

基于手机导航轨迹数据的城市大规模人群出行模式分析

吴杭彬¹, 陈茜茜¹, 靳慧玲², 傅琛³, 黄炜¹, 刘春¹

(1. 同济大学 测绘与地理信息学院, 上海 200092; 2. 自然资源部超大城市自然资源时空大数据分析应用重点实验室, 上海 200063;
3. 北京大学 地球与空间科学学院, 北京 100871)

摘要: 手机导航轨迹数据具有多种交通方式, 反映大规模人群的活动情况, 适合开展不同交通方式下的出行模式研究。基于手机导航数据, 利用 LightGBM 模型实现出行交通方式分类, 得到步行、机动车和非机动车 3 种交通方式下的人群轨迹。基于这 3 类交通方式, 给出人群出行在周末和工作日下时间、空间和距离的分析指标, 并对上海 4 天数百万条手机导航数据开展了实验分析。结果表明: 在时间分布上, 上海居民的周末出行高峰比工作日更晚并且持续时间更短, 出行方式主要以机动车和步行为主; 在空间分布上, 机动车主要集中在高架区域, 步行主要集中在地铁站附近, 高架路和地铁站的引导标志不充足, 周末交通枢纽和商圈类热点区域比工作日多; 在距离分布上, 导航出行距离符合截断幂律分布, 人群导航出行以中短距离为主, 并随距离增长迅速衰减。研究结果可以为城市规划、城市交通管理政策的制定提供理论依据和技术支撑。

关键词: 手机导航轨迹数据; 交通方式识别; 出行模式
中图分类号: P311 **文献标志码:** A

Analysis of Urban Mass Crowd Traveling Patterns Based on Mobile Phone Navigation Trajectory Data

WU Hangbin¹, CHEN Qianqian¹, JIN Huiling², FU Chen³, HUANG Wei¹, LIU Chun¹

(1. College of Surveying and Geo-Informatics, Tongji University, Shanghai 200092, China; 2. Key Laboratory of Spatial-Temporal Big Data Analysis and Application of Natural Resources in Megacities of the Ministry of Natural Resources, Shanghai 200063, China; 3. School of Earth and Space Sciences, Peking University, Beijing 100871, China)

Abstract: Mobile phone navigation trajectory data has a variety of traffic modes, reflecting the activities of mass crowd, which is suitable for the study of traveling

patterns in different traffic modes. Based on mobile phone navigation data, the LightGBM model for traffic mode classification is first proposed to obtain the trajectories of the population in three transportation modes: walking, motorized, and non-motorized mode. Based on these three types of modes, the analysis indexes of time, space, and distance traveling patterns on weekends and weekdays are given, and an experimental analysis is conducted with the mobile phone navigation data of millions of people in Shanghai for four days. The results show that in terms of time distribution, the weekend traveling peak of the residents in Shanghai is later and shorter than that of weekdays, and traveling modes are mainly motor vehicles and walking. In terms of spatial distribution, motor vehicles are mainly concentrated in elevated areas, walking is mainly concentrated near subway stations, the guidance signs of elevated roads and subway stations are not sufficient, and there are more traffic hubs and shopping hotspots on weekends than on weekdays. In terms of distance distribution, navigation traveling distance conforms to the power-law distribution of intercepted segments, and the navigation traveling of the crowd is dominated by short and medium distances, and decays rapidly with distance growth. The research results can provide theoretical basis and technical support for urban planning and urban traffic management policy formulation.

Key words: mobile phone navigation trajectory data; transportation mode recognition; traveling patterns

大规模人群出行可以揭示人类活动规律, 在人类日常活动性研究、交通状况预测、城市规划等方面

收稿日期: 2023-04-26

基金项目: 国家自然科学基金(42271429, 42171452)

第一作者: 吴杭彬(1983—), 男, 副教授, 博士生导师, 工学博士, 主要研究方向为激光扫描数据处理、轨迹数据挖掘。

E-mail: hb@tongji.edu.cn



论
文
拓
展
介
绍

具有广阔的应用前景。目前国内外研究学者已经开展了大量研究,根据数据源,相关工作可分为基于单一数据源的人群出行模式分析以及基于多源数据的出行模式分析。

基于单一数据源的大规模人群出行模式研究所使用的数据一般包括手机信令数据、出租车轨迹数据、网约车订单数据等。例如,丁亮等^[1]利用手机信令数据识别上海市内手机用户的居住地和办公地,在此基础上提取用户的通勤信息,并研究了中心城区的就业中心的能级、腹地 and 势力范围。Wu等^[2]分析了上海市长距离出租车轨迹的时空模式,探索了地铁对长距离出行出租车客流的分流效应以及程度。Tang等^[3]将城市中出租车轨迹的OD矩阵转换为网络模型,利用网络分析的方式研究城市中的人类活动性,并且使用社区发现算法划分了交通区域。Liu等^[4]使用上海市出租车轨迹研究了城市内的人类活动性,在传统Lévyflight模型的基础上引入了人口分布以体现地理空间异质性,实验结果与真实出租车轨迹模式相符,验证了模型的有效性。Liu等^[5]基于开放的GPS轨迹数据,使用马尔科夫链和聚类分析了游客微观运动时空行为模式。Liu等^[6]使用社交媒体签到数据研究城市间的人类活动性,发现距离衰减服从幂律分布的重力模型可以很好地拟合城市之间的交互强度,并用社区发现算法对城市群进行划分,得到了与省级行政边界接近的结果。

