

基于手机信令数据的居民出行调查扩样模型

陈小鸿¹, 陈先龙^{1,2}, 李彩霞², 陈嘉超²

(1. 同济大学 道路与交通工程教育部重点实验室 上海 201804; 2. 广州市交通规划研究院信息模型所, 广东 广州 510030)

摘要: 系统分析了居民出行调查的误差成因, 简要介绍了传统加权扩样模型并分析了其局限性, 特别是对沉默出行需求揭示不足的缺点。结合手机信令数据用于出行行为研究的优势及特点, 提出了利用手机信令数据进行居民出行调查扩样及对挖掘沉默出行需求的方法。设计了包含5个主要步骤的居民出行调查样本数据、交通运行监测数据和手机信令数据相结合的扩样模型结构。以驻点分类技术为基础, 建立了基于手机信令数据的出行活动分布模型。在获取长周期手机信令数据推断人口职住分布的基础上, 基于居民出行调查RP(revealed preference)数据对交通方式划分模型参数进行了标定。从出行目的组成、时间分布和出行距离分布3个方面, 对比分析了调查样本、传统加权扩样模型和新模型的计算结果的差异, 结果表明新模型在揭示因瞒报、漏报而产生的沉默出行需求方面作用显著。

关键词: 居民出行调查; 扩样; 加权扩样; 手机信令数据; 沉默出行需求; 驻点分类

中图分类号: U491

文献标志码: A

Sample Expansion Model of Household Travel Survey Using Cellphone Data

CHEN Xiaohong¹, CHEN Xianlong^{1,2}, LI Caixia², CHEN Jiachao²

(1. Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, Shanghai 201804, China; 2. Information and Modelling Department, Guangzhou Transport Planning and Research Institute, Guangzhou 510030, China)

Abstract: In this paper, the causes of errors in household travel surveys were analyzed, the traditional weighted sample expansion model was introduced, and its limitations, especially the lack of unreported trip records were analyzed. Combining the advantages and characteristics of the research method using cellphone

data on travel behaviors, a novel model of using cellphone data to expand the sample of household travel surveys and mining unreported trip records was proposed. A sample expansion model was designed including five steps, combining the household travel survey data, traffic operation monitoring data, and cellphone data. Based on the stationary point classification technology, a travel behavior distribution model based on cellphone data was established. The parameters of the model split model were calibrated based on the revealed preference data of household travel surveys. The difference between the results of the survey sample, the traditional weighted sample expansion model, and the proposed model was analyzed from the aspects of travel purpose composition, time of day, and travel distance distribution. The results show that the model proposed reveals the unreported trip record caused by false reports and omissions.

Key words: household travel survey; sample expansion; weighted sample expansion; cellphone data; unreported trip record; stationary point classification

居民出行调查(household travel survey, HTS)是对城市居民家庭属性、个人特征、出行日志活动和意愿等的专项调查, 是交通模型开发与城市交通需求中长期的预测基础。居民出行调查最早起源于1930年代的美国^[1], 并在随后的底特律都市区交通研究^[2-3]和芝加哥交通研究^[4-5]中得以完善。由于调查数据可以关联出行者属性和出行活动属性, 迄今为止没有其他调查或数据可以替代。但居民出行调查是一种抽样调查, 必须通过扩样来推断总体特征指标。既有居民出行调查扩样^[6]模型主要包括基于抽样率的简单加权扩样和基于多种母体数据多重加

收稿日期: 2020-05-23

基金项目: 国家自然科学基金重点项目(71734004)

第一作者: 陈小鸿(1961—), 女, 教授, 博士生导师, 工学博士, 主要研究方向为区域与城市综合交通规划方法、交通政策、公共交通规划与慢行交通规划。E-mail: chenxh@tongji.edu.cn

通信作者: 陈先龙(1978—), 男, 教授级高级工程师, 博士生, 主要研究方向为交通模型及交通大数据开发与应用研究。E-mail: 1811104@tongji.edu.cn



论文
拓展
介绍

权扩样。从扩样结果来看看,多重加权扩样由于引入了更多的约束条件,更能够接近城市实际,但样本的缺陷及误差,特别是沉默出行需求(unreported trip record)难以在扩样过程中得到根本上的修正。沉默出行需求主要是指调查过程中因调查员漏填或者被调查对象漏报或者瞒报出行活动记录。一般而言,调查数据采集过程中,上下班、上学/放学、日常生活部分的刚性出行记录填报通常比较完整,其他非日常活动的弹性出行活动则容易缺失,特别是涉及个人私密信息的出行活动基本无法获取。居民出行调查本身是一项高投入、高成本的综合性调查,难以重复实施,且一旦调查实施完成就无法更改。所以对沉默出行需求的处理办法不多,目前使用最广泛的技术为开展志愿者调查,利用GPS辅助技术获得志愿者全轨迹数据,对出行调查记录表格进行补充,但通常志愿者样本量有限。总体来说,由于沉默出行需求的占比高,数据挖掘难度大,这也是扩样分析过程中公认的重点和难点。

从国际上来看,全美居民出行调查^[7,8](National Household Travel Survey, NHTS)所采用的扩样方法均为加权扩样,主要考虑了家庭和个人2种母体要素及无响应率(none-response rate)。2010/2011纽约居民出行调查^[9]及HOBBS^[10]、Richardson^[11]等所采用的扩样方法也同样是加权扩样模型。英格兰全国居民出行调查^[12]开始于1988年,每年度开展一次,2018年调查^[13]扩样仍采用的是加权扩样模型。其他一些国家也开展全国性的居民出行调查,如新西兰^[14],爱尔兰^[15],南非^[16]等均采用的是加权扩样模型。国外的研究主要采用加权扩样模型,依赖较为详细的家庭属性和个人属性分类数据,在此基础上非常重视对沉默出行需求的研究,如Argiropoulos^[17]、Forsman^[18]、Heathcote^[19]等,DVRPC^[20]还采用了志愿者GPS轨迹数据来进行修正。

中国大陆城市开展居民出行调查起步较晚,天津^[21]于1981年率先在国内开展居民出行调查,随后北京^[22]、上海^[23]、广州^[24]等超大城市陆续开展了该项工作。住房和城乡建设部2014年发布的《城市综合交通体系规划交通调查导则》^[25]和2018年推出的《城市综合交通调查技术标准(GB/T51334—2018)》^[26]进一步促进了中国大陆城市开展居民出行调查的规范化。在扩样方法方面,国外城市基本采用的都是按照人口的分类来构造扩样母体进行加权扩样的方法,主要考虑的因素包括家庭特征(家庭组

