

文章编号:1005-9679(2018)04-0089-07

基于 GPS 数据的城市居民多日出行活跃度分析

张 蕾 隽志才 高晶鑫

(上海交通大学 安泰经济与管理学院, 上海 200030)

摘 要: 通过提出居民出行活跃度的概念和测度方法,分析居民多日出行行为的活跃程度,丰富多日出行行为分析的理论体系,为城市空间优化和交通政策引导提供参考。居民出行的活跃程度是出行频率、出行距离、出行时长、出行方式和出行目的等出行要素的综合体现。基于上海市 2014—2015 年智能手机 GPS 辅助的居民一周活动-出行调查,通过分析居民的出行频率、出行距离、出行时长、出行方式和出行目的等多个出行活跃度指标,运用 K-均值聚类的方法定义四种出行活跃度类型——多方式、多目的、长距离三种活跃出行者类型和较不活跃出行者类型。数据显示,不同活跃度类型人群在平日、工作日及周末拥有的出行特征具有明显差异,同时通过多项 Logistic 回归模型分析得出居民的个人及家庭社会经济属性对活跃度类型的显著影响。出行活跃度概念和测度方法的提出有助于分析城市居民出行行为的复杂性,便于对居民出行的活跃程度形成综合全面的考量,有助于分析居民出行需求对城市空间与交通政策的反馈作用。

关键词: 交通运输系统工程;多日出行活跃度;K-均值聚类;多项 Logistic 回归模型;手机 GPS 数据;出行调查

中图分类号: U-9 文献标志码: A

A Multi-day Activity Level Study of Citizens' Travel Behavior Based on GPS Data

ZHANG Lei JUAN Zhicai GAO Jingxin

(Antai College of Economics & Management, Shanghai Jiao Tong University, Shanghai 200030, China)

Abstract: This paper comes with the concept and measure method of multi-day activity level of citizens' travel behavior, in order to enrich the theory system of multi-day travel behavior and provide reference to the layout of urban space and travel demand management. Individual activity level of travel behavior reflects the combination of several important indicators, such as travel frequency, distance, time, mode and purpose. Based on a GPS-facilitated activity-travel survey dataset collected in Shanghai in 2014 and 2015, this paper analyzes the travel indicators mentioned above and defines four groups of different travel behavior activity levels by means of k-means clustering, which are "multi-mode", "multi-purpose", "long-distance" and "inactivity". The study finds that there are intergroup differences in activity-travel behavior between weekdays and weekends among the four groups. Furthermore, this paper analyzes the impact of activity level of citizens' travel behavior by examining the relationship between travel activity level and socio-demographic factors using a multinomial logistic model. Taking most important indicators of travel behavior into consideration, the concept of multi-day activity level of citizens' travel behavior provides an effective

收稿日期:2017-06-05

基金项目:国家自然科学基金项目“基于智能手机的个体活动链信息获取及挖掘方法”(51478266)。

作者简介:张蕾(1991—),女,湖北襄阳人,上海交通大学硕士研究生,研究方向为出行行为分析,E-mail: zyzfy003@sjtu.edu.cn;隽志才(1954—),男,吉林长春人,上海交通大学教授,工学博士,博士生导师。

tive and reasonable way to understand complicated activity- travel behavior of urban residents and the travel demand's feedback effect to the urban space and transportation policy.

Key words: transportation systems engineering; multi-day activity level of travel behavior; k-means clustering; multinomial logistic regression; GPS data of smartphones; travel survey

1 研究的数据来源

本调查以智能手机为载体,是基于 GPS 轨迹数据的活动-出行调查,研究区域选在上海市。调查过程中每位志愿者需要提供自己的社会经济属性,然后在自己的手机上安装相关定位-上传软件,连续上传一周的 GPS 定位数据。研究人员通过数据挖掘的可视化处理后,通过带提示的电话回访调查确认志愿者的行程和活动-出行信息。调查流程如图 1 所示,带提示的电话回访调查界面如图 2 所示。

