

# 北京居民活动与出行行为时空数据采集与管理

柴彦威<sup>1</sup>, 申悦<sup>1</sup>, 马修军<sup>2</sup>, 赵莹<sup>1</sup>

(1. 北京大学城市与环境学院, 北京 100871; 2. 北京大学信息科学技术学院, 北京 100871)

**摘要:** 本研究对传统活动日志调查与基于GPS、LBS的移动数据采集在居民活动—移动数据的获取和应用方面进行对比, 并以2010年7月在北京进行的居民活动与移动调查为例, 探讨了个体行为时空数据采集的方法、存在问题和处理方式。北京市的实验调查采取活动日志调查与基于GPS、GSM两种不同定位方式的移动数据采集相结合的方法, 以定位设备为基础、以互动式调查网站为平台、以面对面和电话访谈为补充, 对天通苑和亦庄两个郊区居住区的样本居民进行了为期一周的行为时空数据采集。针对时空轨迹和活动日志存在的问题分别进行处理, 并对数据质量进行管理, 旨在为城市活动—移动系统研究中的精细化的数据采集与管理提供理论、方法和实践经验。

**关键词:** 时空数据; 活动—移动行为; GPS; LBS; 北京市

## 1 引言

随着现代交通技术的不断发展, 交通设施的不断完善, 人类的出行方式越来越多样化, 移动能力不断增强, 整个地球处于“时空压缩”的状态<sup>[1]</sup>。同时, 现代生活方式也增强了人类移动能力的复杂性, 多居住地与多工作地情况的出现, 旅游、散步等以出行为目的的出行的普及, 以计算机技术为核心的信息与通信技术 (ICT) 对居民的出行行为的替代、促进、改变等影响<sup>[2,3]</sup>, 人类在虚拟空间中移动能力的增强, 越来越复杂的人类移动性引发了地理、物理、生物、心理等领域的关注。20世纪70年代初, 城市规划学者蔡平提出城市活动系统概念<sup>[4,5]</sup>, 城市居民的活动与出行行为作为城市活动-移动系统的重要组成部分, 成为西方城市地理学、交通、和城市规划等领域关注的焦点。

微观个体行为时空数据是城市活动—移动系统分析与研究的核心, 西方不少学者对其采集过程进行了研究与实践<sup>[6,7]</sup>。面对中国城市发展所处的阶段、具有特色的文化与制度背景、居民日常生活方式, 柴彦威等在1995-1998年分别对深圳、天津和大连三个城市居民进行活动日志调查<sup>[8]</sup>, 2007年再次对北京、深圳居民进行了活动日志调查, 并对微观个体行为时空数据的生产过程与质量管理进行了探讨<sup>[9]</sup>, 阎小培和周素红等<sup>[10]</sup>、张文忠等<sup>[11]</sup>也分别从出行、居民行为、满意度等方面有所侧重地进行了行为时空数据采集的实践。然而, 随着相关研究主题的多样化、分析的深入化和模型构建的复杂化, 对数据量、时空精度的要求也逐渐提高。以GPS、LBS为代表的定位技术的快速发展和广泛应用为个体时空数据的采集提供了新的契机, 使数据更加真实化、精确化、规模化, 也使相关研究更加科

收稿日期: 2012-08-05; 修订日期: 2012-12-09

基金项目: 2009年IBM共享大学研究项目; “十二五”国家科技支撑计划项目(2012BAJ05B00)课题(2012BAJ05B04); 国家自然科学基金项目(40801046); 高等学校博士学科点专项科研基金项目(20090001110054)。

作者简介: 柴彦威(1964-), 男, 甘肃会宁人, 教授, 博士生导师。主要研究方向为时间地理学、社会地理学、城市与区域规划。E-mail: chyw@pku.edu.cn

学化、精细化、过程化<sup>[12]</sup>。

国内已有利用定位技术进行居民移动行为数据采集的探讨与基于此类数据的实证研究,但大多基于海量的、汇总的移动数据,忽略了微观个体的差异。技术探讨方面,杨飞等进行了基于手机定位的实时交通数据采集技术的研究<sup>[13]</sup>;方案设计方面,邓中伟等将GPS与传统居民出行调查相结合进行了调查方案的设计<sup>[14]</sup>;实证研究方面,胡小文等利用深圳出租车的GPS数据进行了出租车运行特征的研究<sup>[15]</sup>;应用展望方面,柴彦威等<sup>[16]</sup>、黄潇婷等<sup>[12]</sup>等分别探讨了基于定位技术的移动数据采集在城市规划和旅游者研究中的应用。由于汽车导航和交通管理较早开始使用GPS,基于定位技术的数据采集在交通领域的应用相对领先,而在城市规划、旅游规划等领域仅仅停留在对国外经验的学习和研究应用展望阶段。但在交通领域,技术与理论探讨多于具体实践,实证研究也多基于从相关机构或部门直接获取的基于车辆的定位数据。而如何将定位技术与活动日志调查相结合,在考虑微观个体差异性的同时,更加精确的采集居民个体的时空间行为数据,从而进行更加科学化、精细化的对于整个城市活动—移动系统的研究,国内相关研究与实践基本处于空白。

本研究首先对传统活动日志调查与基于GPS、LBS的移动数据采集在居民活动—移动数据的获取和应用方面进行对比,然后以2010年7月在北京市郊区天通苑、亦庄进行的居民活动与移动的实验调查为例,进行基于定位技术的个体行为时空数据采集和管理的探讨,旨在为城市活动—移动系统研究中的精细化的数据采集提供理论、方法和实践经验。

