

文章编号: 1003-1421(2026)09-0129-12

引用格式: 任文硕, 张健钦, 程歆玥, 等. 建成环境约束下的城市轨道交通出行需求预测方法研究[J]. 铁路运输与经济, 2026, 48(9): 129-140.

REN Wenshuo, ZHANG Jianqin, CHENG Xinyue, et al. Prediction Method for Urban Rail Transit Travel Demand under Constraint of Built Environment[J]. Railway Transport and Economy, 2026, 48(9): 129-140.

建成环境约束下的城市轨道交通出行需求预测 方法研究

任文硕, 张健钦, 程歆玥, 文 政

(北京建筑大学 测绘与城市空间信息学院, 北京 102616)

摘 要: 建成环境显著影响轨道交通客流量, 而现有的轨道交通出行需求预测方法多基于历史客流数据, 没有充分考虑建成环境的多维空间功能属性的约束条件。提出一种基于XGBoost-SHAP的改进预测模型, 并通过贝叶斯优化进行超参数调整, 在城市轨道交通800 m集水区内中心性、路网密度等建成环境要素的约束下, 构建非线性预测模型, 借助XGBoost的集成学习优势提升预测精度, 并通过SHAP对建成环境变量的特征进行重要性分析。以北京市轨道交通为例进行计算实验, 结果表明: XGBoost-SHAP模型在3个时段预测结果的模型性能均优于其他对比模型; 停车场密度、介数中心性和商业区密度等因素对出行需求的影响效果明显。研究验证了建成环境与机器学习结合的出行需求预测方法的有效性和科学性, 为以后的城市轨道交通选址和规划提供了智能决策方法。

关键词: 轨道交通; 客流量预测; 建成环境; XGBoost; Shapley Additive Explanations

中图分类号: U12; U239.5 **文献标识码:** A **DOI:** 10.16668/j.cnki.issn.1003-1421.20250415001

Prediction Method for Urban Rail Transit Travel Demand under Constraint of Built Environment

REN Wenshuo, ZHANG Jianqin, CHENG Xinyue, WEN Zheng

(School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, Beijing 102616, China)

Abstract: The built environment significantly affects the passenger flow of rail transit. However, the existing prediction methods for rail transit travel demand are mostly based on historical passenger data without fully considering the multi-dimensional spatial functional attributes of the built environment. An improved prediction model was proposed based on XGBoost-SHAP, and the hyperparameters were adjusted through Bayesian optimization. Under the constraints of built environment elements such as centrality and road network density within the 800-meter catchment area of urban rail transit, a nonlinear prediction model was constructed. The ensemble learning advantage of XGBoost was utilized to improve the prediction accuracy. SHAP was employed to analyze the importance of built environment variable characteristics. A computational experiment was conducted using Beijing's rail transit as an example. The results show that the model performance

收稿日期: 2025-04-15 修订日期: 2026-07-20

通信作者: 张健钦(1977—), 男, 河北保定人, 北京建筑大学测绘与城市空间信息学院教授。

基金项目: 国家自然科学基金项目(42371416); 北京建筑大学博士研究生科研能力提升项目(DG2025036)

of the XGBoost-SHAP model in the three time periods is superior to other comparison models. Factors such as parking lot density, betweenness centrality, and commercial area density have a significant impact on travel demand. The study verifies the effectiveness and scientific nature of the travel demand prediction method combining the built environment and machine learning, providing an intelligent decision-making method for future site selection and planning of urban rail transit.

Keywords: Rail Transit; Passenger Flow Prediction; Built Environment; XGBoost; Shapley Additive Explanations

0 引言

随着城市化进程加速,轨道交通成为公共交通系统的重要组成部分,是缓解城市交通拥堵的重要载体。其不仅具有高运载量和节能环保等优势,还在缓解城市交通拥堵、促进经济发展和提升城市形象等方面发挥着关键作用^[1]。另外,由于准时性、安全性高、速度快等优势^[2],城市轨道交通也成为人们出行必不可少的工具。然而,由于客流量的逐年增长,给轨道交通带来了不小的承载压力。交通运输部速报数据显示,2024年全国城市轨道交通客运量322.4亿人次,较2023年增加28亿人次、增长9.5%^[3]。因此,精准预测城市轨道交通的出行需求就变得尤为重要^[4],这不仅影响居民出行体验,还是优化线网布局与城市交通规划的核心前提。通过准确把握轨道交通站点的未来出行需求,决策者能够制定科学合理的规划方案,提升城市交通系统的整体承载能力和公共交通服务水平^[5]。

近年来,随着极限梯度提升树模型(XGBoost)、图神经网络(GNN)、Transformer等机器学习和深度学习算法的发展,这些方法被广泛应用于交通客流量需求预测领域。在交通预测方面,自回归积分滑动平均(ARIMA)、多元线性回归等传统方法依赖历史数据的线性假设,难以处理复杂交互效应,例如,ARIMA模型假设时间序列数据是平稳的,并且其变化仅由时间上的线性关系所驱动,这在复杂城市交通环境中往往不成立,多元线性回归也有类似问题。而XGBoost、长短期记忆(LSTM)等机器学习和深度学习算法通过自适应学习机制,可以捕捉到特征之间的复杂交互关系,其优秀的泛化能力和计算效率使其在交通流量预测中得到了广泛应用。如Wang等^[6]利用卷积(CNN)层对图像的时间序列数据提取有价值信息,以及长LSTM层对交通流进行预测。Pan等^[7]通过物理引导时空图神

经网络(PG-STGNN),将交通流物理原理与先进的时空图神经网络算法相结合,用于交通流量的预测。韩皓等^[8]基于轻量级梯度提升机(LightGBM)算法,结合复杂网络与机器学习理论对轨道交通短时客流量进行预测。徐明明等^[9]利用烟花粒子群算法优化BP神经网络,使用优化后的权重和偏置来对轨道交通客流量进行预测。尽管这些研究在预测精度和实时性方面表现优异,能够较好地满足现有轨道交通站点的客流量预测需求,但其高度依赖站点客流量历史数据的局限性不容忽视。尤其在缺乏充分历史数据的情况下,例如对于在建轨道交通站点的客流量预测,该方法的预测效果将难以充分发挥。

城市轨道交通站点周边土地利用混合度、容积率、路网密度等建成环境要素,显著影响客流的时空分布特征^[10]。土地利用混合度的提升能够缩短居住距离,影响通勤客流^[11]。区域的容积率较高,其人口与商业活动也会相对集中,轨道交通潜在的需求量也较大^[12],路网密度反映了城市轨道交通站点的可达性,高可达性街区路网可以提升居民使用城市轨道交通的便利性,进而提升客流量^[13]。然而,现有研究可能会忽视城市建成环境的相关性,只基于轨道交通客流量历史数据进行回归分析。为此,本研究将建成环境作为约束条件,提出一种轨道交通客流量预测方法。通过构建多维特征提取轨道交通站点周边的建成环境要素,融合XGBoost-SHAP算法构建轨道交通出行需求智能预测模型,同时利用贝叶斯优化算法对XGBoost超参数进行优化,以期提升模型构建效率与预测准确性,并将其应用在北京市轨道交通站点的客流量预测中。研究成果可为站点选址规划、站点周边公共交通导向开发(Transit-Oriented Development, TOD)规划等提供理论依据与技术支持。

