doi: 10.3969/j.issn.1672-6073.2025.05.008

基于多因素分析的地铁客流预测方法实证研究

郑宣传[1]，秦　勇[2]，郭建媛[1]，佟　鑫[3]，黄小林[4]
（1. 北京交通大学交通运输学院，北京 100044；2. 北京交通大学先进轨道交通自主运行全国重点实验室，北京 100044；3. 北京城建设计发展集团股份有限公司，北京 100037；4. 西安市轨道交通集团有限公司运营分公司，西安 710016）

摘　要：面向多因素下地铁线路客运量预测需求，通过特征挖掘、相关性分析，从时间、天气、线网规模、城市发展、节假日及活动等因素选取 12 项影响特征指标，针对 XGBoost 模型超参数优化难题，提出基于斑马优化算法(ZOA)的超参数优化算法；构建 XGBoost-ZOA 组合预测模型，设计时序数据预测的交叉验证方案，并在西安地铁开展实例验证。经 ZOA 算法优化后的预测模型，在 1 号线测试集上的预测误差平均误差百分比 MAPE 为 3.85%，与经典的智能优化算法相比其预测效果更佳，证实该算法的先进性。XGBoost-ZOA 组合预测模型在 9 条线路的平均 MAPE 达到 5.18%，最大误差不超过 7%，预测值与真实值曲线吻合度较高，同比 SARIMA 模型及 LSTM 模型，MAPE 分别减少了 17.8%及 21.9%，均方根误差 RMSE 分别减少了 10.8%及 13.6%。最后，基于模型特征重要度，将 9 条线路分为趋势客流导向、假日活动客流导向和多因素均衡导向 3 类，发现假日活动主导的线路客流预测误差较大，周期趋势主导的线路客流预测误差较小，证实了该模型的解释性及实用性较好。

关键词：城市轨道交通；多因素客流预测；XGBoost；斑马优化算法；超参数优化；特征重要度
中图分类号：U231　　**文献标志码：**A　　**文章编号：**1672-6073(2025)05-0057-07

Empirical Study of a Multi-Factor Subway Passenger Flow Forecasting Method

ZHENG Xuanchuan[1], QIN Yong[2], GUO Jianyuan[1], TONG Xin[3], HUANG Xiaolin[4]
(1. School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044; 2. State Key Laboratory of Advanced Rail Autonomous Operation, Beijing Jiaotong University, Beijing 100044; 3. Beijing Urban construction Design and Development Group Co., Ltd., Beijing 100037; 4. Operating Branch, Xi'an Rail Transit Group Co., Ltd., Xi'an 710016)

Abstract: To address multi-factor subway line passenger volume forecasting needs, 12 influential features are selected (time, weather, network scale, urban development, holidays, and events) via feature mining and correlation analysis, and—targeting the difficulty of XGBoost hyperparameter tuning—a hyperparameter optimization approach based on the Zebra Optimization Algorithm (ZOA) is proposed. An XGBoost-ZOA integrated forecasting model is developed, a cross-validation scheme tailored for time-series prediction is designed, and a case study is conducted using Xi'an Metro data. The ZOA-optimized model achieves a MAPE of 3.85% on the Line 1 test set; compared with representative intelligent optimization algorithms, it yields the best performance, confirming the algorithm's superiority. Across nine lines, the XGBoost-ZOA model attains an average MAPE of

收稿日期：2025-04-08　修回日期：2025-07-07
第一作者：郑宣传，男，博士研究生，高级工程师，从事城市轨道交通客流预测及仿真，20114089@bjtu.edu.cn
基金项目：北京市自然科学基金–丰台轨道交通前沿研究联合基金(L221006，L231003)
引用格式：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.

5.18%, with the maximum error below 7%, and the predicted and actual curves show a high degree of agreement. Compared with SARIMA and LSTM, the MAPE decreases by 17.8% and 21.9%, and the RMSE decreases by 10.8% and 13.6%, respectively. Finally, using feature importance, the nine lines were grouped into three categories: trend-driven, holiday/activity-driven, and multi-factor balanced passenger flows. Lines dominated by holidays/activities exhibit larger prediction errors, whereas those governed by periodic trends show smaller errors, confirming the model's interpretability and practical utility.

Keywords: urban rail transit; multi-factor passenger flow forecasting; XGBoost; zebra optimization algorithm; hyperparameter optimization; feature importance

0 引言

准确的客流数据是制定地铁运输组织方案及开展客运组织管理的基础。自 2023 年以来，各城市节假日旅游客流回升，演唱会活动频率突增，加之极端天气频发，导致地铁线路客流变化趋势难以精准预测，给运营组织带来极大压力。研究各类因素对客流变化的影响特征，建立多因素下的精准客流预测模型，具有重要的学术价值与工程应用价值。

从影响因素分析，地铁客流波动受趋势性、周期性及临时性 3 类因素影响。趋势性因素与城市人口、GDP、线网规模[1]、建成环境[2-3]等宏观因素相关，导致客流发生趋势性变化；周期性因素指随月份、星期变化客流呈周期性变动；临时性因素包括节假日[4]、天气[5]、大型活动[6]等，客流发生临时突变。已有模型仅针对单一因素的预测问题，但实际客流是受多因素综合影响，因此如何剖析客流与各因素的关系，构建多因素下的客流预测模型是亟待解决的难题。