## 2 几种数据采集方式的比较

### 2.1 传统活动日志调查

传统日志调查的核心内容为调查日中居民一天24小时内所有的与研究相关的活动与出行及其时空信息,通常采取问卷调查的方式。典型的传统活动日志调查如2007年北京居民活动日志调查。由于活动日志数据内容丰富,可进行基于不同空间、不同属性居民的不同活动的研究<sup>[17-19]</sup>。

活动日志调查的优势在于一天内活动与出行信息的连续性、完整性和全面性。然而,与基于现代技术的调查方式相比,传统活动日志调查也存在一定不足。首先,由于活动日志全部由被调查者填写,其在回忆过去的活动移动情况时无法避免时空信息误差的出现,时间精度受到调查规定和被调查者习惯的影响,空间精度受到被调查者填写精度和数字化过程中空间定位方式的影响;其次,活动移动信息的真实性和详细程度也会受到被调查者问卷填写态度的影响,主观性较强;最后,由于调查内容复杂、填写过程繁琐,传统活动日志的调查规模受到人力、物力、财力等调查成本的限制。

### 2.2 基于GPS的移动数据采集

GPS是是新一代以卫星为基础的电子导航系统,以其全天候、实时、高精度和自动测量的特点融入了经济建设和社会发展的各个领域。1996年美国联邦高速公路管理局(FHWA)委托Battelle公司在Lexington开展了试验性的基于GPS的出行调查<sup>[20]</sup>,此后,澳大利亚、瑞士等国家先后在出行调查中利用GPS作为辅助<sup>[21]</sup>,Shoval等人利用GPS进行旅游者移动性调查<sup>[22]</sup>。北京市在2010年进行的第四次全市交通综合调查中也实施了抽样居民的GPS设备发放,上海市、深圳市、青岛市等也开展了基于GPS的出租车移动数据采集<sup>[15,23]</sup>。GPS数据在中国的交通领域已得到广泛应用,如交通实时监测<sup>[24]</sup>、交通状态分

析<sup>[25]</sup>、交通设施评估<sup>[23]</sup>、居民出行行为研究、出租车行驶方式<sup>[15]</sup>研究等。

基于GPS的移动数据采集在人文地理调查中的优势包括:①较高的时间精度,可进行实时数据的采集;②较高的空间精度,GPS的空间精确度在20米以内;③主动获取时空信息。基于GPS的移动数据采集在人文地理调查中的劣势包括:①GPS在室内没有信号,多用于居民出行调查,而对于居民在某室内地点停留的活动时空信息只能间接获取;②需要为被调查者提供定位设备,调查者需要承担设备损坏或丢失的风险,调查规模也受到设备成本的限制;③无法直接获取移动或活动的详细信息,如交通方式、活动类型等。

### 2.3 基于LBS的手机移动数据采集

移动位置服务(location-based services, LBS)是移动运营商根据移动用户的当前位置,为个人提供特定的信息增值服务,其关键技术是手机移动定位,主流定位技术包括蜂窝基站定位、观察角度、观察时差和辅助全球卫星定位系统(assisted global positioning system, A-GPS)等<sup>[16]</sup>。1996年,美国联邦通信委员会要求电信运营商为手机用户提供紧急救援(E911)服务,最终促成了LBS的出现。此后,德国、法国、瑞典、芬兰等国家纷纷推出各具特色的商用位置服务,最终为基于LBS的手机移动数据采集提供了可能。

由于对手机使用者的隐私保护,目前最常见的LBS数据为手机通话数据,既当用户拨打电话或接发短信时通过手机基站获取用户位置信息,该类数据可以从电信运营商处获得,国外很多学者和机构利用此类数据进行人类移动性、居民日常行为、旅游者游憩行为、城市空间结构、城市景观、社会网络等方面研究。爱沙尼亚塔图大学的Ahas等人利用爱沙尼亚的手机通话数据进行了本国居民的日常活动行为和国外旅游者的旅游行为研究<sup>[26-28]</sup>,美国麻省理工学院的感知实验室(MIT SENSEable City Lab)的Ratti等人利用意大利米兰市的手机通话数据进行了城市“移动景观”研究<sup>[29]</sup>,Marta等人利用LBS数据进行了人类移动性的研究<sup>[30]</sup>。国内对于LBS在交通、企业管理等方面应用的理论探讨较多,实证研究较少,北京大学的康朝贵等利用黑龙江省的手机通话数据进行了人类移动性方面的研究<sup>[31]</sup>。

随着智能手机的普及,国外学者开始尝试开发基于智能手机操作系统的居民时空间行为数据采集软件,在用户准许的情况下,利用辅助全球卫星定位系统(A-GPS)结合手机基站与GPS进行定位,并辅以短信询问或数据采集软件界面获取活动内容、出行交通方式等重要信息<sup>[32]</sup>。