1 研究范围与研究数据

1.1 研究范围

本研究选取北京市五环内区域作为研究范围,研究范围如图1所示。该区域兼具高密度居住区、商业综合体与交通枢纽等功能,且受早晚高峰潮汐效应及大型活动影响显著^[14],适合检验模型时空适应性与鲁棒性。另外,Lei等^[15]在研究中提到轨道交通步行集散区域为400~800 m, Li等^[16]对以前研究分析后也提到轨道交通附近800 m的距离被认为是标准步行距离。因此,研究将轨道交通集水区范围划定为800 m。

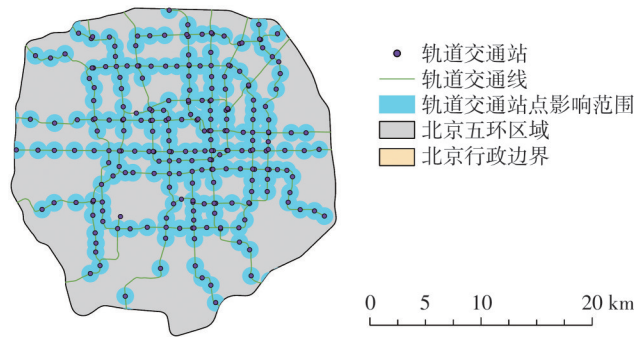


图1 研究范围
Fig.1 Range of study

1.2 数据来源

本研究所用到的数据包括智能卡数据、建筑数据、路网数据、兴趣点(Point of Interest, POI)数据、公共交通站点及线网信息数据和国土空间规划数据。其中,智能卡数据来源于北京智慧交通发展中心,该数据记录了2019年5月1—12日的城市轨道交通智能卡(SSC)数据,可以较好地表征轨道交通出行需求情况;北京市建筑数据和路网数据来源于OSM(OpenStreetMap)官网;北京市POI数据通过高德地图应用程序编程接口(API)爬取;公共交通站点及地铁线网信息数据通过北京市公交和地铁官网以及TransBigData获取;国土空间规划数据来源于北京市规划和自然资源委员会官网。本研究所需的原始数据如表1所示。

1.3 建成环境与客流量特征提取

研究采用多源异构数据融合方法,构建时空特征矩阵以支撑模型训练。整合智能卡数据,对异常值进行剔除,提取有效城市轨道交通智能卡数据

表1 原始数据

Tab.1 Raw data

数据名称	数据类型	数据来源
Smart card data (智能卡数据)	.shp	北京智慧交通发展中心 (https://jtw.beijing.gov.cn/)
北京市建筑数据	.shp	OSM(https://www.openstreetmap.org/)
北京市POI数据	.csv	高德地图 API
北京市路网数据	.shp	OSM(https://www.openstreetmap.org/)
北京公交站点及线网信息数据	.shp	北京公交官网(https://www.bjsubway.com/)
北京轨道交通站点及线网信息数据	.shp	北京地铁官网(https://www.bjbus.com/) TransBigData
北京市国土空间规划数据	.shp	北京市规划和自然资源委员会 (https://ghzrzyw.beijing.gov.cn/)

29 997 237条,分别统计各城市轨道交通站点早高峰、平峰、晚高峰1 h时间窗的出站量作为因变量。基于“5D”理论^[17]选择建成环境指标,然后利用GIS软件实现对各个城市轨道交通站点建成环境要素的提取。①基于北京市建筑数据,量化站点周边800 m缓冲区内建筑密度(单位面积建筑面积)及容积率(总建筑面积与用地面积比值)。②整合公交智能卡数据,剔除异常值,最终得到公交智能卡数据60 242 310条,计算公交进出站量。③融合POI数据计算商业、居住、工作的空间集聚特征、停车场密度及功能混合度(基于熵指数的土地利用多样性)。④整合城市轨道交通线路分布数据,通过TransBigData获取每条地铁的线路和站点,然后处理成总的地铁邻接表,并构建Network网络,利用Python进行城市轨道交通站的中心性分析(包括度中心性、中介中心性、接近中心性和特征向量中心性)与站点换乘线路数的提取。⑤整合路网数据、公交站点数据,对路网密度、公交站台密度等进行提取。将这些要素提取作为自变量,构建多维度指标体系。获取自变量后,引入VIF对自变量的多重共线性进行检验,对VIF值极高(VIF>10)的变量进行筛选,发现PageRank的VIF值过高,因此将其排除在分析之外,最终确定了19个候选变量。变量统计数据如表2所示。其中Mean为轨道交通站点变量平均值;Std.Dev为标准差;Min为最小值;Max为最大值。

2 研究框架与研究方法

2.1 研究框架

本实验的研究框架如图2所示。首先进行数据

表2 变量统计数据
Tab.2 Variable statistical data

变量	统计			
	Mean	Std.Dev	Min	Max
因变量				
早高峰	4 725.525 8	3 604.762 4	87.750 0	26 745.166 7
平峰	3 150.350 6	2 403.174 9	58.500 0	17 830.111 1
晚高峰	1 181.381 5	901.190 6	21.937 5	6 686.291 7
自变量				
抵达城市中心的直线距离	8.278 5	3.676 4	0.307 7	17.040 7
度中心性(Degree Centrality)	0.005 7	0.002 0	0.004 5	0.011 3
中介中心性(Betweenness Centrality)	0.045 2	0.027 5	0.000 2	0.130 8
接近中心性(Closeness Centrality)	0.076 9	0.009 4	0.049 5	0.094 7
特征向量中心性(Eigenvector Centrality)	0.000 5	0.005 8	0.000 0	0.072 5
换乘线路数	0.487 8	0.796 6	0.000 0	3.000 0
路网密度	8.073 0	2.261 8	1.498 9	14.153 0
停车场密度	27.046 8	17.419 9	0.407 4	76.394 4
公交站台数量	3.044 3	5.022 7	0.000 0	22.000 0
公交线路数量	19.550 6	16.602 3	0.000 0	80.000 0
容积率	1.472 4	0.398 9	0.151 1	2.245 5
建筑密度	0.231 9	0.052 9	0.045 8	0.365 1
功能密度	2 001.829 1	1 224.117 6	110.000 0	6 576.000 0
功能混合度	3.888 2	0.233 3	2.975 9	4.260 7
商业区密度	519.158 2	343.630 1	4.000 0	1 996.000 0
工作区密度	583.816 5	453.920 0	34.000 0	2 431.000 0
住宅区密度	102.379 7	58.400 6	7.000 0	308.000 0
公交出站客流量	3 283.754 7	6 134.975 1	0.000 0	47 649.916 7
公交进站客流量	3 275.773 7	6 250.137 4	0.000 0	49 138.916 7