成、收入、车辆拥有)及个人属性(年龄、性别、就业就学等)。本世纪之前,国内城市基本沿用了此类加权扩样方法^[27-28]。香港^[29-30]开展居民出行调查工作相对较早,1973年开展了第一次居民出行调查,并在1981、1992、2002和2011年相继开展该项工作,并配合完成了五次交通模型的开发和维护工作。在扩样模型方面,香港也采用的是加权扩样方法。其他方面,李炬^[31]设计了分数据源、分对象、分方式的三步式组合扩样模型;彭泽宇^[32]在组合扩样模型的基础上引入了Bootstrap重抽样原理构建扩样评价指标体系;李元^[33]同样使用加权扩样方法来设计居民出行调查扩样分析系统。同样,国内也比较重视对沉默出行需求的挖掘工作,天津在2011年第五次综合交通调查^[21]中在国内率先使用GPS辅助设备来挖掘沉默出行需求,随后广州^[6,34]也进行了应用,取得了较好的效果。

总体来说,采用加权校核扩样模型推断的居民出行特征指标,虽然能够比样本指标更加符合城市实际,但无法弥补抽样调查的局限性。首先是调查母体可靠性问题,特别是城市人口规模的准确性;其次是漏报率问题,国内外研究的结论居民出行调查漏报率通常在30%以上;最后是小样本调查所得结果的空间分布误差影响。本文通过对居民出行调查误差成因解析,尝试通过手机信令数据和交通模型相结合的方法对母体特征、出行强度和空间分布等进行研究,以修正居民出行调查的原生缺陷。

1 居民出行调查误差成因及加权扩样的问题分析

1.1 居民出行调查误差成因分析

正如Weisberg^[35]所描述的,调查误差分为涵盖误差(样本母体的准确性)、抽样误差、测量误差、未响应误差和后处理误差。作为一种典型的抽样调查,居民出行调查的误差来源包含3个阶段:抽样阶段、调查阶段和分析阶段。抽样误差主要是因为不能做到完全随机抽样或者均匀抽样,造成样本分布与母体特征不能完全一致,如年龄结构、从业情况、车辆拥有等特征的差异,此外作为抽样母体的人口数据不准确也会带来抽样率失真,特别是中国城市仍然处于快速城镇化过程中、人口流动性大。第二阶段是调查误差,主要包括漏报/瞒报,错填等,特别是当前市民对隐私保护意识不断增强,瞒报、漏报普

遍存在,这也是后续扩样分析难度最大的沉默出行需求的主要组成部分;同时,调查员填写错误也难以完全避免。第三阶段则是扩样分析误差,原始调查数据校验、母体数据可靠性甄别以及扩样模型选择等。三个阶段的调查误差叠加、最终有可能导致推断的居民出行调查指标失真。

1.2 加权扩样模型

加权扩样是指通过样本量乘以权重系数(样本数/总体),居民出行调查中所采用的直接扩样方法为最简单的加权扩样模型。

$$G' = S \times W \quad (1)$$

式中: G' 为扩样所得总体指标; S 为样本指标; W 为权重,取值为1/抽样率。

在居民出行调查过程中,单因素扩样所得结果通常与实际偏差较大,这就要求引入更多的约束条件,如出行主体的年龄段、车辆拥有、从业情况等进行分类。理想情形是按照交叉分类进行扩样,但由于居民出行调查抽样率较低,更详细分类(图1所示,人口可以细分为多达8 640类)相应的样本数量会比较少,且对应母体可获得性差,同时由于样本数量过少对母体特征覆盖更差,扩样结果可能会产生更大的偏差。此外,采用单一变量逐步扩样方法只能满足当前变量的扩样约束条件,例如满足了年龄段结构的要求而车辆拥有可能发生了改变,为此嵌套了迭代循环的多重加权扩样模型被提出。由于多重加权扩样的各层之间互相独立,必然产生无法同时满足各重约束的问题,基于循环的多重加权扩样模型被用于增强对多重约束条件的符合度。国内2015年之前开展的居民出行调查扩样工作大多采用此类方法,计算流程见图2。