从预测方法分析，通常可分为 3 类：一是时间序列预测模型，如自回归移动平均模型(ARIMA)[7]、季节性差分自回归滑动平均模型(SARIMA)[8]，基于时间序列统计特征进行单步及多步滚动预测；二是基于机器学习的集成树模型，如随机森林(random forest，RF)[9]、极限梯度提升算法(eXtreme gradient boosting，XGBoost)[10]等，基于多棵决策树可实现多因素影响下回归预测，在较少的样本数据可取得较好的初始预测精度，如何优化超参数是提升该类模型预测精度的关键；三是基于深度学习的预测模型，如长短时记忆神经网络(LSTM)[11]、Transformer[12]等，此类模型可输入多因素，在大量的训练数据集可取得很好的效果，其“黑箱预测”属性导致模型可解释性较低。

综合考虑预测模型应具备多因素输入、多步预测、可解释性的能力，选择 XGBoost 作为预测模型，首先通过特征相关性分析及探讨选取 12 项影响特征；其次，针对原始 XGBoost 模型预测精度低，超参数优化难的问题，结合斑马优化算法(zebra optimization algorithm，ZOA)[13]实现超参数自动优化，建立 XGBoost-ZOA 组合预测模型；最后，结合西安地铁客流数据开展实例验证，对比分析不同模型的预测精度、预测值与真实值的吻合度，分析模型特征的重要度与实际情况的匹配度，探讨本文提出的 XGBoost-ZOA 模型的优点及实用价值。

1 影响特征分析及选取

1.1 影响特征分析

本文初步选取影响地铁客运量的 15 项特征指标，包含时间(F1～F3)、天气(F4～F6)、线网规模(F7～F8)、城市发展(F9～F10)、节假日(F11～13)及活动(F14～F15) 6 类因素的综合影响，如表 1 所示。

表 1 地铁客运量的影响特征

Table 1 Influential features of subway passenger flow

序号	特征名称	特征说明
F1	日期	数值，以天数表示
F2	月份	1～12 代表月份
F3	星期	1～7 表示星期一至星期日
F4	最高气温	当日最高气温
F5	最低气温	当日最低气温
F6	天气值	含晴、雨、雪、大风等
F7	车站数	全网车站数
F8	换乘站	全网换乘站数量
F9	常住人口	城市常住人口
F10	GDP	城市 GDP
F11	节日类型	如元旦、五一、十一等，转为数字编码
F12	节假日	是否为节假日，0-否，1-是
F13	调休日	考虑假期轮休，0-工作，1-休息
F14	活动类型	0-无赛事、1-体育赛事、2-演唱会
F15	活动地点	1-体育场、2-体育馆

1.2 特征相关性分析及指标选取

对线路客运量及上述 15 项特征量做相关性分析，剔除冗余特征，如图 1 所示。车站数与换乘站数、常

住人口与 GDP、最高气温与最低气温、活动地点及活动类型四组的相关系数超过 0.9，相关性较高，其余特征的相关性系数未超过 0.9，具有一定独立性，故删除最低气温、GDP、活动地点 3 项特征。因全网车站数与换乘站数对客运量的影响具有一定独立性，保留 12 项指标作为自变量，如表 2 所示。

