

城市轨道交通出行的时空特征及 影响因素非线性机制 ——基于梯度提升决策树的成都实证

崔叙¹ 喻冰洁¹ 杨林川¹ 梁源² 张凌菲³ 方翰¹¹

(1. 西南交通大学 建筑与设计学院, 中国四川 成都 611756;

2. 同济大学 城市交通研究院, 中国 上海 201800;

3. 重庆大学 建筑城规学院, 中国 重庆 400030)

【摘要】: 以成都市为例, 基于地铁刷卡、POI、建成环境等多源大数据, 运用GIS、Python、Spark等进行数据处理, 分析156个站点客流量、出行时间的空间分异特征, 并通过梯度提升决策树(GBDT), 解析站域设施数量、建成环境、经济属性等15个因素对客流量与出行时间的非线性影响机制。研究发现: (1) 成都市轨道交通客流量具有中间高一两边低的倒U型分布特点, 每站日均客流量为1.6万人次; (2) 出行时间随距CBD距离增加逐渐增大, 每站乘客平均出行时间为32min, 出行时间的概率密度呈Gamma分布特点; (3) 路网密度、容积率、办公设施、交通设施等对客流量具有非线性正向影响, 而房价、距CBD距离与客流量则分别为“凸”形与“凹”形的非线性关系; (4) 距CBD距离、购物数量、办公数量等与出行时间非线性正相关, 而交通设施、房价等与其负相关。最后, 提出考虑“流量”与“时间”阈值效应的建成环境优化、构建“出行+生活”的轨道通勤圈等规划策略。

【关键词】: 城市轨道交通 轨道通勤圈 路网密度 房价

【中图分类号】: F511.99 **【文献标志码】:** A **【文章编号】:** 1000-8462 (2021) 07-0061-12

当前, 城市蔓延和过度依赖机动化带来了交通拥堵、空气污染等“大城市病”问题^[1]。面对日益严峻的交通和环境问题, 城市轨道交通作为一种高效率、大运量的交通方式, 已经成为大城市落实公交优先、引导绿色出行的重要选择^[2-3]。随着轨道交通网络结构复杂化, 轨道交通出行效率也面临着出行时耗过长、高峰时段人流过度聚集等问题。城市轨道交通客流量、出行时间的时空格局是居民地铁出行的重要表征。地铁站域作为城市居住、就业、商业等设施的主要聚集区域, 其建成环境、设施布局与人群活动的关系“失衡”会阻碍城市的可持续发展^[4-5]。因此, 研究城市轨道交通的人群出行时空特征, 分析其影响因素的作

作者简介: 崔叙(1974-), 男, 黑龙江佳木斯人, 博士, 教授, 博士生导师, 研究方向为城市交通、枢纽站域规划设计。

E-mail: cuixu@home.swjtu.edu.cn

喻冰洁(1991-), 男, 湖南长沙人, 博士研究生, 研究方向为城市轨道交通出行大数据。E-mail: 270433051@qq.com

基金项目: 国家自然科学基金区域创新发展联合基金重点支持项目(U20A20330); 国家自然科学基金项目(51778530); 西南交通大学2019年博士创新基金项目(2017310253)

用机制，有利于认知大城市公共交通需求规律，进而提高城市轨道交通的出行效率，实现公交导向（TOD, Transit oriented Development）^[6]的城市发展目标。

近年来，随着大数据与互联网技术的快速发展，大数据应用于人群时空行为研究日趋多样化与复杂化。手机信令数据、公交刷卡数据、蓝牙数据、GPS 数据、共享交通数据等多源数据，为城市人群出行行为的可计算提供了新的数据基础与分析方法^[7]。城市公共交通系统需要满足人们日常通勤、通学以及购物等多样化出行需求。基于公交智能刷卡数据，国内外研究者从职住通勤圈、行为特征聚类、线网效率、接驳特征、客流预测等方面挖掘了城市居民出行的时空特征^[8-10]。例如，Huang 等利用北京市连续 7 年的轨道交通智能刷卡数据，提出 45min 通勤忍受阈值定律和地铁通勤者的波动周期为 4 年，并剖析了四类人群的通勤时间与住房成本的博弈过程^[11]。王宇宁等构建“时间—距离”模型，基于天津地铁的现状 & 规划线网，开展了主要节点的通勤时距仿真，分析轨道交通通勤圈构建的可行性^[12]。黄洁等计算了北京市 431 万条地铁刷卡数据的出行时间分布，并基于多尺度解析了地铁 OD 的网络特征^[13]。另外，部分学者通过随机森林^[14]、K-均值聚类^[15]等方法进行客流时序变化的轨道车站类型划分。

轨道交通客流出行强度、出行时间与站点周边建成环境、社会群体属性、天气环境等因素存在关联作用，国内外学者对轨道交通客流量影响因素及机制开展了广泛的研究。在土地利用、建成环境方面：金昱基于 logit 模型分析了上海市轨道交通客流时变特征与 500m 范围内的社会属性、站点特征及交通方式等要素的作用关系^[16]。李国强等基于客流 AFC 数据与 POI 数据，分析了城市轨道交通客流时序变化过程，并从土地利用、公共交通等方面挖掘了影响客流的主要因素^[17]。Shi 等基于 GTWR（时空地理加权回归）分析了地铁站点地区建成环境、线网拓扑结构对地铁出行强度的影响，发现站域居住、商业、停车设施的数量以及站点中心度与紧密度对客流量具有正向影响^[18]。在社会经济因素方面：Chava 等通过研究 TOD 地区的房价与人口结构的变化，提出班加罗尔 TOD 站域绅士化对地铁客流量有显著影响^[19]。Becky 等对比了香港与纽约的城市轨道交通客流量的影响因素差异，结果表明就业人口密度、居住人口密度以及居民拥车率均与客流量显著正相关，且以上 3 个社会经济指标对香港的影响更强^[20]。Pan 等基于手机信令数据、地铁刷卡数据解析了站域人口特征、商业规模与地铁客流量的关系，结果显示就业人口数量与上盖商业开发量对客流量影响最大^[1]。

综上所述，国内外学者对城市轨道交通出行特征与影响因素进行了探讨，并取得了丰硕成果，但已有研究在以下方面略显不足：(1) 国内外相关研究均是从“量”的角度探寻轨道交通站点地区建成环境、人群结构等因素对客流量的影响作用，未同时考虑站区属性对出行时间与客流量的影响差异性。从站点地区“流量”与“时间”综合分析其影响机制，对认知轨道交通客流行为、站域空间特征具有重要意义。(2) 大部分学者采用传统的多元回归^[19]、结构方程^[17]、地理加权回归^[18]等模型研究地铁客流量与影响因素之间的线性关系，但也有部分学者对预先设定的空间因素与客流量之间的线性关系提出质疑^[21]。而机器学习则可以反映出响应变量（客流量、出行时间）与预测变量（影响因素）之间的非线性关系^[22]。(3) 以往研究多停留于描述性统计与模型构建层面，未结合影响因素作用机制提出相应的规划策略与阈值范围。