本研究数据来源于 2014 年 11 月至 2015 年 12 月上海交通大学交通运输研究所实施的上海市居民

活动-出行调查,共包含 251 个志愿者,覆盖了 7 个中心城区与 9 个郊区县,包括每位接受调查居民的社会经济属性、一周的活动日志及 GPS 轨迹数据。本研究选取了年龄在 18~60 岁、有正式工作且有效出行记录在三天以上的 160 个志愿者作为有效样本,共包含 881 天的 2 202 条出行记录,平均每位志愿者上传了 5.5 天的出行轨迹。有效样本的基本情况如表 1 所示。有效样本中女性比例高于男性,年龄最小为 19 岁,最大为 49 岁,平均年龄为 30.28 岁,61.9%的居民拥有本科及以上学历,19.4%居民的个人月收入在一万元以上,平均家庭规模为 3.12 人,平均家庭拥有小汽车数量为 0.46 辆。

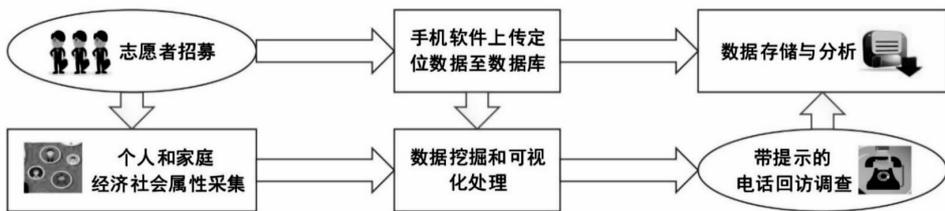


图 1 基于智能手机的调查流程

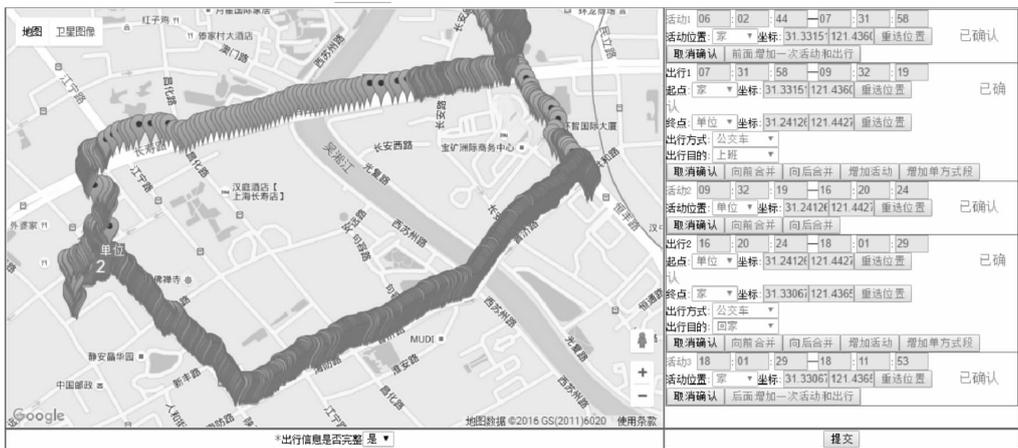


图 2 带提示的电话回访调查界面示意图

2 居民出行活跃度的概念与测度

居民的出行特征包含出行频率、出行方式、出行目的、出行距离和时长等多个方面,现有文献也从不同的维度解读了 GPS 轨迹数据中反映出的这些居民日常出行的行为特征。然而,参照组织行为学领域中企业绩效等相关概念,居民出行是否活跃也应该是对其出行特征各个方面进行的综合全

面的考察。一般认为,出行频率较高的居民出行比较活跃,但就拥有相似出行频率的居民而言,出行距离的远近、出行时间的长短、出行方式和出行目的是单一还是多样,这些因素同样影响着我们对于居民出行是否活跃的判断。因此,笔者综合了以上五个广受认可的居民出行特征,提出居民出行活跃度的概念。本文所指的居民出行活跃度泛指居民日常出行的活跃程度,是对居民的出行

频率、距离、时长、出行目的和出行方式进行的综合全面的考量。每位居民的出行活跃度可以看作一个五维向量,五个指标的大小、分布共同决定了居民出行活跃度类型。

表 1 被调查者社会经济属性

变量		样本数	频率/%
性别	男性	62	38.8
	女性	98	61.2
年龄	平均值	30.28	
教育程度	高中及以下	9	5.6
	大专	52	32.5
	本科	82	51.3
	硕士及以上	17	10.6
婚姻状况	未婚	62	38.8
	已婚	98	61.2
个人月收入	五千及以下	64	40.0
	五千到一万	65	40.6
	一万及以上	31	19.4
职业	政府机关事业单位	16	10.0
	教育科研院所	3	1.9
	生产制造企业	37	23.1
	金融业	4	2.5
	商业服务业	33	20.6
	大学生	9	5.6
其他	58	36.3	
家庭总人数	平均值	3.12	
私家车数量	平均值	0.46	