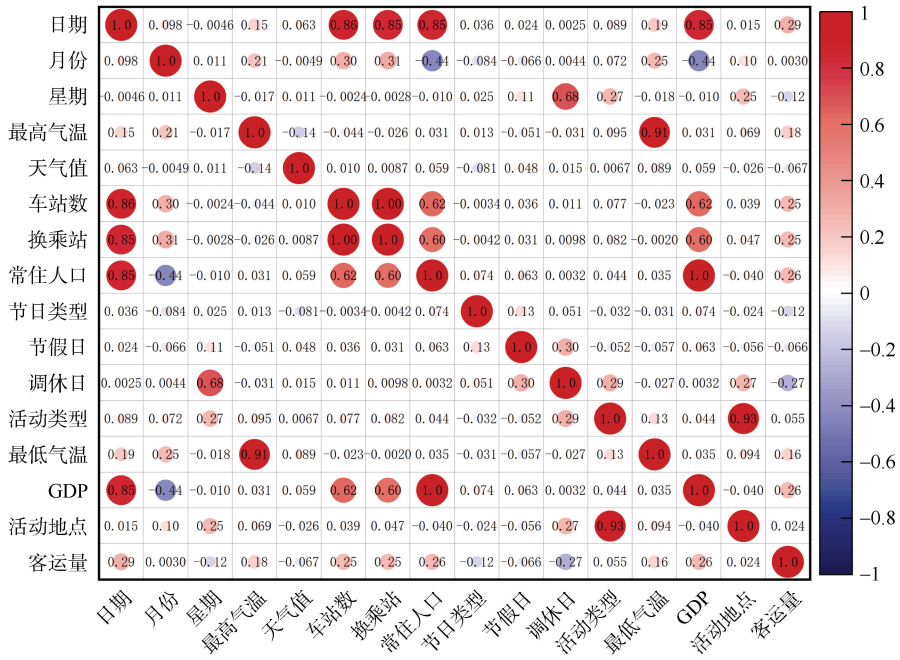


图 1　特征的相关系数热力图

Figure 1　Feature correlation heat map

表 2　模型输入的特征量样表

Table 2　Sample of model input features

日期	月份	星期	最高气温/℃	天气值	车站数/个	换乘站/个	常住人口/万人	节日类型	节假日	调休日	活动类型
2023-04-25	4	2	24	2	173	20	1 307.82	4	0	0	0
2023-04-26	4	3	24	3	173	20	1 307.82	4	0	0	0
2023-04-27	4	4	27	6	173	20	1 307.82	4	0	0	0
2023-04-28	4	5	21	2	173	20	1 307.82	4	0	0	0
2023-04-29	4	6	24	2	173	20	1 307.82	2	1	1	0
…							…				

2　XGBoost–ZOA 组合预测模型

2.1　XGBoost 预测模型原理

XGBoost 是一种改进的梯度提升树算法，其核心思想就是基于特征分裂来生长新的树，每添加新的一棵树就是学习一个新函数，旨在拟合前一棵树的预测残差值。本文期望预测线路客运量，其是根据输入的多维特征，查找落在每棵树的叶子节点上的分数，并在每棵树对应的分数累加得到该样本的预测值 $\hat{y}_i$，其预测模型为

$$\hat{y}_i = g(\boldsymbol{x}_i) = \sum_{k=1}^{K} f_k(\boldsymbol{x}_i), f_k(\boldsymbol{x}_i) = w^k_{q(\boldsymbol{x}_i)} \tag{1}$$

式中，$\hat{y}_i$ 为预测值；$g(x_i)$为预测模型；K 为树的总数量；$\boldsymbol{x}_i$ 为样本 i 对应的特征向量；$f_k(\boldsymbol{x}_i)$为在第 k 棵树的得分；$w^k_{q(\boldsymbol{x}_i)}$ 为 $\boldsymbol{x}_i$ 落在第 k 棵树的叶子节点 $q(\boldsymbol{x}_i)$的得分值。

XGBoost 模型的目标函数是损失最小化，损失函数为

$$L = \sum_{i=1}^{m} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{2}$$

式中，$\sum_{i=1}^{m} l(y_i, \hat{y}_i)$ 为误差项；$l(y_i, \hat{y}_i)$ 为在每次迭代中，样本 i 的真实值 y_i 与预测值 $\hat{y}_i$ 的偏差；m 为样本数量；$\Omega(f_k)$ 代表模型的复杂度，其公式为

$$\Omega(f_k)=\gamma T+\frac{1}{2}\alpha\sum_{j=1}^{T}\left|w_j\right|+\frac{1}{2}\lambda\sum_{j=1}^{T}w_j^2 \tag{3}$$

式中，γT 用于控制树的结构；T 为叶子节点数量；$\frac{1}{2}\alpha\sum_{j=1}^{T}\left|w_j\right|$ 为 L1 正则项；$\frac{1}{2}\lambda\sum_{j=1}^{T}w_j^2$ 为 L2 正则项；w_j 为叶子节点权重；γ、α、λ 均为超参数。

2.2 ZOA 算法原理

ZOA[13]是一种新型启发式优化算法，通过模拟自然界斑马的捕食行为(阶段 1)及防御行为(阶段 2)实现参数优化求解，其算法步骤如下：

①输入待优化模型的数据；②设置种群数量 N 和最大迭代次数 T；③创建初始种群；④基于初始种群计算目标函数适应度；⑤更新先锋斑马位置；⑥阶段 1：基于公式(4)计算 $x_{i,j}^{\text{new,P1}}$：

$$x_{i,j}^{\text{new,P1}}=x_{i,j}+r\cdot(PZ_j-I\cdot x_{i,j}) \tag{4}$$

式中，$x_{i,j}^{\text{new,P1}}$ 为斑马 i 在阶段 1 新解的第 j 维分量；$x_{i,j}$ 为斑马 i 当前解的第 j 维分量，PZ_j 为先锋斑马解的第 j 维分量；r 为介于[0，1]的随机数；I=round (1+rand)，rand 为 0～1 的随机数，round 为四舍五入函数，故 $I\in\{1，2\}$。

⑦阶段 1：基于公式(5)计算 X_i：

$$X_i=\begin{cases}X_i^{\text{new,P1}},F_i^{\text{new,P1}}<F_i\\X_i,\quad\text{else}\end{cases} \tag{5}$$

式中，$X_i^{\text{new,P1}}$、$F_i^{\text{new,P1}}$ 为斑马 i 阶段 1 新解及目标函数值，X_i、F_i 为斑马 i 当前解向量及目标函数值。