基于LBS的手机移动数据采集在人文地理调查中的优势包括:①不需要为被调查者额外提供定位设备;②在有手机信号的地方即可进行定位,室内外均可进行数据采集;③相对于传统的活动日志调查,时空精度有所提高。基于LBS的手机移动数据采集在人文地理调查中的劣势包括:①调查对象为手机用户,会产生一定的样本偏差;②数据的获取通常需要依托电信运营商;③数据的时空精度取决于定位方式,而当利用手机基站进行定位时,空间精确度受到基站密度和手机信号情况的影响。两种手机定位数据中,手机通话数据通常为海量数据,样本量较大,但仅有居民进行手机通话时的时空间信息,而不关注被调查者之间的差异,用以进行城市尺度的汇总研究效果较为理想,但在进行居民时空间行为特征刻画时说服力不足;利用智能手机应用软件进行的数据采集,调查对象仅为愿意下载调查软件参与调查的智能手机用户,样本偏差相对较大,且调查效果与软件的功能、界面等设计密切相关,但由于其能够获取居民个体社会经济属性以及活动移动信息,便于进行非汇总的时空间行为与城市空间互动机理研究,具有较好的推广前景,是未来居民时空



间行为数据采集的重要发展方向。

#### 2.4 国内外移动数据采集方法比较

国外较早开始利用GPS和网络辅助传统的问卷调查获取居民行为时空数据<sup>[20,33]</sup>，目前已实现基于多种定位技术的移动数据采集，除GPS和LBS外，还包括基于蓝牙、Wi-Fi定位技术以及射频识别技术实现的移动数据获取，并且综合多种定位方式和软件应用的智能手机调查正逐渐成为新的发展方向，受到学者们的关注<sup>[34]</sup>。而国内的居民行为时空数据获取仍以传统的问卷调查为主，交通领域有部分基于车辆GPS数据的研究，但居民个体移动数据获取方面的研究聊无几。

目前，对于国内而言，利用GPS进行个体行为时空数据采集的技术已经较为完善，而在欧美等国家近期较为热门的智能手机调查时机尚不成熟，一方面国内智能手机的普及性尚不能较好的满足调查需要；另一方面，由于国内社会调查规范性和标准性的不足，居民出于对个人隐私的顾虑，利用个人手机参与调查的可能性较低。因此，适合在国内进行的时空间行为调查方式为：向居民提供综合多种定位方式的定位设备，并辅以问卷或网上调查平台获取居民个体行为时空数据。

本研究介绍了于2010年在北京进行的居民活动与移动调查，此次调查利用活动日志调查与基于GPS、GSM两种不同定位方式的移动数据采集相结合的方法，以定位设备为基础，以互动式调查网站为平台，以面对面和电话访谈为补充，进行居民一周活动与移动数据的采集。此次调查在数据获取方式上突破了传统基于问卷的日志调查中根据被调查者的记忆和估计进行的回忆式记录，而通过定位技术获取居民的实际轨迹，并利用调查平台使居民根据实际轨迹以及通过轨迹识别出的活动与移动进行活动日志的填写，革新了传统的活动日志调查方法；在调查期限方面突破了工作日和休息日各一日的两日调查，将调查期限延长至一周，为居民一周时空间行为的研究提供了数据基础。

### 3 行为时空数据采集

#### 3.1 调查内容

本研究的调查内容包括居民社会经济属性、居民一周内活动的完整轨迹和活动日志。考虑到本研究调查方式下被调查者的负担与调查有效性之间的关系，本研究结合北京2007年活动日志调查与研究的经验<sup>[9]</sup>，在内容设计时只覆盖活动日志的核心与关键问题。其中，居民社会经济属性包括被调查者个人及家庭基本信息、居住信息、日常交通信息、通讯与网络使用信息；轨迹即被调查者一周内每天24小时的时空信息；活动日志包括连续一周内每天24小时连续的活动和出行情况，活动信息包括每次活动的起始时间、终止时间、活动所在地的设施类型、活动类型、同伴以及满意度评价，出行信息包括每次出行的起始时间、终止时间、交通方式、同伴以及满意度评价。

#### 3.2 调查方式

由于需要获取移动轨迹数据、活动移动信息、地理空间信息、个人属性信息、个人偏好及感受信息等多源数据，本研究采取定位设备、互动式调查网站、面对面及电话访谈相结合的方式进行调查。

虽然GPS、GSM两种定位方式各有不足，却可以实现互补，GPS可以弥补GSM相对较差的定位精度，GSM也可以弥补GPS在室内无信号的不足，因此，本研究选取GPS、

GSM两种方式结合, 数据能够自动上传的定位设备。在定位的时间精度方面, GPS每三分钟定位一次, GSM每小时定位一次。最终以GPS定位点为主, GSM定位点作为辅助。

为了使被调查者能够及时的根据自己的轨迹情况填写活动及出行信息, 本研究进行了互动式调查平台的开发。由定位设备获取的定位点信息可以实时上传至调查平台, 由平台对被调查者的轨迹点进行处理, 根据一定时间内的移动速度对轨迹进行分割, 区分活动和移动, 并结合地理空间信息进行二维显示。被调查者根据自己移动设备对应的账号和密码进行网站登录, 即可观察自己调查期限内任意一段时间的活动轨迹, 并可根据轨迹对当时的活动和出行情况进行回忆, 填写相应信息, 也可以自己添加调查平台未识别的重要活动和出行, 还可以直接在网上完成个人社会经济属性调查问卷的填写(图1)。此外, 调查者也可以通过平台的调查员管理系统对被调查者的当前设备状态(是否开机、是否有信号)、所在位置、轨迹情况、活动日志填写情况、社会经济属性问卷填写情况进行实时监测, 便于调查的顺利进行(图2)。