因此，本文使用成都市 2019 年 3 月的城市轨道交通刷卡数据，结合 POI、建成环境等多源数据，利用 GIS、Python 解析轨道交通站点客流出行量、出行时耗以及出行模式。基于 GBDT（梯度提升决策树）算法^[22]，从站区建成环境、设施数量、经济属性三个层面（15 个因素），揭示其对站点客流量与平均出行时间的影响。采用 GBDT 建模不仅能够探究城市轨道交通出行的主要影响因子的特征度，且能精细刻画影响因素的非线性作用机制。本研究可为大城市轨道交通出行需求管理，以及公交都市下的 TOD 站域建成环境与设施布局优化提供有效阈值与理论参考。

1 研究区概况

成都市是目前全国城市轨道交通发展与建设最快的城市之一。截至 2019 年 3 月，成都市地铁开通里程为 226km，共开通 6 条线，分别为地铁 1、2、3、4、7、10 号线，共计 156 个地铁站点（图 1）。城市轨道交通覆盖 10 个区县，包括核心五区（成华区、武侯区、锦江区、青羊区、金牛区）55 个站点，以及外围区（温江区、郫都区、新都区、双流区、龙泉驿区）101 个站点。

2 数据来源与研究方法

2.1 数据来源

2.1.1 城市轨道交通刷卡数据

客流量与出行时间反映了城市轨道交通出行的重要时空特征。因此，本研究选取客流量与出行时间两个指标作为因变量。

(1) 客流量：采用成都市 2019 年 3 月轨道交通智能刷卡数据 (Smart Card Data, SCD)，使用的数据主要包括成都地铁工作日、非工作日进出站量，全时段 (1h 间隔) 进出站量数据以及 OD 数据等，采用 Spark、Python 对地铁刷卡数据进行处理，获取 156 个站点的日均客流量；(2) 出行时间：地铁刷卡数据记录了卡号、检入站、进站时间与出站时间、检出站，基于刷卡数据可获取 156 个站点的进站旅客的轨道平均出行时间，通过 GIS 对各站的日均客流量以及乘客平均出行时间的时空格局进行分析。

2.1.2 站区设施数量

POI (Point of Interest) 数据作为新出现的数据来源，能反映城市各类服务设施的空间布局特征^[23]。POI 数据的粒度和分类较为精细^[24]，可以结合 SCD 和 POI 数据进行交互分析，解析轨道交通使用人群的出行目的，探测城市轨道交通客流量、出行时间的影响因素。因此，本研究结合参考文献^[17-18]，利用 Python 爬取了高德地图 (<https://www.amap.com/>) 成都市域范围内的 POI 数据，并通过 GIS 对地铁站域 1000m^[25]缓冲区数据进行了筛选和清洗。本次 POI 数据类型主要包括休闲娱乐、医疗保健、金融、居住 (含社区服务)、办公、商业、娱乐、餐饮、生活服务以及文化教育等。

2.1.3 站区建成环境数据

建成环境体现了站点地区城市设计特点，对人群出行行为具有重要影响^[26]。POI 数据主要反映了站区设施的数量与密度，忽略了各类型设施的建筑面积或占地面积。因此，本研究在站区设施数量的基础上，结合 Cervero 等提出的设计、密度等 3D 原则^[27]，选取了路网密度、容积率、距 CBD 距离 3 个变量作为建成环境变量，其中路网密度、容积率数据利用成都市建筑轮廓、道路数据在 GIS 中进行二次处理，分别解析 156 个站点 1000m 范围内的道路边界、用地性质与建筑开发强度等矢量数据。距 CBD 的距离主要通过 Python 计算每个站点距离天府广场的空间距离得到。

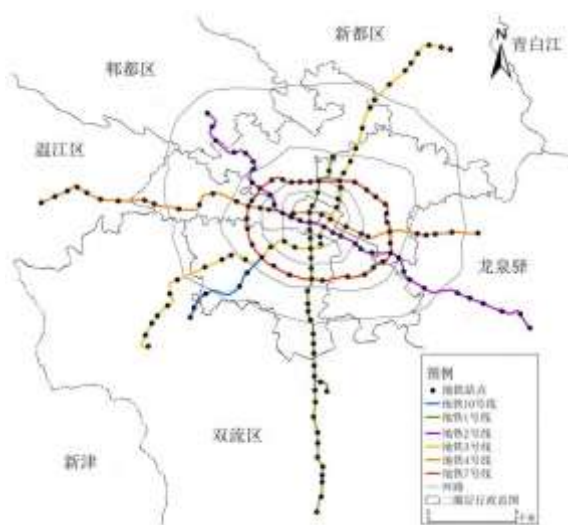


图 1 研究区地铁线路与站点分布图

2.1.4 站区经济属性数据

房价、平均岗位工资数据来源于“链家网 (<https://cd.lianjia.com>)”“智联招聘 (<https://landing.zhaopin.com>)”等网站的数据爬取,结合成都市 2018 年统计年鉴、各区县的 2018 年度统计公报对数据进行修正。

2.2 研究方法

2.2.1 梯度提升决策树 (GBDT)

采用梯度提升决策树 (GBDT) 模型分析客流量、出行时间与影响因素的非线性关系。该方法最初是为了预测和解释计算机中的数据而开发的机器学习方法,由 Friedman 开发^[22]。近年来已逐步应用于交通工程、交通地理、经济地理等研究方向^[28-31]。

相比于传统的线性回归模型,GBDT 具有许多优势^[32]。(1)GBDT 可以解释变量之间的非线性关系。与线性回归不同,GBDT 不假设预测变量与响应变量之间具有线性的关系。该方法通过最小化损失函数来预测响应变量,在控制了模型中的所有其他预测变量之后,生成偏依赖图 (Partial Dependence Plot),以解释预测变量和响应变量之间的关系^[29]。(2)GBDT 提高了模型预测的准确性,其通过阶段性学习数据对预测变量进行权重调整^[21]。(3)GBDT 解决了多重共线性问题,该算法考虑了预测变量之间的相互作用,对预测变量的响应取决于较高层节点的预测变量的值^[33]。

GBDT 属于 Boosting 算法^[34],结合了回归树与提升树 (图 2) 的思想。主要解决一般损失函数的优化问题,其核心是利用损失函数的负梯度作为前一轮学习器的残差近似值,在残差减小的梯度方向建立新的决策树,使模型加速收敛到局部或全局最优解。GBDT 具有很强的预测能力^[35],模型可表示为:

$$F_M(x) = \sum_{m=1}^M T(x; \theta_m) \quad (1)$$

式中: $T(x; \theta_m)$ 表示决策树; θ_m 表示树的参数; M 为树的个数。

决策树 $T(x; \theta_m)$ 的损失函数用 $L(\cdot)$ 表示,在 GBDT 中,损失函数为平方误差函数。用 $T_{m-1}(x_i)$ 表示当前决策树,GBDT 通过最小化损失函数来确定下一棵决策树的参数 $\hat{\theta}_m$ 。

$$\hat{\theta}_m = \arg \min \sum_{i=1}^N L[y_i; T_{m-1}(x_i) + T(x; \theta_m)] \quad (2)$$

