

居民出行活动特征与收入水平的关系 ——以上海市为例

郭思慧^{1,2,3}, 文聪聪^{3,4}, 何云^{1,2,3}, 裴韬^{1,2*}

(1. 中国科学院地理科学与资源研究所, 北京 100101; 2. 中国科学院资源与环境信息系统国家重点实验室, 北京 100101; 3. 中国科学院大学, 北京 100049; 4. 中国科学院遥感与数字地球研究所, 北京 100094)

摘要:居民出行活动与居民的收入水平关系是公共交通、城市地理研究的重要问题。传统获取居民出行活动信息主要基于问卷调查的方式, 不仅成本高、样本量有限, 且研究局限于定性讨论, 研究结果易因受访者的主观意识而产生偏颇。随着信息技术的革新, 传感器记录的大规模人类活动信息为研究居民出行活动特征与居民收入水平关系提供了可能性。本文利用上海市居民时空轨迹数据, 从居民出行活动的角度出发, 首先构建居民出行活动指标, 并利用主成分分析法提取居民出行活动特征的主要成分; 然后对主成分进行K-Means聚类, 并针对不同出行活动特征的类别, 分析居民出行活动特征与居民收入水平的关系, 结果表明: ①居民出行地点多样性与居民出行范围大小是反映居民出行活动特征的主要成分; ②移动范围越小、移动地点多样性越低的居民类别, 其平均工资水平越高; ③不同移动性特征的类别平均收入水平差异与各类别居民工作地的产业发展有关。研究结论可为城市规划及相关经济政策制定提供参考。

关键词:出行活动; 移动性指标; 收入水平; 主成分分析; K-Means均值聚类; 上海市

1 引言

居民的出行活动与居民收入的关系是城市地理学研究的重要内容。传统获取居民出行活动信息的方式主要基于问卷调查(张文尝等, 2007; 丁威等, 2008; 周素红等, 2010; 陆锡明等, 2011), 不仅成本高、获取效率低, 且有效问卷的样本量小, 给居民出行活动研究增加了难度。随着信息技术的革新, 与居民经济活动有关的行为数据被无处不在的传感器实时记录, 通过分析这些来源于社会经济系统的大数据, 以较低的调查成本实时高效地了解居民微观经济属性成为可能(高见等, 2016)。例如, Soto等(2011)利用手机数据中抽取的移动模式(移动半径、移动熵等)、通话模式(通话量、短信条数等)和社交网络模式(通话联系人比例、短信联系人比例等)

特征, 来预测居民的个体经济水平, 预测精度超过80%; Smith-Clarke等(2014)对2个发展中国家的实证研究发现, 手机用户集群行为中的固有模式(如通话次数、通话网络等)特征与普查数据中的贫困指数显著相关; Blumenstock等(2015)结合卢旺达手机网络数据提取居民通话模式特征(如通话次数、通话时长)、通话联系网络特征及移动模式特征(移动范围、出行地点)来预测居民的社会经济水平(收入水平、住房情况、受教育程度等), 并通过预测的居民社会经济水平重建了整个国家的财富分布状况, 结果与政府普查数据高度一致。这些研究都表明, 居民的社会经济活动与居民收入状况之间存在一定的关系。然而, 上述研究所涉及的社会经济活动含义较广, 通话模式、社交模式等都包含在内, 城市居民的出行活动模式与收入水平间的关系实际

收稿日期: 2017-06; 修订日期: 2017-08。

基金项目: 国家自然科学基金项目(41525004, 41421001) [Foundation: National Natural Science Foundation of China, No.41525004, No.41421001]。

作者简介: 郭思慧(1994-), 女, 湖北鄂州人, 博士生, 主要从事空间大数据挖掘研究, E-mail: guosh@lreis.ac.cn。

通讯作者: 裴韬(1972-), 男, 江苏扬州人, 研究员, 主要从事空间大数据挖掘和地统计学等的研究, E-mail: peit@lreis.ac.cn。

引用格式: 郭思慧, 文聪聪, 何云, 等. 2017. 居民出行活动特征与收入水平的关系: 以上海市为例[J]. 地理科学进展, 36(9): 1158-1166. [Guo S H, Wen C C, He Y, et al. 2017. Relationship between travel behavior and income level of urban residents: A case study in Shanghai Municipality[J]. Progress in Geography, 36(9): 1158-1166.]. DOI: 10.18306/dlkxjz.2017.09.012

未能得到充分反映。为研究居民活动与其收入水平是否存在、以及存在何种关系,本文基于上海市居民的时空轨迹数据,从居民出行活动特征的角度出发,根据居民出行活动特征的差异对上海市居民进行聚类,研究各类居民工作地的空间分异,探寻居民出行活动特征与收入水平的关系。通过上述研究,一方面可为决策者针对不同经济水平的居民实施有效的经济决策提供参考,另一方面可为利用居民出行活动特征推断居民收入水平研究提供理论依据。

2 研究区域与方法

2.1 研究区域

上海作为中国的经济中心城市,2015年市域总人口2415万,其中市区常住人口1443万,居民出行空间具有多样性。此外,上海是长江经济带的龙头城市,产业齐全,不同产业人员收入水平不均。因此以上海市为例进行研究具有意义。图1为上海市行政区划,市辖15区1县,包括:金山区、虹口区、杨浦区、青浦区、奉贤区、闵行区、长宁区、宝山区、嘉定区、松江区、普陀区、徐汇区、浦东新区、黄浦区、静安区和崇明县,图中放大显示的部分为上海中心