当调查员利用管理平台进行监测发现问题后, 会对被调查者进行电话访谈, 而被调查者在整个调查过程中也可以随时与调查员联系。同时, 调查员在发放和回收设备的过程中有两次机会对被调查者进行面对面访谈。这些访谈资料既对活动日志数据进行了补充, 也便于了解被调查者的个人偏好与生活状态, 有利于对个体时空间行为的深入解剖。

对于被调查者, 其参与调查的整个流程如图3, 在与调查员签署调查协议后领取GPS设备, 每天定时充电并由本人随身携带, 每天晚上或第二天空闲时间以设备对应的账号和密码登陆调查网站, 查看轨迹, 并根据轨迹填写活动和出行信息。在整个被调查过程中遇到任何问题, 或者因个人原因无法按时上网填写活动与出行情况时, 可直接与调查员进行电话沟通。一周的调查完成后, 上交设备并领取相应报酬。

### 3.3 调查抽样

空间抽样方面, 城市空间范围内的随机抽样为理想的抽样方式, 但本次调查受到成本

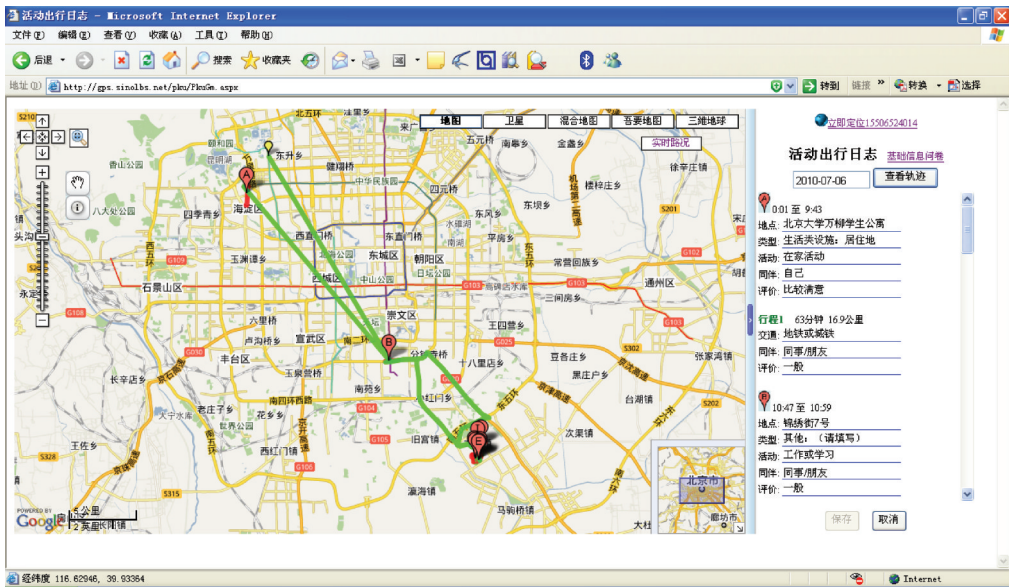


图1 被调查者活动与出行信息填写界面

Fig. 1 Activity and travel information filling interface for respondents

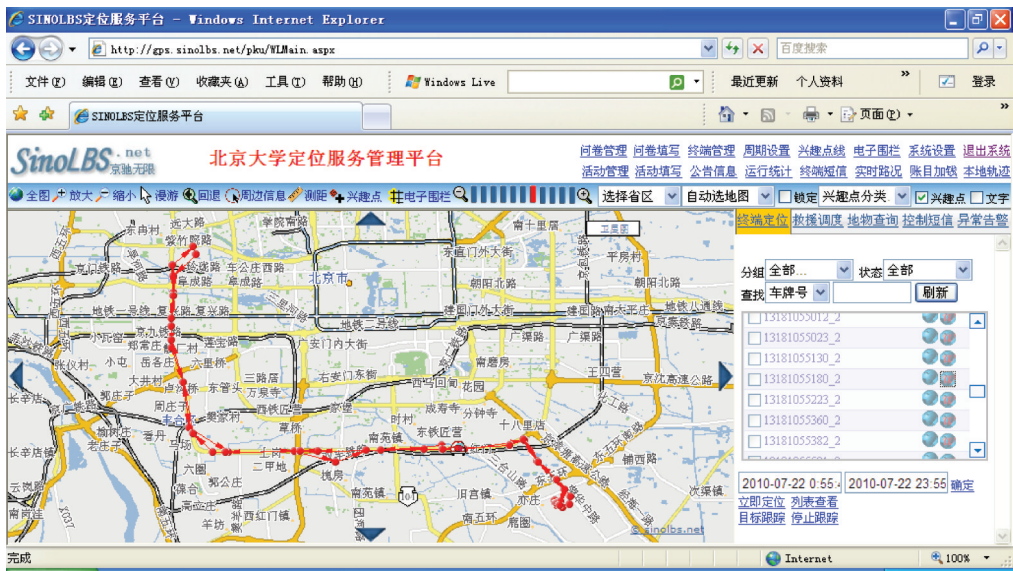


图2 调查员管理界面

Fig. 2 Monitor interface for investigators

和规模的限制无法实现。北京市郊区巨型社区居民的通勤行为是引起城市内部人流移动、交通拥堵的最为突出的因素，因此调查选取两个具有代表性的郊区巨型居住区，并在内部进行空间随机抽样。时间抽样方面，传统的活动日志多采取工作日和休息日各一天的两日日志调查，而本研究的调查方式为较长期数据的获取带来了可能。北京城市居民以一周为生活的基本时间单元，因此本调查对每个居民进行为期一周的跟踪调查。

