引文格式：徐月娟,赵金宝,李明星,等. 时空异质性视角下轨道站点客流影响研究[J]. 地理空间信息, 2024, 22(10): 1-4. doi:10. 3969/j.issn. 1672-4623.2024.10.001

时空异质性视角下轨道站点客流影响研究

徐月娟[1]，赵金宝[1,2*]，李明星[1]，姜嘉伟[1]，刘文静[1]，赵胜利[1]
（1. 山东理工大学 交通与车辆工程学院，山东 淄博 255000；2. 东南大学 交通学院，江苏 南京 210009）

摘　要：轨道站点客流预测及其影响因素的互动关系是交通规划中的关键问题，准确识别影响不同站点客流特征的建成环境因素对组织客流具有重要意义。针对地理空间视角下建成环境对站点客流的时空影响程度的问题，以杭州市城市轨道交通为研究对象，汇集刷卡数据、POI建成数据等多源数据，考虑客流的空间异质性与时空特性构建了混合变化系数和固定系数的S-LGWR影响分析模型，识别不同地理空间位置下影响因素相关系数时空变化特征，为线网空间布局优化、运营管理措施制定和高效安全出行引导提供参考。

关键词：城市轨道交通；客流预测；变量筛选；空间异质性；土地利用
中图分类号：P208　　**文献标志码：**B　　**文章编号：**1672-4623（2024）10-0001-04

Study on the Influence of Passenger Flow on Railway Station from the Perspective of Spatio-temporal Heterogeneity

XU Yuejuan[1], ZHAO Jinbao[1,2], LI Mingxing[1], JIANG Jiawei[1], LIU Wenjing[1], ZHAO Shengli[1]
(1. School of Transportation and Vehicle Engineering, Shandong University of Technology, Zibo 255000, China;
2. School of Transportation, Southeast University, Nanjing 210009, China)

Abstract: The prediction of passenger flow at railway stations and the interaction of its influence factors are key issues in transportation planning, and the accurate identification of construction environment factors affecting the characteristics of passenger flow at different stations is of great significance in organizing passenger flow. In this paper, taking Hangzhou urban rail transit as the research object, collecting multi-source data such as credit card data and POI completion data, we constructed the S-LGWR impact analysis model with mixed variation coefficients and fixed coefficients considering the spatial heterogeneity and spatio-temporal characteristics of passenger flow, and identified the spatio-temporal variation characteristics of impact factors for different geospatial locations. This study can provide reference for optimizing the spatial layout of network, formulating operation management measures and guiding efficient and safe travel.

Key words: urban railway transit, passenger flow prediction, variable screening, spatial heterogeneity, land-use

近年来，我国城市轨道交通系统发展迅速，是缓解城市交通拥堵的有效手段，也是决定城市轨道交通站点行车组织规划以及设施规模的依据[1]。国内外研究大多为基于典型的四阶段法从宏观的角度进行线路和网络客流评估预测，这种方法预测时间较长且对站点周边环境变化反应不敏感。在直接模型中，基于最小二乘法回归(OLS)的回归模型也常被用于相关研究，其以站点客流量作为因变量、影响因素作为自变量构建模型进行预测[2]。尽管OLS模型可以提供相对准确的预测，但其基本假设是因变量与自变量之间的回归系数估计值不随样本空间位置的变化而改变，忽略了不同类型站点的客流特征受到其站域建成环境的影响。地理加权回归（GWR）模型通过建立局部回归方程[3]，在不同的轨道交通站点的待估参数随着空间位置不同，对客流结果的预测具有更高的准确性。张旭[4]等通过构建OLS和GWR的对比预测模型显示GWR模型预测精度更高。随着空间计量理论的发展，GWR的演化模型得到不断发展。钟季汝[5]等利用GTWR模型分析客流量与建成环境的关系。然而目前的研究仍存在一定程度的不足：一是由于站点客流预测作为一个复杂的社会问题，影响因素数量较多且之间可能存在严重的多重共线性特征，需要建立一个有效的变量筛选和复杂度调整模型；二是由于各站点周围环境和经济水平不同，并不是所有影响变量