⑧阶段 2：产生 0–1 的随机数 P_s；若 $P_s\leqslant0.5$，则转入步骤⑨，否则转入步骤⑩。

⑨阶段 2：基于公式(6)的逃跑策略 S_1 计算 $x_{i,j}^{\text{new,P2}}$，转入步骤⑪。

⑩阶段 2：基于公式(6)的进攻策略 S_2 计算 $x_{i,j}^{\text{new,P2}}$，转入步骤⑪。

$$x_{i,j}^{\text{new,P2}}=\begin{cases}S_1:x_{i,j}+R\cdot(2r-1)\cdot\left(1-\dfrac{t}{T}\right)\cdot x_{i,j},P_s\leqslant0.5\\S_2:x_{i,j}+r\cdot(AZj-I\cdot x_{i,j}),\quad\text{else}\end{cases} \tag{6}$$

式中，$x_{i,j}^{\text{new,P2}}$ 为斑马 i 新解的第 j 维分量；R 为常数 0.01；AZj 为被攻击斑马的第 j 维分量；t 为当前迭代次数；P_s 为 0～1 的随机数。

⑪阶段 2：基于公式 7 更新 X_i

$$X_i=\begin{cases}X_i^{\text{new,P2}},F_i^{\text{new,P2}}<F_i\\X_i,\quad\text{else}\end{cases} \tag{7}$$

⑫若当前斑马序号 i=N，则保留目前的最优解，否则遍历下只斑马 i=i+1，转入步骤⑤。

⑬若当前轮次 t=最大迭代次数 T，则输出最优解，算法结束；否则 t=t+1，转入步骤⑤。

2.3 基于 ZOA 的 XGBoost 模型超参数优化

XGBoost 模型中的主要超参数及初始范围如表 3 所示，主要包含 8 项，优化复杂度及难度较大。

表 3 XGBoost 模型的超参数

Table 3 Hyperparameters of the XGBoost Model

参数指标	超参数含义	取值范围
P1	决策树的数量，式(1)的 K	[5，500]
P2	学习率	[0.01，0.5]
P3	树的最大深度	[3，10]
P4	每棵树的样本采样比例	[0.5，0.8]
P5	每棵树的特征采样比例	[0.5，0.8]
P6	L1 正则化项权重，式(3)的 α	[0.0，1.0]
P7	L2 正则化项权重，式(3)的 λ	[0.0，1.0]
P8	分裂最小增益，式(3)的 γ	[0，10]

主流的超参数优化方法包括：网格搜索、随机搜索、贝叶斯优化方法，以及基于智能优化算法(如粒子群算法(PSO)、遗传算法(GA)、差分进化算法(DE)、灰狼优化算法(GWO))，本文将 ZOA 算法的优化效果与前 5 种算法对比。ZOA 算法的模型参数少，除公式(6)的 R 为常数 0.01，待优化参数为种群数量 N 和最大迭代次数 T，本文所有算法中均设置 N=20，T=100，贝叶斯优化方法的计算次数设为 20×100=2 000 次，便于对比同等计算轮次下的优化结果。

ZOA 算法的目标函数适应度为 XGBoost 预测模型的平均误差百分比(MAPE)，解集为 XGBoost 模型的超参数。超参数的优化过程在训练集中开展，将训练集划分为 M 组的交叉训练集-验证集，将 M 组验证集的预测误差 MAPE 的均值，作为 ZOA 算法中每次迭代计算的适应度函数值，实现验证集上的平均 MAPE 值最小化，提升超参数的泛化能力，最后在测试集上检验预测结果。

3 算例分析

3.1 数据来源及验证方案

本文以西安地铁 9 条线路为例开展算例验证，数据样本包含 560 条客运量数据，时间范围为 2023 年 2 月 3 日至 2024 年 8 月 15 日，如图 2 所示。

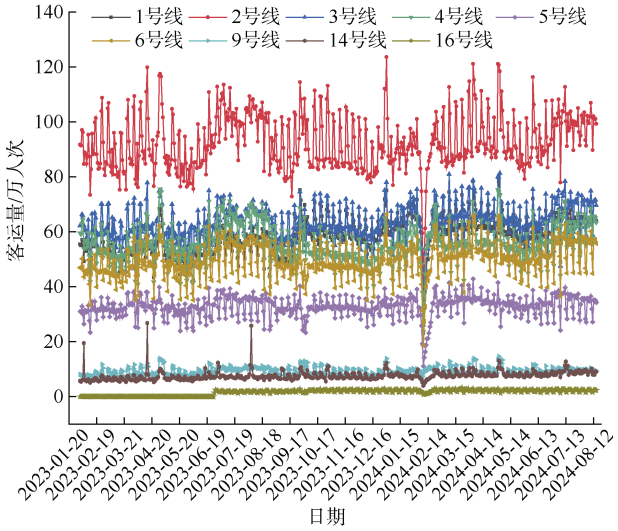


图 2　地铁线路客运量的分布

Figure 2　Distribution of passenger volume on subway lines

本文按 80%、20%的比例划分训练集与测试集。在训练集用于超参数寻优，测试集上检验模型预测效果。在训练集中设计了图 3 所示的交叉验证方案，共 4 组，第 1 组前 336 d 数据为交叉训练集，后 28 d 数据为验证集，后 1 组依次向后滑动。