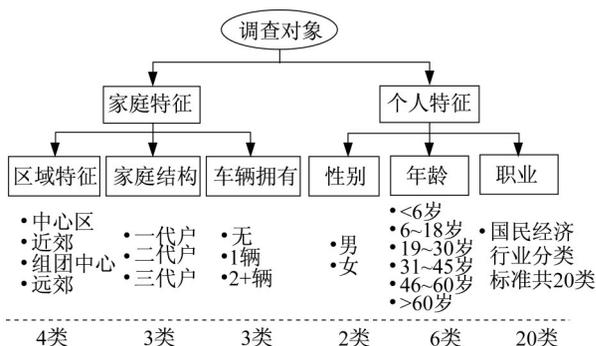


图1 调查对象分类示意

Fig. 1 Population classification for HTS

1.3 加权扩样模型存在的问题

虽然多重循环加权扩样能够避免单因素加权扩

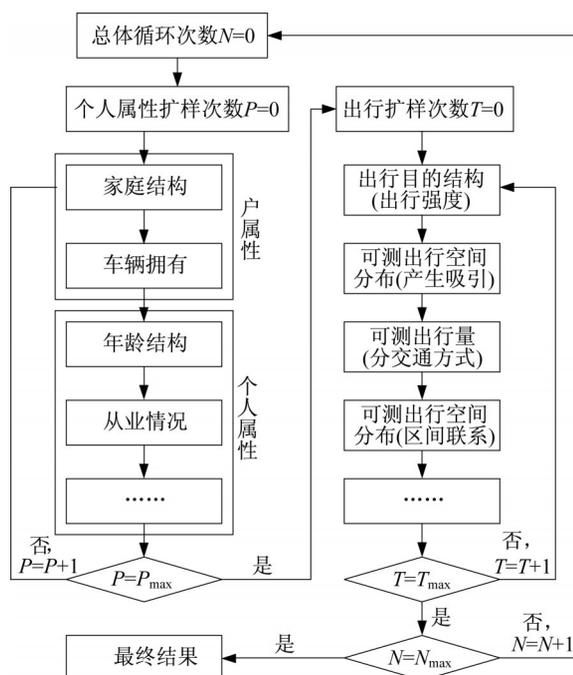


图2 多重循环加权扩样计算流程

Fig. 2 Procedure of multilevel expansion with loop

样误差,同时增加对多重约束条件的符合性,但还是存在一定问题。首先,加权扩样完全依赖调查样本数据,而沉默出行需求部分的出行几乎没有记录,只能通过漏报率指标强行补充和修正,难以对其目的、方式、空间分布等指标进行校正。其次,出行主体人具有多重属性,包括家庭特征(家庭区位、人口组成、车辆拥有等)和个人特征(年龄、性别、从业情况、收入等),进行完备分类并获取母体指标的难度较大。第三,出行本身是多维指标,包括强度、目的、方式、空间和时间等要素,由于居民出行调查基于家庭开展,是一种发生端调查,所得出行强度和目的指标相对比较可靠,交通方式、空间联系和时间及耗时分布则受样本影响较大。第四,加权扩样结果受样本指标的影响较大,一方面抽样率相对较小无法覆盖所有类型的样本,另一方面小概率出行记录会影响整体指标。第五,居民出行调查的本质为特征调查,前面几种问题的影响在加权扩样中会产生叠加甚至放大,所得城市总体出行指标与实际存在偏差,不能直接应用于城市交通运行评价,只能作为交通模型的输入参数间接使用。最后,多重循环加权扩样模型的收敛条件难以确定,只能是在实际应用过程中根据具体情况取值,一般经验是内层循环次数为3~5次、外层循环8~10次计算结果即可基本稳定。

2 手机信令数据分析出行行为特点分析

随着智能手机的普及,2010年以来手机话单数据、信令数据和各类互联网数据被逐渐应用到出行行为^[36-38]分析中。一方面使得对人的行为与活动追踪成为可能,另一方面也为交通模型众多不确定的输入提供了可能边界。首先,人口母体更加明确。以往人口数据更多依赖统计局、公安局和民政局等政府部门提供的统计数据,不仅数据因统计口径不同差异较大,活动人口数量与真实住地也不能充分反映。其次,就业岗位分布更加准确。经济普查所得的就业岗位数为机构注册地,注册地和实际工作地分离较为普遍。长周期的移动通信大数据有助于推算更加准确的岗位分布情况。第三,职住之间的对应关系得以建立。传统统计口径人口和岗位数据之间是互相独立的,通过长周期移动通信数据分析,可以建立人口和岗位之间的对应关系,这为解析城市通勤出行打下更坚实的基础。第四,移动通信数据可以用于人的空间移动识别。传统调查对人的出行来自受访对象的填报,存在可能涉及隐私出行瞒报和遗忘出行漏报的情形,而移动通信数据可以还原全日出行活动,有助于沉默出行需求^[29]的挖掘。第五,手机信令数据的样本量大,能够减少抽样调查样本代表性不足带来的误差。

手机信令数据的特点能够较好地弥补抽样调查的不足,但也存在一定的局限性。首先,难以实现活动人口完全覆盖。即便是用户数最多的中国移动公司其市场占有率也只是70%左右,总有非手机用户的存在,即使获得了所有运营商的数据也不能覆盖所有人口,而且一人多号的问题普遍存在,所以使用手机信令数据也需要进行扩样和校核。其次,手机信令数据的信息量相对有限,并不能开展准确的出行行为及影响要素分析。手机信令数据提供的位置信息依赖于的基站地理位置坐标,精度受限;移动过程中触发基站位置漂移频发,带来驻点识别难度大且存在不确定性;出行活动的目的识别难度大且所用交通方式获取困难;缺少出行主体的职业等用户特征信息。而这些恰是居民出行调查的优势所在。

综上所述,充分发挥手机信令和抽样调查数据的特点,将二者的优点融合、有助于更准确地解析居民出行特征和城市交通系统运行状态。

3 基于手机信令数据的居民出行调查扩样模型设计与算法

3.1 数据基础与模型结构总体设计

本文所使用的数据基础主要包括居民出行调查数据、手机信令数据、交通运行状态监测数据、公交IC卡、出租车GPS数据和高速公路收费数据、卡口车牌识别数据和其他统计数据,详见表1所示。

传统方法中,居民出行调查所得特征指标是交通模型的核心输入参数。需求预测建模基本思路是基于交通小区的人口、就业、车辆拥有等属性和交通设施网络,结合居民出行调查指标和城市交通运行指标及统计数据、对交通模型进行标定和校核,完成初步交通模型开发。在此基础上,利用城市交通运行指标,使用包括OD矩阵估计等方法,完成交通模型的进一步校正,得到能够与城市交通运行状况基本匹配的交通模型。从结果来看,出行分布是基于数学规划模型最优化求解结果,满足 modelled trip length distribution (MTLD) 和 observed trip length distribution (OTLD) 最佳拟合,可以理解为无穷解中与设定约束和收敛条件匹配最好的一组。但是否能够真实反映城市交通活动实际状况仍然存疑,主要包含两个方面的问题,一是观测所得OTLD为抽样调查结果,存在误差是必然的;二是城市通常是多中心的,数学上的最优解只是可行解而必定不是现实的唯一正确解。为了解决传统加权扩样及交通模型中存在的这些问题,本文提出了基于手机信令数据的居民出行调查扩样模型,核心思想是基于手机信令数据开展出行分布研究,获得分时段分目的出行分布矩阵,在此基础上完成基于居民出行调查记录RP(revealed preference)数据的交通方式划分模型标定、并利用交通运行数据进行校验,最后获得城市居民出行特征指标,模型结构见图3。

模型主要包含5个步骤:

(1)基于手机信令数据的分目的分时段OD矩阵计算:通过手机信令数据推断用户的职住分布,结合嵌入职住位置信息的驻点分析模型,推断出行目的(分通勤、基家生活、基家其他和非基家),进而获得分时段、分目的OD矩阵,这是模型的第一个难点。

(2)居民出行调查数据整理:对调查所得出行记录进行整理,获得每次出行的基本属性,包括交通方式、出行距离、出行耗时、出行费用及其他指标,并通过加权扩样获得初始权重系数。

表1 主要数据源及说明
Tab. 1 Description of datasets

数据	数据内容及说明	采集周期
居民出行调查数据	8.2万户属性,包括户籍属性、人口数、住房情况、车辆拥有等 20.7万人属性,包括年龄、性别、职业、受教育程度等 38.6万条出行记录信息,包括出发/到达时间、出行目的、出行方式、换乘等	2017-9-11—2017-9-24
手机信令数据	中国移动、中国联通手机信令数据 用户属性标签(职住地、是否流动人口、年龄结构、...) 信令位置数据(以基站为坐标的时间-空间位置数据序列)	2017年9月
交通运行监测数据	基于腾讯位置数据测算的广州市、佛山市和东莞市共5145个交通小区分时段(早高峰、平峰和晚高峰)、分方式(小汽车、公交车、轨道交通)的出行指标矩阵	2017年9月工作日平均
公交IC数据	公交车刷卡流水数据 公交车GPS轨迹数据 地铁刷卡流水数据	2017年9月
出租车GPS数据	出租车轨迹数据 出租车上下客位置数据	2017年9月
高速公路收费数据	广东省高速公路收费站位置 各收费站收费流水数据	2017年9月
卡口车牌识别数据	2379个道路卡口位置 车牌识别流水数据(脱敏后)	2017年9月
其他统计数据	2015年广州市1%人口抽样数据,包括人口数、年龄结构等信息 2017年社区/村上报人口数据 2017年广州市统计年鉴 ...	