城区(《上海市城市总体规划(1999年-2020年)》)。

2.2 数据及方法

2.2.1 数据来源

研究数据(表1)主要有:①上海市居民时空轨迹数据;②上海市各区县平均工资数据。其中,数据①来自2016年上海联通“沃数据开放应用大赛”组委会,通过对上海市642548名联通用户的手机信令数据按小时进行等时间间隔采样,获取用户各整点时刻的位置,得到用户的时空轨迹。时间为2015年12月27日-2016年1月6日,每一条记录含“日期”、“用户ID”及24小时的经纬度坐标共50个字段,如表2所示。数据②来自“搜狐新闻网”报道(<https://m.sohu.com/a/114893948-391481/?pvid=000115-3w-a>)。该数据以“职友集”网站对分布在上海市16个区县共8208个企业近一年内(2015年9月-2016年9月)发布的薪酬数据统计为依据;各区县内参与统计的企业个数及由此得到的平均工资,如表3所示。

2.2.2 研究方法

为分析居民出行活动特征与居民收入水平的关系,一方面需要用定量化指标刻画居民的出行活动特征,另一方面需要辅助数据反映居民收入水平。结合本文使用的数据,对居民出行活动特征的刻画,通过从居民时空轨迹数据中提取一系列移动

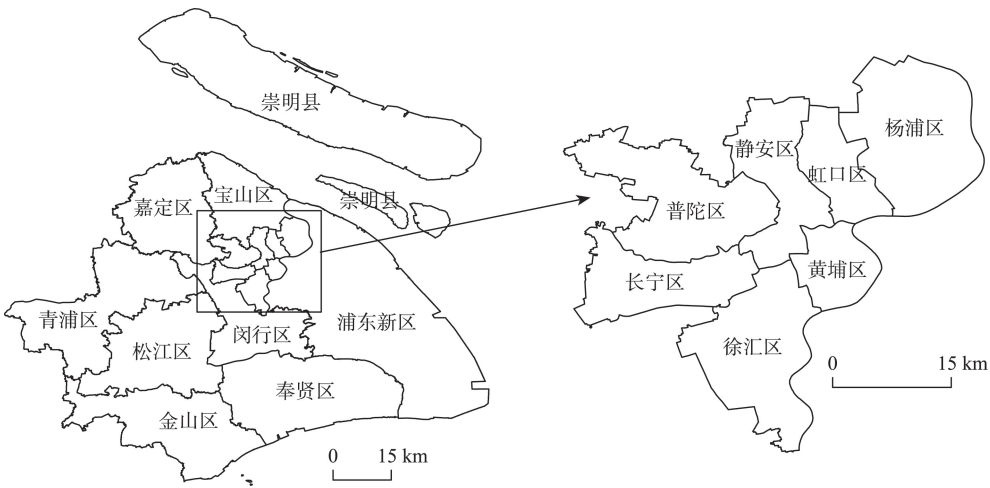


图1 上海市区域图
Fig. 1 Administrative division of Shanghai Municipality

表1 研究数据说明
Tab.1 Data used in the research

数据	时间	数据量	详细说明
居民时空轨迹数据	2015年12月27日-2016年1月6日	642548名用户5776605条轨迹	用户在24小时内的每整点的经纬度坐标
上海各区平均工资数据	2016年	16条	据16个区县内8208个企业发布薪酬统计而得

表2 居民时空轨迹数据示例

Tab.2 Sample data of residents' activity trajectories

日期	ID	经度	纬度	...	经度	纬度
20151230	ff7cfb0e717cc3a48af443209168ef92	121.28	31.35	...	121.28	31.34
20160101	bf3f04034044fd5c9d5f9445c3a3eda3	121.38	31.22	...	121.38	31.21

表3 各区县月平均工资

Tab.3 Average monthly salary in the districts of Shanghai Municipality

地区	月平均工资/元	样本企业个数	地区	月平均工资/元	样本企业个数
金山区	4590	80	宝山区	5340	214
虹口区	4890	250	嘉定区	5560	258
杨浦区	5200	282	松江区	5680	1584
青浦区	5230	166	普陀区	5790	432
奉贤区	5240	159	徐汇区	5960	477
闵行区	5280	422	浦东新区	5980	617
长宁区	5300	290	黄浦区	6100	258
崇明县	5310	103	静安区	7240	380

性指标来实现;居民收入水平取自居民工作单位所在行政区的月平均工资。由于居民出行活动特征是在个体层次上利用每一位居民的一系列移动性指标加以表达,而居民的收入水平状况则是在区域尺度上利用宏观统计平均值表示,因此无法直接通过传统的相关性分析方法来探究二者的相关关系。为此,本文考虑从居民出行活动特征的角度出发,根据居民出行活动特征的差异对居民进行聚类,通过分析各类别居民的收入水平变化,验证居民出行活动特征与居民收入水平的关系,方法流程如图2所示。本文主要处理步骤为:①针对居民时空轨迹,一方面通过到访频度法(Kang et al, 2010; Yuan et al, 2012; Calabrese et al, 2013)识别居民工作地;另一方面根据出行轨迹特征提取出行活动特征指标;②对出行活动特征指标进行主成分分析,从中选取刻画居民出行活动特征的主要成分;③对主成分进行K-Means聚类,得到具有不同移动性特征各类别居民;④以各类居民为研究对象,分别求取其平均移动性指标和月平均工资,探讨各类居民月平均工资随居民出行活动特征的变化规律。