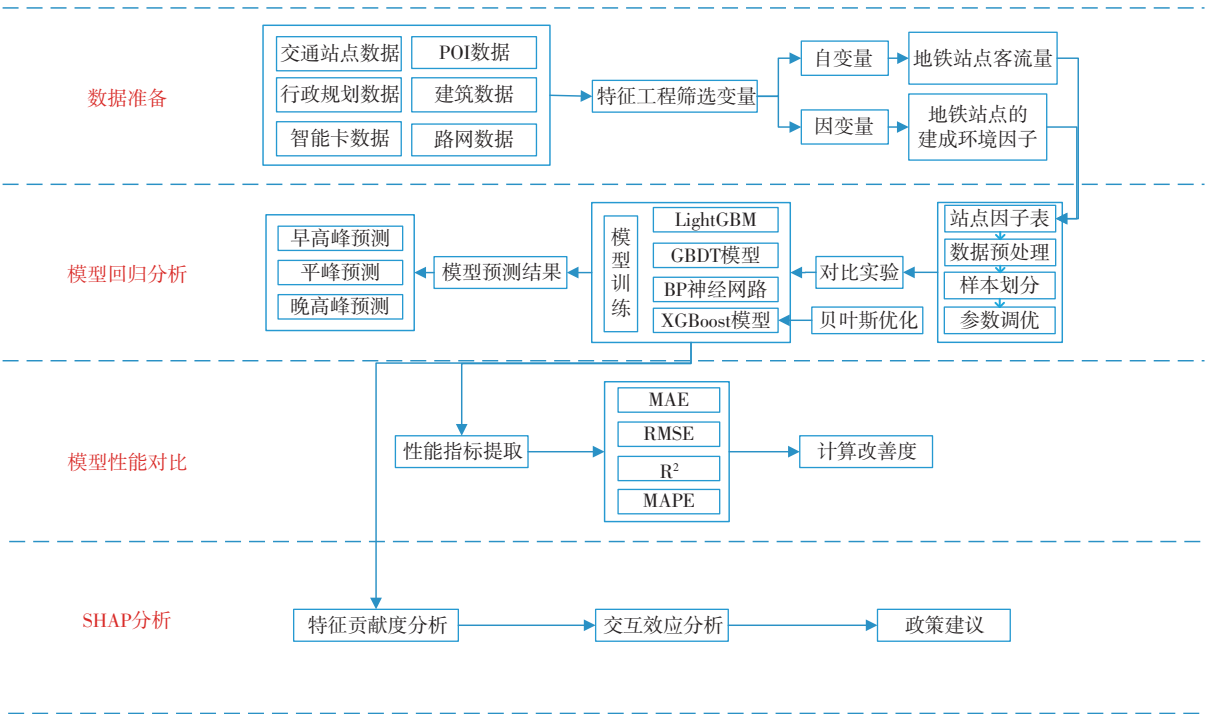


图2 研究框架

Fig.2 Framework of study

预处理,并通过构建空间特征筛选关键变量,生成站点因子表;随后进行样本划分和参数调优,将调优结果导入模型并对模型进行训练,得到回归预测结果;接着,通过选取模型性能指标对模型性能进行对比;最后进行特征贡献度和交互效应分析,并给出相应的政策建议。

2.2 XGBoost模型

轨道交通需求受建成环境多维变量的非线性交互作用驱动,传统线性模型难以捕捉此类复杂关系。XGBoost是一种基于梯度提升决策树(Gradient Boosting Decision Tree, GBDT^[18])的集成学习算法,它包含多个决策树,各决策树之间存在明确的先后顺序,且后一个决策树会考虑前一个决策树的偏差,并重点关注前一个树做错的训练样本,逐步优化预测结果,可以显著提升预测精度。

XGBoost的第 t 个决策树的目标函数公式如下。

$$L^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t)}) + \Omega(f_t) \quad (1)$$

式中:损失函数 $l(y_i, \hat{y}_i^{(t)})$ 衡量预测值 $\hat{y}_i^{(t)}$ 与真实值 y_i 的差异(如均方误差、交叉熵);正则化项 $\Omega(f_t)$ 表示第 t 棵树的模型复杂度,防止过拟合。

在经过 t 次迭代后,第 t 棵树对第 i 个轨道交通站点客流量的预测值就是前 $t-1$ 棵预测树的预测值+第 t 棵树的预测值,即

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x_i) \quad (2)$$

式中: $\hat{y}_i^{(t)}$ 为第 i 个地铁客流量的预测输出值; $\hat{y}_i^{(t-1)}$ 为前 $t-1$ 棵树对第 i 个轨道交通站点的预测值; $f_t(x_i)$ 为第 t 棵树对第 i 个轨道交通站点的预测值。

XGBoost算法在目标函数中加入正则化项(L1/L2),同时与二阶梯度优化结合,在提升预测精度的同时有效抑制过拟合;内置缺失值处理机制可自适应学习稀疏特征的最优分裂方向,最终在复杂非线性关系建模中展现出卓越的泛化能力与工程适用性,可以更好地对城市轨道交通站点的建成环境指标进行回归分析。

2.3 贝叶斯优化

贝叶斯优化(Bayesian Optimization, BO)是一种用于优化黑盒函数的迭代算法,广泛应用于高维复杂的超参数调优任务^[19]。其核心思想是利用之前试探的结果来指导未来的搜索,避免冗余的无效计算,从而提高搜索效率。贝叶斯优化通过贝叶斯推

断对目标函数进行概率建模,根据当前模型的预测结果选择具有最大潜在改进的参数组合,并不断更新模型,最终在有限预算内找到全局最小值或最大值,因此被广泛应用于超参数的调整优化中。

2.4 SHAP模型

XGBoost模型虽然能很好地对出行需求进行预测,但其对自变量的可解释性较为有限;而SHAP^[20]是一种基于博弈论的合作贡献分配方法,可以很好地解释机器学习模型的预测结果。相对于XGBoost模型,SHAP具有以下优势。①可以提供单一预测的详细解释,还可以量化每个特征在不同情景下的贡献,有助于更深入地理解模型的预测机制。②能够揭示特征之间的交互效应,有些特征在单独考虑时对预测影响较小,但与其他特征交互时却影响显著。

对于特征 j 的SHAP值 φ_j ,其计算公式为

$$\varphi_j = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} [f(S \cup \{j\}) - f(S)] \quad (3)$$

式中: φ_j 表示第 i 个轨道交通站点的第 j 个建成环境因子对预测值 y_i 的贡献度, $\varphi_j > 0$ 表明其提升了预测值, $\varphi_j < 0$ 表明其降低了预测值; F 表示特征全集,即模型使用的所有建成环境因子的集合; S 表示不含建成环境因子 j 的特征子集; $f(S)$ 表示基于特征子集 S 的模型预测值; $\frac{|S|!(|F|-|S|-1)!}{|F|!}$ 为权重项,对建成环境因子 j 在所有可能子集组合中的贡献进行加权平均; $[f(S \cup \{j\}) - f(S)]$ 为特征 j 的边际贡献,表示建成环境因子 j 加入子集 S 后对预测值的增量贡献。