(3)基于交通模型的分时段交通系统运行指标:分早、晚高峰和平峰、夜间4个时段,各OD点对之间的出行距离、耗时、费用等指标,一方面用于检验校核对应的样本数据,同时为无样本OD点对提供出行指标数据。

(4)基于RP数据的交通方式划分模型标定:通过居民出行调查所获得出行记录和初始权重系数,对交通方式划分模型进行标定,并计算模拟所得分时段分目的的分方式OD矩阵,这是模型第二个难点。

(5)交通系统运行监测数据校验:将第四步所得调查OD矩阵与交通系统运行监测所得矩阵进行校验,作为模型收敛条件,如满足要求则循环停止,否则重新计算出行记录的权重系数,并返回第二步开始新的循环。所使用的公交OD和轨道OD矩阵数据源自公交IC卡数据分析结果;出租车OD矩阵源自出租车GPS轨迹及载客数据分析结果;小汽车OD矩阵则来自2379个道路卡口及高速公路收费流水分析结果。

3.2 基于手机信令数据的驻点识别与出行分布模型研究

利用手机信令数据进行职住分析已经有较为

成熟的解决方案,但用于分析出行活动尚处于探索过程中。究其原因,手机信令数据的位置信息依赖基站且在活动过程中漂移频发,造成结果不稳定。为此,对于常住人口笔者设计了嵌入职住的时空核聚类初始驻点分类模型^[6],模型结构见图4,研究的基本要素主要包括时间、距离、速度,基于中国移动PaaS平台工作环境,先对用户进行时间坐标排序,并将初始点标识为驻留点,采用启发式搜索方法,从初始点往下搜索每一个时间点的坐标是否满足时空出行的时间、距离和速度要求,满足时空出行要求的坐标点,暂定为位移点,直至搜索到不满足出行要求的坐标点,则上一刻坐标点为驻留点,再基于此刻驻留点,再循环向下搜索满足条件的驻留点。驻点识别主要参数阈值如下:①由于信令为模糊位置数据,短距离出行较为敏感,经测试将最小可识别出行距离设为700m计算结果趋于稳定;②对于出行距离小于5km的短距离出行,出行时间大于5min,出行速度大于步行速度 $1\text{km}\cdot\text{h}^{-1}$;③对于出行距离大于5km的长距离出行,出行速度大于 $5\text{km}\cdot\text{h}^{-1}$ 。

以驻点分类为基础,研究结合出行者的驻点序列、职住地及驻点时空联系,提出了驻点与出行目的

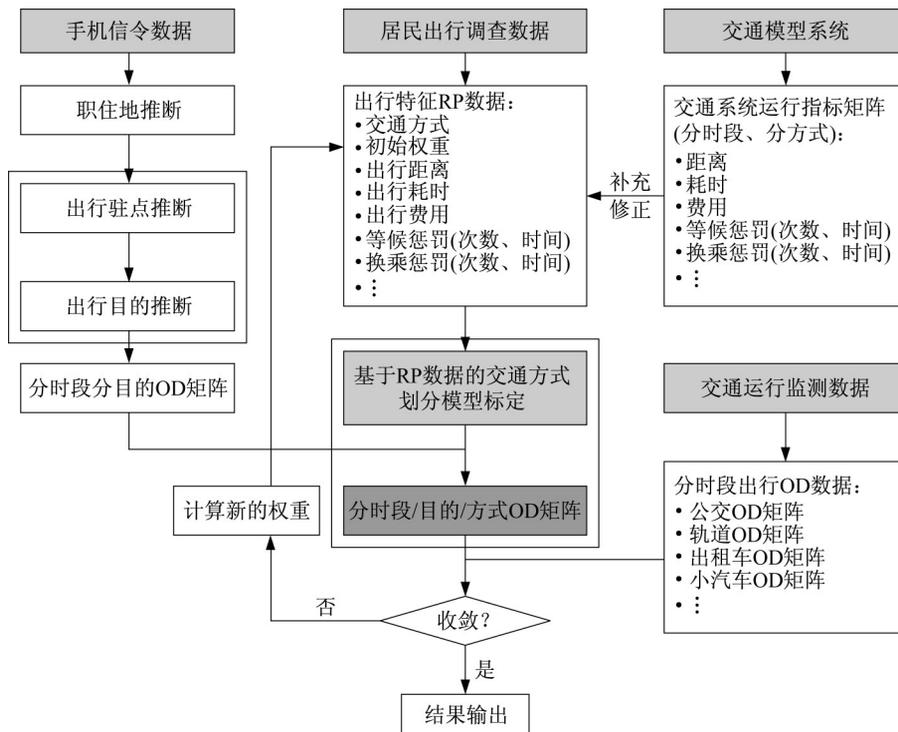


图 3 基于手机信令数据的居民出行调查扩样模型系统结构

Fig. 3 Expansion model framework of HTS based on cellphone data

分类准则(见图 5):①通勤(含通学):与居住地高频联系,且活动呈周期性;②生活出行:高频联系驻点,经常居家往返或返家过程中逗留,活动时间较为固定,且呈区域一致性(区域中较多用户访问该驻点的频率较高);③商务出行:和工作地高频联系;④其他出行:其他无规律出行。

根据以上分类准则,最终获得各类出行活动组成见表 2,由于商务出行比例较低(约 1.5%),在后续应用中将商务出行并入其他出行来处理。从数据中不难发现,城市日常出行活动中与三类具有稳定性特点的驻点(家、工作地和生活驻点)相关的出行活动约占 2/3,其他偶然出行则约占 1/3,这也在一定程度上说明了城市的相对稳定性。

3.3 基于居民出行调查样本 RP 数据的方式划分模型

与 SP 调查数据不同,居民出行记录只包含本次出行所使用的交通方式的特征指标信息,缺少其他可供选择交通方式的特征指标,需要进行补充。一般可以使用交通模型进行特征矩阵计算(Skim Matrix)获得。关于交通方式划分模型的标定 William Greene^[39]在其综述性文献《Discrete Choice modelling》中对多元 logit 模型(Multinational Logit Model, MNL)、混合 logit 模型(Mixed Logit Model)、巢式 logit 模型(Nested Logit Model)等离散