(1) 居民工作地识别及移动性指标提取

在进行居民工作地识别和移动性指标提取之前,首先对原始轨迹数据进行清洗,包括剔除24小时轨迹数据皆为空值的记录、删除同一用户在同日的轨迹出现多种记录的错误数据、消除重复记录等。针对清洗后的居民时空轨迹数据,利用其中7个普通工作日(非周六、日且非元旦假期)背景下的

轨迹来识别居民工作地:统计每个居民于上午10点至下午15点时间段内出现在不同地点的次数,将7个工作日内累积出现次数最多的坐标视为该居民的工作地坐标。

考虑到节假日(元旦)对居民出行活动的影响,

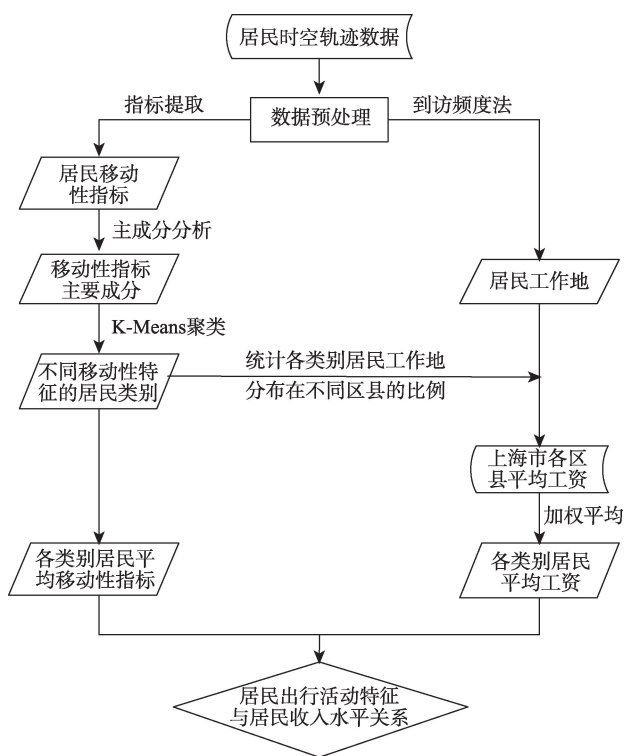


图2 研究方法框架图

Fig.2 Flowchart of the research method

将居民轨迹分为节假日和工作日2大类,采用相同的方法对2类轨迹分别提取移动性指标。由于人类出行规律的复杂性与多样性,刻画人们移动行为的指标尚未统一,但有关人们移动模式的研究主要关注出行地点的多样性及出行范围大小两方面的特征(Brockmann et al, 2006; González et al, 2008)。因此,本文选取的7个移动性指标(地点个数、移动熵、空间多样性、回旋半径、最远距离、平均出行距离、职住距离)也围绕居民到访地点及出行范围两方面特征展开。其中,地点个数(Soto et al, 2011; Frias-Martinez et al, 2013)、移动熵(Yuan et al, 2012)、空间多样性(Eagle et al, 2010)用来描述人们到访地点的多样性,例如,移动熵或空间多样性指标越大,表明居民到访不同地点的概率越均一,即居民访问不同地点的随机性较大,没有明显到访频繁的场所(Song et al, 2010)。回旋半径(Giannotti et al, 2011; Blumenstock et al, 2015)、最远距离、平均出行距离(Frias-Martinez et al, 2013)、职住距离(Yuan et al, 2012)用于反映人们出行范围的大小。各项指标具体说明如下:

1) 地点个数:指研究时段内用户经过的所有不同地点个数。例如,假设用户X在研究时段内共有2条轨迹记录,分别为1月4日与1月6日的轨迹。若用户在1月4日的轨迹数据中,只有上午9点和下午15点2个时刻下有坐标记录(2个坐标不同,分别记为A点、B点),其他时刻都为空值;在1月6日的轨迹数据中,只有上午6点与上午8点有坐标记录(2个坐标相同,但不同于A、B,记为C点),其他时刻都为空值,则该用户在研究时段内的地点个数为3(A、B、C点)。

2) 移动熵:计算公式为:

$$Entropy(a) = -\sum_{i=1}^N p_{ai} \log_2 p_{ai} \quad (1)$$

其中: $Entropy(a)$ 为用户 a 的移动熵; p_{ai} 为用户 a 经过第 i 个地点的频率; N 为用户 a 的地点个数。仍以用户 X 为例,若 $N=3$, $p_{x1}=1/4$ (A点出现的频率), $p_{x2}=1/4$ (B点出现的频率), $p_{x3}=2/4$ (C点出现的频率)。移动熵刻画了个体出行地点的异质性(González et al, 2008),移动熵越大,表示个体出行到不同地点的倾向性越不明显。

