h}}_t \quad (14)$$

$$\mathbf{h}_t^G = \text{GRU}(\mathbf{h}_t^{\text{KAN}}) \quad (15)$$

GRU 层的输出 H^G 经过全连接层处理, 得到进站客流预测结果 Y_{pre} , 如式(16)所示。

$$Y_{\text{pre}} = \text{MLP}(H^G) \quad (16)$$

1.4 KDC 模块

在时间序列预测类任务中, 模型预测结果往往会存在一定的系统偏差, 特别是在进站客流量复杂变化的场景下。因此, 本文通过引入时间加权的滑动平均误差与动态自适应修正系数, 设计了 KDC 模块。假设当前得到的进站客流预测值为 y_t^{pre} , 真实客流数据序列为 y_t , 则两者的误差为

$$e^t = y_t - y_t^{\text{pre}} \quad (17)$$

考虑到客流预测误差具有时间相关性, 采用 KDC 对过去 L 个时刻的误差进行加权平均, 以获得时间加权滑动误差项 e'_t

$$e'_t = \frac{\sum_{i=1}^L w_i e_{t-i}}{\sum_{i=1}^L w_i} \quad (18)$$

式中, L 为滑动时间窗口长度; w_i 为时间权重; e_{t-i} 代表当前时刻 t 之前第 i 个时间步的预测误差。

此外, 为防止出现修正过度导致结果出现震荡, 设计基于方差的动态自适应系数 k_t

$$k_t = \frac{k_0}{1 + \text{Var}(e_{t-L:t-1})} \quad (19)$$

式中, k_0 为基准修正增益; $\text{Var}(e_{t-L:t-1})$ 表示最近 L 个时刻误差的方差。

当误差波动较小时, k_t 值增大, 修正作用增强; 当误差波动较大时, k_t 自适应调整, 从而抑制剧烈修正。综合时间加权误差与自适应系数, 修正后的预测结果 $\bar{y}_t$ 可以表示为

$$\bar{y}_t = y_t^{\text{pre}} + k_t e'_t \quad (20)$$

因此,在 AFT-K-GRU 的基础上,通过设计 KDC 机制所构建的新型 AFT-K-GRU⁺模型进一步实现了基于残差反馈的动态加权修正,使模型预测结果能够实时根据误差分布自适应调整,有效地减少了系统性偏差与滚动累积误差,具有良好的稳定性与可扩展性,为地铁站进站客流时序预测提供了一种高鲁棒性的处理方案。

2 模型评估指标

为评估模型对进站客流量的预测性能,本研究使用平均绝对误差(MAE)、均方误差(RMSE)、平均绝对百分比误差(MAPE)、决定系数(R^2)作为预测结果评价指标,如式(21)一式(24)所示^[12]。

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - y_i^{\text{pre}}| \quad (21)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y_i^{\text{pre}})^2} \quad (22)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - y_i^{\text{pre}}}{y_i} \right| \times 100\% \quad (23)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - y_i^{\text{pre}})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (24)$$

式中, y_i 是真实数据值; y_i^{pre} 是模型预测结果; $\bar{y}$ 是真实数据的均值。

MAE 衡量的是预测结果的平均偏差,易于理解,但对异常值不太敏感。RMSE 对误差较大的结果更加敏感。 R^2 可以衡量模型解释方差的能力,用于评估模型的整体拟合程度。为了更全面的评估预测效果,我们引入动态时间规整(dynamic time warping, DTW)^[21],作为比较 2 个时间序列结果相似度的量化方法。DTW 通过逐个对比找到 2 列数据的最佳路径,从而计算出它们的相似度,如图 2 所示。

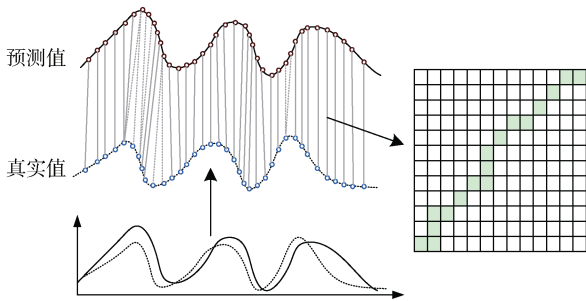


图 2 动态时间规整

Figure 2 Dynamic time warping

假设 2 个时间序列 $P = \{p_1, p_2, p_3, \dots, p_n\}$ 与 $Q = \{q_1, q_2, q_3, \dots, q_m\}$, 定义一个距离度量 $Dist$ 计算对应点之间的距离, Ct 作为累计距离, 最终 DTW 计算过程如式(25)一式(27)所示。

$$Dist(i, j) = (p_i - q_j)^2 \quad (25)$$

$$Ct(i, j) = Dist(i, j) + \min \begin{cases} Ct(i-1, j) \\ Ct(i, j-1) \\ Ct(i-1, j-1) \end{cases}, Ct(0, 0) = 0 \quad (26)$$

$$DTW(P, Q) = Ct(n, m) \quad (27)$$

综上所述, 4 种指标结合分析可以弥补各自的局限性, 帮助进行模型预测效果对比与参数调优。

3 实例分析

为了评估 AFT-K-GRU⁺模型的预测性能, 本研究选择中国北京市某地铁站 6 月至 7 月间 3 个星期的乘客进站闸机刷卡数据构建数据集, 进行模型的训练与评估。数据集采样间隔为 1 min, 采样时间段为每天上午 5 点到 12 点, 乘客进站数量 1 周内的变化如图 3 所示。

