

手机信令数据的出行测度准确性分析 ——基于与居民出行调查数据的比较

王 德¹, 韩滨鹂², 张天然³, 游智敏¹, 李心恬¹

(1. 同济大学建筑与城市规划学院, 上海 200092; 2. 成都市青白江区委社会工作部, 成都 610300;

3. 上海市城市规划设计研究院, 上海 200040)

摘 要:传统的居民出行调查是城市交通规划的基础数据,是城市居民出行行为的最权威记录。随着信息技术发展,以手机信令数据为代表的位置大数据,在时空行为研究领域展现出巨大的潜力。然而,手机信令数据能否取代出行调查数据却存在较大的争议。焦点在于:手机信令数据的出行准确性难以评估,其识别结果在多大程度上可信、能否较为准确地测度居民出行特征等问题有待回答。论文利用上海居民出行调查数据及时间相近的手机信令数据,比较两类数据在出行强度、出行时间分布、出行空间分布、出行目的等出行特征的差异,并分析归纳两类数据产生偏差的程度及原因。研究揭示手机信令数据在描述出行人口人均出行次数及街道以上尺度的出行OD分布等方面准确度较高,在出行时耗、通勤出行比例方面准确度较低。研究认为手机信令数据产生偏差的原因主要由手机信令数据位置记录不完整等数据质量、停留时间阈值、基站密度不均与空间阈值的影响,以及职住地识别偏差引起。研究认为可通过样本筛选、数据链修复、识别方法改进等手段提高手机信令数据识别的精度。

关键词:出行指标;出行链;出行距离;偏差;校核;上海

信息与通信技术的快速发展,特别是手机信令数据(mobile signaling data, MSD)等位置大数据的涌现,为居民出行活动研究提供了丰富的数据资源^[1-4]。相较于传统的出行调查数据,手机信令数据为研究人员提供了大规模、高覆盖率、长时段的轨迹数据,有利于深入研究个体行为规律,同时避免了传统调查的主观偏差^[5-6]。然而,手机信令数据也存在一些固有限制:作为通信运营商运营过程中的副产物,手机信令数据需经筛选和转化为居民的出行数据,这意味着不同参数和算法可能导致相异的识别结果;手机信令数据的空间精度与基站的分布有关,通常在500~2000 m之间,不能获取用户准确位置,并可能存在连接“非最近基站”误差;手机信令数据时间记录间隔通常超过1 h,难以完整记录移动用户的连续位置;用户信息的匿名性导致手机信令数据无法提供个体社会经济属性、出行目标和

方式等方面的准确信息。因此,为了更好地应用手机信令数据,应进一步精确量化其数据误差,了解其在时空行为研究中的适用性,以及采取措施避免潜在误解,从而加强基于手机信令数据研究的可靠性和有效性。

已有对MSD结果准确性进行分析的研究大多利用居民出行调查数据(household travel survey, HTS)作为参考数据源。HTS一直被视为传统居民出行研究中最常见和最有效的数据之一,通过问卷调查、入户访谈、电话询问等方式获得,能够采集居民一段时间内的活动或出行信息^[7],并采集MSD所缺少的城市居民家庭属性、个人特征和出行活动属性等信息^[8]。Calabrese等^[9]利用波士顿MSD识别的出行次数与美国HTS数据进行了比较,是对MSD结果进行验证的较早尝试之一。此后,研究展开对两类数据源得到的出行OD(origin-destination)分

收稿日期:2023-08-17;修订日期:2024-03-01。

基金项目:国家自然科学基金项目(52378069)。[Foundation: National Natural Science Foundation of China, No. 52378069.]

第一作者简介:王德(1963—),男,教授,博士生导师,主要从事城市规划方法论、空间与行为、城市大数据、城市模型领域的教学与研究。E-mail: dewang@tongji.edu.cn

引用格式:王德,韩滨鹂,张天然,等.手机信令数据的出行测度准确性分析:基于与居民出行调查数据的比较[J].地理科学进展,2024,43(5):854-869. [Wang De, Han Binli, Zhang Tianran, et al. Accuracy analysis of mobile signaling data in measuring travel indices: Based on the comparison with household travel survey. Progress in Geography, 2024, 43(5): 854-869.] DOI: 10.18306/dlkxjz.2024.05.002

布、出行时间分布、出行距离等特征的比较。如Schneider等^[10]在巴黎和芝加哥对人们工作日的日常出行模式进行分析,发现出行调查数据和手机信令数据在揭示人们出行次数、目的地数量、出行链等日常出行模式特征上具有高度一致性;郭莉等^[11]对深圳出行OD量级、距离及时间差异研究,发现两类数据出行距离分布的一致性高;陈小鸿等^[12]利用MSD对居民出行调查扩样方法进行探索,同样得到MSD推算结果在空间、距离分布上具有较高可靠性。