3) 空间多样性:计算公式为:

$$D(a) = \frac{Entropy(a)}{\log(N)} \quad (2)$$

其中: $D(a)$ 为用户 a 的空间多样性; $Entropy(a)$ 和 N 与移动熵中的变量意义相同。空间多样性越大

表示用户出行到不同地点的概率越均一(Eagle et al, 2010),即没有明显到访率更高的地点。

4) 回旋半径:计算公式为:

$$Radius(a) = \sqrt{\frac{1}{N} \sum_{i=1}^N (\overrightarrow{r_{ai}} - \overrightarrow{r_{ag}})^2} \quad (3)$$

其中: $Radius(a)$ 为用户 a 的回旋半径; $\overrightarrow{r_{ai}}$ 为用户 a 在第 i 个地点的坐标; $\overrightarrow{r_{ag}}$ 为用户 a 经过的所有地点的坐标中心; N 同前。

5) 最远距离:用户轨迹中相隔最远的2个轨迹点之间的欧氏距离。

6) 平均出行距离:用户平均每天出行的距离大小。

7) 职住距离:用户工作地与居住地之间的欧氏距离。

8) 此外,由于本文使用的移动轨迹是通过手机信令数据转换而得,而手机信令数据记录的用户坐标连续性不一致,导致有些用户有较完善的轨迹点坐标记录(每天24小时的所处的位置都有坐标记录),而有些用户的轨迹点坐标记录则存在某些时刻缺失,因此增设“地点频次”指标来描述居民的平均签到次数(平均每天有坐标记录的次数)。仍以用户 X 为例,该用户共有2天轨迹记录,第1天有坐标记录的次数为2(上午9点和下午15点),第2天有坐标记录的次数也为2(上午6点与上午8点),则该用户的地点频次为 $\frac{2+2}{2}=2$ 次,即平均每天有2次坐标记录。综上,共8个指标用于刻画居民出行活动特征,其中7个为描述移动性特征的指标,1个为衡量签到次数的指标。

(2) 主成分提取及K-Means聚类

本文旨在从居民的出行活动角度出发,对不同出行活动特征的居民类别进行分析,因此利用移动性指标大小对居民进行聚类。考虑到移动性指标间可能存在一定的相关性,在聚类之前,首先对8个指标进行主成分变换,提取主要影响成分。为确定主成分个数,利用碎石图显示特征值随主成分个数的变化,根据 Kaiser-Guttman 准则(Cliff, 1988; Kaiser, 1991),选取特征值大于1的值为成分。在此基础上针对提取的主成分对居民进行K-Means值聚类,根据 Calinski-Harabasz 指标(Caliński et al, 1974)确定聚类个数,利用R软件中的 cascadeKM 包进行聚类,得到不同出行活动特征的居民类别。

(3) 各类别居民的平均移动性指标和平均工资计算

针对各类居民进行分析:对各类居民的移动性

指标求均值,得到各类别居民的平均移动性指标值。根据各类别内工作地在不同行政区的居民比例,求取该类别居民的平均工资,计算公式为:

$$C_k = \sum_{i=1}^n p_{ki} s_i \tag{4}$$

其中: C_k 为第 k 个类别居民的平均工资; $n = 16$ 为上海市行政区县的个数; p_{ki} 为第 k 个类别中工作地属于第 i 个行政区内的居民比例; s_i 为第 i 个行政区的月平均工资。

3 结果分析

3.1 居民出行活动特征的主要成分

对8个指标采用方差极大旋转(Rencher, 1992)提取主成分,每个成分仅由几个主要指标解释(即每个成分只有少数几个很大的荷载,其他都是很小的荷载)。根据 Kaiser- Guttman 准则(Cliff, 1988; Kaiser, 1991)选择特征值大于1的成分为主成分,节假日与工作日的主成分个数都为3,各主成分与指标间的相关系数如表4所示。

由表4可知,节假日与工作日背景下8个活动特征指标对3个主成分的影响特征为:①相同点:居民出行地点多样性指标(如地点个数、移动熵、空间多样性)及居民的出行范围指标(如回旋半径、最远距离、平均出行距离及职住距离)分别为成分1与成分2的主要影响因素;居民出行地点频次、出行地点倾向性(移动熵或空间多样性)为成分3的主要影响因素。②不同点:节假日背景下,若居民签到次数越多(地点频次越大)、存在频繁签到的地点(移动熵小),则成分3值越大;而工作日则相反,居民签到次

数越少(地点频次越小)、无明显到访率较高的地点(空间多样性大),成分3值也较大。具体分析如下:

(1) 节假日中3个主成分的方差解释度分别为34.2%、32%、16.5%,其中成分1主要由移动熵、地点个数及空间多样性3个移动性指标来解释,且与这3个指标都成正相关,说明成分1主要刻画居民到访不同地点的多样性,居民出行地点多样性越高,成分1值越大;成分2主要解释与居民出行范围大小有关的4个指标(如回旋半径、最远距离、平均出行距离及职住距离),出行范围越大,成分2值越大;成分3与地点频次成正相关、与移动熵成负相关,说明如果居民平均每天记录的签到次数很多,且存在频繁签到的地点(到访其中几个地点的频率明显高于其他地点),则成分3的值较大。