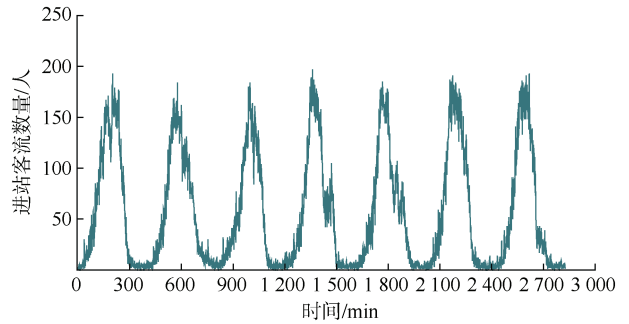


图 3 进站乘客数量变化

Figure 3 Variation in the number of entering passengers

从图 3 可以观察到, 进站乘客数量在日内变化方面存在一定的周期性变化规律, 日内客流量具有明显的先上升再下降的趋势, 但在具体短时变化中存在一定的差异; 在周内时间尺度上, 乘客数量变化呈现稳定的变化规律, 乘客进站的峰值与最低值基本相同。

每分钟乘客进站量占比如图 4 所示。对每天瞬时进站客流数据的统计结果呈现高度偏斜的分布特征, 大多数时间(47.1%)位于最低人数区间(0-25 人)内, 表明该地铁站的运行负载较为安全。同时, 每分钟超过 150 人进站的高负载情况相对罕见, 这些短暂但强烈变化的数据对模型的客流预测提出了挑战。此外, 中等

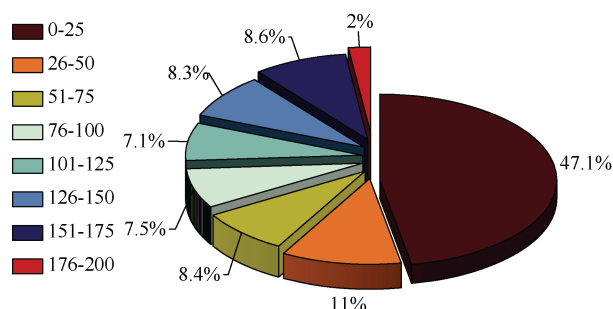


图4 进站乘客瞬时数量分布

Figure 4 Instantaneous distribution of entering passengers

范围的流入间隔(26-50人)占数据集剩余的53%,各子范围之间的时间占比相对平衡。这表明,对于长期内进站客流的预测需求是持续的,模型不仅要预测客流的整体趋势,还需要准确捕捉进站人数的过渡变化和中等负载波动。

3.1 数据预处理

在模型进行客流预测前,对于原始的进站客流数据需要进行预处理操作,原始数据集包含带时间的进站刷卡记录,每个记录对应于单个乘客的进站事件。为解决记录缺失或数据异常的问题,我们需检查数据集中记录中断或超出统计预期的时间间隔。对于这些异常数据,利用采样线性插补进行插值处理。为减少时间不稳定性及尺度差异对预测结果的影响,我们对数据集使用 Min-Max 缩放方法进行归一化处理,将原始输入 x_t 转换为 $x'_t \in [0,1]$,如式(28)所示。

$$x'_t = \frac{x_t - \min(x)}{\max(x) - \min(x)} \quad (28)$$

此外,为增强模型学习顺序模式的能力,本文使用滑动窗口方法对数据进行划分,采用固定长度大小为 P 的窗口预测 $t+1$ 时刻的进站客流量。同时,按照时间顺序将数据集划分为训练集与测试集,训练集包含80%的时间序列数据,其余20%作为测试集进行模型性能评估。

3.2 模型参数设置

本文所提出的 AFT-K-GRU⁺模型采用端到端的反向传播机制,通过误差反馈实现模型各模块参数的动态自适应调整,具体的模型反馈调参流程如下:首先,输入序列经线性投影与位置编码后进入 AFT 模块,用于捕获长时依赖关系和全局特征表示;然后, AFT 输出的序列特征输入到 KAN 层,通过核函数映射实现特征的非线性重构与增强,再由 GRU 模块对时间序列的局部动态特征进行建模,经输出层生成预测结果;