3 研究结果及分析

3.1 数据预处理

提取建成环境多维变量后,需要对数据进行预处理,这是模型构建的重要环节,直接影响预测精度和稳定性。首先处理异常值和缺失值。对于缺失值,本实验采用基于空间相似站点的插值方法进行填充,以保证数据的连续性和完整性;对于异常值,通过统计分析方法识别,剔除显著高于或低于历史均值的极端客流数据,从而降低噪声对模型的干扰。然后对数据进行标准化,将不同量级的特征

映射到相同尺度，以减少特征之间的量纲差异对模型训练的影响，同时加快优化过程的收敛速度。

另外，由于研究区域内站点数据有限，为增强模型的鲁棒性和泛化能力，本实验通过数据增强^[21]手段扩展训练样本量，在特征值周围按照一定比例添加随机噪声，设置扰动比例为1%，保证扰动幅度微小且物理合理，从而生成多个相似但不完全相同的新样本。另外为避免噪声在训练过程中产生的过度学习，在训练过程中采用早停法。为了保证客流量等非负特征的合理性，增强过程中对扰动后的数据进行了截断处理，避免出现负值。最后，数据增强的所有统计量均从训练集计算，测试集保持原始状态，从而不影响模型的预测。通过这种方式，可以在保持原始数据分布特征的前提下，增加训练样本的多样性，使模型能够更好地学习潜在规律^[22]。

3.2 贝叶斯优化算法寻找 XGBoost 最优超参数

贝叶斯优化算法广泛应用于高维复杂的超参数调优任务，本研究利用贝叶斯优化算法寻找 XGBoost 最优超参数。贝叶斯优化相较于传统的网格搜索优化方法，具有显著的优势。网格搜索通过遍历预设的所有参数组合来寻找最优解，但其计算量开销大，会导致大量计算；相比之下，贝叶斯优化则是一种基于概率模型的全局优化算法，通过构建目标函数的概率模型并逐步选择有可能提供最优结果的超参数组合，从而减少了不必要的运算，显著提升了效率。

本实验采用 Python 中的 Optuna 库进行贝叶斯优化，首先需要定义目标函数，在该函数中通过设定不同的超参数搜索空间，然后通过 Optuna 的 study.optimize() 函数进行自动化的超参数调优，其最大迭代次数 n_trials 设置为 100。Optuna 会根据搜索过程中的反馈调整下次搜索的超参数区间，最终找到最优的超参数配置，贝叶斯优化损失函数下降图如图 3 所示。

从图 3 可知，损失函数呈现出不规律变化特征。这一现象源于贝叶斯优化算法的核心机制：其目标并非确保每轮试验均优于前一轮，而是通过在探索与利用之间动态权衡，以有策略的试错方式，并且动态调整策略，在有限的试验次数内高效逼近

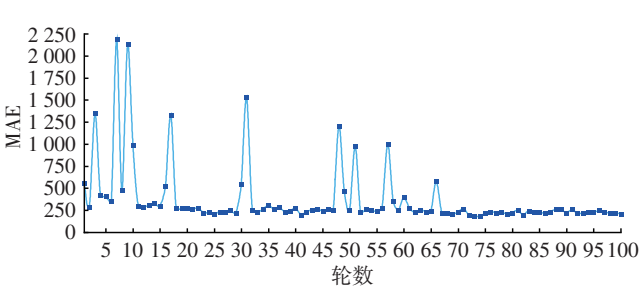


图3 贝叶斯优化损失函数下降图
Fig.3 Bayesian optimization loss function descent

全局最优超参数配置。
在优化过程早期，算法倾向于探索参数空间中尚未充分采样的区域，即便部分区域可能导致模型性能显著下降，从而造成损失曲线剧烈振荡。随着迭代进行，当算法识别出若干高性能的超参数组合后，搜索策略逐渐转向在这些高潜力区域进行精细化采样，使得模型性能趋于稳定。而在第 66 轮后，loss 值进入收敛阶段，波动幅度显著减小，表明优化过程已逐步聚焦于性能较优的局部区域，并接近最优解。

贝叶斯优化后不同时段 XGBoost 超参数如表 3 所示。

表 3 XGBoost 超参数 Tab.3 XGBoost hyperparameters			
名称	说明	时段	值
learning_rate	学习率	早高峰	0.029 9
		平峰	0.027
		晚高峰	0.029
max_depth	树的最大深度	早高峰	4
		平峰	3
		晚高峰	3
n_estimators	树的数量	早高峰	106
		平峰	191
		晚高峰	269
subsample	训练数据采样比例	早高峰	0.807
		平峰	0.831
		晚高峰	0.855
colsample_bytree	每棵树特征采样比例	早高峰	0.988
		平峰时段	0.921
		晚高峰时段	0.937

3.3 不同时间下各模型预测结果

为验证模型的有效性与先进性，同步建立 LightGBM、GBDT 梯度提升树与 BP 神经网络作为对比基准。各模型均基于同一训练集(80% 样本)进行参数优化：LightGBM 选定的超参数组合为：学习率 0.05、最大深度 6、决策树数量 100；GBDT 设置树深度为 4 以防止过拟合，同时学习率设置为 0.15，

决策树数量设置为 87。XGBoost 则使用 3.2 节中得到的超参数,对实验区域内各轨道交通站点出站客流量进行回归分析预测,另外,在模型训练过程中采用早停法,以避免对噪声的过度学习。三时段客流量预测结果对比图如图 4 所示。

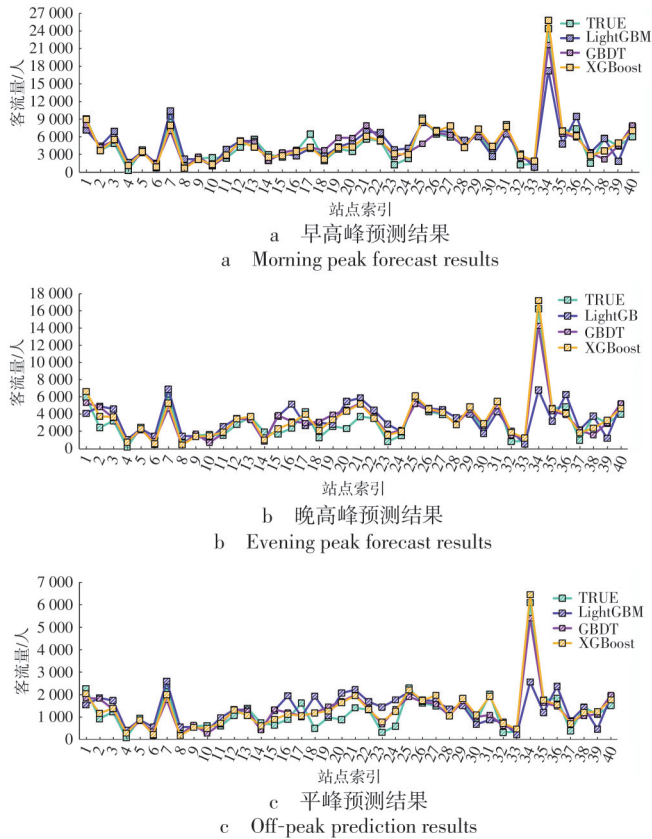


图4 三时段客流量预测结果对比图

Fig.4 Comparison of passenger flow prediction results in three time periods

3.4 模型性能对比

实验以均方根误差(RMSE)、平均绝对误差(MAE)、决定系数(R^2)与平均绝对百分比误差(MAPE)为评估指标,对比 3 个模型在不同时段的预测表现。模型预测性能对比如表 4 所示。