收稿日期：2023-05-23。
项目来源：山东省自然科学基金资助项目（ZR2021MF109）；山东省研究生教育质量提升计划建设项目（SDYJG19110）。
第一作者简介：徐月娟（1999—），硕士研究生，研究方向为交通大数据分析与管理，E-mail：xuyuejuan1220@163.com。
通信作者：赵金宝（1987—），博士，讲师，硕士生导师，研究方向为交通大数据分析与管理、城市多模式交通优化，E-mail：jinbao@sdut.edu.cn。

会因为空间位置变化明显变化，变量参数在不同地理位置（站点位置）有可能是会发生变化的，也有可能是固定的。

针对上述问题，本文以杭州市轨道交通站点工作日全日集散量作为因变量，采集站点吸引范围内土地利用、POI建成数据等多源数据集，同时建立全局Lasso模型、局部LGWR模型与混合变化系数和固定系数的S-LGWR模型，为提高轨道交通运营效率提供参考。

1 研究区域与数据来源

1.1 研究区概况

杭州地处中国华东地区，自2012年杭州地铁1号线开通截至2022年9月，运营线路共12条，共设车站259座，线路运营里程共计516 km[6]。本文研究区域主要涉及上城区、西湖区与滨江区等杭州市主城行政区，杭州地铁站点空间分布如图1所示。

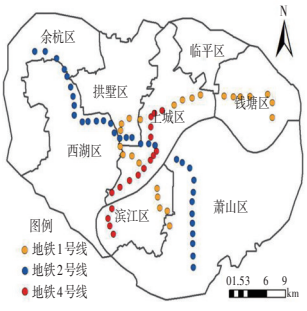


图1 研究区域地铁站点空间分布

审图号：GS(2020)4619

1.2 数据来源与处理

本次研究对象为杭州市一期、二期地铁线路，包含3条轨道交通线路和75个站点（换乘站不重复计）。

1）客流数据。采用杭州市2019年1月连续25 d城市轨道交通智能刷卡数据记录。日均接近260万条记录，利用Python语言处理轨道交通进出站客流数据。

2）建成环境。利用POI（point of interest）点数据来反映城市用地功能分布。通过编写Java爬虫程序，在高德地图开发者平台结合行政区划区域所围成的矩形区域范围进行检索。为了避免过期或者未来数据对模型结果的影响，抓取的POI数据与刷卡数据保持一致。杭州市域内爬取约120万条有效数据。

3）路网数据。数据来源于OpenStreetMap数据库。

4）站点位置数据。利用高德平台抓取75个地铁站点的经纬度坐标。

5）交通运行。利用高德地图平台获取轨道交通站点周围的公共交通出行信息，主要包括公交线路与公交停靠站的数量。考虑杭州市实行的停车+换乘地铁（$P+R$）政策对轨道交通出行的影响，同时获取了各站点周围800 m范围内的停车场数量。

2 研究方法与模型构建

2.1 GWR模型基本原理

地理加权回归（GWR）模型是普通全局回归模型的扩展。传统的多元线性回归采用最小二乘法线性回归模型（OLS）进行参数估计。GWR模型是一种局部变参数回归模型，在不同的轨道交通站点的待估参数随着空间位置不同，是解决变量空间自相关和空间异质性问题的重要方法，对未来结果的预测具有更高的准确性。其模型结构为：

$$f_i=\beta_0(u_i,v_i)+\sum_{j=1}^{m}\beta_j(u_i,v_i)x_{ij}+\varepsilon_i\ \ (i=1,2,\cdots,n) \tag{1}$$

式中，(u_i,v_i)为轨道交通站点i的经纬度；f_i为站点i的客流量；x_{ij}为站点i的第j个局部自变量；ε_i为随机误差项；$\beta_j(u_i,v_i)$为车站i的第j个局部自变量的估计参数。