(2) 工作日中成分1主要由地点个数、移动熵和平均出行距离3个移动性指标来解释,且与3个指标都成正相关,说明如果居民出行距离大,到访地点个数多,且没有频繁到访的地点,则成分1的值较大;成分2与节假日下的解释一致,居民出行范围越大,成分2值越大;成分3与地点频次成负相关、与空间多样性成正相关,说明如果居民平均每天记录的签到次数很少,且不存在到访概率高的地点(每个地点签到的频率都相当),则成分3的值较大。

上述分析表明,无论节假日还是工作日,从8个居民出行活动特征指标中可提取3个主成分:成分1主要描述出行地点多样性,成分2主要描述出行范围大小,成分3则与居民签到频率有关。

3.2 各类别移动性指标与收入水平关系

以3个主成分的欧氏距离为相似性大小度量指标,对3个主成分进行 K-Means 聚类,即相似性大

表4 方差极大旋转的主成分分析表
Tab.4 Principal component analysis by maximum variance rotation

相关系数		节假日			工作日		
		成分1	成分2	成分3	成分1	成分2	成分3
居民活动特征指标	地点个数	0.843	0.191	0.345	0.947		-0.118
	移动熵	0.946	0.193	-0.551	0.885	0.130	0.395
	空间多样性	0.730			0.490		0.812
	回旋半径	0.283	0.893		0.179	0.930	
	最远距离	0.398	0.872		0.383	0.869	
	平均出行距离	0.555	0.741		0.698	0.627	
	职住距离	-0.220	0.604	-0.155		0.622	0.131
	地点频次			0.925		-0.167	-0.912
方差解释	方差解释度	0.342	0.320	0.165	0.326	0.307	0.211
	累计方差解释度	0.342	0.662	0.827	0.326	0.633	0.845

(主成分间欧氏距离小)的聚为同一类,相似性小(主成分间欧氏距离大)的划分到不同类,由此得到不同出行活动特征的居民类别。在工作日,居民根据其出行活动特征可划分为10类,各类别居民的平均移动性指标大小与月平均工资水平关系如表5所示。对各类别按照月平均工资大小由高至低进行排序后发现:不同类别居民的移动性指标变化趋势与月平均工资变化趋势相反,即到访地点多样性指标(地点个数、移动熵、空间多样性)越高且出行范围指标(回旋半径、最远距离、平均出行距离、职住距离)越大的类别,其月平均工资也越高。节假日背景下按出行活动特征进行划分得到9类居民(表6),各类别居民的平均移动性指标与月平均工资的变化规律同工作日相同,即各项平均移动性指标越大,月平均工资越少。为验证各类别居民的收入差异,对各类别居民的工作地分布进行分析。

3.3 工作地空间分析

各类别居民的工作地分布如图3(工作日)、图4

所示(节假日)。工作日背景下,不同类别居民的工作地分布呈现“中心—外围”的圈层结构,且平均移动性指标越小的类别越趋于分布在中心,平均移动性指标越大的类别越趋于分散在市郊周边。这种圈层结构的空间分布与上海市城市发展的总体规划(《上海市城市总体规划(1999年-2020年)》)的空间功能具有一致性:从中心内层到近郊直至远郊村镇,分别以现代服务业、第二产业、第一产业为发展重点,因此从城中心到远郊的工作人员工资水平出现递减规律。①中心城区类别:如第3类、第10类,其移动性特征为:到访不同地点的个数少(小于10个),移动熵和空间多样性指标小(表明居民会频繁到访某些地点),职住距离较短(5 km以内),平均移动范围最小(最远出行距离小于15 km);其工作地特征为:主要分布在上海的中央商务区(如外滩)和主要公共活动中心(如豫园、人民广场等),这些地区是上海市一、二级商业中心集中之地(丁亮等,2017),以发展金融、商贸信息等现代服务业为主,

表5 非节假日平均移动性指标与各类别月平均工资水平关系

Tab.5 Average mobility indicator values and average monthly salary of each clustering group during weekdays								
类别	地点个数	移动熵	空间多样性	回旋半径/m	最远距离/m	平均出行距离/m	职住距离/m	月平均工资/元
3	6.12	1.38	1.89	1235.65	3265.69	1447.18	1449.56	5841.57
10	9.71	1.92	2.05	3804.35	10562.14	5270.43	3961.99	5821.59
4	11.24	2.16	2.15	6341.63	18256.65	9729.46	6304.11	5780.97
9	12.20	2.31	2.22	9265.10	27781.88	14510.48	7496.84	5757.01
5	16.04	2.75	2.41	14630.49	46858.75	28486.24	10598.36	5749.18
1	13.98	2.53	2.32	11889.93	36941.71	20799.34	8860.73	5745.92
8	19.29	3.08	2.53	16699.62	55862.56	39102.57	12831.57	5730.38
2	22.57	3.37	2.62	18836.34	65044.98	52204.91	16626.65	5720.75
7	24.97	3.58	2.68	21438.91	76405.06	69622.45	20768.52	5698.25
6	26.47	3.70	2.69	25747.73	94385.03	96397.05	27879.74	5637.16