由表 4 可以看出,在不同出行时段的预测实验中, XGBoost 在各个时段的预测精度均优于 LightGBM 与 GBDT,尤其在早高峰、晚高峰与平峰时段,其 R^2 分别达到 0.84, 0.86 与 0.87,说明该模型的预测值与真实值的差异更小,模型能更好地拟合数据。同时, XGBoost 的 MAPE 保持在 10% 以内,低于 LightGBM 与 GBDT,体现了其在相对误差控制上的优势。综合而言, XGBoost 在捕捉轨道交通客

表4 模型预测性能对比

Tab.4 Comparison of model prediction performance

时段	模型	性能指标			
		MAE (↓)	RMSE (↓)	R^2 (↑)	MAPE/% (↓)
早高峰	LightGBM	439.33	581.45	0.68	20.5
	GBDT	288.44	348.62	0.79	12.40
	XGBoost	113.23	222.66	0.84	8.63
平峰	LightGBM	209.58	392.67	0.71	19.03
	GBDT	150.94	238.17	0.81	10.74
	XGBoost	146.75	207.70	0.87	8.69
晚高峰	LightGBM	450.68	541.21	0.73	18
	GBDT	197.38	282.94	0.82	11.9
	XGBoost	267.16	297.45	0.86	9.13

流的复杂非线性关系方面效果最佳。

需要注意的是,从整体趋势来看,早晚高峰的 MAE 与 RMSE 普遍高于平峰,且 R^2 相对较低,这主要是由于高峰时段客流波动剧烈,增加了预测难度^[23]。

另外,还对比了仅考虑历史出站量的时序预测模型,使用 ARMIA 模型基于 2019 年 5 月 1—12 日历史数据对客流量进行回归分析, ARMIA 模型预测性能如表 5 所示。

表5 ARMIA 模型预测性能

Tab.5 Prediction performance of ARIMA model

时段	性能指标			
	MAE (↓)	RMSE (↓)	R^2 (↑)	MAPE/% (↓)
早高峰	467.07	986.93	0.44	61.67
平峰	95.04	162.90	0.76	19.70
晚高峰	240.34	462.33	0.51	40.92

由表 5 可以看出,在仅利用有限历史数据的情况下,基于 ARIMA 模型的回归预测精度明显低于根据建成环境因子进行回归的预测精度。这表明,当历史观测数据不足时,单纯依赖历史数据进行回归预测的方法,其效果不及充分利用建成环境特征的回归预测方法。

上述实验结果表明,基于建成环境因子的轨道交通客流量预测方法具有可行性和显著优势。此外, XGBoost 模型不仅能够有效捕捉数据变化规律,具备较强的解释能力,而且预测结果误差较小、精度较高。在轨道交通出行需求预测中,该模型表现出良好的适应性与可靠性,能够满足实际应用中的精度要求。

3.5 SHAP 分析

3.5.1 特征贡献度分析

精准解析建成环境特征对轨道交通需求的贡献度，是优化城市空间布局与交通资源配置的科学基础。基于 SHAP 值解析模型预测逻辑，可以揭示各指标特征的贡献。影响因子 SHAP 值如图 5 所示。

图 5a 是根据平均 SHAP 值对这些变量的重要性进行排序，可以看出，影响轨道交通客流量的建成环境因子中，停车场密度和介数中心性是 2 个最重要的因素，商业区密度、功能密度、公交线路数量也是本研究的其他重要特征；图 5b 中，停车场密度和介数中心性不仅具有较高的重要性，而且高取值对应的 SHAP 值主要分布于正值区域，但局部样本存在负向影响；商业区密度和功能密度同样表现出较为显著的正向影响，高特征值对应较大的正 SHAP 值；公交线路数量也呈现出相似规律。

停车场密度、商业区密度、功能密度等对轨道交通客流量影响大说明服务设施对客流量有很大的吸引作用，同时也表明了高密度的服务设施布局可以提高地铁的利用率^[24]。此外，从图 5b 可以看出，停车场的分布可能对城市轨道交通站点的客流量有正向影响，但在部分样本中，高停车场密度对应负的 SHAP 值，说明停车设施对客流的影响具有一定的非线性，不同区域可能表现出不同的作用机制。在 Henry 等^[25]的研究中提到，部分居民采用“P+R” (Park and Ride) 模式，即先驾车至城市轨道交通站点附近停车，然后换乘城市轨道交通进入市中心，因此若停车设施较为集中，可能会增加客流。反之，高停车场密度也可能意味着居民更倾向于使用私家车出行，也可能减少部分城市轨道交通客流。

另外，介数中心性指标在客流预测中的重要性也较为突出，表明网络结构特征在轨道交通流量预测中起到了关键作用。介数中心性较高的区域往往处于交通网络的关键节点，承担着较大的客流吞吐量^[26]。

而站点距离市中心距离这一因子贡献度较低，这一结果可能由于北京市为多中心布局，天安门虽然是传统意义上的市中心，但并非唯一的核心区域。国贸、中关村、望京这些城市副中心、金融商务区以及其他功能区同样具备强大的客流吸引力，

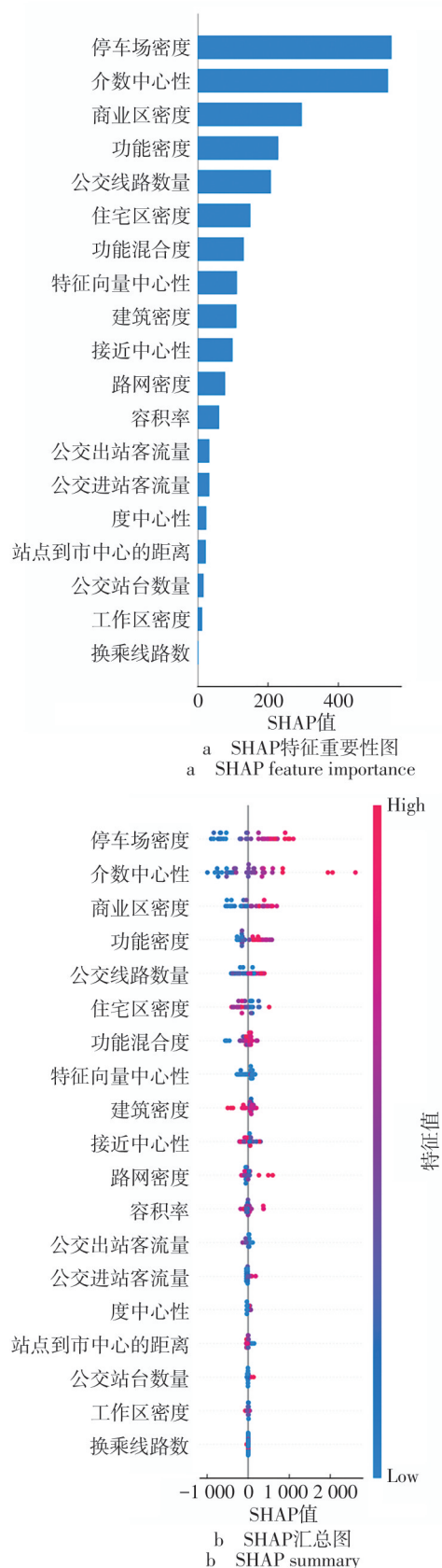


图 5 影响因子 SHAP 值

Fig.5 SHAP values of influencing factors

使得站点与天安门的直线距离无法充分反映其通勤需求和人流密度,这一点在 Wang 等^[27]的研究中也有所提及。工作区密度的 SHAP 重要性远小于住宅区密度和商业区密度,表明在功能分区规划中,住宅和商业区对轨道交通客流的吸引力更为显著,应重点关注。而换乘线路数的重要性最小,说明其影响力有限,仅能作为其他要素的补充。