居民出行活跃度的测度是指利用量化的指标区分具有相似出行活跃类型的人群,方法为 K-均值聚类。本文使用的指标包含以下五个:出行频率指数等于该志愿者平均每天的出行次数与总样本平均每天的出行次数之比,出行距离指数等于该志愿者平均每天的出行距离与总样本平均每天的出行距离之比,出行时长指数等于该志愿者平均每天的出行时长与总样本平均每天的出行时长之比,非强制性指数指的是该志愿者非强制性出行(除去目的为上班、上学、工作外出、回家之外的出行)的出行次数与总出行次数之比,复杂性指数则为该志愿者平均每

次出行所包含的片段数量现总样本平均每次出行所包含的片段数量(一次出行中每换一次交通方式视为进入下一个出行片段)之比。

本文通过调查得到出行数据,对 160 名参与调查的志愿者的五个居民出行活跃度指标分别进行运算,再运用 K-均值聚类的方法对志愿者进行聚类。通过设定不同的类别、多次重复聚类过程并对聚类结果进行比较发现,将志愿者的出行活跃度分为四类所得到的结果具有较强的代表性,每种类型的样本数和聚类中心指标参数见表 2。多方式活跃出行者的出行频率、距离和非强制性出行比例均远低于平均值,出行时长指数和复杂性指数远大于 1;出行次数少但换乘多、时间长的出行者,占总体样本的 7.5%;多目的活跃出行者的出行频率和非强制性指数较高,代表此类人群休闲娱乐类出行较多,出行意愿较强,这类人群占总体样本的 39.4%;长距离活跃出行者的出行距离和出行时长远高于平均值,非强制性指数远低于平均值,代表了职住距离较远、出行以通勤为主的出行者,占总体样本的 4.4%;较不活跃出行者的五个活跃度指标都小于 1,即低于总体平均值,是总体中出行较不活跃的群体,占总体样本的 48.7%。四种居民的出行活跃度类型特征见图 3。

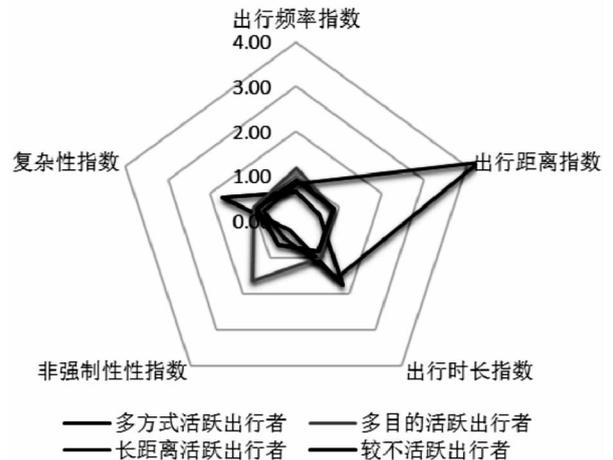


图 3 居民出行活跃度类型特征

表 2 居民多日出行活跃度指标

居民出行活跃度类型	出行频率指数	出行距离指数	出行时长指数	非强制性指数	复杂性指数	样本数
多方式活跃出行者	0.68	0.52	1.79	0.25	1.74	12(7.5%)
多目的活跃出行者	1.20	0.84	0.99	1.68	0.99	63(39.4%)
长距离活跃出行者	0.84	4.16	1.56	0.47	0.77	7(4.4%)
较不活跃出行者	0.94	0.86	0.86	0.66	0.93	78(48.7%)

3 基于居民出行活跃度的描述性统计

3.1 出行频率差异显著,一周内日间分布不平均

本次调查样本总体的平均每日出行频率为 3.23 次,其中周末出行频率高于工作日出行频率,而不同

出行活跃度类型的出行者在出行频率方面具有较为明显的差异,如图 4 所示。平均来看,多目的活跃出行者的出行频率最高,达到 3.84 次/日,较不活跃出行者、长距离与多方式活跃出行者的出行频率依次递减。同时,多方式与多目的活跃出行者在周末的