然而,研究表明,MSD在识别短距离出行活动方面存在局限性。如López-Ospina等^[13]在对西班牙西部城市群的研究中发现,相较于HTS数据,MSD往往更容易忽视短距离或高速的出行活动;李明高等^[14]对枣庄市居民出行率、出行OD分布、出行时间、出行距离和时耗等特征进行对比分析,同样发现MSD对短距离出行的识别存在较大偏差;Alexander等^[15]在对美国出行调查MSD与HTS数据比较中,认为汇总统计单元是影响不同数据源人均出行次数、出行时耗、出行距离等指标的相关性的重要因素。

总体而言,一方面,相关研究仍较少且仅聚焦两类数据出行特征的比较层面,缺乏系统性的总结偏差产生的原因及其修正方法。这使我们难以对误差性质形成清晰的认识,从而影响后续研究中对数据准确性和可信度的判断。另一方面,现有研究多基于市级、区县等较大行政尺度的汇总单元对MSD和HTS数据进行整体对比,较少涉及基于更小尺度研究单元或基于个体出行的分析^[16-17]。个体出行信息是MSD在居民出行研究中的关键组成部分,是构建居民出行模型和推断行为规律的不可或缺的要素,迫切需要我们对其进行更深入的分析 and 校准,以更细致地挖掘MSD中潜在的一致性和差异,为规划和决策提供更准确的基础。

因此,本文从原始MSD及底层HTS数据出发,在居委会、街道、区县等三个空间尺度分析居民出行特征,全面系统地分析了两类数据在测度出行强度、出行时间分布、出行空间分布、出行目的等出行特征的异同,并探讨数据之间偏差产生的原因、大小以及校核方式。研究旨在更加全面地认知两类数据在居民出行研究方面的准确度和适用范围,还原居民出行的真实特征和规律,并提供修正方法,以支持手机大数据在城市空间与规划领域的科学性应用。

1 研究方法与数据来源

1.1 研究数据与区域

本文采用的MSD是上海移动2014年3月连续两周时间内基站记录到的信令数据,作为对比的是和MSD同一年份秋季的上海市HTS数据。虽然两类数据存在半年的区别,但一般认为春秋两季的出行特征类似,具备可比性。HTS数据一方面从上海市城乡建设和交通发展研究院《上海市第五次综合交通调查成果报告》中获得全市出行指标^[18-19],另一方面分别选取城市中心区(杨浦区)与郊区(松江区)的研究范围获取个体出行数据。上海市范围内共有3.7万个基站,且基站具有“中心城区密集、郊区稀疏”的分布特征,选择杨浦和松江区作为典型区,可以更好地观察基站密度对数据误差的影响。MSD和HTS的原始个体出行信息在居委会、街道、区县等三个空间尺度汇总后进行比较分析。出行调查数据的汇总分析与上海市城市规划设计研究院交通规划部门合作进行。

全市每天的原始MSD记录量约7亿条,共累计识别到2314万手机用户。按照出行调查对应的人口分类方式,识别到全市常住人口1319万。重点研究区域杨浦区和松江区居住人口分别为77万、114万。居民出行调查数据抽样率为0.9%,调查对象涵盖本地常住人口以及外来常住人口^[18]。其中,杨浦区和松江区调查样本分别为0.91万、1.32万人,经扩样后得到的样本分别为123万、148万人。

1.2 数据信息构成及可对比指标筛选

本文使用的MSD包括6个字段,分别为已匿名处理的移动终端唯一识别号(MSID)、精细度到秒的信令事件记录产生时间戳(TStamp)、基站的位置区编号(LAC)、基站小区编号(CellID)、信令数据事件类型(EventID,每个编号代表不同事件类型)、数据的有效性代码(Flag,0为有效数据),其中LAC字段和CellID字段可以组成小区基站的唯一编号(LACCell)。

HTS数据分为户基本信息、个人基本信息以及个人出行信息3个部分。户基本信息表包含家庭住址、家庭结构、家庭交通工具、牌照及购车意愿等信息;个人基本信息表包括年龄、性别、户籍、职业、职住地等信息;个人出行信息表详细记录了受访居民从凌晨2:00到次日凌晨2:00出行距离400 m以上的出行情况,包括出行的起止时间、起讫点位置、出行目的、出行交通工具、出行费用和同行人数等内容。