在未来对于北京市轨道交通选址规划中,应重点考虑选址位置的停车场密度、中心性、商业区密度和公交线路等 SHAP 贡献度大的建成环境因子。另外,还需要重视规划站点与已建成站点的联系,更好地发挥其在轨道交通网络上的作用。相比之下,对于换乘线路数和距离市中心的距离等 SHAP 贡献度较低的因子,可适当降低其权重,侧重将资源和注意力集中在对整体网络效能具有更大影响的关键因素上,从而实现更为科学、合理和可持续的轨道交通规划。

3.5.2 交互效应分析

通过 SHAP 交互效应与主效应分离,可以获得额外的信息^[28]。本研究对不同指标的 SHAP 值进行了对比分析。停车场密度和中心性指标对比图如图 6 所示,停车场密度较高的站点,通常在交通网络中表现出较高的中心性。反之,停车场密度较低的站点,其中心性指标也较低,说明停车场密度和中心性两者之间存在较强的正相关关系,进一步验证了停车场资源分布在一定程度上反映了站点对整个交通系统的重要性。

住宅区、工作区、商业区密度对比图如图 7 所示,结果表明,各功能区的密度指标之间存在较高的协同效应,即住宅区密度较高的站点,其周边工作区和商业区密度亦呈现出较高水平,反之亦然。进一步观察发现,当住宅区密度超过 100 时,其对应的 SHAP 值显著上升,对城市轨道交通客流量预测产生较强的正向影响;而当住宅区密度低于 50 时,则会对预测结果产生较大幅度的负向作用。商业区的情况亦呈现类似趋势:当商业区密度大于 750 时,其正向影响明显,而小于 250 时则对预测结果构成较大负向影响。相比之下,工作区密度的 SHAP 值普遍较小,对模型预测结果的影响相对有限。