出行频率远高于工作日,而长距离活跃出行者与较不活跃的出行者出行则较为平均,工作日与周末没有明显差异。

3.2 交通方式各有偏好,长距离类型特征突出

交通方式方面,四种活跃度类型的出行者差异较为明显。多方式活跃出行者以地铁为主要交通方式,

其次为公交车和步行,这与城市公共交通的换乘机制一脉相承。长距离活跃出行者对于交通方式的选择非常集中。小汽车占比为 62%,远远超过其他方式,体现了这类人群职住距离远、通勤出行多的特点。多目的活跃出行者与较不活跃出行者的交通方式分布相对平均,以公共交通为主,辅以小汽车和步行等。

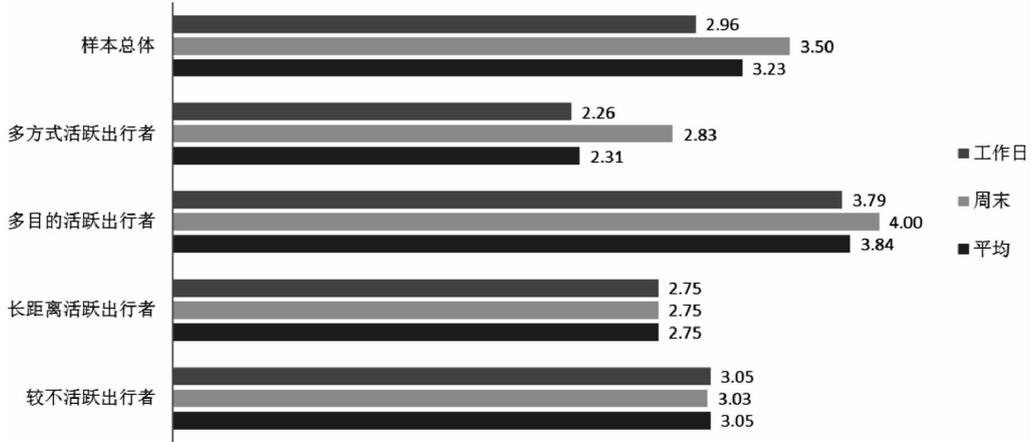


图 4 四种出行活跃度居民出行频率分布

具体到工作日与周末的比较方面,多方式活跃出行者在工作日首选地铁出行,周末则更多使用公交车和步行;长距离活跃出行者依旧选择小汽车为主要的出行方式,周末与工作日相比减少了地铁出行而增加了步行;多目的活跃出行者在周末出行时

除了公交车之外,更多地选择了小汽车和自行车出行,分别对应着他们的家庭出行与休闲出行;较不活跃的出行者也在周末增加了小汽车的出行比例,结合志愿者访谈,这种情况多由于周末家庭共同出行导致。四种出行活路居民交通方式分布见图 5。

