

# 基于职住地关系的城市轨道交通乘客出行 特征挖掘和乘客分类<sup>\*</sup>

周 琪<sup>1</sup> 龚 璐<sup>2</sup> 马明辰<sup>1</sup> 孙守胜<sup>1</sup> 常青青<sup>2</sup>

(1. 同济大学电子与信息工程学院, 201804, 上海; 2. 公安部第三研究所, 200031, 上海)

**摘 要** [目的] 依据城市轨道交通出行轨迹数据挖掘出行特征时, 乘客出行活动模式与其居住地和就职地站点紧密关联, 在出行轨迹上表现为乘客乘车线路固定、进出站站点集中等特征。同时, 关联居住地和就职地的乘客出行轨迹也呈现出一定的波动性, 与职住地强关联的出行数据和其他出行数据间往往表现出较明显的差异。为实现对乘客出行轨迹特征的精准挖掘, 厘清城市轨道交通乘客出行数据中的职住地关系。[方法] 提出一种基于职住地关系的出行特征挖掘和乘客分类机制, 从基于居住地、就职地、社会活动地的乘客轨迹活动热点角度形成初步职住地判断, 探索与职住地关联性强的出行时空规律; 通过基于乘客出行轨迹特征值提取的 K-means (K 均值) 聚类算法对职住地判断后的乘客聚类结果, 形成精准匹配乘客出行特征的乘客分类方法。[结果及结论] 仿真验证结果表明, 职住地关系明确的乘客出行模式分为两点式、两点两线式、三点式及多点式等 4 类, 该出行模式分类结果是构建精细化乘客画像与制定智能安检策略的重要技术基础。

**关键词** 城市轨道交通; 乘客出行特征; 乘客分类; 职住地关系

中图分类号 U293.13

DOI:10.16037/j.1007-869x.2024.08.019

## Urban Rail Transit Passenger Travel Characteristics Mining and Passenger Classification Based on Workplace-residence Relationship

ZHOU Qi<sup>1</sup>, GONG Lu<sup>2</sup>, MA Mingchen<sup>1</sup>, SUN Shousheng<sup>1</sup>, CHANG Qingqing<sup>2</sup>

(1. School of Electronics and Information Engineering, Tongji University, 201804, Shanghai, China; 2. The Third Research Institute of the Ministry of Public Security, 200031, Shanghai, China)

**Abstract** [Objective] When mining travel characteristics from urban rail transit travel trajectory data, passenger travel activity patterns are found closely related to the stations near their residence and workplace. This is reflected in their travel

trajectories by fixed boarding/alighting routes and concentrated entry/exit stations. Meanwhile, passenger travel trajectories associated with residence and workplace exhibit certain fluctuations, and the travel data strongly related to workplace-residence often differ significantly from other travel data. To accurately extract passenger travel trajectory characteristics, it is essential to clarify the residence-workplace relationship within urban rail transit passenger travel data. [Method] A mechanism for mining travel characteristics and classifying passengers based on residence-workplace relationship is proposed. The preliminary workplace-residence judgments are formed by identifying the hotspots of passenger trajectory activities related to residence, workplace and social activity locations. Then, the spatio-temporal patterns strongly associated with workplace-residence are explored. A K-means clustering algorithm extracted on the basis of passenger travel trajectory characteristic values is used to cluster passengers according to their residence-workplace judgment, resulting in a passenger classification method that precisely matches the passenger travel characteristics. [Result & Conclusion] Simulation validation results indicate that passengers with clearly defined residence-workplace relationship exhibit four distinct travel patterns of two-point, two-point-two-line, three-point, and multi-point. These travel pattern classification results make up a crucial technical foundation for constructing detailed passenger profiles and developing intelligent security check strategies.