从规划角度出发,基于上述分析结果,本研究

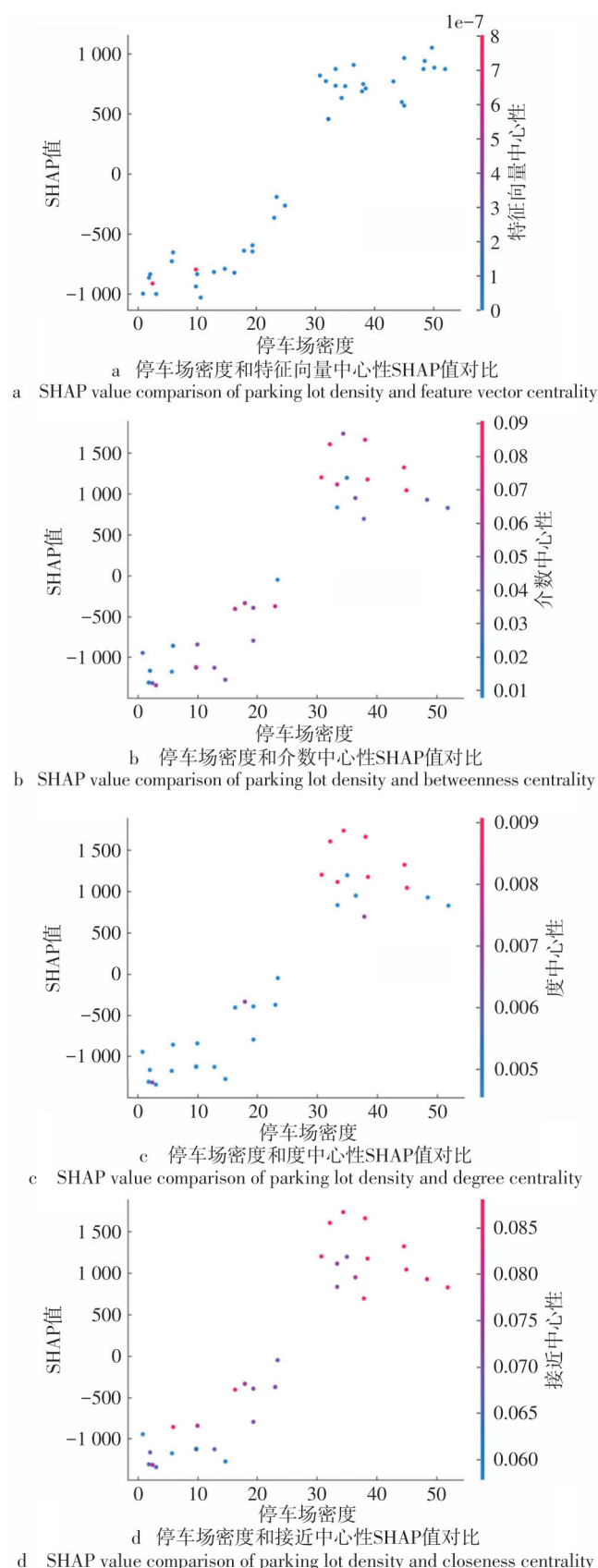


图6 停车场密度和中心性指标对比图

Fig.6 Comparison of parking lot density and centrality index

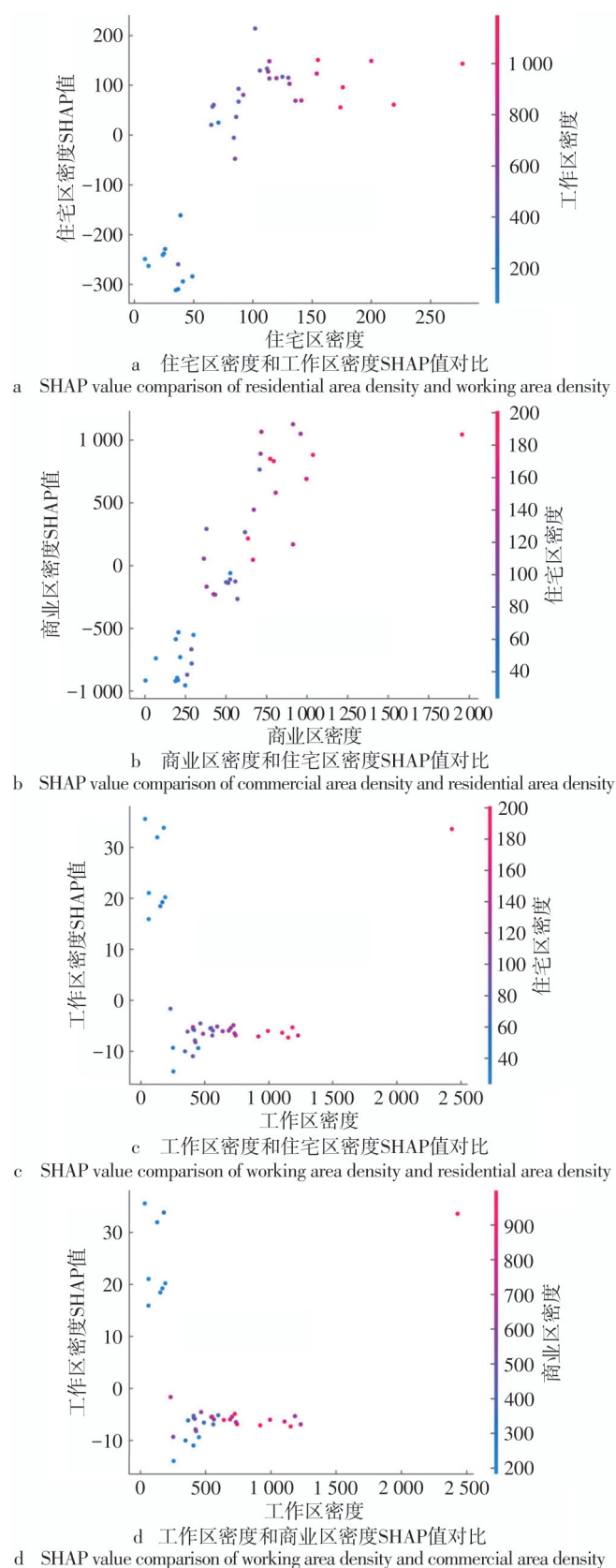


图7 住宅区、工作区、商业区密度对比图

Fig.7 Density comparison of residential area, working area, and commercial area

提出以下3点启示。①在SHAP分析中,住宅区密度与商业区密度对客流量有较强的影响,另外住宅区与商业区的密度协同效应表明,区域内不同功能区之间的联系对客流有显著影响。因此,优化站点周边的职住平衡和商业布局可以提升区域整体客流效益,通过调整站点周围住宅区与商业区的密度差距,使得这些站点客流量有更为合理的分布,从而形成良性的人流、车流及经济流互动效应,为城市轨道交通运营提供稳固的客流支撑。②公交线路数量也对客流量的特征贡献度较高,通过优化公交线路布局,例如增加公交线路覆盖的站点数,发挥公交在疏解和分流城市轨道交通客流中的作用,能够有效减缓城市轨道交通站点的交通需求压力。③停车场密度与网络中心性之间的正相关关系表明,停车场的分布直接影响站点的交通需求压力,必须合理优化停车场分布,以减轻由“P+R”引发的客流集中问题。特别是在中心性较低的城市轨道交通站点附近,适当增加停车场密度,以吸引部分驾车出行者转而利用低负荷站点的换乘服务,可以显著增加客流;而在中心性较高的站点则应适度减少停车设施供给,以避免过量停车引发的客流集中问题,从而实现核心区域客流压力的有效缓解。

4 结论与展望

本研究在建成环境的约束条件下,以北京五环内城市轨道交通站点为对象,通过分析城市轨道交通智能卡数据提取客流量作为因变量,同时利用多源空间数据分析技术提取建成环境特征作为自变量,构建了多维度指标体系。采用XGBoost算法对轨道交通出行需求进行预测,结合SHAP分析建成环境变量的重要性,构建了XGBoost-SHAP预测模型,同时利用贝叶斯优化算法对XGBoost模型超参数进行优化,深入探索不同建成环境要素对轨道交通客流的影响。

在XGBoost-SHAP模型构建阶段,采用贝叶斯优化算法与5折交叉验证优化XGBoost超参数,并使用SHAP解析特征贡献度的非线性机制。为验证模型有效性,将XGBoost模型与LightGBM,GBDT,BP神经网络进行对比实验。另外,基于SHAP分析结果,本研究定量评估了各建成环境因子对城市轨道交通

交通客流变化的贡献度。实验结果表明：①XGBoost模型在3个时段预测结果优于LightGBM, GBDT和BP神经网络模型。②停车场密度和介数中心性对轨道交通客流的影响最为显著；与此同时，商业区密度、功能混合度和公交线路数量等指标也对客流变化具有不可忽视的影响作用。上述研究结果可为轨道交通站点选址规划提供方法支撑。

鉴于我国城市发展规模和阶段的不一致性，本研究得到的空间效应与规划启示是否适用于其他地区仍需进一步验证。北京市具有独特的交通网络结构和人口分布特征，因此，模型在其他城市尤其是中小型城市中的应用，可能会受到城市规模、功能区分布、交通密度等因素的影响。未来的研究可以考虑在其他城市尤其是中小城市进行测试和调整，以验证其跨城市的泛化能力，还可以选择其他时段来验证模型的时间泛化能力。另外，本研究未能整合实时感知数据及共享单车、出租车等多源数据，这些数据的动态变化能够为交通需求的预测提供更加精准的支持。未来研究可进一步对共享单车、出租车等多源数据进行整合，探索多模态交通需求协同预测的综合框架，为智慧城市建设和可持续发展提供更为精细化、动态化的技术支撑。