最后,通过基于残差反馈的知识驱动修正控制器 KDC 对结果进行优化,这一部分采用独立的网络搜索进行参数调整。需要指出的是在模型训练阶段,预测输出与真实值之间的误差由 MSE 计算得到;通过反向传播算法将该误差信号沿网络逆向传播,依次传递至输出层、GRU、KAN 与 AFT 模块的可训练参数;优化器 Adam 根据梯度信息自动更新各模块的参数,以最小化整体损失函数,实现全模型的联合优化。模型的参数配置见表1。

表1 模型参数设置

Table 1 Model parameter settings

模块	符号	参数值	作用
AFT	L_{AFT}	2	AFT 层数
GRU	L_{GRU}	2	GRU 层数
GRU	$Drop$	0.2	防止过拟合
GRU	$hidden_size$	128	神经元数
KDC	L	10	记忆长度
KDC	k_0	0.8	增益初值
KDC	k_m	0.3	修正下界
KDC	k_n	1.2	修正上界

3.3 模型预测结果对比

为了全面评估所提出的 AFT-K-GRU⁺模型对不同场景下进站客流的预测性能,采用 ARMA、SVR、MLP、反向传播神经网络模型(Backpropagation neural network, BP)^[22]、LSTM、GRU、TCN、Transformer 以及 CNN-BiLSTM 组合模型^[23]等时序预测模型作为对比方法,模型在训练过程中采用相同的参数设置,评估指标选择 MAE、RMSE、 R^2 与 DTW,且在计算 DTW 前对数据进行了归一化处理。各模型预测结果如图5所示,评估指标结果见表2。

表2 预测结果对比

Table 2 Comparison of prediction results

模型	MAE	RMSE	DTW	R^2
ARMA	36.164 5	42.067 9	11.221 3	0.444 5
SVR	32.197 0	43.641 7	11.348 5	0.403 0
MLP	7.353 2	10.450 5	0.613 3	0.966 2
BP	7.602 2	11.257 0	0.655 9	0.960 8
LSTM	7.468 5	10.057 1	0.638 6	0.970 4
GRU	7.611 9	10.089 4	0.645 4	0.968 5
TCN	7.466 1	10.053 9	0.551 7	0.968 7
Transformer	8.137 7	11.523 5	0.681 7	0.958 9
CNN-Bi-LSTM	8.444 8	11.491 9	0.631 1	0.958 6
AFT-K-GRU	6.391 0	9.207 1	0.475 1	0.974 1
AFT-K-GRU ⁺	6.074 3	9.034 4	0.447 2	0.974 5

从表 2 中可以看出,传统的时序预测模型如 ARMA、SVR 等在处理进站客流变化等复杂非线性数据时表现较差,拟合效果难以满足实际预测需求;MLP、BP、LSTM、Transformer 等在预测精度上显著优于传统模型, R^2 均达到了 0.95 以上,显示出了较强的拟合能力;但在 MAE 与 RMSE 评估中没有达到最优结果,在捕捉细微波动和突发变化方面仍有提升空间。本文所提出的 AFT-K-GRU⁺模型在所有评估指标上均取得了最优结果, R^2 达到了 0.974 5,表明模型在预测结果的拟合精度和稳定性方面具有一定的优势;与最新方法 CNN-BiLSTM 相比,AFT-K-GRU⁺模型 R^2 提升了约 1.6%,在模型泛化程度上具有一定的优势。此外,如图 5 所示,AFT-K-GRU⁺模型预测曲线在变化幅度上接近于真实值曲线,且能够有效拟合小范围的数据变化,表现出对客流数据中复杂模式的良好捕捉能力。

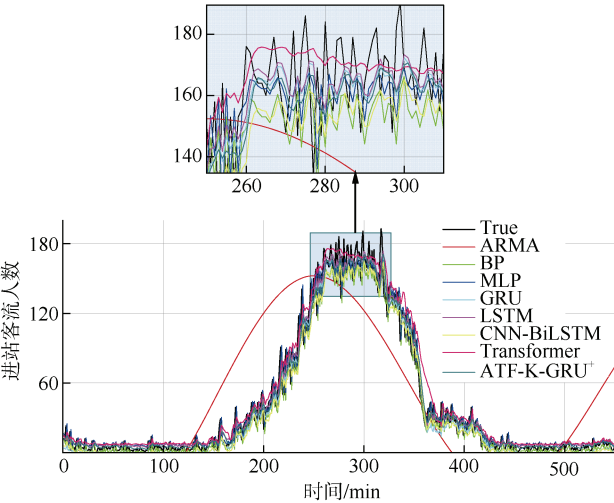


图 5 不同模型预测结果
Figure 5 Prediction results of different models

3.4 时间颗粒度分析

在实际场景的进站客流预测中,不同的时间颗粒度的预测对应不同的运营管理需求策略。短时间内的进站客流预测(如 15 min)可以应用于站内引导设施的实时调度,应对一些突发事件的发生;而长颗粒度的进站客流预测(如 60、120 min)则可以应用于地铁站的日常运行规划以及客流趋势分析。为了全面评估模型在不同场景下的适用性与时间颗粒度,本节选取 15、30、60、120 min 共 4 种时间颗粒度,对模型的性能进行对比分析。评价指标选择 MAE、RMSE 与 MAPE,不同颗粒度下的模型预测结果见表 3。