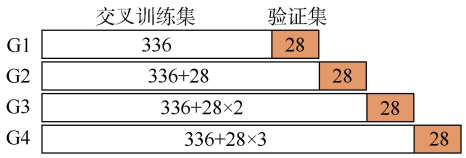


图 3　交叉验证方案

Figure 3　Cross-validation scheme

3.2　模型超参数优化结果对比

以 1 号线为例，图 4 为在训练集上采用不同智能优化算法得到的 XGBoost 模型超参数优化结果，经过

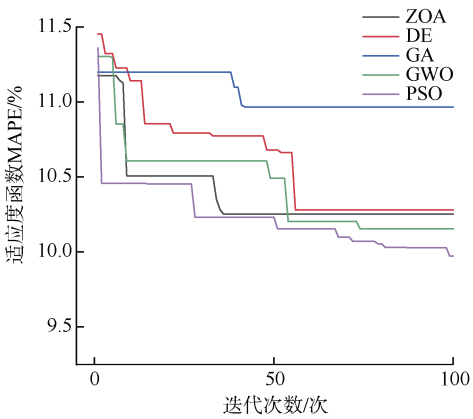


图 4　不同优化算法的超参数优化过程

Figure 4　Hyperparameter optimization process of algorithms

100 轮迭代，算法均趋于稳定，其中 ZOA 优化算法在训练集的预测误差 MAPE 从 11.17%下降至 10.25%。

基于训练集上优化后的模型参数，观测模型在测试集上的预测表现，由表 4 发现 ZOA 优化算法在测试集 MAPE、MAE(平均绝对值误差)、RMSE(均方根误差) 3 项误差指标分别为 3.85%、24 554、38 349，在对比模型中表现最优，同比 GWO 优化算法误差指标分别减少了 4.2%、7.7%、8.1%，证实了 ZOA 算法在 XGBoost 模型超参数调优的适用性。

表 4　不同超参数优化模型的预测误差

Table 4　Prediction errors of different hyperparameter optimization models

数据集	误差项	ZOA	GWO	PSO	DE	GA	贝叶斯
训练集	MAPE/%	10.25	10.15	9.97	10.28	10.97	10.63
测试集	RMSE	38 349	40 047	41 118	38 985	42 000	39 167
	MAE	24 554	26 598	28 700	28 234	29 270	26 949
	MAPE/%	3.85	4.19	4.55	4.52	4.65	4.28

基于 ZOA 优化算法优化得到各线路 XGBoost 模型的超参数如表 5 所示。

表 5　优化后的 XGBoost 模型超参数

Table 5　Optimized XGBoost model hyperparameters

线路	P1	P2	P3	P4	P5	P6	P7	P8
1	221	0.22	4	0.50	0.50	5.94	0.26	3
2	42	0.09	3	0.54	0.54	2.69	0.73	1
3	74	0.46	3	0.65	0.60	5.65	0.26	3
4	74	0.09	3	0.60	0.57	1.61	0.43	1
5	78	0.33	3	0.51	0.50	1.06	0.43	1
6	85	0.28	3	0.58	0.60	2.86	0.80	5
9	46	0.10	3	0.63	0.53	0.05	0.57	1
14	317	0.31	7	0.75	0.56	4.34	0.59	3
16	398	0.26	4	0.77	0.71	1.83	0.54	3

3.3　模型预测误差分析

基于 XGBoost-ZOA 模型预测得到各线路的预测结果如表 6 所示，该模型在 9 条线路的 MAPE 指标均值为 5.18%，同比 SARIMA 模型及 LSTM 的模型，预测误差指标 MAPE 分别减少了 17.8%及 21.9%；RMSE 分别减少了 10.8%及 13.6%，且各线路的 MAPE 指标均控制在 7%以内，即预测准确率达到 93%以上，证实了预测结果的鲁棒性及可靠性。

图 5 为 3 种模型在 1 号线测试集上的预测结果，与其他模型对比，XBGoost-ZOA 模型的预测值与真实值的吻合度更好，准确预测其周期性变化趋势。

表 6 预测模型的预测误差对比

Table 6 Comparison of prediction errors of prediction models

线路	SARIMA		LSTM		XGBoost-ZOA	
	RMSE	MAPE/%	RMSE	MAPE/%	RMSE	MAPE/%
1	43 856	5.12	47 439	5.71	38 349	3.85
2	82 282	6.08	77 436	5.52	80 013	5.81
3	48 012	5.34	59 385	6.96	43 451	4.58
4	53 073	6.31	49 996	6.37	45 349	5.56
5	23 101	4.82	26 409	6.15	19 736	4.18
6	40 838	5.65	45 151	6.58	34 823	4.61
9	12 176	8.11	9 222	6.33	9 271	6.68
14	11 167	7.98	9 597	7.8	10 285	6.72
16	2 247	7.25	2 510	8.25	1 363	4.61
平均	35 195	6.30	36 349	6.63	31 404	5.18