单一的数据源通常只能代表部分群体的行为,而多源数据融合能更全面地反映整体的活动模式,增强分析结果的代表性。Ma等^[7]使用北京公交和地铁数据,基于DBSCAN聚类检测出行模式,并结合K-means聚类分析出行模式规律。Liu等^[8]使用GPS数据、公交和地铁卡数据,探索了居民日常出行模式,并基于聚类分析了不同地点之间的关系以及出行与社会经济、土地类型之间的关系。Tang等^[9]使用GPS轨迹和网络数据来研究时空行为,提出了一个基于信息素的融合模型,模拟线上和线下的娱乐活动模式,基于现实世界和网络上的过去行为对

未来行为进行预测。Xia等^[10]基于地铁卡数据和GPS轨迹数据,利用行程位移、行程持续时间和行程间隔3个指标定性和定量地分析2种数据集,最后使用最大似然估计和贝叶斯信息准则,定量地探索了周末和工作日人类活动模式。岳梦雪^[11]利用深圳市出租车轨迹、公交车及地铁刷卡数据,通过基于邻接关系的时空谱聚类方法挖掘出行模式,并分别从交通方式、社会背景和出行移动方面对居民的出行模式进行分析。

综上所述,使用单一来源数据或多源数据都可以满足一定场景下的大规模人群出行模式分析,但前者存在时空分辨率低、交通方式单一、用户群体组成单一、采集成本高、样本量小等问题^[12];后者可以得到多种交通方式下的人群出行模式,但数据获取和融合难度较大。为此,本文研究采用具备多出行模式的手机导航数据,利用LGBM(Light Gradient Boosting Machine)分类方法和时空出行模式分析方法,以上海市为例,分析大规模人群的出行模式,比较不同模式下的城市人群出行特征。

1 实验区及数据

使用的上海GIS数据来源于OpenStreetMap网站。上海市位于长江入海口,常住人口2 489.43万人,行政区划面积6 340.5km²,下辖16个区(县)。

所用的手机导航数据时间范围为2017年9月16日(周六)至2017年9月19日(周二),原始数据格式如表1所示。数据采样间隔为1s,4天的总数据量约400G,共54.9亿个GPS点,日活跃用户约145至169万人,如表2所示。

表1中日期为包含年、月、日的8位字符串;时间为包含时、分、秒的6位字符串;用户ID为加密后的16位字符串,可用于对用户进行识别;经度、纬度为GCJ02坐标系下的点坐标;速度为坐标点的瞬时速度;方位角为点的前进方向,以正北为参考方向。

表1 导航轨迹原始数据

Tab. 1 Raw navigation trajectory data

日期	时间	用户ID	经度/(°)	纬度/(°)	速度/(km·h ⁻¹)	方位角/(°)
2017-09-18	06:54:45	794c6f51f4d51074	121.355 547	31.290 222	86	278
2017-09-18	06:54:45	Cd1ae0429f488249	121.403 642	31.286 996	51	103
...
2017-09-18	06:54:45	a6a91e7567acd6cf	120.910 564	31.658 986	0	0

表2 导航用户分布统计

Tab. 2 Distribution statistics of navigation users

日期	GPS点数	不同ID数
09-16(周六)	1 470 631 299	1 690 704
09-17(周日)	1 391 451 332	1 561 675
09-18(周一)	1 267 941 832	1 454 796
09-19(周二)	1 359 194 630	1 451 457
4天总数	5 489 219 093	6 158 632

2 基于手机导航数据的城市大规模人群出行分析方法

2.1 导航数据预处理

原始导航数据中存在脏数据、轨迹未分割等问题,因此需要进行预处理。具体包括:

(1)冗余数据清洗:主要包括同一用户在同一时刻记录的多个不同或相同坐标点。记录时间不在合理范围内的坐标点。

(2)轨迹分割:采用时间间隔分割法,提取单一用户的所有导航数据并按时间排序,若前后两数据点时间间隔超过时间间隔阈值,则在中断点上将原始轨迹划分为2条轨迹。手机导航轨迹采样间隔为1s,但城市环境中存在高楼、高架桥、隧道等遮挡,可能导致定位中断一段时间,因此时间间隔阈值设置为5min。

(3)异常轨迹清洗:冗余数据清洗仅针对数据点

层面的异常,但有些异常或无效点无法根据单个点识别,因此在完成轨迹分割后再进行轨迹清洗。包括两方面:①误触数据清洗,对于持续时间不足10s、位移不足10m的轨迹,认为是用户误触产生,删除整条误触轨迹。②定位漂移数据清洗,部分轨迹中出现时间相邻的点在2个相距较远的位置来回跳动的情况,影响了导航轨迹长度的计算,这种情况认为出现了定位漂移,需删除整条轨迹。