表 3 不同时间颗粒度实验结果

Table 3 Experimental results under different temporal granularities

时间颗粒度	模型	MAE	RMSE	MAPE
15 min	LSTM	4.908 7	6.291 8	30.403 8%
	TCN	6.233 5	7.888 3	28.367 2%
	Transformer	6.117 9	7.313 0	28.593 0%
	AFT-K-GRU	4.486 9	5.755 9	27.684 5%
	AFT-K-GRU ⁺	2.541 4	5.663 1	27.301 0%
30 min	LSTM	4.523 3	5.462 8	27.059 4%
	TCN	5.648 2	6.884 4	27.456 4%
	Transformer	5.577 3	6.417 0	27.694 5%
	AFT-K-GRU	4.182 5	4.953 9	25.137 3%
	AFT-K-GRU ⁺	4.096 7	4.104 8	22.897 0%
60 min	LSTM	3.903 4	4.786 1	33.077 5%
	TCN	4.909 2	5.986 0	30.864 7%
	Transformer	4.640 2	5.466 6	32.007 9%
	AFT-K-GRU	3.662 8	4.325 8	24.034 1%
	AFT-K-GRU ⁺	3.369 3	4.139 6	23.755 5%
120 min	LSTM	3.765 0	4.412 9	25.404 0%
	TCN	4.532 0	5.372 6	29.143 5%
	Transformer	4.592 7	5.175 5	30.139 4%
	AFT-K-GRU	3.526 4	4.003 5	24.406 7%
	AFT-K-GRU ⁺	2.827 4	3.545 7	22.235 2%

从表 3 可以看出,本文所提出的 AFT-K-GRU⁺模型在不同时间颗粒度下均取得了最佳表现。15 min 时间条件下,进站客流数据波动较大,模型的预测结果易受突发事件影响,AFT-K-GRU⁺模型捕捉短期波动能力相较于其他基线模型更为出色;在 30 与 60 min 条件下,LSTM、TCN 模型的预测性能出现一定程度的波动,AFT-K-GRU⁺模型的表现基本稳定,说明模型对于中长时间数据也有较好的拟合效果;在 120 min 条件下,所有模型的预测误差均有下降,但 AFT-K-GRU⁺模型依旧保持领先优势。如图 6 所示,实验结果表明,AFT-K-GRU⁺在不同时间颗粒度下均展现出稳健性和优越性,验证了其在多尺度客流预测任务中的有效性和泛化能力。

3.5 不同场景下预测结果对比

为了验证模型在不同场景下的预测适应性与泛化能力,本研究选取工作日、双休日、节假日 3 种场景下的地铁进站客流数据进行对比试验分析。其中,工作日场景数据为周一到周五范围内的客流数据,双休日