**Key words** urban rail transit; passenger travel characteristics; passenger classification; workplace-residence relationship

关联居住地和就职地的城市轨道交通出行轨迹呈现出一定的波动性, 如由通勤乘客轨迹构建的职住地出行链出现断裂或分裂出额外的出行热点时, 由于信息被模糊化, 给准确分析职住地出行轨迹带来挑战<sup>[1-3]</sup>。在出行模式挖掘方面, 城市轨道交通出行轨迹可通过 GPS (全球定位系统)<sup>[4-6]</sup>、拓

<sup>\*</sup> 上海市科委社会发展科技攻关项目 (21DZ1200702)

扑图构建等可视化手段进行分析,了解其所在位置区域的特征,用于出行预测等方面<sup>[7-8]</sup>。目前,针对轨迹数据的不同时空属性对出行模式划分的影响程度和权重配置研究较少。本文提出从轨迹数据隐含的职住地关系角度,研究与职住地强关联的城市轨道交通出行数据的时空规律,进一步实现出行特征挖掘和乘客分类,可为城市轨道交通智能安检策略制定等提供依据,具有重要的实用意义。

## 1 基于城市轨道交通出行热点划分的职住地判断

居住地站点周边主要以住宅、居住类用地为主,客流潮汐特征显著,早高峰相对晚高峰更集中,即此类站点通勤特征明显;就职地站点与居住地站点进出站客流呈相反趋势。为研究关联职住地的城市轨道交通出行方式,本文依据早、晚高峰时段通勤乘客进站频率将乘客的活动地分成居住地站点、就职地站点和社会活动站点(除去居住地站点和就职地站点以外的其他站点,包括购物、旅游等活动场所)。

为衡量基于热点分类的职住地判断后乘客通勤规律程度的有效性,采用工作日刷卡总数、日均出行次数、进出站点相似度、每日首次刷卡时间的平均值和首次刷卡时间加权标准差5个指标进行分析。其中,进出站点相似度表征乘客在一个月所有工作日内出行轨迹数据的起讫点的相似程度。计算乘客所有出行站点的概率 $p$ , $p=1$ 表示此卡号有出行记录的工作日内每天都在此站点进站或出站。根据站点出现次数和出行站点概率制定站点相似度特征值赋值规则,如表1所示。

表1 站点相似度特征值赋值规则

Tab.1 Assignment rules of site similarity eigenvalues

站点出现次数 $n$	$p > 0.9$ 站点个数	$0.7 < p \leq 0.9$ 站点个数	特征值赋值
$n < 10$	$\forall$	$\forall$	0
	$\geq 2$	$\forall$	3
	1	1	3
$n \geq 10$	0	2	3
	1	0	2
	0	1	1
	0	0	0

## 2 基于出行模式的乘客特征分类算法

考虑到刷卡数据规模大且聚类特征属性为数值型,本文采用  $K$ -means ( $K$  为簇的个数) 算法进行城市轨道交通出行模式聚类。度量聚类模型性能的指标采用 SSE(簇内平方和),并基于此判断职住地识别后乘客分类的类别数,提取相应的特征变量值作为  $K$ -means 聚类算法的特征变量输入,并根据赋值规则对乘客数据进行量纲一化处理。

1) 社会活动站点最大频率。依据乘客所有出行站点筛选出社会活动站点,并计算最高频率。

2) 社会活动站点距离。将出行轨迹途经站点个数作为计算起讫点间距离依据,计算每两个站点间的最短路径。社会活动站点距离为1)中筛选出的最大频率社会活动站点距离乘客居住地站点的最短路径距离。

3) 居住地占比。居住地站点在所有出行起讫点出现次数占比。

4) 就职地占比。就职地站点在所有出行起讫点出现次数占比。

## 3 基于职住地关系的出行空间标签仿真验证

### 3.1 仿真环境与数据预处理

试验仿真环境使用 Python3.7 编程,可实现数据预处理、乘客刷卡数据匹配、统计站点客流分析及使用聚类算法对乘客基于职住地关系精细分类。数据集采用上海轨道交通公开的 AFC(自动售检票)系统中2015年4月1日—30日的上海公共交通卡真实刷卡数据,单日数据条数约为1 000万条,数据为 csv 格式。首先进行数据预处理,包括进出站时间由24 h 计时统一转换为秒计时的时间戳形式、删除非运营时间的错误数据等。然后进行轨迹匹配处理,并进一步随机筛选出在工作日有出行记录的10 000名乘客轨迹数据以便后续分析通勤乘客的出行模式规律,得到完整有效的刷卡数据如表2所示。表2中展示了4月23日刷卡数据中前3条经过轨迹匹配及随机筛选后的有效轨迹数据。