2.2.2 Gamma 函数

人群出行研究中往往用幂律分布、泊松分布、正态分布以及 Gamma 分布等对其规律进行描述。利用 Minitab 的概率分布拟合模块,将乘客的出行时间的客流量分布概率进行拟合,并结合 Brockmann^[36]、黄洁^[13]、金昱^[16]等人的研究,发现乘客出行时间的客流量概率密度分布符合 Gamma 函数特征。其函数可表达为:

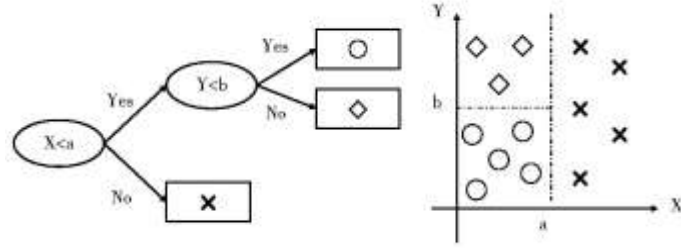


图 2 分类树示意图

$$f(t) = \frac{1}{\Gamma(k)\theta^k} t^{k-1} e^{-\frac{t}{\theta}}, t \in (0, \infty), k > 0, \theta > 0 \quad (3)$$

式中：k 表示出行时间分布的形态参数； θ 表示分布的占比参数；t 表示轨道出行时间。

3 轨道交通出行的时空特征

3.1 客流量时空分布

3.1.1 客流量空间分布格局

客流量前十的站点分别为春熙路（10.6 万/d）、成都东客站（8.5 万/d）、犀浦站（6.4 万/d）、天府三街（5.9 万/d）与天府广场（4.7 万/d）等，主要为换乘站、首末站以及枢纽站，是成都市轨道交通网络中的核心集散站点；后十位站点主要分布在城市四环外待开发的外围区域，如麓湖（0.02 万/d）、武汉路（0.06 万/d）、红石公园（0.16 万/d）等。客流最集中的区域为天府广场—天府五街 15.6km 核心段，平均每站流量达到了 3.4 万/d，远超全市的每站 1.6 万/d，该区段也是成都市多个大型就业地、商业中心的集中区域，站点地区开发强度大，具有较强的客流吸引能力，凸显了成都市双核心发展模式的城市结构（图 3）。

各站客流量空间分布趋势由东—西与北—南，均呈现两边低中间高的倒 U 型特征（图 4）。城市中心区的客流量显著大于城市外围地区，三环内日均客流量为 1.8 万人/站，而三环外为 1.2 万人/站，这主要是由于城市用地开发强度的空间异质性、站区功能业态混合的多样性、以及轨道线网密度的片区差异性等原因造成的。

3.1.2 客流量线路分担率

从图 5 可以看出 1、2 号线的工作日客流压力明显，分别占到了全日总客流量的 27.37% 与 22.75%。特别是晚高峰，1 号线客流断面分担率达到了 35.5%；周末客流的分线占比相对均衡，2、3 号线的全日客流量占比超过了 1 号线，三线分担率分别为 26.62%、21.32%、20.75%，且高峰时段与全日分担率占比相似。最早开通的 1 号线（2010 年开通）与 2 号线（2012 年开通）客流量较大，这说明早期开通的地铁线路主要目的是解决城市通勤结构与通道问题，并以适应城市空间结构、站区建成环境为总体目标。成都市大量的新兴就业地集聚在 1 号线附近，如天府三街、世纪城等；而传统商业中心则主要分布在 2 号线周边，如春熙路、一品天下、牛王庙等。因此，工作日与非工作日线路分担率的差异也体现了站点地区服务设施空间分异下的轨道出行需求规律。

3.1.3 客流量时序变化特征

工作日客流 1 小时时序曲线属于双峰型（图 6），早高峰的峰值接近 20 万客流量，大于晚高峰的 15 万。而非工作日客流分散，高峰不显著，属于无峰型，客流量最大峰值仅为 8 万，全天客流分布均衡系数高，8:30~19:30 客流量均大于 5 万，而工作日的高峰与平峰的差值达到了 15 万，客流时序曲线的峰度较高。主要原因为：工作日通勤客流具有早晚高峰潮汐现象，而非工作日客流出行目的、出行时段选择较为多样，因此两者会出现双峰型与无峰型的差异。近年来，随着成都市人口的快速聚集、轨道建设的不断加快，工作日高峰段分担率呈逐年上升趋势，城市远郊区的通勤者通过轨道交通获得中心区就业机会得到提升。因此，通过站点就业地与居住地的结构调整，减少轨道交通极限通勤与高峰过载的现象具有重要意义。



图 3 各站点客流量空间分布格局图

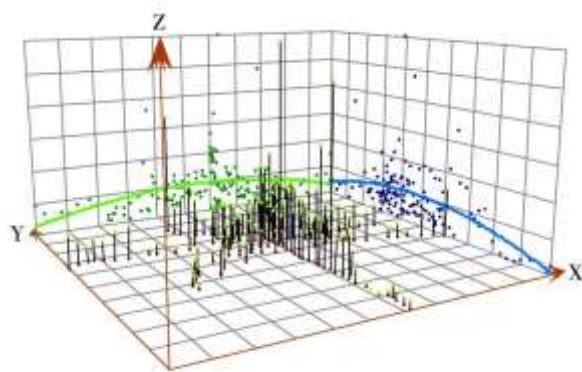


图 4 站点客流量空间分布趋势图

3.2 出行时间空间格局与概率密度分布

3.2.1 出行时间的空间格局

基于地铁刷卡数据的乘客进出站时间与站位信息，统计 156 个站点进站乘客的平均出行时间（以下简称为出行时间），平均每站出行时间约为 32min。图 7 可以看出出行时间分布的空间异质性较强。出行时间较长的站点主要为两类：一是分布在城市三环外的典型就业站点，大型就业站点辐射能力强，客流腹地大，与城市多个圈层的站点联系强度高，大跨度通勤较多，因此轨

道出行时间较长，如天府五街、金融城、天府三街等，平均出行时间约为 38min；二是位于城市边缘区的枢纽型、郊区型站点如成都西、犀浦站、龙泉驿等，其具有连接周边区县、附近乡镇的中转枢纽作用，平均出行时间约为 40min；乘客出行时间较短的站点集中在靠近传统或新兴就业集聚地附近的居住型站点，这类站点居住群体区位的选择主要考虑距就业地距离、公共交通便利度，以及城市未来的发展方向，如城市传统核心区的通惠门、人民公园等，天府新区南部边缘的广都、海昌路等，平均出行时间约为 25min。