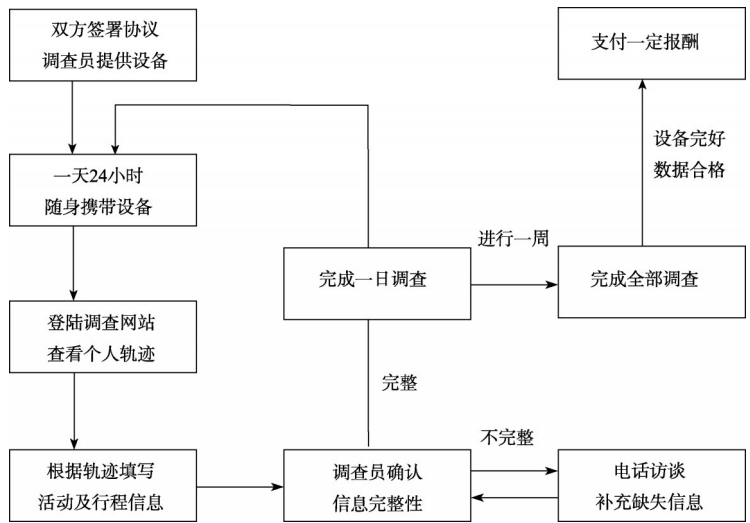


图3 被调查者参与调查流程图

Fig. 3 The survey participation progress of respondents

本研究的调查时间为2010年7月9日至7月25日，共进行两周，选择北京市天通苑和亦庄两个社区作为研究区域，通过当地居委会及企业联系被调查居民，要求被调查者具有一定的通勤距离、尽量可以上网，并在这两个条件的基础上保证样本在空间和社会经济属性上的随机性，最终得到有效样本100个<sup>①</sup>。

① 研究选择居住在天通苑的样本，覆盖五个大区，通过6个居委会选取51个样本，其中1人中途放弃调查，最终得到50个有效样本，天通苑本区21个、西区7个、北区9个、东区11个、中区2个。由于亦庄功能的复合性，本研究分别通过7个居住社区的居委会和开发区内两个公司选取调查样本51个，其中1个中途放弃调查，最终得到50个有效样本，37个居住在亦庄，13个在亦庄就业在外居住。



天通苑位于北京市北五环至北六环之间,是一个大型居住社区,也是北京市经济适用房的重点建设社区之一。1999年由顺天通房地产开发集团建设,占地面积约8 km<sup>2</sup>,规划人口40万,现已入住人口30多万。职能方面,天通苑以居住职能为主,区内及周边就业岗位较少,“有居无业”情况严重,每天大量居民需向外通勤,给交通带来了巨大的压力。交通方面,天通苑交通拥堵现象严重。地铁13号线经过天通本苑,地铁5号线横贯南北,但因乘客众多,上下班高峰需排长队进站。快速公交3线的开通以及立汤路的拓宽使交通状况得到了一定的缓解,但在上下班高峰时仍不容乐观。此外,天通苑还存在公共设施贫乏,教育设施、医疗机构严重不足,黑车、流动摊贩泛滥,城市管理落后等问题。

亦庄位于北京市东南五环与六环之间,其功能定位经历了“开发区—产业基地—卫星城—设立行政区—新城”的过程。2005年初国务院审议通过的《北京市城市总体规划(2004-2020年)》中,提出了构建“两轴—两带—多中心”的新城市空间格局,位于东部发展带上的亦庄是北京重点发展的三个新城之一。目前已基本形成了电子信息、生物医药、装备制造、汽车4大主导产业,并将数字电视、绿色能源、文化创意、生产性服务业作为新兴的主导产业。在亦庄发展的初期,由于入驻企业较少,当地开发了一批中高档房地产项目,吸引了一批在北京市区工作、收入水平较高的居民在此购房。随着亦庄的发展,大量外资及企业入驻的同时也吸引了一批居民到此就业,但由于本地所提供的住房可支付性较差,大部分在亦庄就业的居民不具有在本地购房的能力,存在职住不匹配、空间错位的现象。交通方面,目前地铁轨道交通L2号线已于2010年年底开通,为解决亦庄交通问题、促进职能升级、商业服务业发展、形象提升等方面提供了新的机遇。作为正在飞速发展的新城,亦庄仍面临着职住空间错位的问题,居民的活动与出行、以及对本地和区域外服务设施的利用和满意度情况尤为值得探索。

从区位条件看,两个社区都位于北京市的五环和六环之间,一南一北,遥相呼应;从交通情况看,天通苑交通便利,附近有两条轨道交通线路,而亦庄公交站点稀疏,目前仅一条轨道交通线路经过;从职能方面看,天通苑主要是居住职能,内部就业岗位极少,而亦庄作为北京市重点发展的新城,职能混杂,有大量的就业岗位。两个社区都存在的问题,且具有典型性,有较强的研究价值。

## 4 行为时空数据管理

整个调查共获取以下4方面的数据:被调查者社会经济属性、一周内时空轨迹和活动日志,调查员对每个被调查者的感性认识。其中,时空轨迹和活动日志数据受到调查方式的影响存在一定的误差,需要进行校正。