表6 节假日平均移动性指标与各类别月平均工资水平关系

Tab.6 Average mobility indicator values and average monthly salary of each clustering group during holidays								
类别	地点个数	移动熵	空间多样性	回旋半径/m	最远距离/m	平均出行距离/m	职住距离/m	月平均工资/元
1	3.35	0.89	1.58	734.18	1772.68	1347.97	3118.66	5824.24
8	5.92	1.58	2.17	3415.76	8578.54	6808.19	6871.25	5793.82
7	6.98	1.86	2.38	6580.40	16868.21	14122.77	7912.96	5758.87
4	7.75	2.04	2.47	10134.06	26699.50	22858.12	8794.92	5743.65
2	8.86	2.25	2.57	13336.10	36107.52	33590.36	10166.70	5739.76
6	10.04	2.46	2.65	16609.67	46314.27	46325.09	12064.22	5738.16
9	11.37	2.67	2.71	19539.39	56282.00	63042.27	15180.39	5722.99
5	13.09	2.93	2.78	22379.24	67260.64	87637.96	18822.01	5700.55
3	14.26	3.10	2.83	26726.16	83624.84	128193.00	22658.60	5669.64

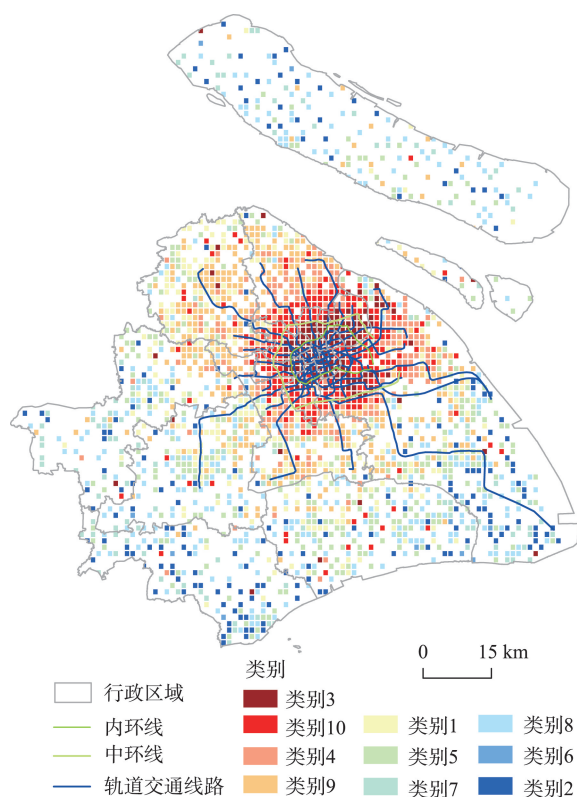


图3 工作日聚类结果工作地空间分布图

Fig.3 Spatial distribution of workplaces of the clustering groups during weekdays

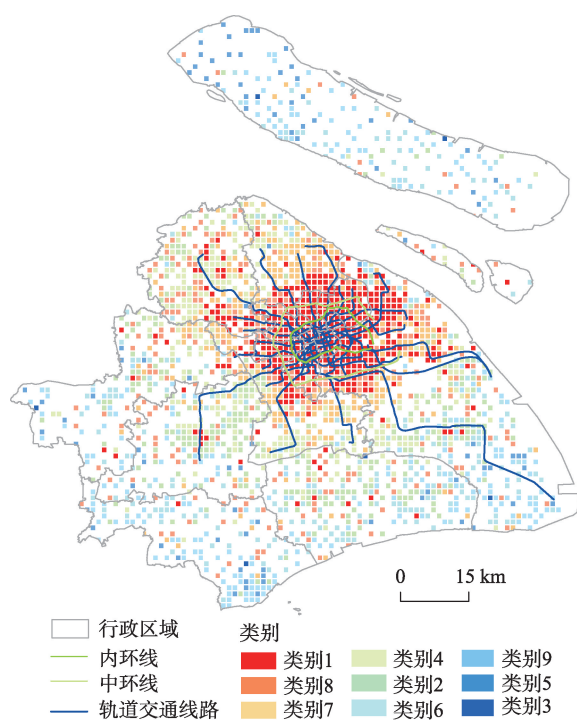


图4 节假日聚类结果工作地空间分布图

Fig.4 Spatial distribution of workplaces of the clustering groups during holidays

因此工作地在中心城区的第3、10类居民的月平均工资水平较高。②近郊区类别:第4、9、1、5类居民,其移动性特征表现为:到访不同地点的个数居中(5~15个),职住距离在5~10 km之间,平均出行距离在25 km以内;其工作地特征为:主要分布在地铁线路可达的区域范围内,以现代制造业为主,因此这几类居民的月平均工资水平处于中等。③远郊区类别:如第7、8、6、2类居民,其移动性特征表现为:到访不同地点的个数最多(15个以上),职住距离大于10 km,平均出行距离高于25 km;其工作地特征为:主要分布在上海市远郊区的村镇,这些地区以发展现代都市农业及旅游业为主,其居民的月平均工资水平相对较低。

节假日各类别居民的工作地分布也呈类似的圈层结构:工作地在城中心的类别主要为各项移动性指标都较小的类别(如第1、8、7类),这些类别的居民更有可能在陆家嘴等金融中心工作,其收入水平相应较高;而工作地在远郊区的主要为各项移动性指标都较大的类别(如第9、5、3类),其居民更有可能从事农业等相关工作,总体上平均收入水平相应较低。