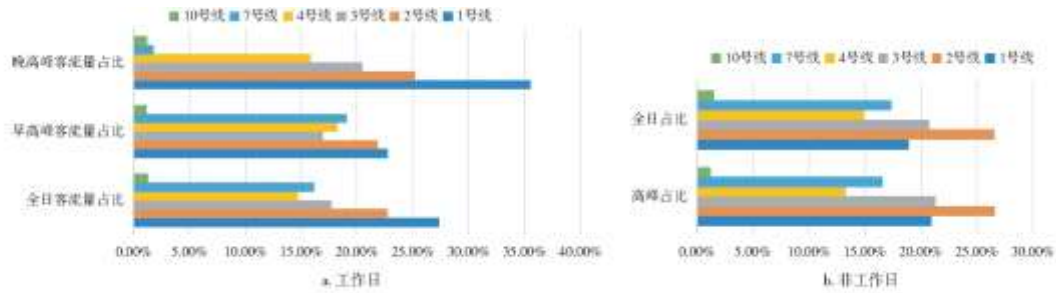


图 5 客流量各线路分担率图

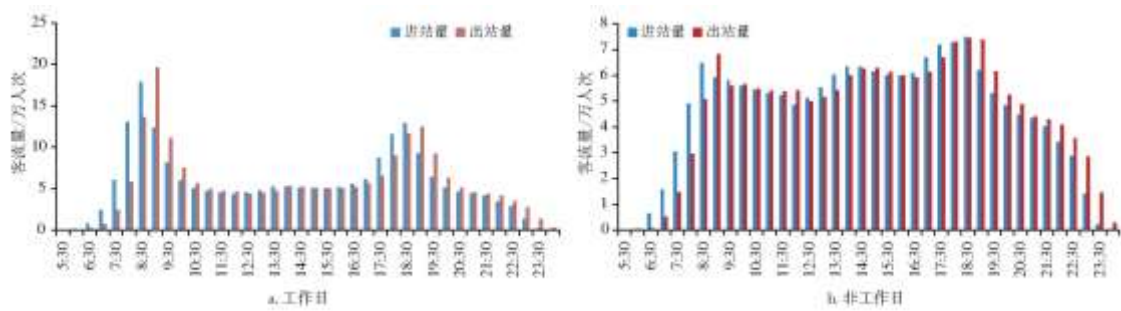


图 6 客流量 1h 时序曲线变化图

出行时间的空间格局呈现外围高、中间低的 U 型分布特征（图 8），由城市中心，沿着环线向外逐渐增大。其中，三环内的乘客平均出行时间为 28min，而三环外的乘客平均出行时间为 36min。主要原因为：三环内地铁站域的公共服务设施较为完善，站点之间的功能形成互补与混合利用，线网密度大，站点间距小，能在较小的范围内完成各类出行目的，因此出行时间要更短。

3.2.2 出行时间的概率密度分布

成都市工作日乘客平均出行时间为 30.2min，而非工作日的乘客平均出行时间为 32.6min。这说明通勤出行的总体出行时间略小于非通勤出行时间。运用 Minitab 对出行时间分布进行概率密度拟合（图 9），发现乘客“出行时间—客流量”概率密度拟合符合 Gamma 分布的形状关系，表现为“中短出行时间”集聚的基本特征。对比工作日与非工作日的形状参数可以看出，工作日的波峰峰度更大，说明其集中出行特征更显著。

根据图 10 可以看出工作日出行时间 45min 以内占据 85.1%的累积概率，30min 以内占据 65.1%的累积概率；而非工作日在 45min、30min 两个出行时间点内的累积概率分别为 81.9%和 61.6%，这说明工作日短途出行占比要高于非工作日，进一步验证了工作日平均出行时间要低于非工作日。另外，无论工作日还是非工作日，15~25min 都是最集中的出行时段，分别占据工作日/非工作日全时段流量的 33.9%和 31.4%，从伦敦^[37]、上海^[16]以及北京^[38]轨道交通的出行时间概率分布规律研究中，也可以观察到

相同的规律。

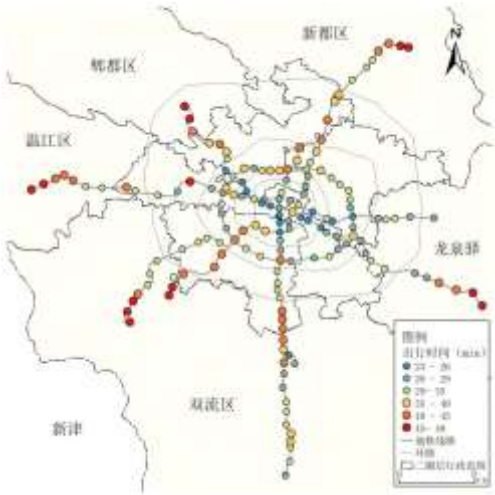


图 7 各站乘客平均出行时间分布图

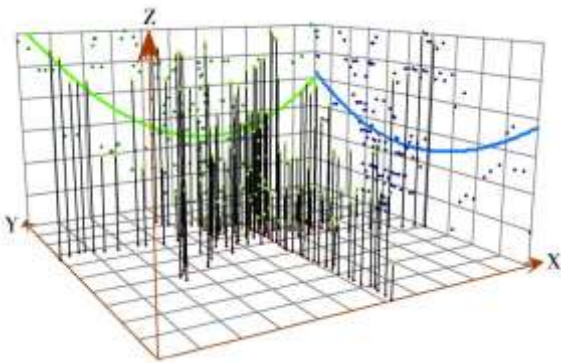


图 8 各站乘客出行时间空间分布趋势图

4 基于 GBDT 的出行量、出行时间影响因素非线性作用机制

4.1 影响因素对客流量的非线性作用机制

利用五折交叉验证（five-fold cross-validation）结合网格搜索算法获得 GBDT 模型的最优参数，得到客流量与影响因素的最终模型（ $R^2=0.91$ ）。在最终模型中，根据 GBDT 的影响因素的特征排序（图略），提取影响地铁客流量变化的最重要的 7 种因素偏依赖图（PD），对其阈值效应、非线性关系进行分析，分别为：路网密度、容积率、距 CBD 距离、办公设施、房价、文化教育、交通设施等，这表明相比站区经济属性与设施数量，建成环境更易影响站点的客流量。