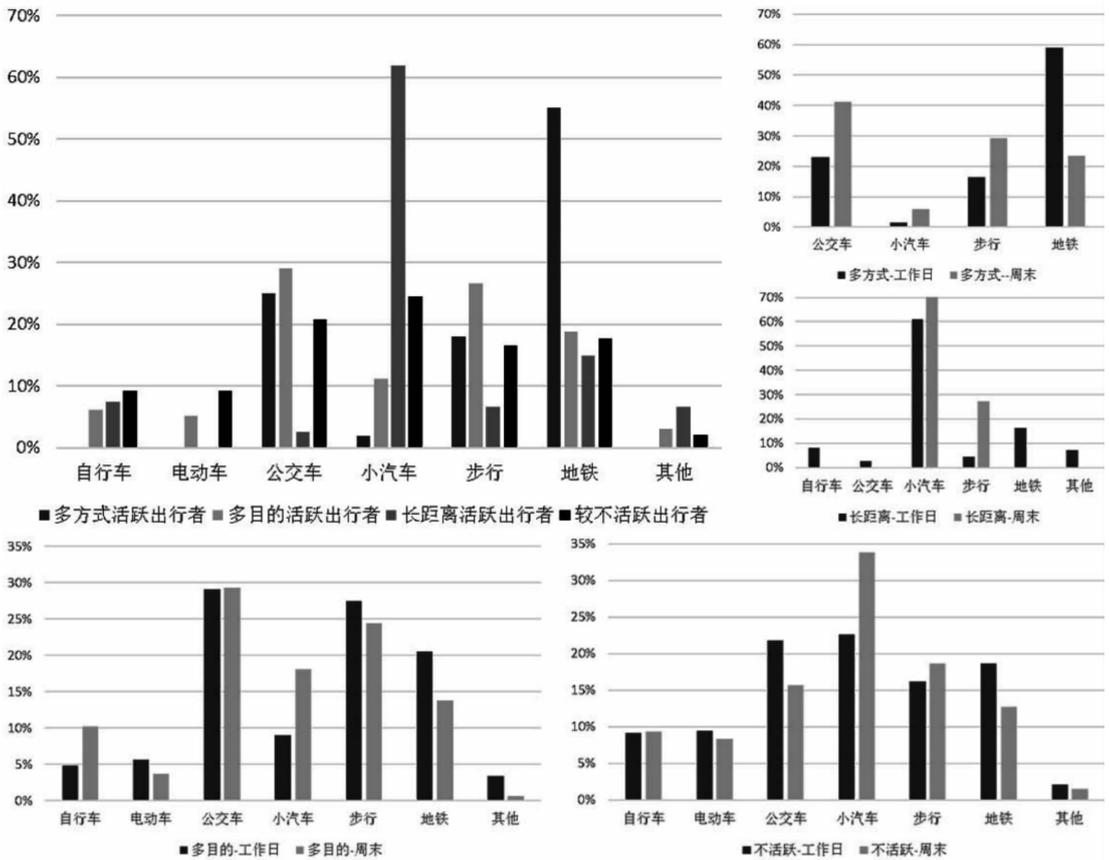


图 5 四种出行活跃度居民交通方式分布

3.3 出行目的分布不均,“多方式”类型通勤比例高

不同出行活跃度类型的出行者在出行目的方面同样具有明显差异。多方式活跃出行者的通勤出行比例超过 80%,可见其多样化的出行方式主要是由于上下班途中换乘导致;多目的活跃出行者顾名思义,出行目的分布较为均匀,购物、外出就餐、休闲娱乐等出行远超其他类型;长距离活跃出行者与较不

活跃出行者与工作相关的出行比例都略超过 40%,然而前者工作外出的比例仅为后者一半,表明长距离活跃出行者的工作地点较为稳定。此外,对四种类型工作日与周末出行目的对比分析发现其变化趋势较为一致,均减少了工作相关出行,增加了购物、外出就餐和休闲娱乐出行的比例,与现实生活情况较为一致。四种出行活跃度居民出行目的分布见图 6。