### 4.1 轨迹数据处理

轨迹数据主要存在数据缺失、数据噪声和人机分离的问题。数据缺失由设备关机和无信号两种原因造成。被调查者的误操作或未按时充电都可能导致设备关机,尽管在调查过程中调查员的实时监测、及时提醒对这种现象有所缓解,但仍然无法完全避免。由于GPS定位方式的缺陷,在室内或乘坐地铁过程中都有可能出现无信号所导致的数据缺失,尤其是每天凌晨初始化后一直到被调查者出门前在家的时段内,数据缺失现象较为普遍,而由于时间精度较粗,GSM定位方式对于数据的补充程度有限。数据噪声可分为远距离临时噪声和近距离长期噪声。近距离临时噪声主要指由于GPS的信号问题导致某一个或几个定

位点瞬间漂移至距离实际位置较远的地方,在本研究调查过程中,曾出现定位点漂移至市内其他区域、其他省份甚至其他国家的现象,这种情况出现次数较少且比较容易识别。近距离长期噪声主要指由于GPS的定位精度导致的对于同一地点的定位不在同一经纬度上的问题。人机分离由被调查者自身原因造成,也是发放定位设备的调查方式存在的不足,调查员的及时提醒对该情况有所缓解,但同样无法避免。

对于轨迹数据的基本的处理包括:通过轨迹点的移动速率识别远距离噪声点并进行删除,通过定位频率和相邻轨迹点位置识别室内缺失轨迹点并进行插补。进一步的处理还包括对家、工作地、其他停留地经纬度进行识别以实现近距离长期噪声的空间校正,对乘坐地铁过程中的轨迹通过最短路径法进行插补,对出行中缺失数据通过周期性规律的学习进行插补等。由于GPS数据量较大,数据处理往往需要在ArcGIS中进行二次开发以实现数据的自动修复,但这种修复不可避免会产生一定的误差,因此需要根据具体研究的需要,在保证数据真实性的基础上进行。

#### 4.2 日志数据整理

活动日志存在的问题有数据缺失、错误的记录和重复的记录,导致这些问题的主要原因包括定位数据存在缺失和噪声问题,调查平台对于活动和出行的识别和区分存在误差,被调查者自己填写和添加信息造成的错误与重复。数据缺失主要包括由于轨迹数据缺失造成的活动日志不完整,持续时间较短致使调查平台未能识别某些活动或出行造成的缺失。错误的记录主要包括轨迹数据噪声致使调查平台识别出实际上不存在的出行,调查平台对于活动和出行的错误分割,被调查者对于活动和出行信息的错误填写。重复的记录主要由系统识别信息和用户自填信息时间上的重复造成的。

对于活动日志数据进行整理的原则是在尽量尊重原始数据的前提下,以轨迹数据为参照,确保一天内活动日志数据的连续性、完整性和可信性。虽然轨迹数据也存在一定的问题,但基本可认为是客观的、真实的、可信的,将每人每天的轨迹数据点在ArcScene环境中以三维的形式显示,通过对相关信息的提取,对原始的活动日志数据进行添加、删除和修改的操作,对每一次处理的操作方式、操作依据进行记录。

#### 4.3 数据质量管理

在对轨迹数据和活动日志进行处理的过程中对数据质量进行评价。轨迹数据质量按照轨迹点数量、活动移动发生变化的关键点的缺失程度以每个被调查者每天的数据为单位进行评价,分为好、一般、差三级。定位点数量较多,关键点数据完整的为质量好的轨迹数据;定位点数量较多,有部分关键点缺失的为质量一般的轨迹数据;定位点数量较少,关键点缺失严重的为质量差的轨迹数据。活动日志按照数据的完整性和可信程度进行评价,分为完整可信、不太完整可信、完整不太可信和不完整不可信四类,其中完整性即活动日志是否完整包含一天24小时内连续的活动信息,可信程度即在正常居民日常活动情况和调查员对被调查者了解的基础上,活动日志是否真实、合理、可信。

对数据的质量进行统计(表1),在所有的轨迹和活动日志数据中,除去较差的活动轨迹数据和不完整不可信的活动日志数据,还有有效的轨迹数据516人·天,有效的活动日志601人·天,基

表1 数据质量统计表

Tab. 1 Statistics of data quality

轨迹数据		活动日志	
数据质量	数量/人·天	数据质量	数量/人·天
好	293	完整可信	242
一般	223	不太完整可信	179
差	189	完整不太可信	180
		不完整不可信	99
总计	705	总计	700



本可以满足研究需要。

## 5 结论与讨论

本文以2010年7月在北京市进行的居民活动与移动调查为例,调查采取活动日志调查与基于GPS、GSM两种不同定位方式的移动数据采集相结合的方法,以定位设备为基础、以互动式调查网站为平台、以面对面和电话访谈为补充,对北京市天通苑和亦庄两个居住区的样本居民进行为期一周的行为时空数据采集。数据库包括被调查者社会经济属性、一周内时空轨迹和活动日志、调查员感受。轨迹数据存在数据缺失、数据噪声和人机分离的问题,结合相关算法对室内缺失轨迹点进行插补,对远距离数据噪声点进行删除。活动日志存在的问题有数据缺失、错误的记录和重复的记录,在尽量尊重原始数据的前提下,以轨迹数据为参照,确保一天内活动日志数据的连续性、完整性和可信性,对原始的活动日志数据进行添加、删除和修改的操作。