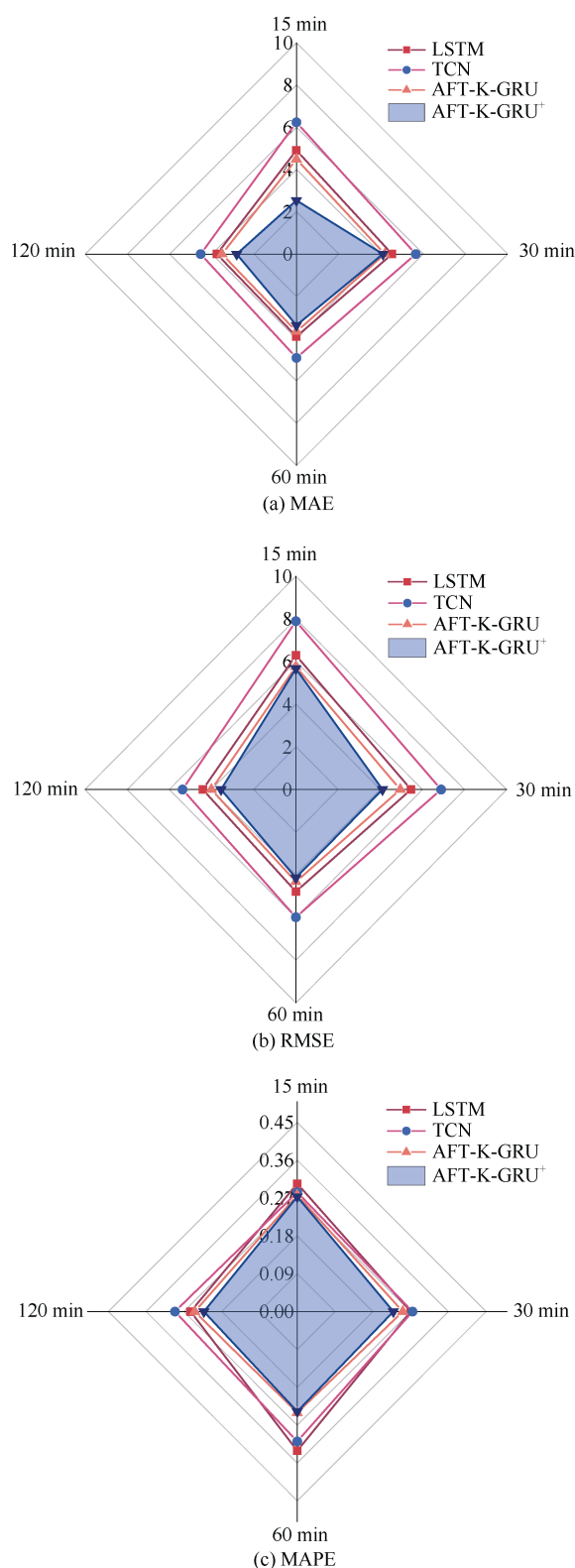


图6 不同时间颗粒度下实验评估指标对比
Figure 6 Comparison of evaluation metrics under different temporal granularities

场景数据采用周六和周天的客流数据,而节假日场景数据选取儿童节当天客流数据。3种场景下客流变化趋势如图7所示。

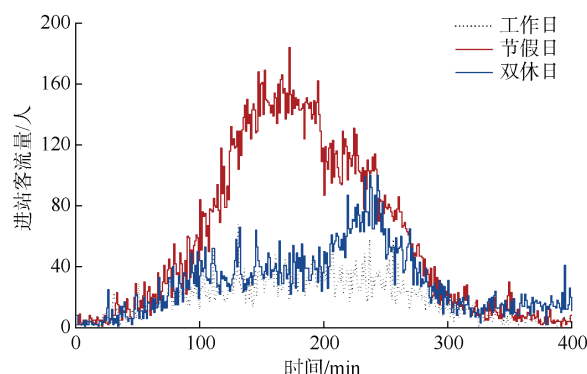


图7 不同场景下进站客流变化趋势

Figure 7 Trends in passenger inflow under different scenarios

图7给出了AFT-K-GRU⁺模型在工作日、节假日和双休日3种场景下进站客流量随时间的演化结果。可以看出,工作日数据体现了通勤模式下规律性强与周期性显著的特点;双休日客流分布无明显变化规律;而节假日数据则反映出客流量明显增多及快速变化的特征。这些差异为多场景下模型预测性能对比提供了重要依据。

AFT-K-GRU⁺模型与LSTM、CNN-BiLSTM这2种对比模型在3种不同场景下的预测结果如图8所示。本研究是从3种场景中随机各选取1天的客流数据进行的试验验证,评价指标采用的是MAE与RMSE。结果表明,相较于对比模型,AFT-K-GRU⁺模型在工作日、双休日、节假日3种不同场景下的预测结果都取得了最佳效果,进一步表明所构建的模型对不同场景下客流预测的强适应性。在工作日场景中,模型所精准识别的早晚高峰规律,能为优化列车运行图、加密高峰时段发车频次等精细化管控措施提供直接且可靠的依据;在双休日场景中,模型可以捕捉关键时间片特征,可为商业区、景点等站点的临时运力调度及安检通道开放数量调整提供参考;而在节假日场景中,模型可为重点车站分级限流设计及应急调度策略等提供数据支持。

3.6 消融试验

为深入分析AFT-K-GRU⁺模型中各模块对于预测结果的贡献度,本研究设计了一系列消融试验对比。试验数据选取1周内乘客进站刷卡数据,通过控制变量法对各模块进行独立移除,评估各模块对于预测精度的影响。

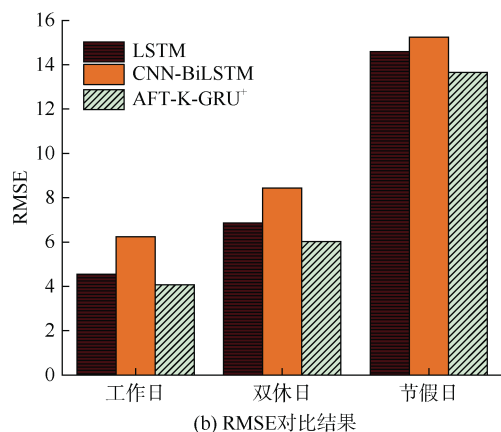
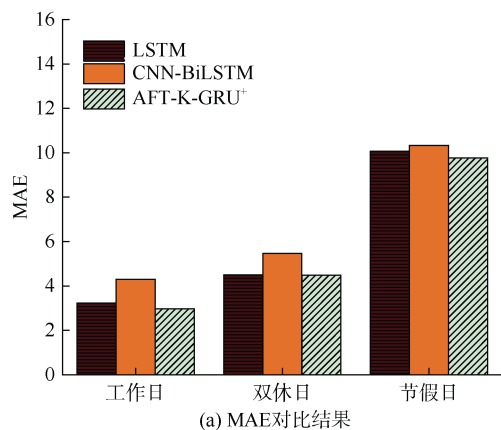