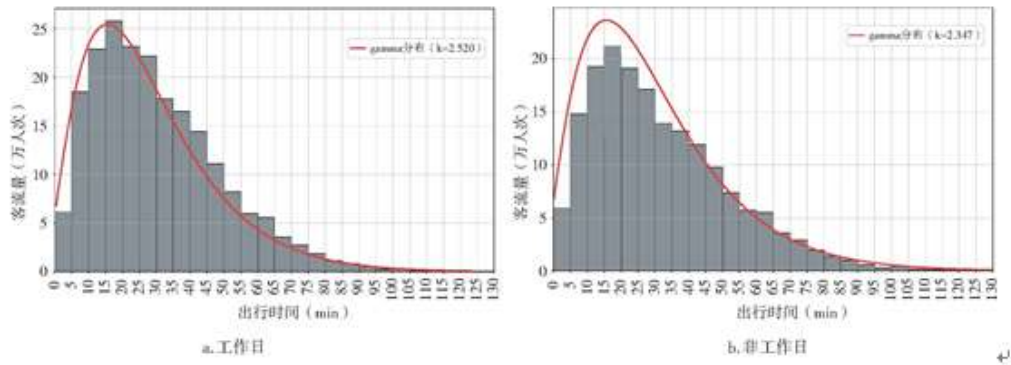


图 9 出行时间 gamma 拟合分布图

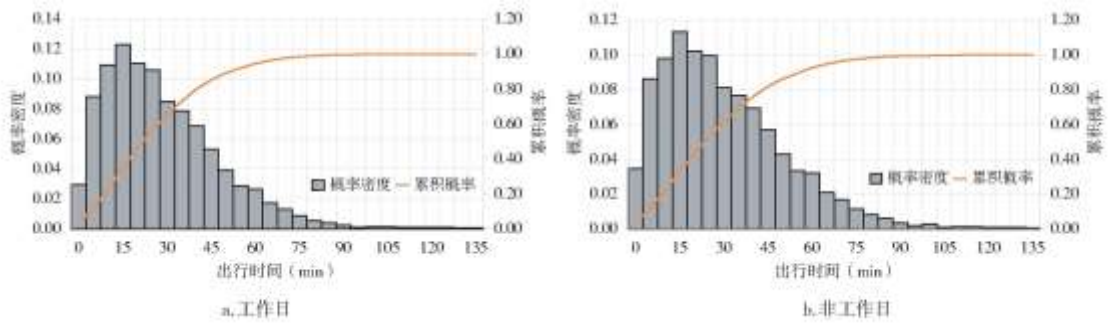


图 10 出行时间累积概率分布图

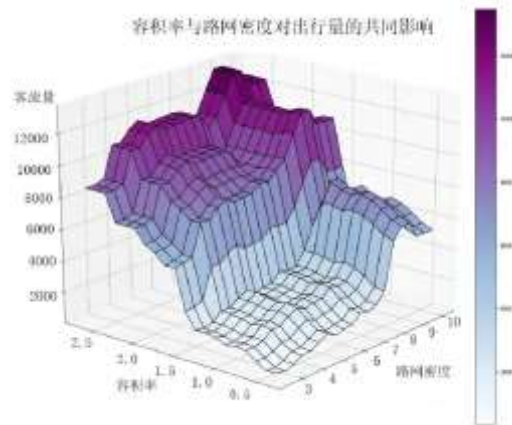


图 11 路网密度、容积率对客流量影响的联合 PD 图

基于路网密度、容积率对客流量联合作用的三维偏依赖图（图 11）可以看出，站点地区合理的街区尺度以及适度的容积率可显著增强区域的地铁客流量。(1) 路网密度特征重要度为 0.099，偏依赖图曲线在 $7.5\text{km}/\text{km}^2$ 之前表现为小范围波动上升趋势，对客流量贡献区间为 $1.7\sim 1.9$ 万，在超过 $7.5\text{km}/\text{km}^2$ 后，客流量迅速提升至 2.3 万，之后基本保持不变。这表明 $7.5\text{km}/\text{km}^2$ 路网密度对站点地区地铁出行具有显著提升作用（图 12）。路网密度高的站区街区尺度较小，步行可达性较高，地铁慢行接驳较方便，有利于提高人群轨道交通使用频率，这与黄晓燕^[39]、金昱^[16]等人的研究是一致的。(2) 容积率对轨道交通出行量影响系数为 0.076，

偏依赖图可以看出容积率在 1.5 和 2.4 左右对客流量影响显著，容积率体现了地块的开发强度与人流的聚集程度，一方面开发强度大的站区经济活力高、出行强度大，另一方面容积率较高的站点地区集中于居住、办公与核心商圈区域，高峰时段轨道出行强度大。

距 CBD 距离与客流量呈波动“凹”形非线性关系（图 12），影响系数为 0.073。当距离从 0km 增加至 18km 时，客流量波动减少约 2500 人，18~26km 对客流量几乎没有影响，而当距离从 26km 增加至 28km 时，客流量又增长约 1500 人。可能的原因是：城市中心区服务设施完善，为获取医疗、就业、教育等服务，大部分地铁出行者会向城市中心区移动，而城市边缘区域的站点客流来源地广，腹地不仅包含站域，也服务城市周边的乡镇地区居民，因此，会形成这种 U 形非线性关系。

办公设施、文化教育设施数量对客流量的影响均为非线性正相关，影响系数为 0.072 与 0.061，分别在超出 300 与 75 范围时对客流量增长最显著。这表明目前通勤、通学仍是城市轨道交通最重要出行目的之一。(1)办公设施较为集中的轨道交通站区客流辐射面大，人流聚集程度与出行频率较高，且该区域停车资源紧张，道路易拥堵。因此，工作人群回家更倾向于使用地铁作为交通工具。(2)而文化教育设施较为集中的站区，以学生与教师群体的通学、通勤为主，学生的出行方式主要为公共交通，因此文化教育设施对客流量的影响也较为显著。

房价对客流量影响系数为 0.071，偏依赖图非线性关系表现为“凸”形特征。在房价 0.8~2 万区间，客流量从 1.65 万人梯度增加至 1.95 万人，之后 2.3~3 万范围，又逐步减少至 1.8 万人，这表明中等房价站区对客流量的影响比高房价与低房价站区更大。主要原因是高收入者与低收入者更倾向于就近居住，而中等收入阶层在房价、收入、通勤三者博弈过程中，更倾向居住于城市中、低房价区域，并使用轨道交通，前往高收入集中区就业；另外，高房价区域通常有车率要高于低房价区域，因此轨道出行率相对较低。

交通设施与客流量呈正相关，影响系数为 0.060。当交通设施（公交站点、共享单车停靠点、停车场、汽车站等）数量从 200 增加至 250 时，客流量增加了约 4000 人。这表明公交车站、共享单车停靠点等交通设施的数量往往会对地铁的乘车人数产生积极影响，交通设施增强了地面公共交通的可达性，提升城市轨道交通接驳的便利性，有利于站点区域以及站区以外的人群使用地铁。

4.2 影响因素对出行时间的非线性作用机制

利用五折交叉验证（five-fold cross-validation）结合网格搜索算法获得 GBDT 模型的最优参数，得到站点乘客平均出行时间（以下简称出行时间）与影响因素的最终模型（ $R^2=0.75$ ）。在最终模型中，根据 GBDT 的影响因素特征重要度排序（图略），提取影响出行时间变化最重要的七种因素偏依赖图（图 13）进行分析，分别为：距 CBD 距离、交通设施、居住及社区服务、生活服务、房价、办公设施以及购物服务等。

距 CBD 距离与出行时间偏依赖图（图 13）呈非线性正相关关系，其影响系数达到了 0.194，这表明站点距 CBD 的距离是影响各站点乘客出行时耗的最重要因素。一方面由于公共服务设施、就业岗位集中于中心城区，部分居住在城市外围的出行者需要向心长距离出行获取就业机会，导致随距离增大出行时间增加；另外，由于轨道线网的建设先后顺序，城市中心区的轨道线网密度要远大于城市外围区，且中心区站点的客流来源去向腹地要小于郊区型站点，因此，中心区站点的换乘便利度高、接驳时间短、出行绕行系数小，乘客平均出行时间也相对较少。从偏依赖图可以看出当距离达到 4km、10km 时，出行时间增加约 6min、5min，偏依赖曲线呈强烈非线性增长趋势。距离城市中心 4km、10km 为成都市 1 环与 2 环的位置，环线与轨交站点的叠合区域可达性较高，这也反映了 2 环内的“环内出行”、2 环外“向心出行”的轨道出行圈特点。