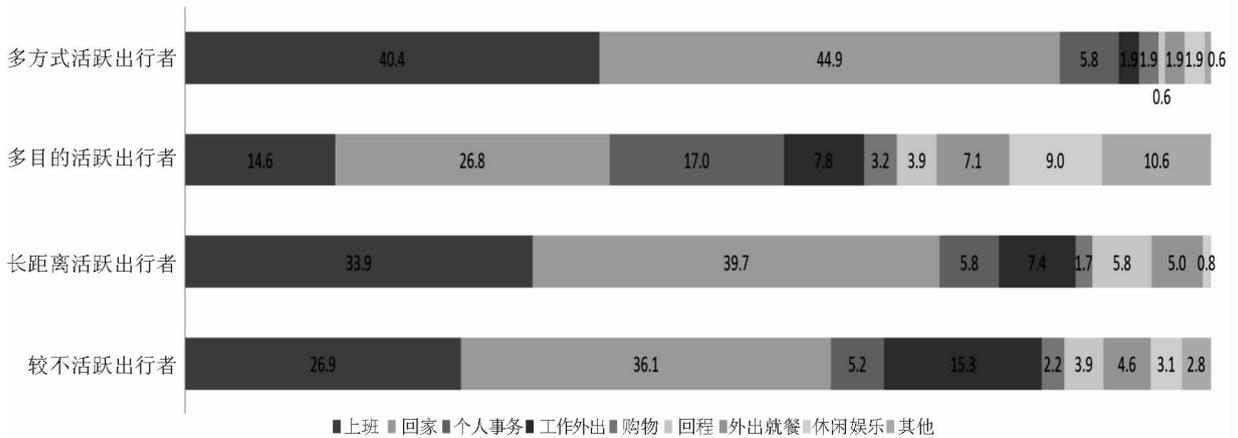


图 6 四种出行活跃度居民出行目的分布

3.4 出发时间分布规律不同,“多方式”与“不活跃”构成早晚高峰期

上海市第五次综合交通调查数据表明,2009—2014 年,居民出行的平峰和高峰差异有所减少。从本次调查出行者的出发时间进行分析,多方式活跃出行者和较不活跃出行者这两种类型居民的出行是早晚高峰的主要构成者,而多目的与长距离活跃出行者的出发时间分布则较为平缓,前者是因为其通勤出行较少,后者则是因为通勤距离长有意避开了交通量最集中的时段。四种出行活跃度居民出发时间分布见图 7。

跃度类型为因变量,使用多项 Logistic 回归模型,探讨各类人群的社会经济属性特征,以及家庭因素对出行活跃度的影响。

4.1 变量

个人社会经济属性包括教育程度、职业、个人月收入、是否拥有驾照以及上下班时间是否固定。家庭社会经济属性包括家庭规模、家庭拥有的交通工具数量以及是否有青少年家庭成员。具体见表 3。

表 3 居民出行活跃度类型影响因素模型解释变量

变量	变量含义	变量类别	备注
1. 个人社会经济属性			
<i>m</i>	受教育程度为硕士及以上	离散变量	1=是,0=否
<i>l</i>	个人月收入为五千元及以下	离散变量	1=是,0=否
<i>h</i>	个人月收入为一万元及以上	离散变量	1=是,0=否
<i>g</i>	职业为政府机关事业单位	离散变量	1=是,0=否
<i>s</i>	职业为商业服务业	离散变量	1=是,0=否
<i>d</i>	拥有驾照	离散变量	1=是,0=否
<i>f</i>	上下班时间固定	离散变量	1=是,0=否
2. 家庭社会经济属性			
<i>n</i>	家庭人口规模	连续变量	
<i>b</i>	家庭拥有自行车数量	连续变量	
<i>m</i>	家庭拥有摩托车数量	连续变量	
<i>c</i>	家庭拥有小汽车数量	连续变量	
<i>k</i>	家庭中有 12~18 岁家庭成员	离散变量	1=是,0=否

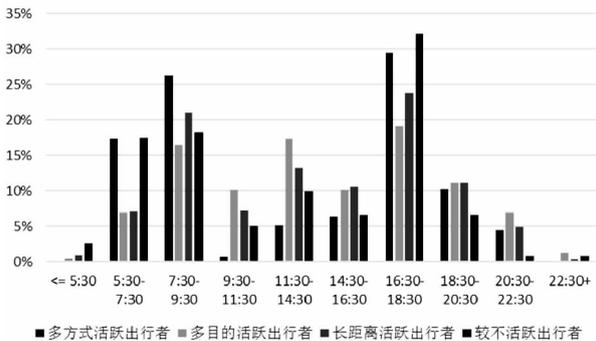


图 7 四种出行活跃度居民出发时间分布

4 上海市居民出行活跃度类型与社会经济属性特征互动解读

从以上分析不难看出,不同出行活跃度类型的居民在出行频率、交通方式、活动参与等活动-出行行为方面都有较为明显的差异。本节以样本出行活

4.2 模型拟合结果

模型分析结果如表 4 所示,通过卡方检验、似然比检验等方式,证明模型具有较高的拟合度,解释力较强。

(1)居民个人社会经济属性对于出行活跃度类型具有显著影响。收入越低的居民越不太可能是长距离活跃出行者,是由于长距离活跃出行者多采用小汽车作为主要的出行方式,他们的出行距离是平均值的 4.1 倍而出行时长仅为平均值的 1.6 倍,而低收入人群更多地采用价格相对低廉的公共交通。同时,较高收入的群体则更可能为多方式活跃出行者,他们不拘泥于交通费用,可以根据出行目的和时间更灵活地采用不同交通方式。在政府机关事业单位工作的居民更可能为多目的或长距离活跃出行者,说明相对于其他职业,这类居民具有更稳定的出行模式,普遍拥有较远的通勤路程,或者由于稳定的