由此可知,根据出行活动特征划分的上海市居民类别,出行地点多样性越小,出行范围越小,越有可能在城中心工作,整体上平均收入水平也越高;而出行地点多样性大,出行范围也大的类别,则更有可能在远郊区工作,整体上平均收入相对较低。

4 结论

本文基于上海市居民时空轨迹数据,构建居民移动特征指标,采用K-Means方法对移动性指标的主成分进行聚类,得到不同移动性特征的居民类别。通过分析各类别居民平均移动性指标与平均工资水平变化规律,得到如下结论:①居民出行地点多样性与居民出行范围大小是反映居民出行活动特征的主要成分,分别可用到访不同地点个数、移动熵、空间多样性指标以及回旋半径、最远距离、平均出行距离、职住距离加以描述。②移动范围越小、移动地点多样性越低的类别居民,其月平均工资水平越高。③不同移动性特征的类别平均收入水平差异与各类别居民工作地分布有关,移动范围小、移动地点多样性低的类别居民,其工作地主要分布在以金融商务业为主的中心商务区,平均工资水平相应较高;相反,移动范围大、移动地点多样性

高的类别居民,其工作地主要分布在以农业发展为主的远郊区,平均工资水平相应较低。

本文因数据方面的误差而存在以下不足:本文使用的居民工资数据以上海市各区县内企业发布的薪酬统计为依据,该数据表征不同行政区内工作人员的平均薪资水平,用其计算的类别平均工资只能从整体上反映不同类别间居民的薪资水平差异,未考虑各类别内居民薪资水平差异对结果造成的影响,使得本文对于居民出行活动特征与居民收入水平间关系的探讨局限于不同类别间居民的宏观比较。因此,本文结论的理论和实践意义为:一方面,从实际应用角度,有助于决策者掌握居民的收入水平状况,从而进行有针对性的决策。例如,根据《上海市城市总体规划(2016-2040)(草案)》,为提高居民幸福感,改善居民通勤状况是其中亟待解决的问题之一。通过本文的研究发现,除少部分工作地在中央商务区的居民外,大部分居民的工作地都分布在城市中心区以外,这些居民的通勤距离大,且收入水平不高,为节约出行费用成本,可能会更多地选择地铁、公交等公共交通工具出行,加大了内外环线之间公共交通需求压力,因此,政府应相应地增加地铁、公交线路,以满足与改善这部分工薪阶层的出行需求。另一方面,在理论层面,本文的结论验证了利用居民的出行活动特征可在一定程度上反映居民收入水平的高低,为进一步预测个体收入水平提供理论依据。例如,若将个体居民的出行规律用移动性指标进行量化,并能采集到部分个体居民的实际收入数据,则可运用回归分析等方法研究个体居民移动性指标与收入间的相应关系,进一步提升居民收入水平的预测水平。

但本文存在以下不足:由于使用的居民工资数据以上海市各区县内企业发布的薪酬统计为依据,该数据表征不同行政区内工作人员的平均薪资水平,用其计算的类别平均工资只能从整体上反映不同类别间居民的薪资水平差异,未考虑各类别内居民薪资水平差异对结果造成的影响,使得对于居民出行活动特征与居民收入水平间关系的探讨局限于不同类别间居民的宏观比较。因此,本文虽论证了居民出行活动特征在一定程度上反映了居民收入水平的高低,但并未深入解释两者的因果关系及相互影响机理。为更全面地揭示居民出行活动特征与其收入水平的关系,未来还需结合居民出行目的、出行方式等出行影响因素,从微观与宏观相结合的视角,对居民出行活动与收入水平的动态影响

机制与模式进行深入的研究。

参考文献(References)