交通设施与出行时间的偏依赖曲线呈负相关，影响系数为 0.134。主要原因是交通设施集中站区公共交通便利，出行方式选择多样，出行者可使用接驳时间更短的公共交通（公交巴士、出租车、网约车等）方式出行，分担了部分地铁中、长距离的出

行。而交通设施缺乏的轨交站域，地铁成为长途出行者的首选交通方式，因此出行时间较长。

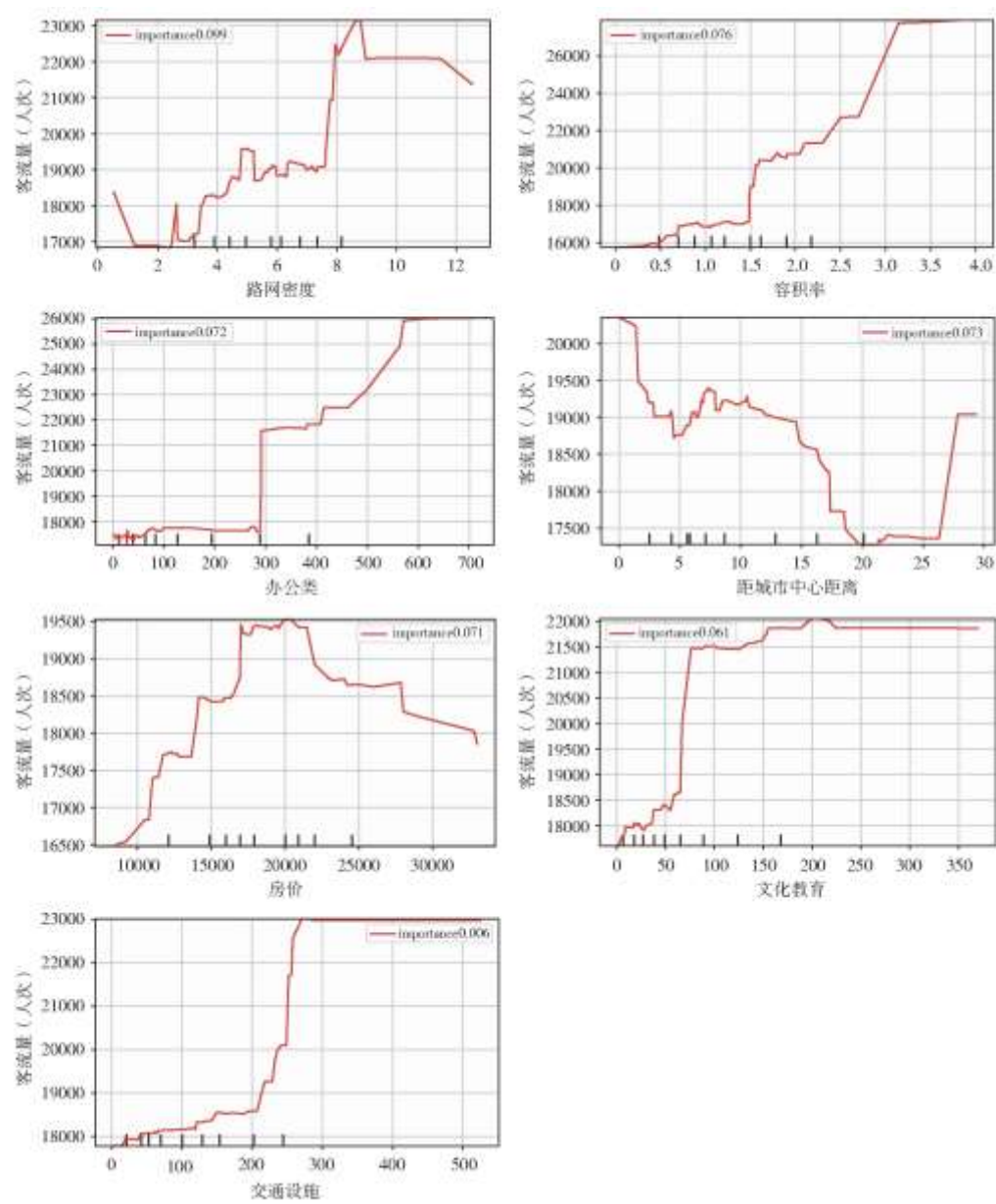


图 12 部分影响因素与客流量的非线性关系图

房价对出行时间的影响系数为 0.076，整体呈非线性负向影响关系。成都高房价地段集中分布于城市核心区以及高新的华阳、广福等大型集中居住区，这些区域靠近金融城、世纪城等集中就业型站点，就业可达性高，通勤距离较短。

办公设施数量与出行时间的偏依赖图呈三段式的非线性变化趋势，影响系数为 0.070。当办公设施数量从 0 增加到 130 左右时，对出行时间的影响较小，而超过 130 范围之后，出行时间从 37.8min 上升至 38.2min，之后直到 400 范围内都为缓慢增长。主要原因可能为：公司数量高度集中的站域就业辐射圈大，导致到此就业及回家的轨道通勤者出行距离相对较长。另外，高密度就业集中区的早晚高峰人流高度集聚，拥挤系数大，站内候车时间长，也进一步增加了出行时间。

购物设施数量对出行时间的影响为正相关,影响系数为 0.061。当购物设施数量从 0 增加至 300,出行时间平缓提升了 0.5min,之后超出 300 范围直到 750,出行时间从约 38min 增加至 39.5min,且在 450 左右出行时间具有显著的阈值效应。

5 结论与讨论

5.1 结论

本文基于成都市轨道交通刷卡数据、POI 数据、建成环境、经济属性等多源大数据,从出行量与出行时间两个方面,解析了城市轨道交通客流的时空分布特点。运用机器学习(梯度提升决策树)方法,探讨了站点地区影响因素对出行量、出行时间的非线性作用机制,研究表明:

第一,成都市地铁客流量空间分布趋势由东一西与北一南均呈现两边低、中间高的倒 U 型特征,城市中心区的客流量显著大于城市外围地区,每站平均客流量为 1.6 万/d。地铁客流最集中的区域为天府广场—天府五街 15.6km 核心段,这与城市双核心结构的发展趋势一致;早期开通的 1、2 号线承载了大部分客流量,两线客流承载率近 50%;工作日客流时序变化规律呈现双峰型,具有早晚高峰潮汐特征,而非工作日则呈无峰型,客流分布较均衡。

第二,站点尺度下的进站客流平均出行时间空间格局呈现外围高、中间低的 U 型分布趋势,三环内的乘客平均出行时间为 28min,而三环外的乘客平均出行时间为 36min;地铁出行时间与出行量概率密度分布呈 Gamma 曲线特征,工作日平均出行时间为 30.2min,而非工作日为 32.6min,且集中出行时间均符合 15~25min 定律。