2.2 LASSO变量选择模型

城市轨道交通站点客流预测作为一个复杂的社会问题，各个影响因素之间可能存在严重的多重共线性特征[7]。LASSO模型是在传统最小二乘估计上L_1正则化加上一个L_1范数惩罚，目标函数表示为：

$$L(b)=\sum(y-Xb)^2+\lambda\|b\|_1=\sum(y-Xb)^2+\lambda|b| \tag{2}$$

式中，$\lambda\|b\|_1$为函数的惩罚项；λ为惩罚系数；$\|b\|_1$为回归系数b的正则。

2.3 模型评价

为了进一步评估模型的准确性，需利用相关评价指标进行量化，选用决定系数（R^2）、Akaike信息指数（AIC）及平均绝对百分比误差（MAPE）对模型进行精度评价。R^2越大，AIC、MAPE越小，模型拟合效果越好，其中，在样本小的情况下，AIC转变为AICc。

$$R^2=1-\sum_{i=1}^{n}(y_i-y_i^*)/\sum_{i=1}^{n}(y_i-\bar{y}_i) \tag{3}$$

$$\mathrm{AIC}=n\ln\left(\sum_{i=1}^{n}(y_i-y_i^*)\right)-n\ln n+2p \tag{4}$$

$$A_{\mathrm{MAPE}}=\frac{1}{N}\sum_{i=1}^{n}\left|\frac{y_i-y_i^*}{y_i}\right|\times 100\% \tag{5}$$

式中，y_i为站点刷卡客流的真实值；y_i^*为站点客流的预测值；n为样本数；p为参数个数。

3　LASSO-GWR预测模型

3.1　因变量选取

对杭州市2019年全网75个轨道交通站点以15 min为粒度对25 d的进行拆分发现，除了个别站点，大部分站点的每日平均进出站人数具有高度一致性，如图2所示，其中横坐标表示时间（以小时为单位），纵坐标表示该时段内进站客流量。因而本文选取站点工作日平均进出站之和全日集散量作为因变量。

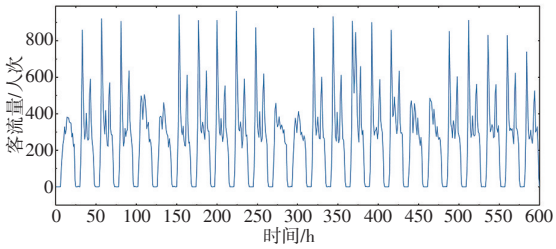


图2　城市轨道交通站点客流分布

3.2　自变量选取

车站直接需求估计模型认为，站点客流量受到该站在交通网络位置以及其他车站的影响[8]。从杭州市轨道交通站点客流的时空分布中可以看出，相邻站点的客流分布较为相似。参考国内外对站点客流量影响因素的研究[9]，初步确立以站点客流吸引范围内的土地利用、交通特征以及站点自身属性因素为自变量。使用莫兰指数（Moran’s I）检验各变量的空间依赖性，具体的描述性统计如表1所示。

表1　自变量的描述性统计

类别	变量	描述	最小值	最大值	均值	Moran’s I	VIF
土地利用	DINE	餐饮服务的数量	1	1 244	345	0.921	3.123
	SCEN	风景名胜的数量	0	30	2	0.676	4.906
	SHOP	购物服务的数量	1	55	18	0.502	6.519
	FINA	金融保险服务数量	0	196	41	0.795	5.213
	RESI	商务住宅的数量	3	222	73	0.804	13.502
	LIFE	生活服务的数量	10	369	154	0.481	7.892
	SPORT	体育休闲的数量	2	234	60	0.586	7.996
	HL	医疗保健的数量	4	237	82	0.686	7.749
	GOVE	政府机构的数量	2	368	102	0.754	9.415
	HOTEL	住宿服务的数量	0	361	58	0.643	4.671
	CORP	公司企业的数量	7	466	170	0.461	3.326
	EDU	科教文化的数量	1	316	111	0.605	6.595
	PARK	停车场的数量	5	281	84	0.778	11.106
交通特征	BUS	公交停靠站的数量	1	9	3	0.163	3.953
	LINE	公交线路的数量	2	73	18	0.246	3.834
站点自身属性	PORT	站点进出口数量	1	19	5	0.134	2.561
	CINE	站点换乘线路数量	0	1	0.09	0.051	2.195
	YEAR	站点投入运营年数	0.64	6.17	3.72	0.748	2.118
	SITE	站点可达性	0.17	19.8	8.13	0.977	2.42