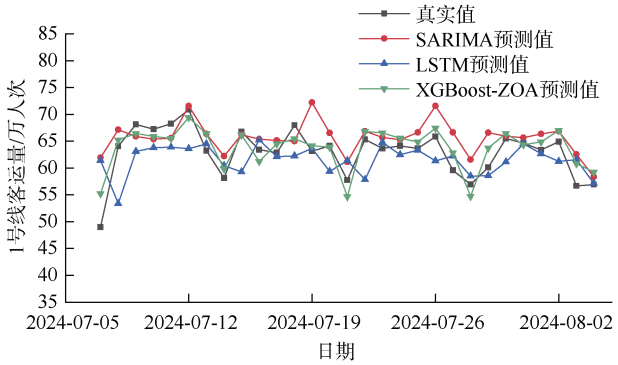


图 5 不同模型的预测结果对比

Figure 5 Comparison of prediction results of different models

图 6 为 XBGoost-ZOA 模型在各线路的测试集上的预测表现，其预测值与真实值吻合度较高，证实了本模型具有良好的泛化能力。

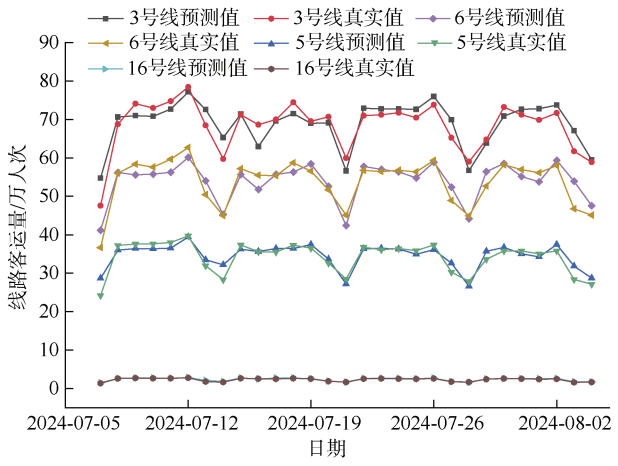


图 6 XGBoost-ZOA 模型在其他线路预测结果

Figure 6 Prediction results of XGBoost-ZOA model for other lines

3.4 不同线路预测结果差异性分析

XGBoost-ZOA 模型的特征重要度是衡量各因素对线路客运量预测值的贡献度(图 7)，按影响特征的客流主导类型，将线路划分为 3 类：

特征	1号线	2号线	3号线	4号线	5号线	6号线	9号线	14号线	16号线
日期	8.6%	12.0%	9.7%	7.8%	10.3%	9.1%	5.2%	4.7%	17.1%
月份	3.7%	9.8%	7.3%	12.4%	7.6%	8.4%	6.1%	3.5%	0.2%
星期	5.3%	20.6%	14.3%	16.5%	9.9%	13.7%	13.8%	1.9%	0.4%
最高气温	3.6%	9.0%	5.1%	9.2%	6.5%	7.1%	6.0%	3.9%	0.2%
天气值	6.9%	7.0%	7.0%	4.8%	5.7%	7.6%	4.2%	3.3%	0.2%
车站数	8.4%	5.8%	5.4%	6.2%	3.0%	6.2%	9.3%	4.7%	25.6%
换乘站	39.0%	3.5%	18.2%	5.2%	2.9%	1.9%	5.1%	19.5%	27.3%
常住人口	9.3%	4.2%	3.5%	8.5%	12.4%	14.4%	10.9%	8.9%	0.8%
节日类型	6.7%	10.9%	6.3%	10.7%	10.2%	7.7%	12.1%	4.6%	0.5%
节假日	4.3%	8.1%	6.3%	9.2%	4.6%	5.9%	13.2%	9.6%	0.0%
调休日	2.4%	8.9%	11.4%	7.0%	24.8%	13.7%	12.4%	4.6%	27.3%
活动类型	1.8%	0.0%	5.6%	2.6%	2.2%	4.3%	1.6%	30.8%	0.3%
MAPE/%	3.9	5.8	4.6	5.6	4.2	4.6	6.7	6.7	4.6

图 7 XGBoost-ZOA 模型的特征重要度

Figure 7 Feature Importance of XGBoost-ZOA Model

1) 趋势客流导向型线路：在研究日期范围内，2023 年 6 月、9 月分别开通 16 号线、1 号线三期，因此 16 号线及 1 号线受新线开通影响的客流显著提升，线网规模特征(含车站数、换乘站)的贡献度分别高达 52.9%及 47.5%。新线开通后，客流呈周期波动，故预测误差较小。

2) 假日活动客流导向型线路：9 号线为假日旅游客流主导的线路，其假日特征(含节日类型、节假日、调休日)的贡献度高达 37.7%，与其服务旅游景点的功能定位高度相关；14 号线为活动客流主导的线路，其活动特征重要度高达 30.8%，源于奥体中心站承接的大型赛事及演出客流，客流激增变化。由于活动及假期诱增客流变化幅度大，导致预测误差偏大。