此次试验调查突破了传统的单纯基于问卷的日志调查,基于网上调查平台综合移动跟踪、问卷调查、访谈等多种方式获取轨迹数据、活动移动信息、地理空间信息、个人属性信息、个人偏好及感受信息多源数据,在国内首次实现了非汇总的定位技术与活动日志的结合,为还原个体的活动移动行为、深入剖析行为决策的影响因素提供了重要数据基础。本次调查将移动数据与日志相结合,革新了依赖被调查者记忆和估计的回忆式日志调查,将定位技术获取的实际轨迹作为日志信息填写的依据;日志与访谈的结合加深了对个体生活状态的理解,深化了对个体行为决策影响因素的解释;在调查期限方面突破了工作日和休息日各一日的两日调查。

本次试验调查也存在一定的局限:①调查成本较高,限制了调查规模。在整个调查过程中,需要耗费大量的人力、物力、财力。调查前需要联系居委会寻找合适样本,对被调查者进行培训,对调查网站进行设计、开发与调试;调查期间需要调查员的实时监测以及与被调查者沟通;调查后需要进行设备回收、报酬发放等。受到单个样本调查成本的限制,无法进行大规模的调查,因此本研究的调查方法在进行推广时需要一定程度的改进。②活动与出行信息的获取受到调查网站算法的限制。本研究建立了互动式调查网站,并利用相关算法实现活动与出行的自动识别与切割。一方面,系统的自动识别影响了被调查者的信息填写精度,如停留在同一地点的时间段自动认为在进行同一项活动造成同一地点进行不同活动的信息缺失,再如接送他人或为汽车加油等活动停留时间过于短暂系统无法识别,即使被调查者可以自行添加未识别的活动与出行,数据补充量仍然有限。另一方面,系统无法识别定位设备产生的数据缺失和数据噪声,在进行活动与出行的自动识别时可能产生错误,数据缺失可能造成活动、出行分类的错误,数据噪声可能导致系统识别出本不存在的出行。网站算法的限制为调查过程中与被调查者的沟通以及调查后的数据处理增加了巨大的工作量;③要求被调查者具有一定的通勤距离、尽量可以上网,对样本的代表性产生了一定的影响。

与传统的活动日志问卷调查相比,此次调查对于居民个体信息的获取更加详尽和精确,但由于调查平台的不完善在整个调查过程中需要大量的调查员进行信息确认,并且GPS设备、调查平台的搭建也增加了调查的成本,采集居民个体行为时空数据所需要的人力及成本均大于传统问卷调查。因此,该实验调查的推广一方面需要减少调查的信息量,

更重要的是对系统进行完善,包括自动过滤数据噪声、提高活动和出行识别的准确度、增加系统对用户的自动提醒功能、增强用户界面的友好性等,在未来还可以结合土地利用与相关算法对居民的活动类型、出行交通方式等信息进行识别,从而降低用户任务量,彻底革新居民行为时空数据采集方法。

本文旨在为城市活动-移动系统研究中的真实化、精确化的数据采集与管理提供理论、方法和实践经验,尽管调查过程以及获取的数据存在一定的缺陷与局限,但本研究填补了国内定位技术与活动日志调查相结合进行精确的居民活动-移动数据采集方面的空白,具有一定启示性作用。而这种基于定位技术的数据除了使研究更加科学化、精细化外,是否能够开拓出新的研究主题,及其在城市规划、交通规划、旅游规划中的切实应用也值得相关学者的关注与探索。