模型初选的自变量有19个，分别对变量进行共线性检验，结果见表1所示。发现变量方差膨胀系数（VIF）的值较大，模型呈现较强的共线性。考虑到城市轨道交通客流预测的复杂性，影响因素间存在多重共线性和不确定性等问题，直接删除共线性较强的变量无法保证模型的合理性和准确性，故选用Lasso算法筛选出主要影响变量，通过交叉验证方法，确定 λ 值。λ 值的选择原则是使得Lasso模型的均方误差（MSE）最小。

数据集MSE随着 λ 值的变化过程中，确定最佳模型的位置。模型系数结果显示弱影响或无关影响的变量为SCEN、LINE、SHOP、HL、GOVE、EDU、RESI和PARK。因此，在建立考虑参数平稳性的全局模型和局部模型时将这8个自变量集中移除。

4　模型结果与分析

4.1　参数平稳性检验

LGWR模型通过对全部变量进行地理变异性测试（geographical variability test）来区分全局影响变量和局部影响变量，测试每一个自变量的空间变异值，当值大于2时，则表明该变量的空间非平稳性较弱，应考虑将其作为全局变量[10]。对全部变量进行参数平稳性检验结果表明，BUS、CLIN、SITE及SPORT变量的空间变异值大于2，故将这4个变量作为全局变量。

4.2　参数估计

为了揭示变量间的空间关系以及对比分析不同模型的参数估计值差异，最终建立的Lasso、LGWR和S-LGWR3个不同的模型进行比较。最终筛选得到的11个影响变量具体的参数拟合结果如表2所示。

表2　模型参数 β 估计结果

变量类型	影响变量	Lasso/β	LGWR/β		S-LGWR/β	
			min	max	min	max
常数	Intercept	8 272.195	7 931.02	12 311.25	6 723.47	11 323.88
局部变量	PORT	363.619	212.57	529.29	175.33	510.11
	YEAR	631.53	610.40	1287.14	644.57	1 187.01
	DINE	126.09	124.19	133.47	123.26	133.08
	FINA	139.143	129.20	148.75	134.99	161.93
	LIFE	−68.556	−80.66	−63.24	−83.01	−66.06
	HOTEL	56.496	44.14	68.29	42.81	67.74
	CORP	−15.781	−19.26	−9.26	−19.08	−8.91
	BUS	444.002	389.37	633.95	605.59	
固定变量	CLIN	−3 370.6	−5 099.79	−2 788.44	−4 312.26	
	SITE	−321.617	−653.71	−169.03	−298.27	
	SPORT	−19.036	−22.15	−12.74	−11.77	

利用ArcGIS 10.3平台将表2中局部影响变量的拟

合参数估计值范围进行分级色彩可视化，分类默认的是自然间断点分级法，类的数量设置为5，即选择分级字段值的分段数量为5段，部分解释变量在空间上的影响程度变化展示如图3所示。

从表2中估计参数值可看出，在局部影响变量中，根据S-LGWR模型得到的局部解释变量的参数估计结果，站点进出口数量PORT、餐饮服务的数量DINE、金融保险服务数量FINA、住宿服务的数量HOTEL和站点投入运营年数YEAR的参数估计值分别在[175,511]、[123,134]、[134,162]、[644,1 188]和[42,68]区间内，这5类变量与站点客流均为正相关，相比较而言站点进出口数量和站点投入运营年数对站点客流量更具有积极影响，即站点开通时长越长则其平均客流量将越大。站点多设置1个出入口，站点客流量最多会增加大约510人次。而站点周边公司企业的数量CORP和生活服务的数量LIFE客流为负相关，参数估计值在[-19，-9]、[-84，-66]区间内，表明生活服务的数量和周边公司企业的数量对客流具有一定程度的消极影响。

在固定影响变量中，公交停靠站的数量BUS有显著的正相关性，公交停靠站的数量每增加一个，客流量会增加大约605人次，其原因可能是扩大了站点对公交绿色接驳出行方式的吸引更多的潜在出行者的范围[11]。而站点换乘线路数量CLIN 、站点可达程度SITE和体育休闲的数量SPORT与站点客流为负相关，增加这三者的数量会对站点客流具有一定消极影响。