2.2 基于LGBM的导航数据模式分类

经过预处理的导航数据无交通方式标注,还无法直接用于人群的出行模式分析,需要进行交通方式分类。本文采用轻量级梯度提升机模型(LightGBM)识别导航轨迹数据的步行、非机动车和机动车3种交通方式^[13]。LightGBM是由微软亚洲研究院开发的一种基于决策树算法的梯度提升框架,支持并行学习,采用直方图算法,节点分类不需要遍历数据,提高了速度;采用分裂增益最大的叶节点的方式(leaf-wise)而非基于层的生长方式(level-wise),减少了不必要的分裂,降低了计算量,同时使用最大深度限制防止过拟合^[14]。

交通方式识别模型中使用60个特征:5%~95%每5分位数的速度特征(19个)、5%~95%每5分位数的加速度特征(19个)、5%~95%每5分位数的角度转变量(19个)、停留率、速度转变率、角度转变率^[15],各特征计算式如表3所示。

表3 交通方式特征计算公式

Tab. 3 Calculation formula for characteristics of transportation mode

特征	公式及描述
速度特征	$v_i = \frac{d}{t}$, 其中 d 为点 p_i 和 p_{i+1} 之间的大圆距离, t 为点 p_i 和 p_{i+1} 之间的时间间隔。
加速度特征	$a_i = \frac{v_{i+1} - v_i}{t}$, 其中 v_{i+1} 和 v_i 分别为点 p_{i+1} 和 p_i 的速率, t 为时间间隔。
角度转变量特征	$h_i = H_{i+1} - H_i $, 其中 H_i 为点 p_i 的方位角。
停留率	$R_s = \frac{ P_s }{D}$, 其中 $ P_s $ 为轨迹中速度小于某一速度阈值的点数量,阈值取 $3.4 \text{ m}\cdot\text{s}^{-1}$ ^[16] , D 为轨迹长度。
速度转变率	$R_{vc} = \frac{ P_v }{D}$, 其中 $ P_v $ 为轨迹中速率变化(P_i)大于某一阈值的点数量, $P_i = \frac{ v_{i+1} - v_i }{v_i}$, 阈值取 $0.26 \text{ m}\cdot\text{s}^{-2}$ ^[16] 。
转角转变率	$R_{hc} = \frac{ P_c }{D}$, 其中 $ P_c $ 为转角转变量大于某一角度阈值的点数量, R_{hc} 阈值取 19° ^[16] 。

GeoLife是微软亚洲研究院提供的开源GNSS(Global Navigation Satellite System)轨迹数据集,收录了2007年4月至2012年8月期间182位志愿者的移动轨迹,其中65人标注了各段轨迹的交通方式,包括步行、自行车、公交车、汽车、火车、地铁、飞机等^[16]。使用GeoLife数据集训练和评价LightGBM

交通方式识别模型,并对训练好的LightGBM模型进行5折交叉验证^[17]来避免模型过拟合。

2.3 人群出行模式

2.3.1 出行时间模式

利用不同交通模式下的导航时间序列来分析导航人群的出行时间模式。以30min为间隔,分步行、

非机动车和机动车3种交通方式,统计每个时间段内总的导航轨迹数量,得到的时间序列代表导航出行时间模式。通过分析时间序列的峰值等特征时间,按照不同的交通方式分析周末和工作日的人群出行时间模式。

2.3.2 出行空间模式

利用不同交通模式下导航终点在城市空间中的热点来分析导航人群的出行空间模式,使用核密度分析法来计算热点。通过核密度分析法分析周末和工作日不同交通方式导航轨迹的空间模式,并结合高架路中心线、地铁站出入口等辅助数据对热点地区进行识别。空间模式分析流程如图1所示。

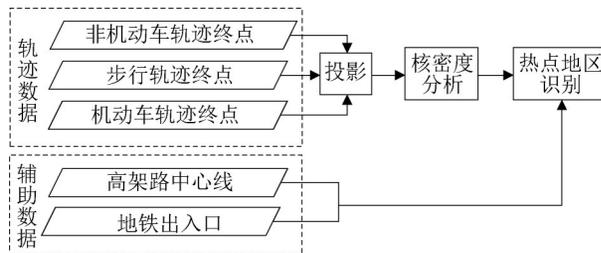


图1 导航轨迹空间模式分析流程

Fig. 1 Flowchart for analysis of navigation trajectory spatial pattern

2.3.3 出行距离模式

人群出行距离的分布反映了人群的活动性,截断幂律分布是一种概率分布模型,可以模拟导航轨迹的出行距离模式。

幂律分布属于重尾分布(heavy-tail),其概率密度函数尾部的衰减比指数函数慢,在双对数坐标系下呈现为一条直线,被广泛用于对人群活动性建模。对于人类活动性而言,虽然大部分人的活动集中在小范围内,但仍有一部分人会进行长距离的出行,例如去离家较远的机场、火车站、旅游景点等。对于城市内部的活动而言,人群的移动受到出行时间、城市边界的限制,移动的距离不是无限的,而是在长距离区间有更快的衰减速度^[18],截断幂律分布很好地体现了这一特性。以往的研究使用指数截断幂律分布模型在手机信令数据^[19]、出租车轨迹数据上^[20]拟合了这一现象,但区分交通方式的研究较少见。本文使用截断幂律分布按照不同的交通方式对手机导航轨迹长度 d 进行拟合,如式(1):