### 参考文献(References)

- [1] Harvey D. The condition of postmodernity: An enquiry into the origins of cultural change. Blackwell Publishers, 1990.
- [2] Kwan M P, Dijst M, Schwanen T. The interaction between ICT and human activity-travel behave. *Transportation Research Part A*, 2007, 41(2): 121-124.
- [3] Andreev P, Salomon I, Pliskin N. Review: State of teleactivities. *Transportation Research Part C*, 2010, 18(1): 3-20.
- [4] Chapin F S. *Human Activity Patterns in the City: Things People Do in Time and in Space*. New York: John Wiley & Sons, Inc. 1974: 21-42.
- [5] Hanson S. Reviews: Human activity patterns in the city: Things people do in time and in space. *Annals of the Association of American Geographers*, 1976, 66 (1):153-155.
- [6] Stecher C (eds.). *Travel Survey Methods: Quality and Future Directions*. Emerald Group Publishing, 2006.
- [7] Stopher P R. Use of an activity-based diary to collect household travel data. *Transportation*, 1992, 19(2): 159-176.
- [8] 柴彦威, 刘志林, 李峥嵘, 等. 中国城市的时空间结构. 北京: 北京大学出版社, 2002.
- [9] 柴彦威, 张文佳, 张艳, 等. 微观个体行为时空数据的生产过程与质量管理: 以北京居民活动日志调查为例. *人文地理*, 2009, 24(6): 1-9.
- [10] 阎小培, 周素红, 毛蒋兴, 等. 高密度开发城市的交通系统与土地利用: 以广州为例. 北京: 科学出版社. 2006: 213-235.
- [11] 张文忠, 尹卫红, 张锦秋, 等. 中国宜居城市研究报告. 北京: 社会科学文献出版社, 2006.
- [12] 黄潇婷, 柴彦威, 赵莹, 等. 手机移动数据作为新数据源在旅游者研究中的应用探析. *旅游学刊*, 2010, 25(8): 39-45.
- [13] 杨飞, 裘炜毅. 基于手机定位的实时交通数据采集技术. *城市交通*, 2005, 3(4): 63-68.
- [14] 邓中伟, 季民河, 陈雯, 等. 耦合被动式GPS与网络调查的居民出行调查. *交通运输系统工程与信息*, 2009, 9(6): 1-6.
- [15] 胡小文, 冯均佳. 基于GPS数据采集的出租汽车交通运行特点研究. *城市交通*, 2007, 5(2): 91~95.
- [16] 柴彦威, 赵莹, 马修军, 等. 基于移动定位的时空行为数据获取方法与地理应用研究. *地域研究与开发*, 2010, 29 (6): 1-7.
- [17] 柴彦威, 张文佳. 基于家庭的购物行为时空间决策模型及其应用. *地理研究*, 2010, 29(2): 338-350.
- [18] 张艳, 柴彦威. 基于居住区比较的北京城市通勤研究. *地理研究*, 2009, 128(15): 1327-1340.
- [19] 申悦, 柴彦威. 转型期深圳居民日常活动的时空特征及其变化. *地域研究与开发*, 2010, 29(4): 67-71.
- [20] Battelle. *Lexington Area Travel Data Collection Test: Final Report*. Columbus, OH: Battelle Memorial Institute, 1997.
- [21] Stopher P C, FitzGerald, Xu M. Assessing the accuracy of the Sydney Household Travel Survey with GPS. *Transportation*, 2007 (34): 723-741.
- [22] Shoval N. Tracking technologies and urban analysis. *Cities*, 2008, (25): 21-28.
- [23] 史文欢, 孔庆杰, 刘允才. 基于出租车跟踪的市区路网交通状态估计方法. *交通信息与安全*, 2009, 27(5): 29-32.
- [24] 李筱菁, 孟庆春, 魏振钢, 等. GPS技术在城市交通状况实时检测技术中的应用. *青岛海洋大学学报(自然科学版)*, 2002, 32(3): 475-481.
- [25] 黄慧. 基于GPS的交通流动态分析与表达[D]. 上海: 上海交通大学, 2007.
- [26] Ahas R, Mark U. Location based services: New challenges for planning and public administration? *Futures*, 2005, 37 (6): 547-561.
- [27] Ahas R, Aasa A, Mark U, et al. Seasonal tourism spaces in Estonia: Case study with mobile positioning data. *Tourism*

- Management, 2007, 28: 898-910.
- [28] Ahas R, Aasa A, Roose A, et al. Evaluating passive mobile positioning data for tourism surveys: An Estonian case study. *Tourism Management*, 2008, 29: 469-486.
- [29] Ratti C, Pulselli R M, Williams S, et al. Mobile landscapes: Using location data from cell phones for urban analysis. *Environment and Planning B: Planning and Design*, 2006, 33(5): 727-748.
- [30] González M C, Hidalgo C A, Barabási A L. Understanding individual human mobility patterns. *Nature*, 2008, 453: 779-782.
- [31] Kang C G, Gao S, Liu Y, et al. Analyzing and Geo-visualizing Individual Human Mobility Patterns using Mobile Call Records. The 18th International Conference on Geoinformatics. 2010年9月. IEEE数据库.
- [32] Licoppe C, Diminescu D, Smoreda Z, et al. Using mobile phone geolocalisation for 'socio-geographical' analysis of co-ordination, urban mobilities, and social integration patterns. *Tijdschrift voor economische en sociale geografie*, 2008, 99(5): 584-601.
- [33] Buliung R N, Roorda M J, Rimmel T K. Exploring spatial variety in patterns of activity-travel behaviour: Initial results from the Toronto Travel-Activity Panel Survey. *Transportation*, 2008, 35(6): 697-722.
- [34] Chen C, Gong H, Lawson C, et al. Evaluating the feasibility of a passive travel survey collection in a complex urban environment: Lessons learned from the New York City case study. *Transportation Research Part A*, 2010, 44(10): 830-840.

## The collection and management of space-time data of individual behavior based on location-based technologies: A case study of activity-travel survey in Beijing

CHAI Yanwei<sup>1</sup>, SHEN Yue<sup>1</sup>, MA Xiujun<sup>2</sup>, ZHAO Ying<sup>1</sup>

(1. College of Urban and Environmental Science, Peking University, Beijing 100871, China;

2. School of Electronics and Computer Science, Peking University, Beijing 100871, China)

**Abstract:** The space-time data of individual behavior is one of the most important quantitative foundations of modeling and analyzing in the research of human geography, transportation and urban planning. The development of location-based technologies as well as information and communication technologies (ICT) makes accurate space-time data collection more feasible and affordable. Some unique conditions in China during the transition make it possible to obtain more detailed and reliable data. This study compares individual behavior survey methods and applications based on traditional questionnaires with those based on location-based technologies. It further explores the collection and management of individual space-time data, using the seven-day activity-travel survey in 2010 Beijing as the dataset. The survey methods contain location-aware devices which are a combination of GPS and GSM, interactive survey website including activity diary filling interface for respondents and monitor interface for administrators, and interview on telephone and face to face. Respondents' socio-economic characteristics, one-week space-time trajectories and activity-travel diaries are included in the database. Moreover, the study demonstrates the data problems that were met such as data noise, data loss, problems associated with the algorithm. We attempt to offer theory, method and application experiences of highly accurate data collection and management to research on urban activity-travel system.

**Key words:** space-time data; activity-travel behavior; GPS; LBS; Beijing