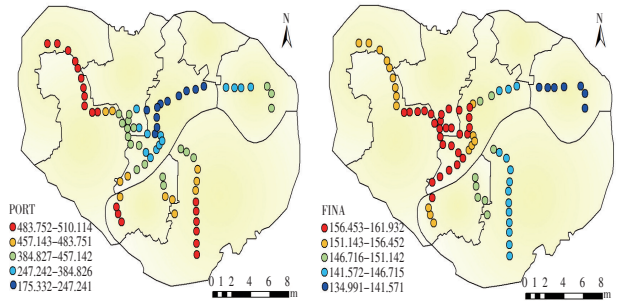


图3 客流拟合系数空间分布

审图号：GS(2020)4619

4.3 模型评估

利用3个评价指标对不同模型的拟合效果进行对比分析，结果见表3。

表3 3种模型的拟合优度评价

评价指标	Lasso	LGWR	S-LGWR
R^2	0.733	0.864	0.902
AICc	1 586.12	1 562.81	1 553.79
MAPE/%	19.56	15.53	11.21

由表3可知，3种模型的解释能力均在73%~91%之间，但S-LGWR模型的误差值均小于其他2种预测模型。从Lasso模型到S-LGWR的R^2值增大了0.169，模型有着显著的改善，3种模型之间AICc值的差额均超过3；从平均绝对百分比误差来看，S-LGWR模型较Lasso模型减少了8.35%，说明S-LGWR模型的预测精度更高，适用性更强。

5 结 论

本文以杭州市3条轨道交通线路为研究对象，建立全局Lasso模型、局部LGWR模型与混合变化系数和固定系数的S-LGWR模型对站点客流量进行预测，主要结论为：①考虑到城市轨道交通客流预测的复杂性，直接删除共线性较强的变量无法保证模型的合理性和准确性。利用Lasso算法能够有效的筛选出主要影响变量，降低影响变量复杂度。②实例结果表明S-LGWR模型的预测精度更高，适用性更强。建模结果可显示出影响变量对客流量影响的空间变化关系，模型稳定性较强，模型拟合结果的精度整体良好。

参考文献

[1] 王冠,肖昶,代琦. 基于时空数据的武汉市绿色交通出行模型研究[J]. 地理空间信息,2023,21(1):20-25

[2] 冯海琦,赵玉梅,王琳琳. 山东省社会消费品零售额的影响因素分析:基于Lasso回归模型[J]. 鲁东大学学报(自然科学版),2021,37(4):310-314

[3] 和云润,陈朝情,殷唐燕,等. 基于地统计模型的林分蓄积量空间异质性研究[J]. 西北林学院学报,2023,38(3):169-175

[4] 张旭. 基于精细化用地数据的城市轨道交通客流预测[D]. 北京:北京交通大学,2019

[5] 钟季汝. 城市建成环境对轨道交通客流特征影响研究[D]. 大连:大连理工大学,2019

[6] 孟兵兵,周健勇. 杭州地铁OD数据客流时空分布特性分析[J].软件导刊,2021,20(11):158-162

[7] 施斌,井发明,张浩宽. 基于GIS的乡镇工业用地空间分布研究[J]. 地理空间信息,2023,21(4):82-84

[8] 殷鹏,周旗,魏倩枝,等. 寒潮灾害预警下公众情感响应与时空分异研究[J]. 地理空间信息,2023,21(3):90-94

[9] 庞磊,任利剑,张哲浩,等.基于乘降客流特征的轨道交通站点分类及客流量影响因素分析[J]. 交通运输系统工程与信息,2023,23(4):184-193

[10] 程国柱,周林芳,徐慧智. 基于GWR的城市轨道交通车站客流预测模型[J]. 武汉理工大学学报(交通科学与工程版),2019,43(2):198-203

[11] 宋秀容,全雪贞,张洁,等. 结合局部与全局特征的点云语义分割卷积网络[J]. 地理空间信息,2023,21(3):25-30