图 8 多场景下进站客流预测结果

Figure 8 Passenger inflow prediction results in multiple scenarios

相比于传统的客流预测方法, AFT-K-GRU⁺模型的特点是结合了注意力机制与 GRU 模型在短时信息依赖提取上的优势, 并通过 KAN 网络的非线性拟合能力进行映射。此外, 模型进一步融合知识驱动修正 KDC 模块, 从而全面提升预测精度。我们通过 Transformer 模块替换 AFT-K-GRM 模型中的 AFT 模块构造变体 T-K-GRU 模型, 通过移除 KAN 网络构造变体 AFT-GRU 模型, 通过移除 GRU 模块构造变体 AFT-KAN 模型, 通过移除 KDC 模块构造 AFT-K-GRU 模型。预测结果见表 4, 各模型拟合结果如图 9 所示。

表 4 消融试验结果

Table 4 Results of ablation experiments

模型	MAE	RMSE	R^2
T-K-GRU	6.592 5	9.757 6	0.970 2
AFT-GRU	7.774 5	9.890 2	0.969 7
AFT-KAN	7.235 1	10.494 2	0.965 6
AFT-K-GRU (无 ResNet)	6.632 4	10.519 7	0.968 4
AFT-K-GRU	6.391 0	9.207 1	0.974 1
AFT-K-GRU ⁺	6.074 3	9.034 4	0.974 5

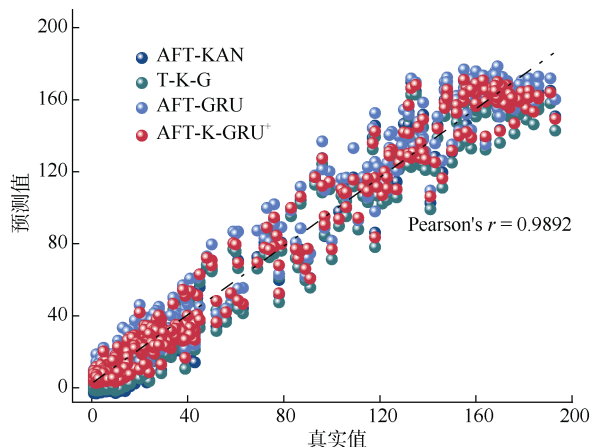


图 9 模型拟合结果

Figure 9 Result of model fitting

从表 4 所示结果可以总结出本研究提出的模型在所有性能指标上均表现最佳, MAE 为 6.074 3, RMSE 为 9.034 4, R^2 为 0.974 5。与变体模型 T-K-GRU 相比, AFT-K-GRU 在 MAE 上降低了约 3.06%, RMSE 降低了约 5.64%, R^2 提高了约 0.39%。这说明引入注意力机制在提高预测效率的同时, 能够帮助模型更高效地捕捉进站客流长期变化信息。相比 AFT-K-GRU(无 ResNet)模型, 存在 ResNet 机制的 AFT-K-GRU 模型 MAE 降至 6.391 0, RMSE 降至 9.207 1, R^2 提升至 0.974 1, 拟合效果显著增强, 模型对真实数据方差的解释能力提升, 表明加入 ResNet 机制后能够有效缓解模型映射过程信息弱化的问题。与 AFT-GRU 相比, AFT-K-GRU 使 MAE 降低了约 17.78%, 使 RMSE 降低了约 6.90%, 使 R^2 提高了约 0.45%, 这进一步验证了 KAN 网络在非线形拟合方面的优势。此外, 相比于变体模型 AFT-KAN, AFT-K-GRU 在 MAE 上降低了约 11.68%, 在 RMSE 上降低了约 12.27%, 在 R^2 上提高了约 0.88%, 这表明采用 GRU 架构可以增强模型对短期客流变化特征的提取, 提高了预测结果的精确度。最后, 与 AFT-K-GRU 模型相比, AFT-K-GRU⁺模型在 MAE 上降低了约 4.9%, 在 RMSE 上降低了约 1.88%, 在 R^2 上提高了约 0.04%, 表明采用 KDC 机制可以进一步提高模型预测结果的精确度。图 9 拟合结果可视化显示出 AFT-K-GRU⁺模型取得了最优皮尔逊相关系数。综上, 消融试验结果清晰地展示了各个组件对模型性能的贡献, 验证了 AFT-K-GRU⁺模型在地铁进站客流预测场景上的有效性。

4 结论

本研究针对地铁进站客流量预测问题, 提出了一

种新型组合深度学习方法 AFT-K-GRU⁺, 实现了对多尺度时间依赖特征的高效捕捉和预测结果的自适应优化。该方法的创新性与核心结论如下:

1) 提出一个深度融合 AFT、GRU 与 KAN 模块的端到端架构。其中, AFT 用于捕捉长期周期规律, GRU 用于提取短期波动特征, KAN 用于增强非线性表达能力, 该设计在保证高效的同时提升了模型的可解释性。