$$p(d) = \alpha d^{-\beta} e^{-\lambda d} \quad (1)$$

式中: p 为轨迹数据长度服从的概率分布; d 为轨迹数据长度; α 为截断幂律分布的系数; β 为幂律分布的指数(形状参数); λ 为指数衰减因子。

导航距离模式分析的流程如图2所示。以100m为间隔,分别统计机动车、步行、非机动车轨迹长度在各区间的比例,作为出行距离等于区间中点时的概率。对长度超过1km的导航轨迹,分别用截断幂律分布的概率密度函数进行拟合。

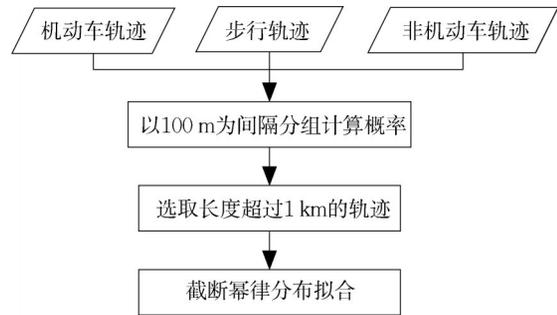


图2 导航距离模式分析流程

Fig. 2 Flowchart for analysis of navigation distance pattern

3 实验结果与讨论

3.1 出行模式分类结果

使用GeoLife数据集训练LightGBM交通方式识别模型,并进行5折交叉验证,模型在测试集上的平均准确率为 0.944 ± 0.001 。用调参后的模型对测试集进行预测,得到的混淆矩阵如表4所示。

表4 测试集预测结果

Tab. 4 Prediction results of test set

交通方式	预测值			精确率	召回率	F1分数
	机动车	步行	非机动车			
机动车	586	29	10	0.954 4	0.937 6	0.945 9
步行	16	574	17	0.932 0	0.958 1	0.944 9
非机动车	12	26	294	0.915 9	0.885 5	0.900 5

将训练好的模型应用于实验导航数据,并将识别出的交通方式附加到轨迹数据库中的相应轨迹上。以导航终止时间为标准进行统计,4天的轨迹总体情况如表5所示,总计轨迹数837万条,其中机动车轨迹数417万条,步行轨迹数374万条,非机动车轨迹数46万条。因为部分用户在一天中多次使用导航,从而导致轨迹总数显著大于表2中的不同ID数。

3.2 上海市导航人群出行模式

3.2.1 出行时间模式

分机动车、步行和非机动车3种交通方式分析导航轨迹的时间模式,结果如图3所示,图中虚线代表的峰值等特征时间如表6所示,以11:00代表区间

表5 轨迹数据分类情况统计

Tab. 5 Statistics of classification of trajectory data

日期	机动车轨迹数/条	步行轨迹数/条	非机动车轨迹数/条	总计轨迹数/条
09-16(周六)	1 110 307	1 081 356	125 658	2 192 921
09-17(周日)	1 064 156	923 640	114 052	2 101 848
09-18(周一)	994 633	863 981	112 985	1 971 599
09-19(周二)	997 554	872 186	113 450	1 983 190
总计	4 166 650	3 741 163	466 145	8 373 958

11:00—11:30,以此类推。

3种导航方式的导航时间模式均呈现明显的早高峰—午高峰—晚高峰周期性,峰值时间基本一致,但峰值的强弱存在明显差异:机动车工作日午高峰较强,而步行、非机动车早晚高峰较强。周末(9月16日、9月17日)与工作日(9月18日、9月19日)的峰值持续时间和峰值时间也有所不同。周末高峰持续时间为11:00到17:30,工作日高峰持续时间为8:30到18:00,周末高峰持续时间比工作日减少3h。周末第一峰值时间比工作日推迟约2.5h,可以看出周末导航用户的生活节奏较为悠闲,上午出行高峰在11:00左右,而工作日则表现出通勤的特征,第一峰值时间在8:30左右。第二峰值在周末和工作日基本相同,分别出现在13:30到14:00,对应午休结束后的出行量回升。第3峰值在周末和工作日也基本相同,分别出现在17:30到18:00,对应外出就餐或通勤。转折点出现在20:00左右,导航数量下降速度放缓,这一点在周末尤其明显,可能与外出休闲娱乐有关。