作息时间比较方便安排其休闲娱乐类出行。拥有驾照的居民更不可能是多目的活跃出行者,是由于拥有驾照的居民较其他人而言更多使用小汽车出行,而由于交通拥堵、停车空间紧缺等问题限制了城市居民对于汽车的使用,多目的活跃出行者更多地采用公共交通和步行作为休闲购物等出行的交通方式,这一现象在前文的交通方式分析中也有印证。上下班时间稳定的居民拥有更固定的出行模式,通勤时更经常采用固定线路和换乘方式,因而与出行时间长、出行方式多的多方式活跃类型相匹配。从分析结果看,出行者是否为高学历对出行活跃度类型影响不显著,可能因为随着公共设施配套的完善和居民受教育程度的普遍提升,高学历在交通平台的知悉和选择上不再具有优势,因而表现出更多样化的特点。

表 4 社会经济属性对居民出行活跃度类型影响的多项 Logistic 模型

变量	多方式活跃出行者			多目的活跃出行者			长距离活跃出行者		
	系数	显著性	概率比	系数	显著性	概率比	系数	显著性	概率比
截距	-2.54	0.15		0.03	0.96		-5.2	0.01	
<i>m</i>	1.33	0.24	3.79	0.74	0.29	2.09	0.5	0.73	1.65
<i>l</i>	1.46	0.15	4.3	-0.11	0.81	0.9	-3	0.08*	0.05
<i>h</i>	1.94	0.07*	6.97	0.48	0.37	1.62	-0.48	0.71	0.62
<i>g</i>	-17.18	1	0	1.18	0.07*	3.25	2.43	0.08*	11.41
<i>s</i>	1.26	0.17	3.53	0.6	0.19	1.82	-16.8	1	0
<i>d</i>	-0.45	0.62	0.64	-0.85	0.06*	0.43	1.72	0.24	5.61
<i>f</i>	2.18	0.06*	8.81	0.32	0.44	1.37	2.03	0.19	7.64
<i>n</i>	-0.64	0.14	0.53	-0.21	0.29	0.81	0.42	0.35	1.52
<i>b</i>	-2.3	0.03**	0.1	-0.2	0.47	0.82	-2.32	0.05**	0.1
<i>m</i>	0.48	0.52	1.62	0.9	0.12	2.47	-17.52	1	0
<i>c</i>	0.34	0.7	1.41	0.54	0.19	1.71	-0.37	0.75	0.69
<i>k</i>	-16.2		0	0.53	0.56	1.7	5.6	0.01***	271.3
模型拟合参数	Chi-Square=59.095			Sig. = 0.009			Pseudo R2: Cox and Snell= 0.309		

注: * 代表在 10% 显著性水平上显著, ** 代表在 5% 显著性水平上显著, *** 代表在 1% 显著性水平上显著;参照类:较不活跃出行者类型

(2)家庭社会经济属性从交通资源和家庭成员构成方面影响居民出行活跃度类型。家庭中拥有越多自行车数量的居民越不可能是多方式和长距离活跃出行者类型,这是由于多方式活跃出行者更多地使用多种公共交通换乘,长距离活跃出行者更多选择小汽车作为出行方式,而拥有自行车数量较多的居民偏好较短的出行距离,出行更加灵活。家庭成员构成方面,家庭中有 12~18 岁青少年的居民更可能是长距离活跃出行者,这类出行者考虑家庭中青少年的活动和出行的影响,更多地居住在近郊区并以小汽车作为主要出行方式,对于这类出行者而言

通勤距离是相对次要的考量因素。

5 结论与讨论

本文基于居民出行的出行频率、出行距离、出行时长、出行方式和出行目的等多个衡量出行活跃程度的维度,将居民出行活跃度划分为多方式、多目的、长距离三种较活跃出行者类型以及较不活跃出行者类型,并基于上海市 2014—2015 年基于手机 GPS 收集的出行数据,分析每种活跃度类型出行者日常出行的行为特征以及工作日和周末的出行差异,并利用多项 Logistic 模型从居民的个人和家