基于职住地关系的乘客分类采用在工作日有出行记录乘客的出行数据。为筛选出通勤需求度高的乘客,基于上述10 000名乘客数据,统计其所有工作日的出行次数总和,绘制嵌套饼状图,如图1所示。内层饼图表示不同出行次数类别中乘客

表 2 经过刷卡数据匹配的乘客刷卡数据

Tab. 2 Passenger card-swiping records after card-swiping data matching

卡号	刷卡日期	刷卡时刻	进站时间/s	进站乘坐线路	进站站点	出站时间/s	出站乘坐线路	出站站点
3003834123	2015-04-23	18:07:44	65 264	7 号线	龙阳路站	67 254	3 号线	上海火车站站
2002182104	2015-04-23	08:07:19	29 239	4 号线	临平路站	30 213	6 号线	源深体育中心站
2002182104	2015-04-23	17:09:03	61 743	6 号线	源深体育中心站	62 838	4 号线	临平路站

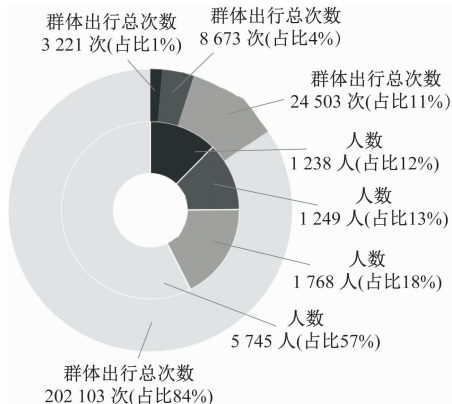


图 1 不同出行频率类别中乘客数量分布以及出行次数总和

Fig. 1 Passenger number distribution and total number of trips in different travel frequency categories

数量分布,外层饼图表示不同出行次数类别乘客对应的出行次数总和。结果显示,在 10 000 名乘客中,工作日总出行次数达到 20 次的乘客有 5 745 名。由以上乘客数量统计数据可知,本文设置工作日总出行次数的出行需求度评价指标是合理有效的。

### 3.2 职住地关系 K-means 聚类结果

为对比职住地关系明确乘客与未处理过职住地关系乘客的工作日通勤特征差异,进行了仿真模拟分析。图 2 为仿真模拟结果中通勤乘客在职住地关系判断前后 5 项属性特征值平均值雷达图。与未处理过职住地关系的乘客的轨迹数据相比,职住地关系明确乘客的出行起讫点主要集中于居住地和就业地,社会活动站点占比较低;乘客更倾向于早高峰时段首次刷卡上班;具有明确职住地关系的乘客,其工作日通勤特征更明显,且出行稳定性和规律性更强。

对聚类后的数据 4 项特征变量进行标准差量纲一化处理,计算不同簇数下的簇内误差平方和,确定最合适的簇数为 4。将社会活动站点最大频率、社会活动站点距离、居住地占比、就业地占比 4 个属性特征分别作为坐标绘图,不同簇的样本用不同深浅色点表示,簇心用较大黄色点区分,

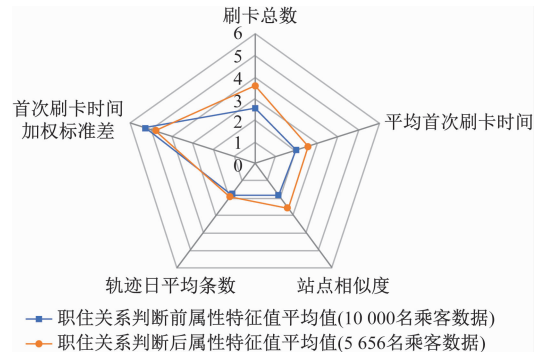


图 2 通勤乘客在职住地判断前后 5 项属性特征值平均值雷达图

Fig. 2 Average value radar chart of five attribute eigenvalues before and after commuting passenger workplace-residence relationship judgements