选择模型均有非常全面的描述,考虑各种交通方式的无关独立性(Independence of irrelevant alternatives, IIA),本文选择 Nested logit 模型作为待标定的模型结构。鉴于手机信令数据在短距离出行识别误差较大的特点,研究结合调查基础数据和既有交通模型,剔除了步行交通方式,设计了包含自行车、助动车、摩托车、出租车、小汽车、公交车和轨道交通在内的 7 种交通方式组成的 Nested Logit 模型(图 6)。

经整理,调查数据库中包含非步行交通方式的从家出发去工作的出行记录 53 336 条,约占全部调查记录的 32.3%。结合不同交通方式的特点选择不同特征变量指标来构建效用函数(utility function),主要包括时间、费用(票价或使用费、停车费、道路收费,……)、换乘次数、车内时间、车外时间等指标。对于单次出行所选择的交通方式使用调查所得指标,对于未选择交通方式则利用高德地图 API 获取出行起终点对应出行时段的起讫点(居民出行调查出行互动的起讫点信息为经纬度坐标)之间各种交通方式的距离、耗时、换乘、票价等指标。

使用通用离散选择模型分析软件 NLOGIT 对该模型进行标定,模型计算结果见表 4,所有特征变量的参数标定结果的特征参数的显著性水平都在 1% 以内,模型结果可以用于后续的计算。

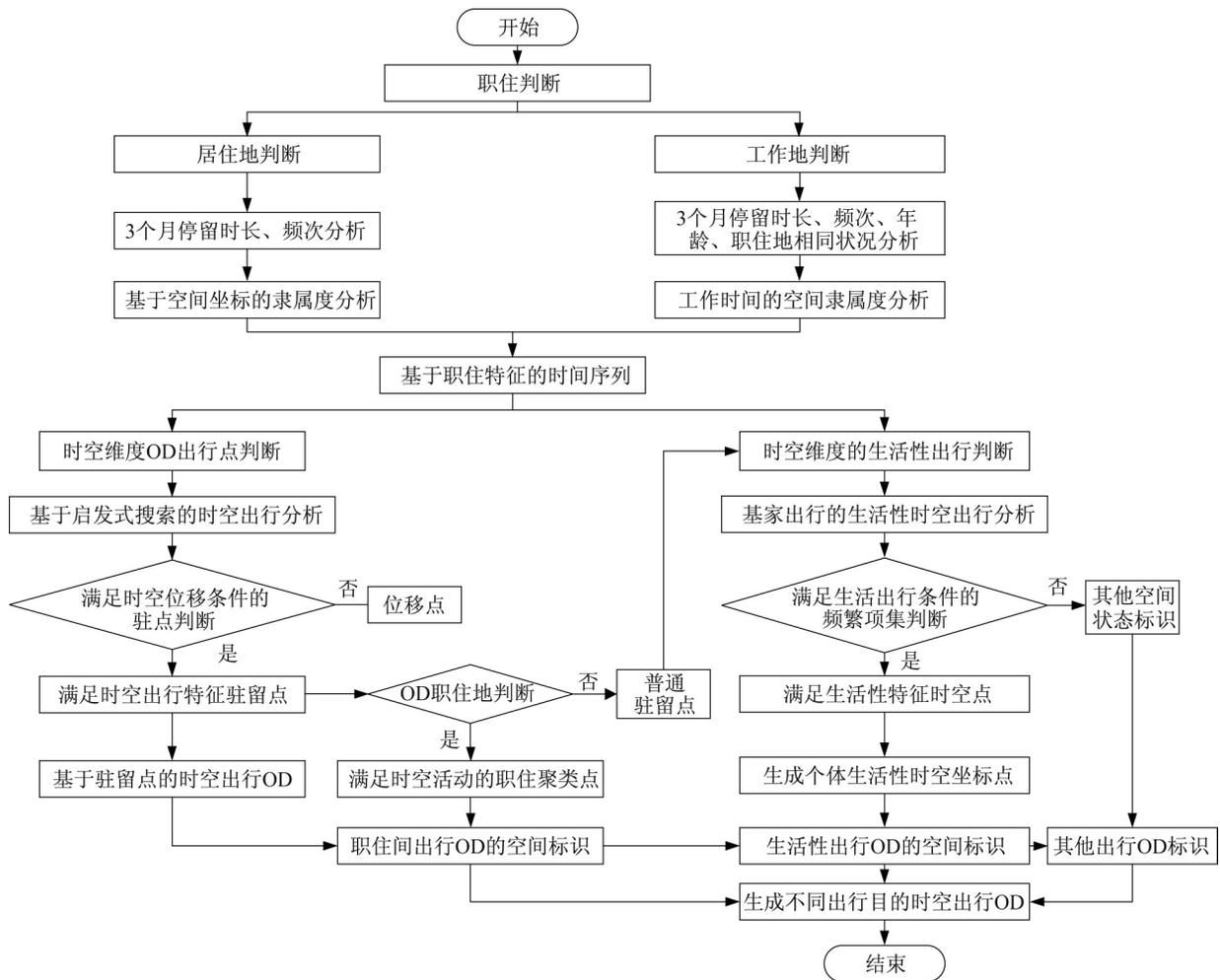


图4 驻点判断及分类模型结构

Fig. 4 Model framework of stay-points identification and classification

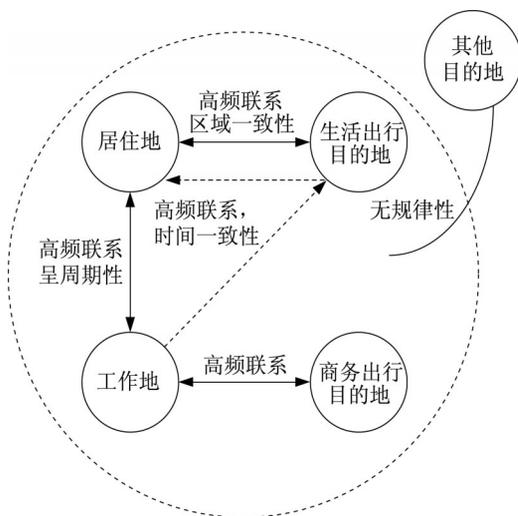


图5 手机信令驻点分类逻辑示意

Fig. 5 Logic diagram for classification of stationary points based on cellphone data

表2 基于手机信令数据的出行活动分类识别结果
Tab.2 Classification of trip purposes based on cell-phone data