通过对两类数据信息构成的分析,本文选取手机信令数据识别方法较为成熟,且出行调查数据表单内包含的指标(表1),包括出行起讫点、出行发生率、出行时间、出行距离、出行OD、出行目的等进行比较。

1.3 分析指标定义与误差分析方法

基于以上可比指标的选择,本文讨论的出行特征要素包括出行强度、出行时间、出行空间分布、出行目的。出行强度方面,其核心的量化指标有:出行人口人均出行次数、出行比例、出行频次等。出行比例值的大小与人均出行次数有较强的关联性。出行时间方面,常用的指标有出行时耗、出发或到达时刻分布。出行空间分布方面,常用的描述和量化指标有出行距离分布、出行OD分布等。出行目的方面,本文只能选取是否通勤作为分类^[20-22],对两类数据出行目的进行比较。各项指标的具体定义如表2所示。

在误差分析校核方法上,本文通过底层数据的反向逐级分解,找到误差源头。从数据采集、数据清洗和识别等环节对信令数据质量、出行链识别、

职住地识别三方面偏差进行分析。在校核方法上,一是以出行调查数据为基准进行校核,二是通过对识别方法的改进、样本质量的控制等方式对手机信令数据进行修正。

1.4 手机信令数据处理

(1) 数据清洗。手机信令数据在产生的过程中不免会产生一些噪音数据或异常数据^[23],为保证研究结果的可信度,需要对手机信令数据异常数据进行处理。处理方法一是去除记录质量较差的用户(记录数据少于3条/d);二是处理重复记录数据、乒乓切换数据、漂移数据等噪音数据。

(2) 出行链识别。上海市居民出行调查中对出行的定义为:利用有路名的街道或公路,完成一次有目的的活动,出行400 m以上或步行5 min以上^[18],本文采用相同定义对出行链进行筛选。目前较为主流的出行链识别方法是时空间聚类法,又称DB-SCAN算法^[24]。该算法的思路是通过时间阈值和空间阈值识别用户的停留,并且在两次不同地方停留之间识别出一次出行。这种时空间聚类的识别方法关键在于时间阈值和空间阈值的设定,空间阈值和时间阈值越小,识别的出行次数越多。已有研究对这两个阈值的取值一般源于出行调查中对出行的定义或经验数值^[25]。张维^[26]在确定空间阈值为500 m的情况下,通过计算认为取值下限为600 s才

表1 数据信息构成及可比较指标的选择
Tab.1 Composition of data and selection of comparable indices

指标	手机信令数据	居民出行调查	可比性判断
个体信息			
家庭结构	—	○	×
交通工具拥有	—	○	×
居住地	○	○	√
就业地	○	○	√
年龄	—	○	×
性别	—	○	×
职业	—	○	×
户籍类型	—	○	×
出行信息			
起讫点位置	○	○	√
起讫点时间	○	○	√
出行距离	○	○	√
出行时耗	○	○	√
出行目的	○	○	√
出行方式	—	○	×
同行人数	—	○	×
出行费用	—	○	×

注:○表示手机信令数据可识别的信息、出行调查包含的信息;—表示不包含相关信息;√表示具有可比性;×表示无法进行比较。

表2 出行指标定义

Tab.2 Definition of travel indices		
出行特征	量化指标	指标定义
出行强度	人均出行次数	某地区或城市居民的单日人均出行次数,即出行次数和总人口的比值
	出行人口	某地区或城市居民当日有出行行为的人口数量
	出行人口出行次数	某地区或城市中所有人口的出行总量与出行人口总数的比值
出行时间	出行比例	是指某地区或城市中有出行的人口总数与该地区人口总数的比值
	出行时耗	指一次出行所耗费的时间
	出发时刻分布	按一天各个时段统计的每个时段出行量占当天出行总量的比例
出行空间	平均出行距离	某地区或城市居民的出行总距离与出行次数的比值
	出行距离分布	不同出行距离的出行次数占出行总次数的比例
	出行OD	出行首末点位置分布
出行目的	通勤出行	如果一次出行的两端停留点分别为家和工作地(学校)而无其他活动,则该次出行称为1次通勤出行

能把低速运动行为剔除出去,并将时间阈值取值1200 s。

在时间阈值层面,本文从手机信令数据自身的特征出发,以空间阈值分别取值100~2800 m,取值间隔100 m,共28种情况。在每种情况下,以时间阈值分别取值600~3200 s,取值间隔200 s,共13个节点。计算每种情况下不同节点出行次数识别的变化率。计算发现,随着时间阈值的增加,出行次数识别量不断平滑减少,但在时间阈值从1200 s增加到1400 s时,出行次数的识别量有相对更明显的下降。这说明在手机信令数据中,有更多居民的停留时长集中在1200~1400 s。因此,结合已有研究对时间阈值的取值经验,并考虑手机信令数据样本自身的特征与基站分布特点,取值为1200 s。