3) 多因素均衡驱动型线路：2 号线、3 号线、4 号线客流受多种特征共同影响，各特征贡献度较均衡。

按照周期(含月份、星期、调休日)，趋势(含日期、车站数、换乘站、常住人口)，天气(含最高气温、天气值)，假日活动(含节日类型、节假日、活动类型)四大类因素对已有特征进行合并得到表 7。从表 7 发现：

表 7 模型特征重要度占比

Table 7 Profortion of model feature importance %

线路	周期因素	趋势因素	天气因素	假日活动因素	误差 MAPE
1	11.4	65.3	10.5	12.8	3.9
2	39.3	25.6	16.1	19.0	5.8
3	33.0	36.8	12.1	18.1	4.6
4	35.9	27.6	14.0	22.5	5.6
5	42.3	28.6	12.1	17.0	4.2

续表

线路	周期因素	趋势因素	天气因素	假日活动因素	误差 MAPE
6	35.8	31.6	14.7	17.9	4.6
9	32.3	30.5	10.3	26.9	6.7
14	9.9	37.8	7.2	45.1	6.7
16	28.0	70.8	0.4	0.8	4.6

线路客运量预测误差与假日活动因素比重呈正相关，如 14 号线、9 号线的比重高达 45.1%及 26.9%，导致这两条线的预测误差明显高于其他线路。线路客运量预测误差与周期趋势因素比重呈负相关，即比重越高，则规律性越明显，其预测误差更低。

4 结论

针对多因素影响下的线路客运量预测需求，构建了 XGBoost-ZOA 组合预测模型，设计了适用于时序数据预测的交叉验证方案，结合西安地铁实际数据开展实例验证，主要得出以下结论：

1) ZOA 优化后的 XGBoost 预测模型在客运量预测中表现优异，以西安地铁 1 号线为例，优化后 XGBoost 模型在测试集上的 MAPE 为 3.85%，同比预测误差均优于其他基准优化算法，证实了 ZOA 优化模型的先进性及合理性。

2) XGBoost-ZOA 组合模型的预测精度高、鲁棒性强，在西安地铁 9 条线路上的平均预测误差 MAPE 为 5.18%，最大 MAPE 均低于 7%，平均预测精度超过 94%，同比优于 SARIMA 模型和 LSTM 模型，且预测值与实际值曲线吻合度高，证实了 XGBoost-ZOA 模型的可靠性。

3) 基于模型特征重要度分布将西安地铁线路分为趋势客流导向、假日活动客流导向和多因素均衡导向 3 类。各线路主导特征类型显著影响预测精度，假日活动因素占比高的线路(如 14 号线、9 号线)客流波动幅度大，预测误差较大；周期趋势因素占比高的线路客流规律性强，预测误差较小。分析结果与实际情况高度吻合，证实了本模型的可解释性及实用性。

后续研究将聚焦假日活动主导的线路特征，拓宽数据样本范围，并结合 POI 等数据，增强线路预测结果的可靠性及模型泛化能力。

参考文献

[1] 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(01): 102-106+142.
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.

[2] 许心越，孔庆雪，李建民，等. 建成环境对轨道交通客流的时空异质性影响分析[J]. 交通运输系统工程与信息, 2023, 23(04): 194-202.
XU Xinyue, KONG Qingxue, LI Jianmin, et al. Analysis of spatio-temporal heterogeneity impact of built environment on rail transit passenger flow[J]. Journal of transportation systems engineering and information technology, 2023, 23(04): 194-202.

[3] 李之红，郝堃，王健宇，等. 建成环境影响下的城市轨道交通客流多步短时预测[J]. 交通运输系统工程与信息, 2025, 25(1): 160-172.
LI Zhihong, QIE Kun, WANG Jianyu, et al. Multistep short-term prediction of urban rail transit passenger flow under influence of built environment[J]. Journal of transportation systems engineering and information technology, 2025, 25(1): 160-172.

[4] 赵利强，李瑞森，唐水雄，等. 基于多维可预知特征的 TCN-LSTM 城轨短期客流预测[J]. 北京化工大学学报(自然科学版), 2024, 51(05): 86-96.
ZHAO Liqiang, LI Ruisen, TANG Shuixiong, et al. Short-term urban rail passenger flow prediction using temporal convolutional network-long short-term memory (TCN-LSTM) based on multidimensional predictable features[J]. Journal of Beijing University of Chemical Technology (Natural Science), 2024, 51(5): 86-96.

[5] 冯树民，刘浩，李来成. 雨雪天气下轨道交通客流预测模型[J]. 哈尔滨工业大学学报, 2022, 54(9): 1-6.
FENG Shumin, LIU Hao, LI Laicheng. Prediction model of rail transit passenger flow in rain and snow weather[J]. Journal of Harbin Institute of Technology, 2022, 54(9): 1-6.

[6] 郑俊锋，方旭峰，王晨阳. 基于时间序列分解方法的城市轨道交通大客流预测方法[J]. 城市轨道交通研究, 2023, 26(8): 163-170.
ZHENG Junfeng, FANG Xufeng, WANG Chenyang. Prediction method for urban rail transit large passenger flow based on time series decomposition method[J]. Urban mass transit, 2023, 26(8): 163-170.