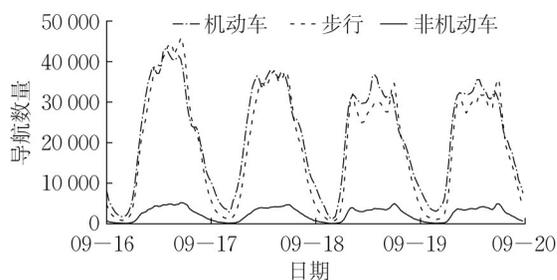


图3 不同交通方式导航轨迹时间模式

Fig. 3 Navigation trajectory time pattern of different traffic modes

表6 导航特征时刻

Tab. 6 Feature time of navigation

日期	第一峰值时刻	第二峰值时刻	第三峰值时刻	晚间转折点
周末	11:00	14:00	17:30	20:00
工作日	8:30	13:30	18:00	20:00

3.2.2 出行空间分布模式

分步行、非机动车和机动车3种交通方式分析

导航轨迹的出行空间分布模式。

3.2.2.1 机动车轨迹空间模式

周末和工作日机动车轨迹终点的核密度图如图4所示,图中清晰呈现了路网的形状,其热点地区主要分为3种情况:

(1)高架道路。图4的右下方小图标示了上海市区4条主要高架路(内环高架、南北高架、延安高架、沪闵高架),与核密度图中密度在2 000点·km²以上的高密度带状区域重合。这一现象表明高架区域车流量大,路网较复杂,而引导标志还不够充足,有较多的用户进入高架区域后需要借助导航驾驶,或通过了高架区域的复杂路段后随即关闭导航。除了高架区域以外,一些地面主要道路的交汇处也形成了热点地区,如图中A5对应的五角场区域是5条道路的交汇点,A6、A7各对应世纪大道—张杨路—东方路、衡山路—虹桥路—肇嘉浜路3条道路的交叉点。这些路口附近有众多商业、商务场所,车流量大、空间构造复杂,在此开启导航和以此为目的地进行导航的轨迹多。

(2)交通枢纽。图4中B1、B2、B3分别对应虹桥枢纽、浦东国际机场、上海火车站,其中虹桥枢纽和浦东机场都在外环以外,离中心城区较远,虽设有地铁站,但驾车或乘出租车前往仍是一种主要的交通方式,因此轨迹终点密度高。

(3)隧道的出入口。图4中C1、C2、C3分别对应外滩隧道(3.3km,位于外滩地下)、延安东路隧道(2.3km,连通浦东和浦西)、上海长江隧道(9.0km,连通浦东新区和崇明岛)的一个出入口。这些隧道长度较长且车流量大,穿越隧道需要的时间超过轨迹分割的时间阈值,GNSS定位在车辆进入隧道后中断,导致驶入隧道前的最后一个点被错误地识别为轨迹终点,并在隧道出入口处形成伪热点区域。

相比周末热点,工作日热点地区不包括A5五角场,这是因为工作日前往五角场商圈娱乐的人相对减少。

3.2.2.2 步行轨迹空间模式

周末和工作日步行轨迹终点的空间分布如图5所示。除五角场—江湾体育场站以外,密度超过10 000点·km²的热点地区都分布在内环以内,比机动车轨迹的热点地区更加集中。外环以内的热点地区、外环以外密度相对较高的区域与地铁站的分布吻合,这些地铁站分别对应交通枢纽(虹桥枢纽、上海火车站、上海南站)、旅游景点(七宝站)和热门商圈(南京东路、静安寺、徐家汇等),反映了大量用户

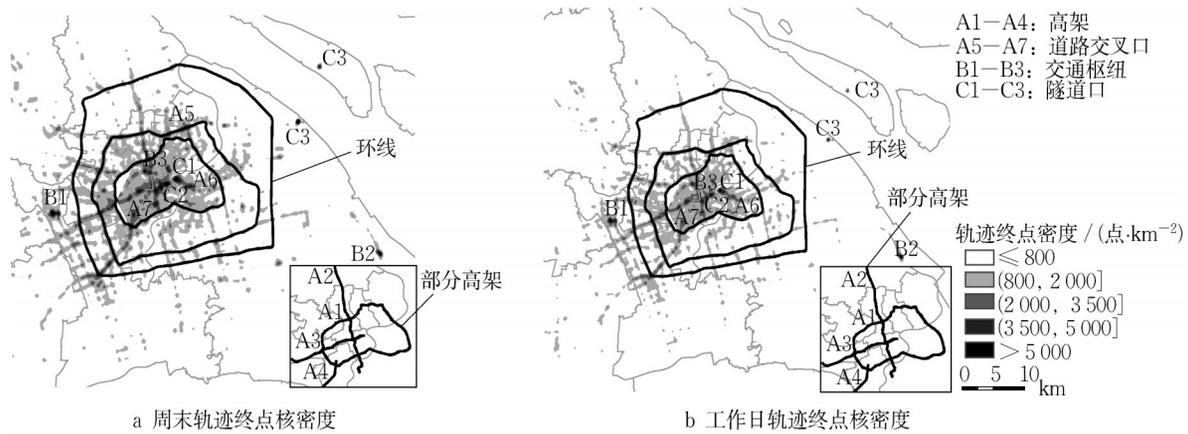


图4 机动车轨迹终点核密度

Fig. 4 Endpoint nuclear density map of motor vehicle trajectory

以这些地铁站为目的地进行导航,或在出了地铁站后使用导航寻找附近场所的行为。

这一现象表明地铁是上海市交通系统的重要组成部分,大量用户通过地铁出行,地铁站周围形成了人流密集区域,在如图5所示的15个热点地铁站出入口内部的醒目位置需要增加指示牌和地图,标注附近主要设施(例如商场、景点)的位置以及应从哪个出站口出站;在上述地铁站出入口附近的地面道路上、主