在空间阈值层面,本文出行链识别采用动态空间阈值,即空间阈值的选用与基站密度相关。选用 $K=6$ 的 K 扇面法作为动态阈值算法,在这种动态阈值算法下,中心城核心区域的空间阈值取值在500 m以下,近郊取值为1000~1500 m,使得在基站密度高的地方空间阈值低,反之阈值高。从而解决固定空间阈值不同区域基站不均匀带来的识别误差,得到更准确的识别结果。

(3) 职住地识别。现有较主流的职住地识别思路是认为个体在夜间经常所在的位置即为居住地,白天工作时段内经常所在的位置即为工作地^[27]。本文采用谢栋灿等^[28]提出的职住地识别算法识别职住地,夜间所取的时间段为20:00~6:00,白天所取的时间段为9:00~18:00,并将识别的职住地结果映射到空间分析单元。本文采用第六次人口普查的普查区单元作为分析空间单元。由于普查区的面积在市中心小、在郊区大,而基站的分布“中心城密集、郊区稀疏”,因此会出现部分中心城的普查区由于面积小而没有基站、远郊的普查区由于基站稀疏

而没有基站;与此相反,部分普查区的基站数量甚至高达100个以上。因此,需要进行基站数据与普查区数据的匹配。部分研究采用泰森多边形作为基站的服务范围,并且把按照普查区单元形心落入泰森多边形分配相对应的基站,这种方法一定程度上符合手机终端连接就近基站的原则,但存在基站之间的服务范围重叠的问题。本文利用基于高斯核函数的用户空间匹配算法,在这种分配函数下,相同面积的地块越远离基站,分配的比例就越低,而且距离越远衰减越快;两个地块与基站距离相同的情况下,地块面积越大,分配的比例就越小。通过以上方法把基站的数据分配到普查区单元,更符合手机信令数据收集特征以及数据自身特征。通过二者的匹配,既可以避免部分普查区没有基站的情况,也可以减少误差。

通过以上方法,共识别出1371万个具有稳定居住地和工作地的用户,样本量达到了第六次人口普查(下称“六普”)上海市常住人口的57%左右。在街道单元层面,手机信令数据识别出的稳定居住人口与“六普”常住人口数之间的相关系数达到0.913,相关性显著($P<0.001$),表明手机信令数据识别的居住人口在空间分布上与实际情况吻合。

2 出行特征指标比较

2.1 出行强度

出行强度特征以出行发生比例和人均出行次数两个指标进行分析。杨浦区MSD计算的2014年3月连续两周日均出行比例和HTS一日出行发生比例分别为66%、80%,松江区分别为64%、78%(表3)。比较可知MSD测算得出的出行比例明显低于HTS,偏低值达17%。一般认为HTS反映的出行比例是真实的,这一结果说明MSD存在较多未被识别

表3 出行强度特征比较
Tab.3 Comparison of travel intensity indices

指标		全市		杨浦区		松江区	
		MSD	HTS	MSD	HTS	MSD	HTS
日均出行比例/%		66	—	66	80	64	78
人均出行次数	常住人口	1.52	2.16	1.56	1.87	1.47	1.89
	/(次/(人·d))	(-30%)		(-17%)		(-22%)	
	出行人口	2.31	—	2.33	2.31	2.24	2.36
				(+1%)		(-5%)	

注:括号中的百分数表示手机信令数据计算得到的人均出行次数相对出行调查数据的偏差。

出的出行人口。

具体分析人均出行次数的精度。杨浦区和松江区MSD测算出的常住人口分别偏低17%、22%。究其原因,是由于MSD出行人口识别率较低,导致计算得到总人口人均出行次数偏少。MSD识别的杨浦区出行人口人均出行次数为2.33次/(人·d);松江区出行人口人均出行次数为2.24次/(人·d)。与HTS数据得到的杨浦区2.31次/(人·d)、松江区2.36次/(人·d)相比,两类数据已十分接近,偏差在5%以内。可见,采用出行人口计算的人均出行指标进行比较更具合理性。