					%
通勤 (含部分通学)	基家生活	商务出行	其他	合计	
48.3	18.1	1.2	32.4	100.0	

注:通勤(含部分通学)包含从家去工作地和从工作地回家;基家生活包括从家去生活性驻点和从生活性驻点回家。

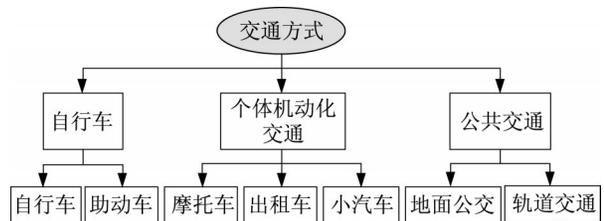


图6 Nested Logit模型结构

Fig. 6 Framework of Nested Logit model

表 3 各交通方式效用函数
Tab.3 Utility function by different modes

交通方式	效用函数
自行车	$U(\text{BIKE}) = B0_BIKE + B1_COST * GC + B1_TIME * TIME$
助动车	$U(\text{EBIKE}) = B0_EBIKE + B1_COST * GC + B1_TIME * TIME$
摩托车	$U(\text{MOTOR}) = B0_MOTO + B1_COST * GC + B2_TIME * TIME$
出租车	$U(\text{TAXI}) = B0_TAXI + B2_COST * GC + B2_TIME * TIME$
小汽车	$U(\text{CAR}) = B0_CAR + B2_COST * GC + B2_TIME * TIME$
公交车	$U(\text{BUS}) = B0_BUS + B3_COST * GC + B1_IVT * IVT + B1_OVT * OVT + B1_NTR * TRANS + B3_TIME * TIME$
地铁	$U(\text{METRO}) = B0_METRO + B4_COST * GC + B2_IVT * IVT + B2_OVT * OVT + B1_NTR * TRANS + B4_TIME * TIME$

注: $B0_*$ 为各交通方式效用函数的常数项; $B1_*$ 、 $B2_*$ 、 $B3_*$ 和 $B4_*$ 为变量参数; GC (general cost)为广义费用; $TIME$ (travel time)为出行耗时; IVT (in-vehicle time)为车内时间; OVT (out-vehicle time)为车外时间; $TRANS$ (number of transfer times)为换乘次数。

表 4 基家工作(HBW)出行交通方式划分模型标定结果指标
Tab.4 HBW parameter calibration result for mode split

变量说明	变量	参数	标准差	z	Prob. $ z > Z^*$	95% 置信区间		
常数项	自行车	B0_BIKE	2.327 51	238 578.00	0	1	*****	*****
	助动车	B0_EBIKE	3.828 08	238 578.00	0	1	*****	*****
	摩托车	B0_MOTO	-0.782 81	194 648.00	0	1	*****	*****
	出租车	B0_TAXI	-0.066 64	194 647.90	0	1	*****	*****
	小汽车	B0_CAR	3.017 67	194 647.90	0	1	*****	*****
	公交车	B0_BUS	13.539 60	765 621.90	0	1	*****	*****
	地铁	B0_METRO	14.277 10	765 621.80	0	1	*****	*****
广义费用参数	自行车/助动车/摩托车	B1_COST	1.737 31***	0.073 39	23.670	0	1.593 46	1.881 15
	出租车/小汽车	B2_COST	0.006 17***	0.001 61	3.840	0.000 1	0.003 02	0.009 32
	公交车	B3_COST	-0.0260 90***	0.006 68	-3.910	0.000 1	-0.039 18	-0.013 01
	轨道交通	B4_COST	0.043 27***	0.010 44	4.150	0	0.022 82	0.063 73
出行耗时参数	自行车/助动车	B1_TIME	-1.203 58***	0.051 40	-23.420	0	-1.304 31	-1.102 85
	摩托车/出租车/小汽车	B2_TIME	-0.004 35***	0.001 10	-3.960	0.000 1	-0.006 51	-0.002 20
	公交车	B3_TIME	-4.865 82***	0.550 06	-8.850	0	-5.943 91	-3.787 73
	轨道交通	B4_TIME	-6.774 18***	0.349 84	-19.360	0	-7.459 86	-6.088 50
公共交通运行指标	公交车内时间	B1_IVT	4.866 17***	0.543 08	8.960	0	3.801 76	5.930 58
	公交车外时间	B1_OVT	4.832 92***	0.540 48	8.940	0	3.773 60	5.892 24
	公交换乘次数	B1_NTR	-0.747 88***	0.071 33	-10.480	0	-0.887 69	-0.608 07
	轨道交通车内时间	B2_IVT	6.687 41***	0.336 66	19.860	0	6.027 57	7.347 26
	轨道交通车外时间	B2_OVT	6.648 73***	0.336 65	19.750	0	5.988 91	7.308 55
	轨道交通换乘次数	B2_NTR	-0.916 69***	0.105 32	-8.700	0	-1.123 11	-0.710 27
巢层参数	电动车/助动车	SM	3.026 38***	0.196 00	15.450	0	2.643 00	3.410 00
	摩托车/出租车/小汽车	PRT	3.709 40***	0.619 00	5.990	0	2.495 00	4.923 00
	公共交通	PT	0.943 06***	0.142 00	6.630	0	0.664 00	1.222 00

注: 麦克法登(McFadden)的伪 $R^2 = 0.374 996 3$; 样本数 $N = 533 36$, 迭代次数 $K = 24$; ***, **, * 分别表示在 99%、95% 和 90% 置信水平上显著。

4 扩样模型验证与改善效果分析

提出的居民出行调查扩样方法, 在传统调查方法可获得出行多要素影响关系基础上, 利用了手机信令数据在职住关系和出行空间分布识别的优势,

获得出行关键指标更可靠的估计。本节主要从出行目的构成、时辰分布和出行距离, 比较传统多维加权扩样与入户调查+手机信令相结合方法的差异, 以及后者的改善效果。

4.1 出行目的

调查样本中通勤及通学出行目的等刚性交通出行记录占比最大,达67.3%,而非基家出行目的仅占1.5%。对照志愿者调查结果,约存在31%的漏报、瞒报出行,而其出行目的主要为非基家的弹性出行活动。如表5所示,结合志愿者调查的加权扩样方法能够有效挖掘弹性出行目的的交通需求,非基家出行比例增加到12.4%。与样本及加权扩样的结果相比,基于手机信令驻点分类技术所获得的结果中基家通勤/通学出行比例进一步减少,而非基家及基家其他出行的弹性出行活动明显增加。通勤及通学占比减少原因,主要是部分中小學生无手机且大部分小学生出行距离较短,难以通过手机信令识别,而非基家弹性出行目的占比的提升,可以解释为手机信令数据对此类出行活动的捕获更加全面。