职住地关系 K-means 聚类结果二维展示、三维展示分别如图 3 和图 4 所示。由图 4 可知,不同簇之间的样本点相互隔离,同一簇样本点则分布较为集中。

根据 1 个月内所有工作日的乘客出行数据判断职住地关系后,根据赋值规则量纲一化处理后,5 656 名通勤乘客各出行模式下聚类变量的平均值以及人数占比如表 3 所示。依据聚类结果将工作日城市轨道交通出行模式分为两点式、两点两线式、三点式及多点式 4 类。

1) 两点式。在居住地和就业地间往返,不同乘客出行时间存在一定的差异,二者在出行轨迹空间位置特征的相似性很高,具有高频通勤出行的特征。

2) 两点两线式。工作日出行轨迹中 80% 以上为固定式居住地与就业地两点通勤轨迹,此类乘客偶尔于非上班时段乘坐城市轨道交通到社会活动站点。

3) 三点式。出行轨迹的通勤特征与前两类相似,乘客社会活动站点最大频率与居住地占比之和约为就业地占比,且社会活动站点距离居住地较近,由此可推测乘客有居住地、就业地之外的相对固定社会活动站点,呈现出较强规律性。

4) 多点式。此类乘客工作日出行轨迹可能是两点、三点或更多点出行,但由于这种不规律的出



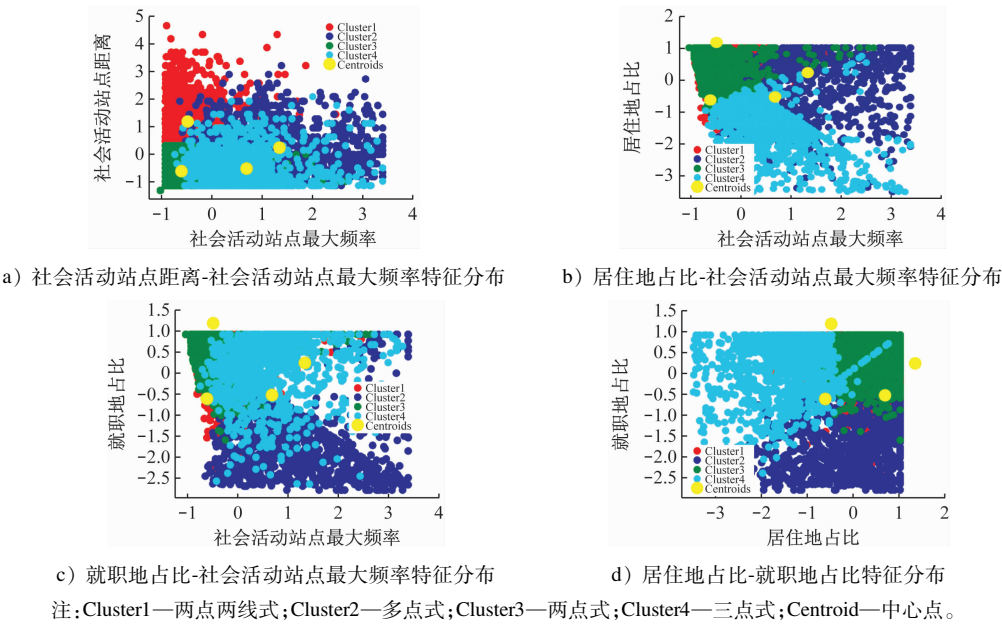


图 3 职住地关系 K-means 聚类结果二维展示

Fig. 3 2D display of K-means clustering results of workplace-residence relationship

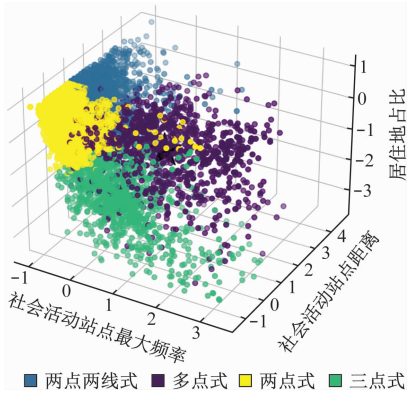


图 4 职住地关系 K-means 聚类结果三维展示

Fig. 4 3D display of K-means clustering results of workplace-residence relationship