参考文献：

- [1] 黄建欣, 龚蔚霞, 张金林, 等. 城市轨道交通站点周边建成环境对地铁客流量的影响研究：以深圳市为例[J]. 上海城市规划, 2023, 6(6): 47-54.
HUANG Jianxin, GONG Weixia, ZHANG Jinlin, et al. Impact of Built Environment of Urban Rail Transit Stations on Metro Passenger Flow: A Case Study of Shenzhen[J]. Shanghai Urban Planning Review, 2023, 6(6): 47-54.
- [2] 袁振洲, 袁晓敬, 杨 洋, 等. 城市轨道交通低碳效应研究综述[J]. 铁道运输与经济, 2024, 46(8): 8-23, 57.
YUAN Zhenzhou, YUAN Xiaojing, YANG Yang, et al. Review on Research of Low-Carbon Effects in Urban Rail Transit[J]. Railway Transport and Economy, 2024, 46(8): 8-23, 57.
- [3] 交通运输部. 全国城市轨道交通年客运量超322亿人次[EB/OL]. (2025-01-22) [2025-08-29]. https://www.mot.gov.cn/fenxigongbao/yunlifexi/202503/t20250310_4165357.html.
- [4] 周庆梅, 何希平. 地铁乘客流量短时预测方法综述[J]. 重庆工商大学学报(自然科学版), 2020, 37(1): 25-32.
- [5] ZHOU Qingmei, HE Xiping. Review of Methods for Short-Term Prediction of Subway Passenger Flow[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2020, 37(1): 25-32.
- [6] LIU C Q, LI L. How Do Subways Affect Urban Passenger Transport Modes? Evidence from China[J]. Economics of Transportation, 2020, 23: 100181.
- [7] WANG J D, OKTOMY NOTO SUSANTO C. Traffic Flow Prediction with Heterogenous Data Using a Hybrid CNN-LSTM Model[J]. Computers, Materials & Continua, 2023, 76(3): 3097-3112.
- [8] PAN Y A, LI F L, LI A R, et al. Urban Intersection Traffic Flow Prediction: A Physics-Guided Stepwise Framework Utilizing Spatio-Temporal Graph Neural Network Algorithms[J]. Multimodal Transportation, 2025, 4(2): 100207.
- [9] 韩 皓, 徐圣安, 赵 蒙. 考虑线网结构的LightGBM轨道交通短时客流预测模型[J]. 铁道运输与经济, 2021, 43(10): 109-117.
HAN Hao, XU Sheng'an, ZHAO Meng. LightGBM Prediction Model of Short-Term Passenger Flow for Rail Transit Considering Network Structure[J]. Railway Transport and Economy, 2021, 43(10): 109-117.
- [10] 徐明明, 唐秋生. 基于烟花粒子群算法优化BP神经网络的轨道客流量预测[J]. 科学技术与工程, 2024, 24(33): 14410-14416.
XU Mingming, TANG Qiusheng. Prediction of Railway Passenger Flow Based on Fireworks Particle Swarm Optimization Algorithm Optimized BP Neural Network[J]. Science Technology and Engineering, 2024, 24(33): 14410-14416.
- [11] 周爱娇, 杨 静, 周浪雅, 等. 城市轨道交通站点与建成环境协调性评估及针对性发展策略研究[J]. 铁道运输与经济, 2025, 47(3): 161-170.
ZHOU Aijiao, YANG Jing, ZHOU Langya, et al. Assessment of Coordination between Urban Rail Transit Stations and Built Environment and Development Strategies[J]. Railway Transport and Economy, 2025, 47(3): 161-170.
- [12] LÓPEZ BAEZA J, CARPIO-PINEDO J, SIEVERT J, et al. Modeling Pedestrian Flows: Agent-Based Simulations of Pedestrian Activity for Land Use Distributions in Urban Developments[J]. Sustainability, 2021, 13(16): 9268.
- [13] 李 鹏. 城市轨道交通对容积率影响的定量分析：以东京

- 为例[J]. 铁道运输与经济, 2024, 46(6): 207–214.
- LI Peng. Quantitative Analysis of the Influence of Urban Rail Transit on Floor Area Ratio: A Case Study of Tokyo[J]. Railway Transport and Economy, 2024, 46(6): 207–214.
- [13] 汪雨菲, 杨皓森, 喻冰洁, 等. 站域建成环境与地铁客流量的非线性关系和协同效应: 可解释机器学习分析[J]. 都市快轨交通, 2024, 37(2): 1–7.
- WANG Yufei, YANG Haosen, YU Bingjie, et al. Nonlinear and Synergistic Effects of Station-Area Built Environments on Metro Ridership: A Shapley Additive Explanations(SHAP) Analysis[J]. Urban Rapid Rail Transit, 2024, 37(2): 1–7.
- [14] 王 晶, 路羨乔, 陆化普. 城市轨道交通车站TOD发展水平与城市活力协同性评价: 以北京市轨道微中心为例[J]. 城市交通, 2024, 22(6): 40–50.
- WANG Jing, LU Xianqiao, LU Huapu. Evaluation of the Coordination between TOD Development Level and Urban Vitality at Urban Rail Transit Stations: A Case Study of Rail Transit Micro-Centers in Beijing[J]. Urban Transport of China, 2024, 22(6): 40–50.
- [15] LEI J Y, HE M, SHI Z B, et al. How Does the Built Environment Affect Intermodal Transit Demand across Different Spatiotemporal Contexts? [J]. Journal of Transport Geography, 2024, 121: 104033.
- [16] LI S Y, LYU D J, LIU X P, et al. The Varying Patterns of Rail Transit Ridership and Their Relationships with Fine-Scale Built Environment Factors: Big Data Analytics from Guangzhou[J]. Cities, 2020, 99: 102580.
- [17] 张 艳, 朱潇冰, 杨培霖, 等. “5D+”视角下城市轨道交通站点的TOD效能评价: 深圳市的实证研究[J]. 现代城市研究, 2022(9): 75–82.
- ZHANG Yan, ZHU Xiaobing, YANG Peilin, et al. TOD Performance Evaluation of Urban Rail Transit Stations from the Perspective of “5D+ ” Dimensions: An Empirical Study of Shenzhen[J]. Modern Urban Research, 2022(9): 75–82.
- [18] HE Z, LIN D, LAU T, et al. Gradient Boosting Machine: A Survey[EB/OL]. [2025–04–15]. arXiv: 1908.06951, 2019.
- [19] FRAZIER P I. A Tutorial on Bayesian Optimization[EB/OL]. [2025–04–15]. arXiv: 1807.02811, 2018.
- [20] LUNDBERG S M, LEE S I. A Unified Approach to Interpreting Model Predictions[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. December 4–9, 2017, Long Beach, California, USA. ACM, 2017: 4768–4777.
- [21] 王启帆. 基于数据增强的小样本学习方法研究[D]. 北京: 北京交通大学, 2023.
- [22] 程国柱, 吕岩峰, 冯天军. 灾害性天气条件下城市轨道交通客流DRN-BiLSTM预测模型[J]. 哈尔滨工业大学学报, 2025(7): 153–161.
- CHENG Guozhu, LYU Yanfeng, FENG Tianjun. Urban Rail Transit Passenger Flow DRN-BiLSTM Combined Forecasting Model under Catastrophic Weather Conditions[J]. Journal of Harbin Institute of Technology, 2025(7): 153–161.
- [23] 李 懿. 基于SSA-LSSVM的短时交通流预测研究[D]. 沈阳: 沈阳大学, 2023.
- [24] XI Y F, HOU Q H, DUAN Y Q, et al. Exploring the Spatiotemporal Effects of the Built Environment on the Nonlinear Impacts of Metro Ridership: Evidence from Xi’an, China[J]. ISPRS International Journal of Geo-Information, 2024, 13(3): 105.
- [25] HENRY E, FURNO A, EL FAOUZI N E, et al. Locating Park-and-Ride Facilities for Resilient On-Demand Urban Mobility[J]. Transportation Research Part E: Logistics and Transportation Review, 2022, 158: 102557.
- [26] LIU J, MENG B, SHI C S. A Multi-Activity View of Intra-Urban Travel Networks: A Case Study of Beijing[J]. Cities, 2023, 143: 104634.
- [27] WANG J, WAN F, DONG C J, et al. Spatiotemporal Effects of Built Environment Factors on Varying Rail Transit Station Ridership Patterns[J]. Journal of Transport Geography, 2023, 109: 103597.
- [28] LUNDBERG S M, ERION G G, LEE S I. Consistent Individualized Feature Attribution for Tree Ensembles[EB/OL]. 2018: arXiv: 1802.03888. <https://arxiv.org/abs/1802.03888>.
- (责任编辑 吴文娟)