2) 设计了基于残差反馈的知识驱动修正控制器 KDC 模块, 实现了基于残差反馈的在线自校正, 模型能够动态调整预测偏差并抑制滚动误差累积。

3) 基于北京市某地铁站实际刷卡数据的测试表明, 在 15~120 min 时间颗粒度下, 模型的预测误差 MAPE 最低达 22.23%, 适用于短期实时调度和长期趋势规划; 在工作日、双休日和节假日场景下, 预测误差 MAE 和 RMSE 均领先于基准模型下的。与 Transformer 和 LSTM 基准模型相比, 平均误差分别降低了 25.3% 和 18.7%。模型在地铁进站客流预测任务中表现出色, 具有较强的泛化能力, 能够为地铁站运营管理提供可靠的客流数据支持。

参考文献

- [1] Johansson A, Helbing D, Al-Abideen H Z, et al. From crowd dynamics to crowd safety: a video-based analysis[J]. *Advances in complex systems*, 2008, 11(4): 497-527.
- [2] Zhang Botao, Xu Junhao, Xie C T, et al. Forecaster as a simulator: Simulating multi-directional pedestrian flow with knowledge-guided Graph Neural Networks[J]. *Computers & industrial engineering*, 2024, 198: 110668.
- [3] 黄凯, 张健钦, 程歆玥, 等. 基于刷卡数据的公交和地铁接驳出行评价研究: 以北京市为例[J]. *都市快轨交通*, 2025, 38(4): 53-60.
Huang Kai, Zhang Jianqin, Cheng Xinyue, et al. Assessment of bus-metro transfer behavior based on card-swiping data: a case study of Beijing[J]. *Urban rapid rail transit*, 2025, 38(4): 53-60.
- [4] 李金海, 窦亮, 张猛, 等. 城市轨道交通初期运营客流预测技术研究[J]. *都市快轨交通*, 2024, 37(1): 102-106.
Li Jinhai, Dou Liang, Zhang Meng, et al. Passenger flow forecasting technology for initial operation of urban rail transit[J]. *Urban rapid rail transit*, 2024, 37(1): 102-106.
- [5] Xie Jiaming, Choi Y K. Hybrid traffic prediction scheme for intelligent transportation systems based on historical and real-time data[J]. *International journal of distributed sensor networks*, 2017, 13(11): 155014771774500.
- [6] Nguyen H T, Nguyen H D, Do T H. An application of vector autoregressive model for analyzing the impact of weather and nearby traffic flow on the traffic volume[C]//2022 RIVF International Conference on Computing and Communication Technologies (RIVF). December 20-22, 2022, Ho Chi Minh City, Vietnam. IEEE, 2022: 328-333.
- [7] 谢臻, 郭建媛, 秦勇. 基于支持向量回归的地铁进站客流短时预测模型[J]. *都市快轨交通*, 2020, 33(2): 82-86.
Xie Zhen, Guo Jianyuan, Qin Yong. Short-term prediction model of subway entry passenger flow based on support vector regression[J]. *Urban rapid rail transit*, 2020, 33(2): 82-86.
- [8] 胡松, 常征, 杨贝, 等. 基于 PSO-SVR 模型的轨道交通客流预测[J]. *都市快轨交通*, 2025, 38(5): 50-56.
Hu Song, Chang Zheng, Yang Bei, et al. Urban rail passenger flow prediction using particle swarm optimization-tuned SVR model[J]. *Urban rapid rail transit*, 2025, 38(5): 50-56.
- [9] 郑宣传, 秦勇, 郭建媛, 等. 基于多因素分析的地铁客流预测方法实证研究[J]. *都市快轨交通*, 2025, 38(5): 57-63.
Zheng Xuanchuan, Qin Yong, Guo Jianyuan, et al. Empirical study of a multi-factor subway passenger flow forecasting method[J]. *Urban rapid rail transit*, 2025, 38(5): 57-63.
- [10] Pang Xiyu, Wang Cheng, Huang Guolin. A short-term traffic flow forecasting method based on a three-layer K-nearest neighbor non-parametric regression algorithm[J]. *Journal of transportation technologies*, 2016, 6(4): 200-206.
- [11] He Yuxin, Li Lishuai, Zhu Xinting, et al. Multi-graph convolutional-recurrent neural network (MGC-RNN) for short-term forecasting of transit passenger flow[J]. *IEEE transactions on intelligent transportation systems*, 2022, 23(10): 18155-18174.
- [12] Ouyang Qi, Lv Yongbo, Ma Jihui, et al. An LSTM-based method considering history and real-time data for passenger flow prediction[J]. *Applied sciences*, 2020, 10(11): 3788.
- [13] Lu Ye, Zheng Changjiang, Zheng Shukang, et al. Multi-spatio-temporal convolutional neural network for short-term metro passenger flow prediction[J]. *Electronics*, 2024, 13(1): 181.
- [14] Yue Min, Ma Shuhong. LSTM-based transformer for transfer passenger flow forecasting between transportation integrated hubs in urban agglomeration[J]. *Applied sciences*, 2023, 13(1): 637.
- [15] Yang Yongjie, Zhang Jinlei, Yang Lixing, et al. Short-term passenger flow prediction for multi-traffic modes: a Transformer and residual network based multi-task learning method[J]. *Information sciences*, 2023, 642: 119144.
- [16] 帅春燕, 谢亚威, 单君, 等. 基于 SSA-SVR 模型的城市轨道交通短时进站客流预测[J]. *都市快轨交通*, 2022,