要设施内部也需要增加醒目的指示牌和地图,便于需要乘地铁的人群找到地铁站出入口的位置。

工作日跨省外出、景区游玩和到市区娱乐的人减少,因此工作日虹桥枢纽(图5中数字2)、上海南站(数字3)、莲花路站(数字13,莲花路站是闵行区连接市区的第一枢纽)和七宝站(数字15,七宝站是商圈站)等区域的热度显著低于周末。

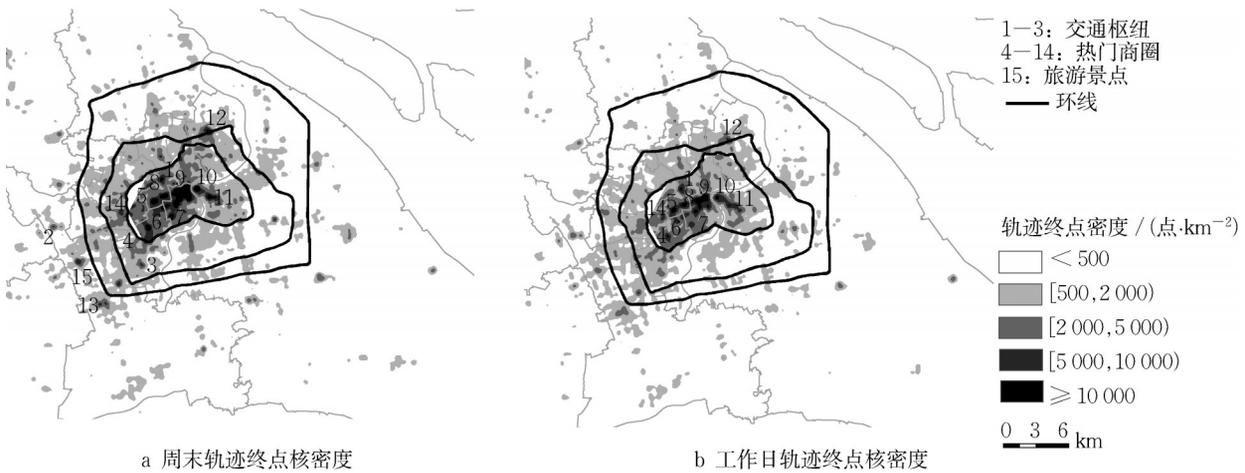


图5 步行轨迹终点核密度

Fig. 5 Endpoint kernel density map of walking trajectory

3.2.2.3 非机动车轨迹空间模式

周末和工作日非机动车轨迹终点的空间分布与步行轨迹类似,集中在内环以内的地铁站附近和商务楼宇密集的地区,如图6所示。这一结果表明,图中14个热点地铁站是很多非机动车用户的目的地,因此地铁站出入口附近需要增加非机动车停放位;同时由上面对非机动车轨迹时间模式的分析可知,工作日的高峰期出现在早晚通勤时段,因此上述地