行在一段时间内频次较低,难以呈现出较强的规律性。如乘客可能早上从居住地出发,然后到达多个地点从事一定的社会活动,最后返回居住地。推测此类乘客可能是销售职业人员,在工作时间段内需要到不同的社会活动点处理事务。

由表 3 可知,两点式、两点两线式及三点式人数占比总和为 82.96%,这三类乘客的出行总数全部为 10 次以上,对通勤需求度极高。若设定轨迹规律授信阈值为乘客规律出行次数 10 次以上,5 656 名乘客中有 4 692 名乘客出行中职场住地关系明确且绝大多数为规律出行。乘客出行模式聚类结果中,出行次数 20 次以上占比 92% 以上,总出行次数占比

表 3 1 个月内所有工作日职住地关系模式下各聚类变量量纲一化的平均值及人数占比

Tab. 3 Average values of each clustering variable normalization under workplace-residence relationship mode for all working days within a month

出行模式	聚类变量量纲一化的平均值				人数占比/%
	社会活动站点最大频率	社会活动站点距离	居住地占比	就业地占比	
两点式	0.098	4.225	0.888	0.898	40.54
两点两线式	0.122	15.373	0.866	0.830	24.79
三点式	0.400	5.013	0.431	0.741	17.63
多点式	0.538	9.527	0.761	0.293	17.04

约为 96%。这表明城市轨道交通乘客的出行模式高度集中,受特定职住地关系的影响较大;乘客出行以短途通勤为主,这与城市轨道交通的短途便捷

性有关,说明乘客更倾向于选择城市轨道交通作为工作日通勤出行方式。

## 4 结语

为实现对城市轨道交通乘客出行轨迹数据中潜在职住地关系的精准特征挖掘,本文提出了基于职住地关系先验划分的乘客分类机制,以及基于时空轨迹先验判断和  $K$ -means 聚类的城市轨道交通出行划分模式。本文研究得出的出行模式分类结果可应用于城市轨道交通安检场景下筛选授信免安检乘客,是构建基于大数据的城市轨道交通高效精准安检系统的重要基础。