- 35(5): 76-83.
- Shuai Chunyan, Xie Yawei, Shan Jun, et al. Prediction of short-term inbound passenger flow of urban rail transit based on the singular spectrum analysis and support vector regression model[J]. Urban rapid rail transit, 2022, 35(5): 76-83.
- [17] Zhai S, Talbott W, Srivastava N, et al. An attention free transformer[J]. arXiv preprint arXiv: 2105.14103, 2021.
- [18] Cho K, Van Merriënboer B, Gulcehre C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[J]. arXiv preprint arXiv: 1406.1078, 2014.
- [19] Liu Z, Wang Y, Vaidya S, et al. Kan: Kolmogorov-arnold networks[J]. arXiv preprint arXiv: 2404.19756, 2024.
- [20] Kolmogorov A N. On the representation of continuous functions of several variables by superpositions of continuous functions of a smaller number of variables[M]. American Mathematical Society, 1961.
- [21] Sakoe H, Chiba S. Dynamic programming algorithm optimization for spoken word recognition[J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 2003, 26(1): 43-49.
- [22] Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors[J]. Nature, 1986, 323(6088): 533-536.
- [23] 赵新辉, 刘斌, 田志强, 等. 基于 CNN-Bi-LSTM 网络的城轨短时客流预测[J]. 兰州交通大学学报, 2024, 43(6): 130-137.
- Zhao Jinhui, Liu Bin, Tian Zhiqiang, et al. Short-term passenger flow prediction for urban railway based on CNN-Bi-LSTM network[J]. Journal of Lanzhou Jiaotong University, 2024, 43(6): 130-137.

(编辑: 傅依萱)

(上接第45页)

- [6] 安睿, 仝照民, 王梓蒙, 等. 武汉市轨道交通 TOD 绩效评估与类型划分: “节点-场所-功能”模型的实证应用[J]. 地理研究, 2024, 43(10): 2684-2701.
- An Rui, Tong Zhaomin, Wang Zimeng, et al. Evaluation and classification of TOD performance for Wuhan railtransit: Empirical application of the “Node-Place-Function” model[J]. Geographical research, 2024, 43(10): 2684-2701.
- [7] Su Shiliang, Zhang Hui, Wang Miao, et al. Transit-oriented development (TOD) typologies around metro station areas in urban China: a comparative analysis of five typical megacities for planning implications[J]. Journal of transport geography, 2021, 90: 102939.
- [8] Liao Cong, Scheuer B. Evaluating the performance of transit-oriented development in Beijing metro station areas: Integrating morphology and demand into the node-place model[J]. Journal of transport geography, 2022, 100: 103333.
- [9] 国务院关于《济南市国土空间总体规划(2021—2035年)》的批复[J]. 中华人民共和国国务院公报, 2024(33): 16-17.
- The reply of the state council on the overall planning of Jinan land space (2021—2035)[J]. Bulletin of the state council of the people's republic of China, 2024(33): 16-17.
- [10] 刘军, 孔令栋. 济南城市轨道交通发展现状研究[J]. 山东交通科技, 2024(5): 153-155.
- Liu Jun, Kong Lingdong. Research on the development status of Jinan urban rail transit[J]. Shandong jiaotong keji, 2024(5): 153-155.
- [11] Ma Jiexi, Shen Zhongwei, Xie Yi, et al. Node-place model extended by system support: Evaluation and classification of metro station areas in Tianfu new area of Chengdu[J]. Frontiers in environmental science, 2022, 10: 990416.
- [12] 夏雪, 盖靖元. 基于 K-Means 聚类算法的城市轨道交通站点分类及客流特征分析[J]. 现代城市轨道交通, 2021(4): 112-118.
- Xia Xue, Gai Jingyuan. Classification of urban rail transit stations and points and analysis of passenger flow characteristics based on K-Means clustering algorithm[J]. Modern urban transit, 2021(4): 112-118.
- [13] 李明星, 赵金宝, 刘文静, 等. 城市轨道交通站点地区土地利用与交通协同性评价[J]. 铁道运输与经济, 2024, 46(8): 222-231.
- Li Mingxing, Zhao Jinbao, Liu Wenjing, et al. Synergy evaluation of land utilization and traffic in urban rail transit station areas[J]. Railway transport and economy, 2024, 46(8): 222-231.

(编辑: 傅依萱)