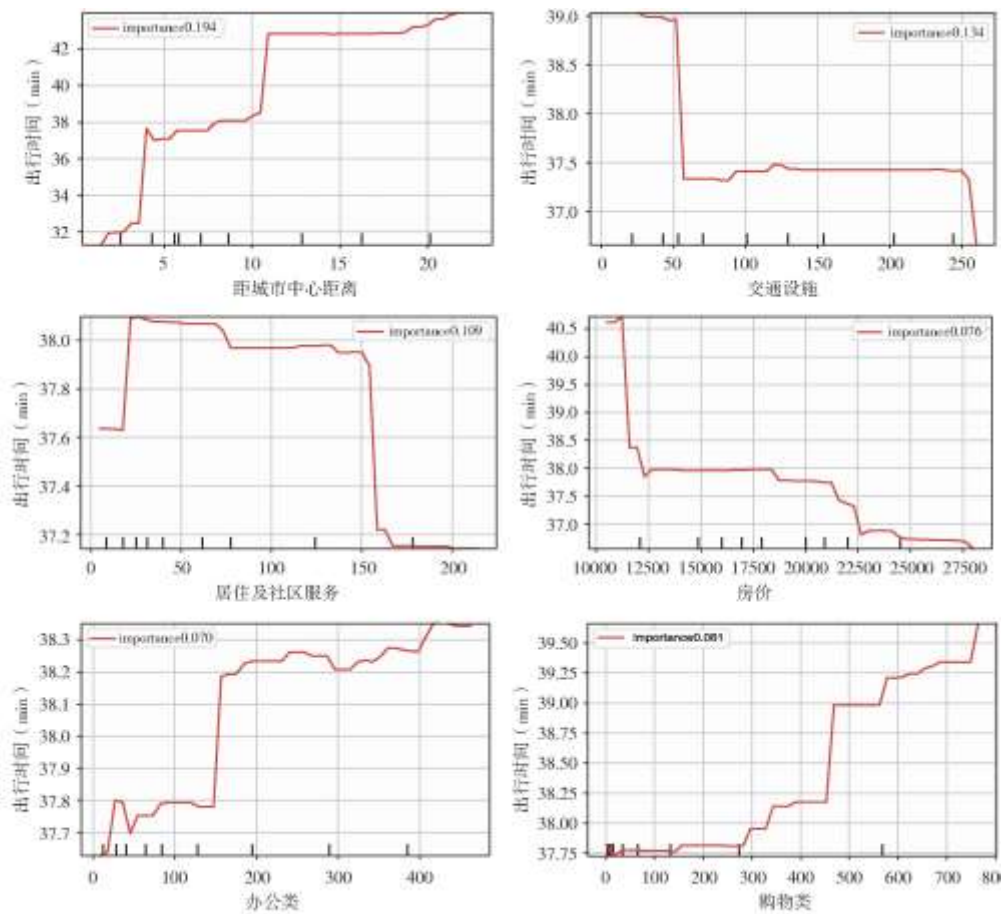


图 13 部分影响因素与出行时间的非线性关系图

第三，出行量、出行时间与 15 个影响因素的梯度提升决策树（GBDT）呈显著的非线性特征与阈值效应，且影响因素与出行量、出行时间的非线性关系差异显著。出行量模型的拟合优度大于出行时间模型，表明相比出行时间，站点地区建成环境、经济属性、设施分布更易影响出行量。

第四，建成环境对地铁出行量的影响较为显著，路网密度、容积率、办公设施、教育文化以及交通设施等与客流量为非线性正相关关系，呈多段式梯度上升特征。这表明高密度开发、可步性强的站区建设，有助于提升地铁的使用率。而房价、距 CBD 距离与客流量关系则分别呈“凸”与“凹”形的非线性曲线，具有典型的圈层效应特征。

第五，距 CBD 距离对站点乘客平均出行时间特征重要度最大（19.4%），呈非线性正相关，主要是由于外围区线网密度过低、职住空间分布不均以及二环内外出行圈分化导致的。另外，购物设施数量、办公设施、居住及社区服务等也与出行时间非线性正相关，交通设施、房价等则呈负相关，说明必要性工作、生活设施对地铁乘客出行时间有重要影响。

5.2 规划启示

第一，考虑“流量”与“时间”阈值效应的建成环境优化。客流量与出行时间是城市轨道交通出行服务研究的重点，而轨道交通站域是乘客使用的主体区，交通设施、办公设施、居住及社区服务设施、路网密度等与出行量、出行时间的阈值效应是站域城市设计综合考虑的重点。合理的设施布局、建成环境控制可有效提高站区地铁使用率，减少通勤时间。以容积率的梯度控制为例，传统的 TOD 开发为中间高、周围低的梯度布局模式^[26]，但是这种布局模式仅仅考虑了核心区高密度开发的人流聚集与经济效应，从容积率的偏依赖图可以看出，1.5 的容积率对出行时间与出行量都具有高敏感性，且当容积率达到 3.2 时，客流量增加不明显。因此，可根据影响因素的阈值效应，结合站点主导类型（就业型、居住型、混合型、枢纽型、综合型）进行容积率差异化定级，在 TOD 核心区和影响区两个圈层分别控制容积率指标上下限，以及圈层开发强度的控制范围，以实现站点地区土地开发的集约化利用与合理控制，引导人群出行与站域空间的良性互动。

第二，结合用户出行规律的职住空间布局与线网密度提升。距 CBD 距离、设施数量对客流量与出行时间的影响都较为显著，主要原因是线路建设的时序关系、城市外围区的轨道交通线网密度过低、边缘区服务设施不足。基于此，提出以下优化建议：(1) 基于轨道交通的“出行+生活”通勤圈构建：成都市部分站点人流量大，且出行时间较长，结合 2 环内的“环内出行”、2 环外“向心出行”的乘客通勤特点，可通过调整不同城市设施的配置与布局，提升站域土地混合使用与产业要素流动效率。实现轨道交通对沿线职住空间的有效链接，构建紧凑高效的轨道通勤圈，并促进站域人群出行与职住空间的合理匹配。(2) 渐进式的线网密度提升：逐步提升外围核心组团区域的轨道线网密度，减少郊区轨道出行者的绕行系数、换乘时间，实现郊区型站点之间的职住区域的快速连接，并进一步压缩长距离通勤者的出行时间，提高整体出行效率。

第三，轨道网络适应城市存量布局向耦合城市空间规划转变。成都城市轨道交通目前处于高速扩张期，在建里程超过 300km，提出了建设陆肖站等 13 个 TOD 的发展目标。基于前述分析可以看出，早期成都市的轨道线网的建设以适应城市的建成空间结构为目标，但也导致了老线路（1、2 号线）站域开发密度过高，客流荷载过大等问题。因此，中后期城市轨道网的建设与规划应该耦合未来空间规划与交通系统，实现公交系统与城市空间的协同发展。

5.3 讨论

本研究是将机器学习的方法应用于城市轨道交通出行强度、出行时间的影响因素及非线性关系的一次有益探索。由于数据获取的限制，考虑的影响因素尚存在不足，如对于各个站区的人群属性变量考虑较少。下一步研究应利用手机信令数据识别站区职住密度与人群属性，增加人群属性指标对站点客流量、出行时间影响的综合分析。另外，地铁客流的时变性规律、站域空

间的圈层差异也是未来城市轨道交通出行研究中需进一步深化的方面。

参考文献:

- [1]Pan H,Li J,Shen Q,et al.What determines rail transit passenger volume? Implications for transit oriented development planning[J].Transportation Research Part D:Transport and Environment,2017,57:52-63.
- [2]赵晖,杨军,刘常平.轨道沿线居民职住分布及通勤空间组织特征研究——以北京为例[J].经济地理,2011,31(9):1445-1451.
- [3]曹小曙,林强.世界城市地铁发展历程与规律[J].地理学报,2008,63(12):1257-1267.
- [4]韩寒.深圳市轨道交通结构与商业活力空间关联分析[J].经济地理,2021,41(3):86-96.
- [5]Guerra E,Cervero R,Tischler,D.Half-mile circle does it best represent transit station catchments?[J].Transportation Research Record,2012,2276:101-109.
- [6]Cervero R,Murphy S,Ferrell C,et al.Transit-oriented development in the United States:experiences, challenges, and prospects[J].Urban Planning Overseas,2005,8(1):1-7.
- [7]Li Y,Yang L,Shen H,Wu Z.Modeling intra-destination travel behavior of tourists through spatio-temporal analysis[J].Journal of Destination Marketing&Management,2019,11:260-269.
- [8]龙瀛,张宇,崔承印.利用公交刷卡数据分析北京职住关系和通勤出行[J].地理学报,2012,67(10):1339-1352.
- [9]Gan Z,Feng T,Wu Y,et al.Station-based average travel distance and its relationship with urban form and land use:An analysis of smart card data in Nanjing City,China[J].Transport Policy,2019,79:137-154.
- [10]付诗航,刘耀林,方莹,等.基于SCD的公共交通换乘时空模式——以武汉市为例[J].武汉大学学报:信息科学版,2020,45(7):1089-1098.
- [11]Huang J,Levinson D,Wang J,et al.Tracking job and housing dynamics with smartcard data[J].Proceedings of the National Academy of Sciences of the United States of America,2018,115(50):12710-12715.
- [12]王宇宁,杨安娜.“时距”视角下城市轨道线网的通勤效率——以天津市中心城区为例[J].经济地理,2019,39(7):67-75.
- [13]黄洁,王姣娥,靳海涛,等.北京市地铁客流的时空分布格局及特征——基于智能交通卡数据[J].地理科学进展,2018,37(3):397-406.
- [14]王子甲,刘海旭,TAKU Fujiyama.基于AFC数据和RF模型的城轨车站服务功能分类[J].交通运输系统工程与信息,2018,18(4):224-230.

-
- [15]尹芹, 孟斌, 张丽英. 基于客流特征的北京地铁站点类型识别[J]. 地理科学进展, 2016, 35(1):126-134.
- [16]金昱. 基于上海轨道交通刷卡数据的乘客出行模式研究[J]. 都市快轨交通, 2019, 32(3):91-96.
- [17]李国强, 杨敏, 王树盛. 基于 AFC 和 POI 数据的轨道交通车站客流影响因素挖掘[J]. 城市交通, 2019, 17(1):102-108, 120.
- [18]Shi Z, Zhang N, Liu Y, et al. Exploring spatiotemporal variation in hourly metro ridership at station level:The influence of built environment and topological structure[J]. Sustainability, 2018, 10(12):4564.
- [19]Chava J, Peter N, Reena T. Gentrification of station areas and its impact on transit ridership[J]. Case Studies on Transport Policy, 2018, 6(1):1-10.
- [20]Becky P Y L, Cynthia C, Eric T H C. Rail-based transit-oriented development:Lessons from New York City and Hong Kong[J]. Landscape and Urban Planning, 2010, 97(3):202-212.
- [21]Gan Z, Min Y, Tao F. Examining the relationship between built environment and metro ridership at station-to-station level[J]. Transportation Research Part D:Transport and Environment, 2020, 82:102332.
- [22]Friedman J H. Greedy function approximation:A gradient boosting machine[J]. Annals of Statistics, 2001, 29(5):1189-1232.
- [23]Eom J K, Choi J, Park M S. Exploring the catchment area of an urban railway station by using transit card data:Case study in Seoul[J]. Cities, 2019, 95:102364.
- [24]Guo Y, He S Y. Built environment effects on the integration of dockless bike-sharing and the metro[J]. Transportation Research Part D:Transport and Environment, 2020, 83:102335.
- [25]申犁帆, 王烨, 张纯, 等. 轨道站点合理步行可达范围建成环境与轨道通勤的关系研究——以北京市 44 个轨道站点为例[J]. 地理学报, 2018, 73(12):2423-2439.
- [26]赵鹏军, 李南慧, 李圣晓. TOD 建成环境特征对居民活动与出行影响——以北京为例[J]. 城市发展研究, 2016, 23(6):45-51.
- [27]Cervero R, Jin M, Mark M. Direct ridership model of bus rapid transit in Los Angeles County, California[J]. Transportation Research Record, 2010, 13:1-7.
- [28]Tao T, Wang J, Cao X. Exploring the non-linear associations between spatial attributes and walking distance to transit[J]. Journal of Transport Geography, 2020, 82:102560.
- [29]Ding C, Cao X, Liu C. How does the station-area built environment influence Metrorail ridership? Using gradient boosting decision trees to identify non-linear thresholds[J]. Journal of Transport Geography, 2019, 77:70-78.
- [30]丁聪, 倪少权, 吕红霞. 基于梯度提升的城市轨道交通客流量预测分析[J]. 城市轨道交通研究, 2018, 21(9):60-63.

-
- [31]翁小雄, 吕攀龙. 基于 GBDT 算法的地铁 IC 卡通勤人群识别[J]. 重庆交通大学学报: 自然科学版, 2019, 38(5):8-12.
- [32]Yin C, Cao X, Sun B. Examining non-linear associations between population density and waist-hip ratio: An application of gradient boosting decision trees[J]. *Cities*, 2020, 107:102899.
- [33]Ding C, Cao X, Wang Y. Synergistic effects of the built environment and commuting programs on commute mode choice[J]. *Transportation Research Part A: Policy and Practice*, 2018, 118:104-118.
- [34]Ma X, Ding C, Luan S, et al. Prioritizing influential factors for freeway incident clearance time prediction using the Gradient Boosting Decision Trees Method[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2017, 18(9):2303-2310.
- [35]钟敏慧, 张婉露, 李有儒, 等. 基于 GBDT 的铁路事故类型预测及成因分析[J/OL]. 自动化学报, 2020, <https://doi.org/10.16383/j.aas.c190630>.
- [36]Brockmann D, Hufnagel L, Geisel T. The scaling laws of human travel[J]. *Nature*, 2006, 439(7075):462-465.
- [37]Roth C, Kang S M, Batty M. Structure of urban movements: polycentric activity and entangled hierarchical flows[J]. *PLOS ONE*, 2011, 6(1):e15923.
- [38]王淑伟. 站点周边用地特性对轨道客流影响机理研究[D]. 北京: 北京工业大学, 2015.
- [39]黄晓燕, 曹小曙, 殷江滨, 等. 城市轨道交通和建成环境对居民步行行为的影响[J]. 地理学报, 2020, 75(6):1256-1271.