[7] 海玲，刘文，刘岩，等. 基于ARIMA算法的地铁乘客流量预测[J]. 计算机与数字工程, 2025, 53(3): 666-670.
HAI Ling, LIU Wen, LIU Yan, et al. Prediction of subway passenger flow based on ARIMA algorithm[J]. Computer & digital engineering, 2025, 53(3): 666-670.

（下转第 102 页）

LIU Yong, ZHOU Yisheng, SUO Xiaoming, et al. Model test on the deformation law of shield tunnel underpassing high speed railway roadbed[J]. Rock and soil mechanics, 2023, 44(4): 941-951.

[10] 魏义山, 钟小春, 刘浩源, 等. 盾构近距离下穿工况下既有隧道钢环预加固弯曲刚度研究[J]. 土木工程学报, 2024, 57(S2): 120-126.
WEI Yishan, ZHONG Xiaochun, LIU Haoyuan, et al. Study on bending stiffness of pre-reinforcement of steel ring in existing tunnel under the condition of shield tunneling at close distance[J]. China civil engineering journal, 2024, 57(S2): 120-126.

[11] 张霄. 高速铁路隧道大断面法施工掌子面稳定性分析及超前支护设计方法研究[D]. 成都: 西南交通大学, 2021.
ZHANG Xiao. Stability analysis of working face in large section construction of high-speed railway tunnel and research on design method of advanced support[D]. Chengdu: Southwest Jiaotong University, 2021.

[12] 姚任行, 邱豪磊, 安然, 等. 宁波地铁窄基坑抗隆起稳定性及变形特性研究[J]. 城市轨道交通研究, 2025, 28(4): 209-216.
YAO Renxing, QIU Haolei, AN Ran, et al. Study on anti-heave stability and deformation characteristics of narrow foundation pits in Ningbo metro[J]. Urban mass transit, 2025, 28(4): 209-216.

[13] 王明年, 杨恒洪, 王旭, 等. 砂土地层隧道掌子面稳定性的物质点法分析[J]. 隧道建设(中英文), 2024, 44(S2): 74-84.
WANG Mingnian, YANG Henghong, WANG Xu, et al. Analysis of tunnel face stability in sandy soil by material point method[J]. Tunnel construction, 2024, 44(S2): 74-84.

[14] 王林枫. 土岩地层无岩肩吊脚墙基坑变形破坏特征及稳定性研究[D]. 重庆: 重庆大学, 2023.
WANG Linfeng. Study on deformation and failure characteristics and stability of foundation pit without rock shoulder hanging wall in soil and rock stratum[D]. Chongqing: Chongqing University, 2023.

（编辑：傅依萱）

（上接第 63 页）

[8] 强添纲, 刘涛, 裴玉龙. 地铁进站客流量 SARIMA 与 GA-BP 神经网络组合预测模型[J]. 铁道运输与经济, 2021, 43(12): 134-142.
QIANG Tiangang, LIU Tao, PEI Yulong. Combined prediction model of SARIMA and GA-BP neural network for subway passenger flow[J]. Railway transport and economy, 2021, 43(12): 134-142.

[9] 许文星, 章玉. 基于随机森林算法的轨道站点短期客流预测[J]. 武汉理工大学学报(交通科学与工程版), 2022, 46(3): 406-410.
XU Wening, ZHANG Yu. Short-term passenger flow forecast of rail station based on random forest algorithm[J]. Journal of Wuhan University of Technology (Transportation Science & Enginerring), 2022, 46(3): 406-410.

[10] 姚明辉, 王威超, 吴启亮, 等. 基于实时数据特征和 XGBoost 算法的城市公共交通枢纽客流量预测[J]. 吉林大学学报(工学版), 2024, 54(11): 3302-3308.
YAO Minghui, WANG Weichao, WU Qiliang, et al. Passenger flow prediction of urban public transportation hubs based on real-time data features and XGBoost algorithm[J]. Journal of Jilin University (Engineering and Technology Edition), 2024, 54(11): 3302-3308.

[11] 张开雯, 何勇, 余家香, 等. 基于麻雀搜索算法和长短期记忆神经网络的轨道交通站点客流预测[J]. 四川师范大学学报(自然科学版), 2025, 48(01): 105-113.
ZHANG Kaiwen, HE Yong, YU Jiaxiang, et al. Passenger flow prediction of urban rail transit stations based on ISSA-LSTM model[J]. Journal of Sichuan Normal University (Natural Science), 2025, 48(01): 105-113.

[12] 禹倩, 张亚东, 郭进, 等. 基于深度集成神经网络的城市轨道交通短时进站客流预测[J]. 铁道学报, 2023, 45(12): 37-46.
YU Qian, ZHANG Yadong, GUO Jin, et al. Short-term inbound passenger flow forecasting for urban rail transit based on deep ensemble neural network[J]. Journal of the China railway society, 2023, 45(12): 37-46.

[13] Trojovská E, Dehghani M, Trojovský P. Zebra optimization algorithm: A new bio-inspired optimization algorithm for solving optimization algorithm[J]. Ieee access, 2022, 10: 49445-49473.

（编辑：傅依萱）