表5 不同分类方法出行目的组成结构比较

Tab.5 Comparison of trip purpose composition for different methods

分类方法	通勤及通学	基家生活	基家其他	非基家
调查样本	67.3	17.9	13.3	1.5
加权扩样	55.9	16.8	14.9	12.4
手机数据驻点分类	48.3	18.1	17.5	16.1

4.2 时间分布

从出发时间分布来看(见图7),调查样本对通勤时段覆盖较好,通过叠加沉默出行需求的加权扩样方法能够在一定程度上实现出行时辰分布的更优化,弥补抽样调查的不足。但从形态来看,手机信令数据的晚间高峰特征更符合广州的实际。

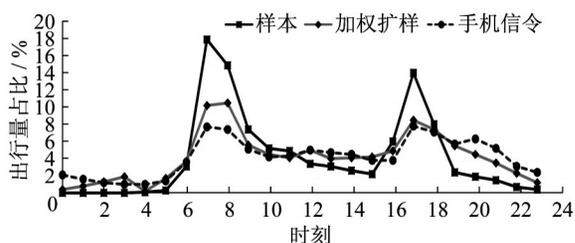


图7 不同扩样模型出行时间分布比较

Fig.7 Comparison of time of day of different models

4.3 出行距离

从出行距离分布来看(见表6),调查样本的短距离出行所占比重较大,长距离出行则相对较少。通过加权扩样能够在一定程度上修正调查样本出行距离分布的偏差,但由于扩样的基础为调查样本,其原生缺陷无法得到修正,具体表现为跨区域超长距离

出行样本扩样系数较大、导致超过50km出行占比达6.1%,这与地铁OD、出租车OD及道路卡口推算所得小汽车OD所获得的长距离出行量存在一定的差异,无疑手机信令推算结果在空间分布和距离分布上具有更高的可靠性。

表6 出行距离分布比较

Tab.6 Trip length distribution of different models

距离/ km	调查样 本/%	加权扩 样/%	手机信 令/%	距离/ km	调查样 本/%	加权扩 样/%	手机信 令/%
0~2	29.6	28.9	28.7	15~20	4.2	5.5	5.0
2~4	21.7	16.2	19.4	20~25	2.1	3.3	3.3
4~6	12.8	10.5	10.3	25~30	1.4	2.6	2.2
6~8	8.8	6.6	7.8	30~40	1.6	3.8	2.8
8~10	6.4	5.4	5.9	40~50	0.7	2.6	1.8
10~15	9.5	8.5	9.3	>50	1.1	6.1	3.6

5 结语

基于手机信令数据的居民出行调查扩样,本质是一种仿真模拟的计算方法。通过长周期的手机信令数据推断人口职住分布及出行OD,并利用居民出行调查所获得RP数据进行交通方式划分模型标定,得到分目的、分交通方式出行OD矩阵,扩样结果能够在一定程度上对样本的原生缺陷进行修正。对比传统的加权扩样方法,新的模型在出行强度及目的组成、时辰分布、距离分布的符合性等方面有明显改善作用,特别是在揭示因瞒报、漏报而产生等的沉默出行需求方面作用显著,这为开展居民出行调查扩样和校核工作提供了一种新的手段。但由于手机信令数据本身为用户的副产品,模糊地址、出行目的存在不确定性,无可靠交通方式信息等局限,也决定了它不能完全替代传统的居民出行问卷调查。此外,由于缺少手机用户属性个体信息,难以按照传统交通模型的有无车辆来开展模型标定工作,有必要结合交通模型技术开展进一步的标定和校验。手机信令数据、交通运行动态监测数据、传统居民出行特征调查与交通模型技术的联合,有助于更好地阐释城市出行活动规律,为城市交通综合治理与规划、运行和管理决策提供科学依据。

作者贡献申明:

陈小鸿:参与研究选题、论文框架设计、论文修订和审核。

陈先龙:参与研究选题、方案设计和实施,并承担了论文的撰写和修改工作。

李彩霞:参与了部分数据整理和统计分析工作。

陈嘉超:参与了部分数据整理和统计分析工作。

参考文献:

- [1] BOYCE D, WILIAMS H. Forecasting urban travel: Past, present and future [M]. Northampton: Edward Elgar Publishing Limited, 2015.
- [2] CARROLL J D. Detroit metropolitan area traffic study part I: Data summary and interpretation[R]. Detroit:[S.n.],1955.
- [3] WEINER E. Urban transportation planning in the united states: History, policy, and practice [M]. 5th ed. [s.l.]: Springer, 2016.
- [4] CARROLL J D. Chicago area transportation study final report volume I: Survey findings[R]. Chicago:[S.n.], 1959.
- [5] 佐佐木纲, 饭田恭敬. 交通工程学[M]. 邵春福, 杨海, 史其信, 等译. 北京: 人民交通出版社, 1994.
SASAKI Tsuna, IIDA Yasukyung. Traffic Engineering [M]. Translated by SHAO Chunfu, YANG Hai, SHI Qixing, et al. Beijing: China Communications Press, 1994
- [6] 马小毅, 龙小强, 金安等. 广州市新一轮交通综合交通调查总报告[R]. 广州: 广州市交通运输委员会, 2018.
MA Xiaoyi, LONG Xiaoqiang, JIN An etc. 2017 Guangzhou comprehensive transportation survey Report [R]. Guangzhou: Guangzhou Municipal Transportation Committee, 2018.
- [7] WESTAT. 2017 NHTS data user guide [R]. Washington D C: Federal Highway Administration, 2018.
- [8] ROTH S B, DEMATTEIS J, DAI Y. 2017 NHTS weighting report [R]. Washington D C: Federal Highway Administration, 2017.
- [9] CASAS J, ZMUD M, LANINI L, et al. 2010/2011 Regional household travel survey: Final report [R]. New York: NYSDOT / NYMTC, 2014.
- [10] HOBBS F D. Traffic planning and engineering [M]. 2nd. [s.l.]: Pergamon Press, 1979.
- [11] RICHARDSON A J, AMPT E S, MEYBURG A H. Survey methods for transport planning [M]. [s.l.]: Eucalyptus Press, 1995.
- [12] EVANS A, KELLY A, SLOCOMBE M. National Travel Survey: England 2018 [R]. London: Department of Transport, 2019.
- [13] CORNICK P, CANT J, BYRON C, et al. National Travel Survey 2018: Technical Report [R]. London: Department for Transport, 2019.
- [14] Ministry of Transport New Zealand. Household travel survey [EB/OL]. [2020-04-12] <https://www.transport.govt.nz/mot-resources/household-travel-survey/>.
- [15] MOONEY R, HEALY S, DULAT E. National household travel survey 2017: Final report [R]. Dublin City: Amárach Research, 2018.
- [16] LEHOHLA P. National household travel survey 2013: Technical report [R]. Pretoria: Statistics South Africa, 2014.
- [17] ARGIROPOULOS I, ANASTASSAKI A, DELOUKAS A. A longitudinal household travel survey: Issues of design, nonresponse and variation in travel behaviour [J]. IFAC Proceedings Volumes, 1997.
- [18] FORSMAN Å, GUSTAFSSON S, VADEBY A. Impact of nonresponse and weighting in Swedish travel survey [J]. Transportation Research Record: Journal of the Transportation Research Board, 2007,1993(1):80.
- [19] HEATHCOTE E A. The role of validation, weighting and expansion in travel surveys [J]. Publication of Australian Road Research Board, 1985:235.
- [20] Delaware Valley Regional Planning Commission. 2012—2013 Household travel survey for the Delaware valley region [R]. Delaware: Delaware Valley Regional Planning Commission, 2014.
- [21] 邹哲, 蒋寅, 朱海明, 等. 天津市综合交通模型框架及关键技术探索 [J]. 城市交通, 2013(5): 28.
ZOU Zhe, JIANG Yin, ZHU Haiming, et al. Comprehensive Transportation Model Framework and Key Technologies in Tianjin [J]. Urban Transport of China, 2013(5):28.
- [22] 李春艳, 郭继孚, 安志强, 等. 城市综合交通调查发展建议——基于北京市第五次综合交通调查 [J]. 城市交通, 2016(2): 29.
LI Chunyan, GUO Jifu, AN Zhiqiang, et al. Comments on urban comprehensive transportation survey: Learn lessons from the 5th Beijing comprehensive transportation survey [J]. Urban Transport of China, 2016(2):29.
- [23] 李娜, 董志国, 薛美根, 等. 上海市第五次综合交通调查新技术方法实践 [J]. 城市交通, 2016, 14(2): 35.
LI Na, DONG Zhiguo XUE Meigeng, et al. The Practice of new survey technology and methodology in the 5th Shanghai comprehensive transportation survey [J]. Urban Transport of China, 2016,14(2):35.
- [24] 马小毅. 居民出行调查数据扩样方法研究 [J]. 交通运输工程与信息学报, 2010(1):14.
MA Xiaoyi. Study on data sampling expansion method of resident trip survey [J]. Journal of Transportation Engineering and Information, 2010(1):14.
- [25] 中华人民共和国住房和城乡建设部. 城市综合交通体系规划交通调查导则 [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. Transportation survey guidelines for urban comprehensive transportation system planning [S]. Beijing: China Architecture & Building Press, 2014
- [26] 中华人民共和国住房和城乡建设部. 城市综合交通调查技术标准 [S]. 北京: 中国计划出版社, 2018.
Ministry of Housing and Urban-Rural Development of the People's Republic of China. Technical standard for urban comprehensive travel survey: GBT 51334-2018 [S]. Beijing: China Planning Press, 2018
- [27] 广州市居民出行调查领导小组. 1984广州市居民出行调查总

- 报告[R].广州:[S.n.], 1985.
Leading Group of Guangzhou Household Travel Survey. 1984 Guangzhou household travel survey report[R]. Guangzhou:[S. n.], 1985.
- [28] 陈声洪. 上海城市交通分析和预测[M]. 上海: 上海科学技术出版社, 1998.
CHEN Shenghong. Shanghai urban transportation analysis and forecasting[M]. Shanghai: Shanghai Science and Technology Press, 1998.
- [29] ARUP. 二零一一年交通习惯调查研究报告[R]. 香港: 香港运输署, 2014.
ARUP. 2011 travel characters survey report[R]. Hong Kong: Hong Kong Transport Department, 2014.
- [30] ARUP. 二零零二年交通习惯调查研究报告[R]. 香港: 香港运输署, 2003.
ARUP. 2002 travel characters survey report[R]. Hong Kong: Hong Kong Transport Department, 2003.
- [31] 李炬. 城市居民出行调查的理论研究与实践[D]. 西安: 长安大学, 2011.
LI Ju. The Study of Theories and Practice in Urban Residents Travel Survey[D]. Xi'an: Chang'an University, 2011.
- [32] 彭泽宇. 居民出行调查数据组合扩样方法及评价研究[D]. 武汉: 华中科技大学, 2016.
PENG Zeyu. Study on sample combined expansion method and evaluation of household trip survey [D]. Wuhan: Huazhong university of science and technology, 2016.
- [33] 李元. 基于多源大数据的居民出行调查校核体系研究[D]. 西安: 长安大学, 2017.
LI Yuan. Research on framework of resident trip survey checking system based on multi-source big data [D]. Xi'an: Chang'an University, 2017.
- [34] 苏跃江, 陈先龙, 吴德馨. 大数据在广州市第三次交通综合调查中的应用[J]. 城市交通, 2019, 17(3): 30.
SU Yuejiang, CHEN Xianlong, WU Dexing. Using big data in the third transportation comprehensive survey in Guangzhou [J]. Urban Transport of China, 2019, 17(3): 30.
- [35] WEISBERG H F. The total survey error approach: A guide to the new science of survey research [M]. Chicago: University Of Chicago Press, 2005.
- [36] 胡永恺, 宋璐, 张健, 等. 基于手机信令数据的交通OD提取方法改进[J]. 交通信息与安全, 2015, 33(5): 84.
HU Yongkai. SONG Lu. ZHANG Jian. *et al.* An improved method of extracting OD information based on mobile phone data[J]. Journal of transportation information and safety, 2015, 33(5): 84.
- [37] 王德, 王灿, 谢栋灿, 等. 基于手机信令数据的上海市不同等级商业中心商圈的比较——以南京东路、五角场、鞍山路为例[J]. 城市规划学刊, 2015(3): 50.
WANG De, WANG Can, XIE Dongcan, *et al.* Comparison of retail trade areas of retail centers with different hierarchical levels: a case study of east Nanjing road, Wujiaochang, Anshan road in Shanghai [J]. Urban Planning Forum, 2015 (3): 50.
- [38] 张天然. 基于手机信令数据的上海市域职住空间分析[J]. 城市交通, 2016, 14(1): 15.
ZHANG Tianran. Job-housing spatial distribution analysis in Shanghai metropolitan area based on cellular signaling data[J]. Urban Transport of China, 2016, 14(1): 15.
- [39] GREENE W. Discrete choice modeling [M]//MILLS T C, PATTERSON K. Palgrave Handbook of Econometrics: Volume 2: Applied Econometrics. London: Palgrave Macmillan UK, 2009:473-556.