铁站在通勤时段需要加强共享单车的调度管理。

相比周末,工作日热点地区不包括五角场站(图6中数字10)—江湾体育场站和金沙江路站(数字13,金沙江路站是商圈站点),这是因为工作日前往商圈的人相对减少。

3.2.3 出行距离模式

使用截断幂律分布拟合3种交通方式的导航轨迹,轨迹长度的概率密度分布散点图和拟合结果如

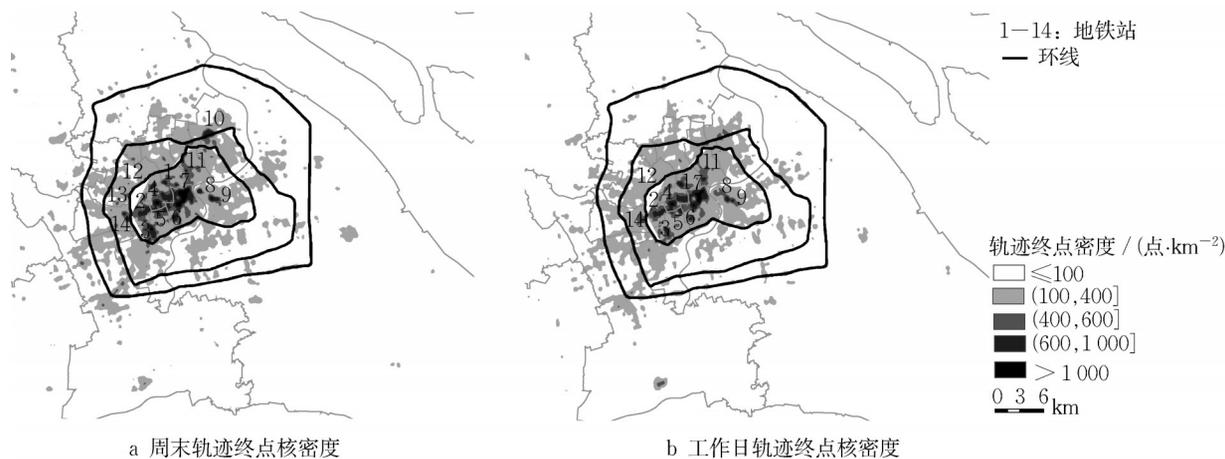


图 6 非机动车轨迹终点核密度

Fig. 6 Endpoint kernel density map of non-motor vehicle trajectory

图7、表7所示, R^2 均超过0.998, 拟合效果较好。机动车轨迹超过30km和步行轨迹超过10km的部分概率密度分布的衰减速度变大, 表现出指数衰减的特征。

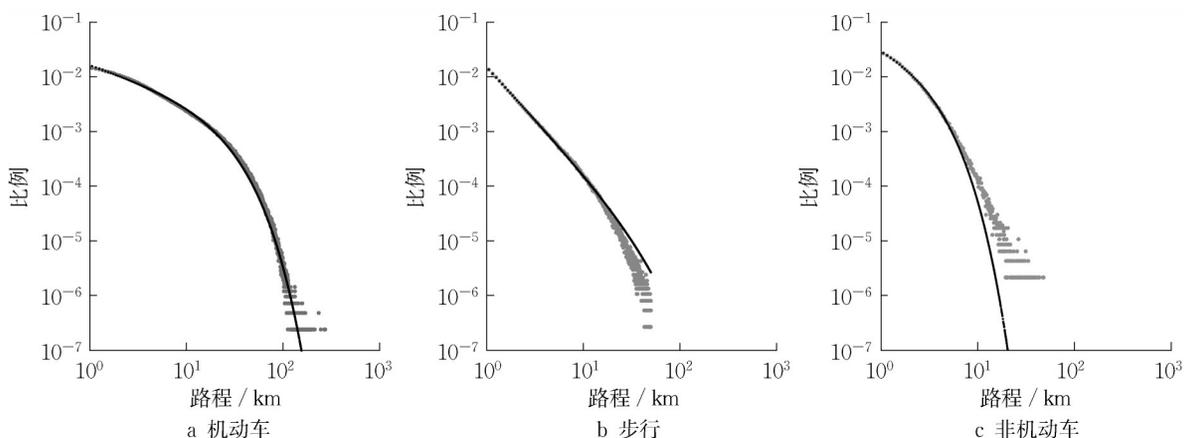


图 7 轨迹长度的概率密度分布散点图与拟合效果

Fig. 7 Scatter plots of probability density distribution of trajectory lengths with fitting effects

表 7 各类轨迹长度分布的截断幂律分布拟合参数

Tab. 7 Truncated power-law distribution fitting parameters for each type of trajectory length distribution

交通方式	α	β	λ	R^2
机动车	0.017	0.580	0.059	0.998
步行	0.015	1.894	0.025	0.999
非机动车	0.051	0.548	0.564	0.999

4 结论

首先基于LightGBM模型构建了轨迹数据的交通方式分类模型, 并在GeoLife测试集上取得了 0.944 ± 0.001 的准确率。随后, 该模型被应用到上海市数百万人口的手机导航轨迹数据上, 实现了机

动车、非机动车和步行的导航轨迹分类。最后, 进一步分析了不同交通方式的导航轨迹数据在时间、空间和距离上的分布, 并得到了以下结论:

(1) 机动车、步行、非机动车3种交通方式的时间模式均表现出早高峰—午高峰—晚高峰周期性, 出行方式主要以机动车和步行为主, 并且周末出行高峰持续时间比工作日少3h, 且周末的早高峰较工作日要晚2.5h, 体现出周末导航出行高峰与工作日相比存在一定的滞后性。

(2) 在出行空间模式方面, 相比工作日, 周末热点地区核密度更高、热点区域更多。机动车轨迹热点的分布与步行、非机动车轨迹有较大差异, 机动车轨迹受路网的影响更大, 热点地区出现在主要道路、

交通枢纽、隧道出口;步行轨迹热点地区与地铁站吻合;非机动车热点地区与步行的分布类似。导航轨迹终点的空间热点提取结果可以为城市规划提供理论依据,热点高架区域需要增加引导标志和路网示意图,热点地铁站出入口内外需要在醒目处增加周边主要设施位置和地铁出入口位置的指示牌,出入口附近的地面上需要增加非机动车停放位并且在早晚通勤时段加强共享单车的调度和管理。

(3)3种交通方式的轨迹都可以用截断幂律分布较好地拟合,导航出行以中短距离为主,概率密度函数按照近似于幂函数的速度衰减,同时有一部分用户会进行长距离的导航,但在长距离出行区间概率密度函数衰减更快,近似于指数函数。