## 参考文献

- [1] 高强, 张凤荔, 王瑞锦, 等. 轨迹大数据: 数据处理关键技术研究综述[J]. 软件学报, 2017, 28(4): 959.  
GAO Qiang, ZHANG Fengli, WANG Ruijin, et al. Trajectory big data: a review of key technologies in data processing[J]. Journal of Software, 2017, 28(4): 959.
- [2] 高伟晴, 潘晓, 吴雷. 一种基于语义轨迹的相似性连接查询算法[J]. 计算机应用与软件, 2020, 37(7): 14.  
GAO Yiqing, PAN Xiao, WU Lei. Similarity join query algorithm based on semantic trajectories[J]. Computer Applications and Software, 2020, 37(7): 14.
- [3] LIU F, BI W, HAO W, et al. An improved fuzzy trajectory clustering method for exploring urban travel patterns[J]. Journal of Advanced Transportation, 2021, 2021(1): 6651718.
- [4] 许华杰, 吴青华, 胡小明. 基于轨迹多特性的隐私保护算法[J]. 计算机科学, 2019, 46(1): 190.  
XU Huajie, WU Qinghua, HU Xiaoming. Privacy protection algorithm based on multi-characteristics of trajectory[J]. Computer Science, 2019, 46(1): 190.
- [5] 蔡莉, 李英姿, 江芳, 等. 面向城市热点区域的不平衡数据聚类挖掘研究[J]. 计算机科学, 2019, 46(8): 16.  
CAI Li, LI Yingzi, JIANG Fang, et al. Study on clustering mining of imbalanced data fusion towards urban hotspots[J]. Computer Science, 2019, 46(8): 16.
- [6] 王培, 江南, 万幼, 等. 应用 Hausdorff 距离的时空轨迹相似性度量方法[J]. 计算机辅助设计与图形学学报, 2019, 31(4): 647.  
WANG Pei, JIANG Nan, WAN You, et al. Measuring similarity of spatio-temporal trajectory using Hausdorff distance[J]. Journal of Computer-Aided Design & Computer Graphics, 2019, 31(4): 647.
- [7] 金盛, 苏弘扬, 张静. 融合出行拓扑与序列分析的车辆时空出行模式挖掘[J]. 交通运输系统工程与信息, 2023, 23(2): 40.  
JIN Sheng, SU Hongyang, ZHANG Jing. Inferring spatial-temporal travel patterns of vehicles combining topology of trips and sequence analysis[J]. Journal of Transportation Systems Engineering and Information Technology, 2023, 23(2): 40.
- [8] WANG W, CHEN J, WANG J, et al. Trust-enhanced collaborative filtering for personalized point of interests recommendation[J]. IEEE Transactions on Industrial Informatics, 2019, 16(9): 6124.
- 收稿日期:2024-01-30 修回日期:2024-03-20 出版日期:2024-08-10  
Received:2024-01-30 Revised:2024-03-20 Published:2024-08-10  
· 通信作者:周琪, 研究生, 2331824@tongji.edu.cn  
· ©《城市轨道交通研究》杂志社, 开放获取 CC BY-NC-ND 协议  
© Urban Mass Transit Magazine Press. This is an open access article under the CC BY-NC-ND license
- [18] 陶亮亮, 曾艳华, 刘振撼, 等. 空气幕对地铁隧道火灾温度及流场的影响研究[J]. 中国安全科学学报, 2021, 31(7): 157.  
TAO Liangliang, ZENG Yanhua, LIU Zhenhan, et al. Study on influence of air curtain on temperature and flow field of subway tunnel fires[J]. China Safety Science Journal, 2021, 31(7): 157.
- [19] 江荷, 朱常琳, 李胜涛, 等. 排烟口对地铁隧道火灾机械排烟效果影响研究[J]. 建筑热能通风空调, 2020, 39(3): 73.  
JIANG He, ZHU Changlin, LI Shengtao, et al. Influence of smoke outlet on mechanical smoke exhaust effect of subway tunnel fire[J]. Building Energy & Environment, 2020, 39(3): 73.
- [20] WENG M C, LU X L, LIU F, et al. Study on the critical velocity in a sloping tunnel fire under longitudinal ventilation[J]. Applied Thermal Engineering, 2016, 94: 422.
- [21] 贡颖, 郭聿, 程洋, 等. 基于 FDS 的火源位置对高压电缆沟道火灾的影响研究[J]. 工业安全与环保, 2023, 49(2): 21.  
GONG Hao, GUO Song, CHENG Yang, et al. Study on the influence of fire source location on high voltage cable trench fire based on FDS[J]. Industrial Safety and Environmental Protection, 2023, 49(2): 21.
- [22] KONG J, YOU W, DENG E, et al. Study of the evolution mechanism of multi-window carriage fire under longitudinal ventilation in a metro tunnel[J]. Tunnelling and Underground Space Technology, 2022(127): 104618.
- [23] 中华人民共和国住房和城乡建设部. 地铁设计规范: GB 50157—2013[S]. 北京: 中国建筑工业出版社, 2014.  
Ministry of Housing and Urban-Rural Development of the People's Republic of China. Code for design of metro: GB 50157—2013[S]. Beijing: China Architecture & Building Press, 2014.
- 收稿日期:2023-05-12 修回日期:2023-06-29 出版日期:2024-08-10  
Received:2023-05-12 Revised:2023-06-29 Published:2024-08-10  
· 第一作者:李宇辉, 副教授, 13645161185@163.com  
通信作者:费瑞振, 高级工程师, feiruizhen@crdc.com  
· ©《城市轨道交通研究》杂志社, 开放获取 CC BY-NC-ND 协议  
© Urban Mass Transit Magazine Press. This is an open access article under the CC BY-NC-ND license