庭社会经济属性的角度分析了居民活跃度类型的影响因素。

(1)使用出行频率、出行距离、出行时长、出行方式和出行目的作为居民出行活跃度的测度,发现不同居民活跃度类型之间存在差异。多方式活跃出行者出行时长和出行链复杂程度较高,表明此类出行者日常出行中常常在不同交通方式间换乘,对于公共交通的熟悉程度和利用率较高,平时出行以通勤出行为主;多目的活跃出行者的出行频率和非强制性指数较高,是一类经常出现休闲娱乐类非通勤出行的人群,在出发时间的选择上也表现得更为平均,通常是城市中拥有较多闲暇时间的居民群体;长距离活跃出行者表现为较长的出行距离和出行时长,居住地与工作点间距离较远,多采用小汽车作为主要的出行方式;较不活跃类型的出行者则在各个指标上都表现得较为平庸,低于样本平均值,代表了城市居民中对交通资源占用较少的人群。

(2)通过对上海地区基于手机 GPS 数据的活动-出行调查数据分析发现,居民出行活跃度类型受到个人和家庭社会经济属性的影响。高收入群体在出行方式的选择上更自由和多样;工作模式固定的居民通勤换乘多,出行链较为复杂;职业为政府机关事业单位的居民或拥有更多休闲类出行或拥有更长的通勤距离;家庭成员中有 12~18 岁青少年的居民更可能居住在近郊,以小汽车为主要交通方式,成为长距离活跃出行者。

(3)本文结合城市居民交通出行的多个特征,对居民出行活跃度进行了定义和测度。居民出行活跃度的概念对居民的出行特征进行综合考量,强调了居民在日常出行与空间互动中形成相对稳定的出行模式,以此透视居民出行活跃度对于社会交通资源使用的影响。对于长距离活跃出行者而言,较长的职住距离导致其通勤多采用小汽车,为引导公共交通的使用、改变现有情形可以从改善城市空间布局、加强多中心建设出发;对于多方式活跃出行者而言,由于通勤过程中换乘较多,可考虑继续采用或保持公交换乘优惠政策以巩固这类人群对于公共交通的使用,并在线路规划时有针对性地听取他们的建议,优化城市交通格局。

(4)本文提出的概念模式偏单一化,未来应进一步细化居民出行活跃度的测度方法,丰满逻辑体系

骨架,构建个体时空行为与活动-出行的互动理论,进一步分析居民出行需求对城市空间与交通政策的反馈作用。同时,文中所使用的分析方法存在一定局限,未能全面考量居民的居住地及工作地区位对于出行活跃度的影响,这些方面将在未来继续研究并加以改善。

参考文献:

- [1] 柴彦威,申悦,肖作鹏,等. 时空间行为研究动态及其实际应用前景[J]. 地理科学进展,2012, 31(6): 667-675.
- [2] BOHTE W, MAAT K. Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands [J]. *Transportation Research Part C: Emerging Technologies*, 2009, 17(3): 285-297.
- [3] SCHLICH R, AXHAUSEN K W. Habitual travel behavior: Evidence from a six-week travel diary [J]. *Transportation*, 2003, 30(1): 13-36.
- [4] WAGNER D P. Lexington Area Travel Data Collection Test: GPS for Personal Travel Surveys [R]. Columbus: Office of Highway Policy Information and Office of Technology Applications, 1997.
- [5] CASAS J, ARCE C. Trip Reporting in Household Travel Diaries; A Comparison to GPS-collected Data [C]//78th annual meeting of the Transportation Research Board, Washington, DC: Transportation Research Board, 1999: 428.
- [6] WOLF J, SCHÖUNFELDER S, SAMAGA U, et al. Eighty weeks of global positioning system traces: approaches to enriching trip information [J]. *Transportation Research Record: Journal of the Transportation Research Board*, 2004, 1870(1): 46-54.
- [7] ZMUD J, WOLF J. Identifying the correlates of trip misreporting-results from the California statewide household travel survey GPS study[C]. proceedings of the 10th International Conference on Travel Behaviour Research, 2003.
- [8] 张俊峰. 基于 GPS 技术的出行 OD 调查研究 [D]. 北京:北京交通大学, 2011.
- [9] ELANGO V V, GUENSLER R, OGLE J. Day-to-day travel variability in the commute atlanta study [J]. *Transportation Research Record Journal of the Transportation Research Board*, 2007, 43 (2014): 39-49.