作者贡献声明:

吴杭彬:提出研究思路、设计研究方案、论文结构把控及修改。

陈茜茜:实验分析、论文撰写。

靳慧玲:数据预处理、协助实验。

傅琛:数据预处理、协助实验。

黄炜:实验与论文撰写指导。

刘春:实验与论文撰写指导。

参考文献:

- [1] 丁亮,钮心毅,宋小冬.上海中心城就业中心体系测度——基于手机信令数据的研究[J].地理学报,2016,71(3):484.
DING Liang, NIU Xinyi, SONG Xiaodong. Measuring the employment center system in Shanghai central city: A study using mobile phone signaling data [J]. Acta Geographica Sinica, 2016, 71(3): 484.
- [2] WU H, FAN H, WU S. Exploring spatiotemporal patterns of long-distance taxi rides in Shanghai [J]. ISPRS International Journal of Geo-Information, 2017, 6(11): 339.
- [3] TANG J, ZHANG S, ZHANG W, *et al.* Statistical properties of urban mobility from location-based travel networks [J]. Physica A: Statistical Mechanics and its Applications, 2016, 461: 694.
- [4] LIU Y, KANG C, GAO S, *et al.* Understanding intra-urban trip patterns from taxi trajectory data [J]. Journal of Geographical Systems, 2012, 14(4): 463.
- [5] LIU W, WANG B, YANG Y, *et al.* Cluster analysis of microscopic spatio-temporal patterns of tourists' movement behaviors in mountainous scenic areas using open GPS-trajectory data [J]. Tourism Management, 2022, 93: 104614.
- [6] LIU Y, SUI Z, KANG C, *et al.* Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data [J]. PLoS ONE, 2014, 9(1): 1.
- [7] MA X, WU Y J, WANG Y, *et al.* Mining smart card data for transit riders' travel patterns [J]. Transportation Research Part C: Emerging Technologies, 2013, 36: 1.
- [8] LIU L, BIDERMAN A, RATTI C. Urban mobility landscape: real time monitoring of urban mobility patterns [C]//Proceedings of the 11th International Conference on Computers in Urban Planning and Urban Management. Hong Kong: Citeseer, 2009: 1-16.
- [9] TANG L, ZOU Q, ZHANG X, *et al.* Spatio-temporal behavior analysis and pheromone-based fusion model for big trace data [J]. ISPRS International Journal of Geo-Information, 2017, 6(5): 151.
- [10] XIA F, WANG J, KONG X, *et al.* Exploring human mobility patterns in urban scenarios: A trajectory data perspective [J]. IEEE Communications Magazine, 2018, 56(3): 142.
- [11] 岳梦雪.多源轨迹数据时空聚类及居民出行模式分析[D].湖北:武汉大学,2017.
YUE Mengxue. Spatio-temporal clustering of multi-source trajectory data and analysis of residents' travel patterns [D]. Hubei: Wuhan University, 2017.
- [12] 陆锋,刘康,陈洁.大数据时代的人类移动性研究[J].地球信息科学学报,2014,16(5):665.
LU Feng, LIU Kang, CHEN Jie. Research on human mobility in big data era [J]. Journal of Geo-information Science, 2014, 16(5): 665.
- [13] WU H, JIN H, XU Z, *et al.* Travel mode classification based on GNSS trajectories and open geospatial data [J]. Transactions in GIS, 2022, 26(6): 2598.
- [14] KE G, MENG Q, FINLEY T, *et al.* Lightgbm: A highly efficient gradient boosting decision tree [J]. Advances in Neural Information Processing Systems, 2017, 30: 3146.
- [15] 王璞,刘洋,黄智仁.一种轻量级梯度提升机的交通模式识别[J].哈尔滨工业大学学报,2019,51(9):96.
WANG Pu, LIU Yang, HUANG Zhiren. Traffic pattern recognition of a lightweight gradient elevator [J]. Journal Harbin Institute Technology, 2019, 51(9): 96.
- [16] ZHENG Y, LI Q, CHEN Y, *et al.* Understanding mobility based on GPS data [C]//In Proceedings of ACM conference on Ubiquitous Computing (UbiComp 2008). Seoul: ACM Press, 2008: 312-321.
- [17] 范永东.模型选择中的交叉验证方法综述[D].太原:山西大学,2013.
FAN Yongdong. Overview of cross validation methods in model selection [D]. Taiyuan: Shanxi University, 2013.
- [18] ZHENG Z, RASOULI S, TIMMERMANS H. Two regime pattern in human mobility: Evidence from GPS taxi trajectory data [J]. Geographical Analysis, 2016, 48(2): 157.
- [19] GONZALEZ M C, HIDALGO C A, BARABASI A L. Understanding individual human mobility patterns [J]. Nature, 2008, 453(7196): 779.
- [20] TANG J, LIU F, WANG Y, *et al.* Uncovering urban human mobility from large scale taxi GPS data [J]. Physica A: Statistical Mechanics and its Applications, 2015, 438: 140.