- 丁亮,钮心毅,宋小冬. 2017. 上海中心城区商业中心空间特征研究[J]. 城市规划学刊, (1): 63-70. [Ding L, Niu X Y, Song X D. 2017. A study on spatial characteristics of commercial centers in the Shanghai central city[J]. Urban Planning Forum, (1): 63-70.]
- 丁威,杨晓光,伍速锋. 2008. 基于活动的居民出行行为研究综述[J]. 人文地理, (3): 85-91. [Ding W, Yang X G, Wu S F. 2008. A review of activity-based travel behavior research [J]. Human Geography, (3): 85-91.]
- 高见,周涛. 2016. 大数据揭示经济发展状况[J]. 电子科技大学学报, 45(4): 625-633. [Gao J, Zhou T. 2016. Big data reveal the status of economic development[J]. Journal of University of Electronics Science and Technology of China, 45(4): 625-633.]
- 陆锡明,顾啸涛. 2011. 上海市第五次居民出行调查与交通特征研究[J]. 城市交通, 9(5): 1-7. [Lu X M, Gu X T. 2011. The fifth travel survey of residents in Shanghai and characteristics analysis[J]. Urban Transport of China, 9(5): 1-7.]
- 张文尝,王成金,马清裕. 2007. 中国城市居民出行的时空特征及影响因素研究[J]. 地理科学, 27(6): 737-742. [Zhang W C, Wang C J, Ma Q Y. 2007. Spatial-temporal characteristics of urban resident trips and influence factors in China [J]. Scientia Geographica Sinica, 27(6): 737-742.]
- 周素红,邓丽芳. 2010. 基于T-GIS的广州市居民日常活动时空关系[J]. 地理学报, 65(12): 1454-1463. [Zhou S H, Deng L F. 2010. Spatio-temporal pattern of residents' daily activities based on T-GIS: A case study in Guangzhou, China[J]. Acta Geographica Sinica, 65(12): 1454-1463.]
- Blumenstock J, Cadamuro G, On R. 2015. Predicting poverty and wealth from mobile phone metadata[J]. Science, 350: 1073-1076.
- Brockmann D, Hufnagel L, Geisel T. 2006. The scaling laws of human travel[J]. Nature, 439: 462-465.
- Calabrese F, Mi D, Lorenzo G D, et al. 2013. Understanding individual mobility patterns from urban sensing data: A mobile phone trace example[J]. Transportation Research Part C: Emerging Technologies, 26(1): 301-313.
- Caliński T, Harabasz J. 1974. A dendrite method for cluster analysis[J]. Communications in Statistics, 3(1): 1-27.
- Cliff N. 1988. The eigenvalues-greater-than-one rule and the reliability of components[J]. Psychological Bulletin, 103 (2): 276-279.
- Eagle N, Macy M, Claxton R. 2010. Network diversity and economic development[J]. Science, 328: 1029-1031.
- Frias-Martinez V, Soguero-Ruiz C, Frias-Martinez E, et al. 2013. Forecasting socioeconomic trends with cell phone records[C]//Proceedings of the 3rd ACM symposium on

- computing for development. Bangalore, India: ACM: 15.
- Giannotti F, Nanni M, Pedreschi D, et al. 2011. Unveiling the complexity of human mobility by querying and mining massive trajectory data[J]. *The VLDB Journal*, 20(5): 695-719.
- González M C, Hidalgo C A, Barabási A L. 2008. Understanding individual human mobility patterns[J]. *Nature*, 453: 779-782.
- Kaiser H F. 1991. Coefficient alpha for a principal component and the Kaiser-Guttman rule[J]. *Psychological Reports*, 68(3): 855-858.
- Kang C, Gao S, Lin X, et al. 2010. Analyzing and geo-visualizing individual human mobility patterns using mobile call records[C]//*Geoinformatics, 2010 18th international conference on Geoinformatics*. IEEE: 1-7.
- Rencher A C. 1992. Interpretation of canonical discriminant functions, canonical variates, and principal components[J]. *The American Statistician*, 46(3): 217-225.
- Smith-Clarke C, Mashhadi A, Capra L. 2014. Poverty on the cheap: Estimating poverty maps using aggregated mobile communication networks[C]//*Proceedings of the SIGCHI conference on human factors in computing systems*. Toronto, Canada: ACM: 511-520.
- Song C M, Qu Z H, Blumm N, et al. 2010. Limits of predictability in human mobility[J]. *Science*, 327: 1018-1021.
- Soto V, Frias-Martinez V, Virseda J, et al. 2011. Prediction of socioeconomic levels using cell phone records[C]//*Proceedings of the 19th international conference on user modeling, adaption and personalization*. Berlin Heidelberg, Germany: Springer: 377-388.
- Yuan Y H, Raubal M, Liu Y. 2012. Correlating mobile phone usage and travel behavior: A case study of Harbin, China [J]. *Computers, Environment and Urban Systems*, 36(2): 118-130.

Relationship between travel behavior and income level of urban residents: A case study in Shanghai Municipality

GUO Sihui^{1,2,3}, WEN Congcong^{3,4}, HE Yun^{1,2,3}, PEI Tao^{1,2*}

(1. Institute of Geographic Sciences and Nature Resources Research, CAS, Beijing 100101, China;

2. State Key Laboratory of Resources and Environmental Information System, CAS, Beijing 100101, China;

3. University of Chinese Academy of Sciences, Beijing 100049, China;

4. Institute of Remote Sensing and Digital Earth, CAS, Beijing 100094, China)

Abstract: The relationship between income and travel behavior characteristics of urban residents is of great concern in urban geography. Income level of residents is an important indicator measuring regional social development, thus understanding this relationship is of great significance for city planning. Before the Big Data Age, due to the lack of residents' travel behavior information, it was difficult to study this relationship. However, along with the innovation of information technology, the use of ubiquitous sensors, such as mobile phones, has produced a large amount of human activity information, enabling the research on the relationship between residents' travel behaviors and income levels. In this study, based on the activity trajectory data in Shanghai Municipality from 27 December 2015 to 6 January 2016, we extracted a series of residents' mobility indicator data to measure mobility characteristics and conducted principal components analyses to extract the major components. We adopted the K-Means clustering method to classify residents into mobility groups and analyzed the feature of each group. Furthermore, the distribution of workplaces is shown to verify the difference in income levels between different mobility groups. Our results show that: (1) diversity of places to travel to and range of travel are two major components measuring residents' travel behavior; (2) residents who have smaller travel range and go to fewer places have higher average salary; (3) between the mobility groups, difference in income levels relate to industrial setup. These results may be useful for city planners to make efficient economic policies.

Key words: travel behavior; mobility indicator; income level; principal component analysis; K-Means clustering; Shanghai Municipality