2.2 出行时间

采用出行时耗和出行时刻两个指标进行出行时间特征比较。

MSD和HTS显示,杨浦区平均出行时耗分别为88.0、34.1 min/次,松江区为93.3、27.2 min/次,均显示手机信令数据的出行时耗被严重夸大。在出行时耗的频度分布中,出行调查中杨浦区和松江区均呈现60、90、120 min出行比例的小波峰,而10 min以内的出行极少(图1)。由于HTS出行时耗是通过受访者回忆进行记录的,受人们主观记忆影

响,10 min以内的出行容易被忽略。相比之下,MSD出行时耗1 h以内的出行比例明显低于HTS,而1 h以上出行比例明显高于HTS。究其原因,原始信令数据中周期性位置更新的信令事件占比高,即较多数据的位置记录更新的时间间隔为1 h或2 h,导致了出行时耗在60、120 min处会出现小高峰,也使得低于1 h的短时间出行时耗被夸大。

在出发时刻分布方面,HTS显示杨浦区平均出发时刻有明显的早晚出行高峰,而MSD日间峰谷差异和早晚高峰差异均不明显。HTS早、晚高峰出行时段出行量占全日出行总量比重分别为16.02%、10.17%,而MSD仅为7.37%、7.89%(图1c)。松江区出行时刻分布特征与杨浦区类似,MSD呈现的出发时刻在一天内分布均匀。本文认为MSD在出发时刻方面的偏差,也主要来源于低精度的位置记录时间颗粒,平均1~2 h的位置更新时间间隔导致出发时刻分布的均匀化。

2.3 出行距离与出行OD

在出行空间特征上,本文比较平均出行距离与出行始末点分布差异。

采用出行首末点之间的直线距离进行计算,

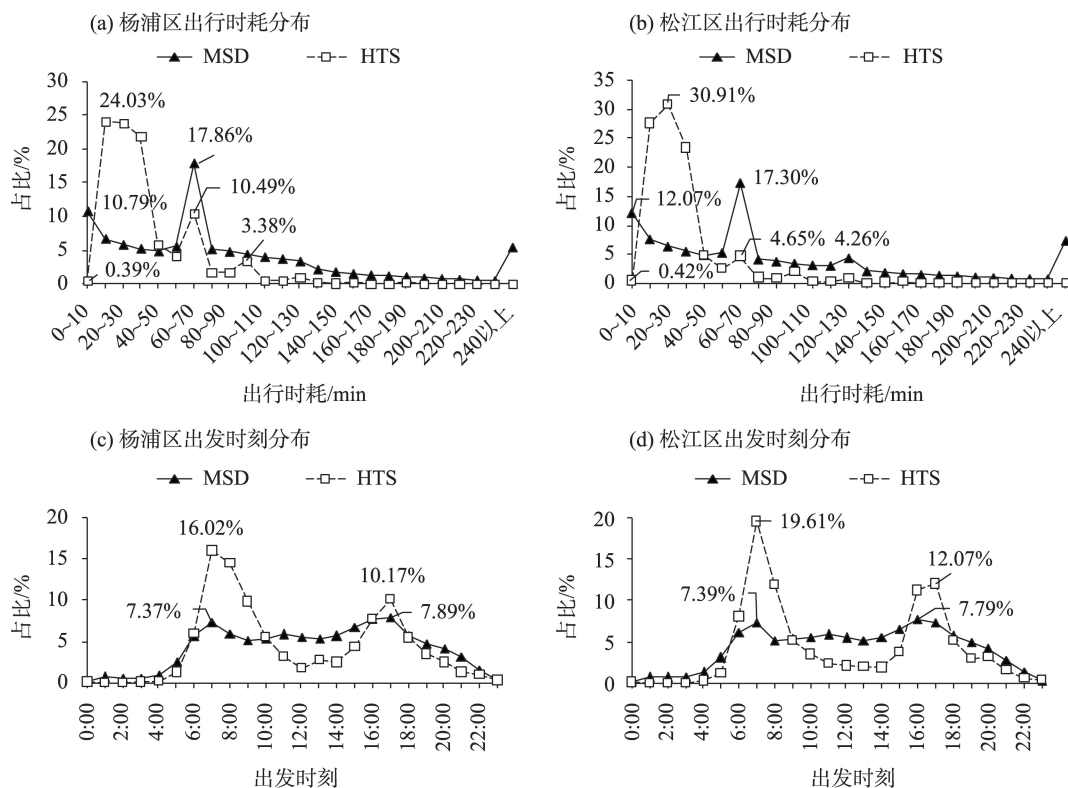


图1 出行时耗、出发时刻分布对比

Fig.1 Comparison of travel time consumption and departure time

HTS显示杨浦区、松江区的平均出行距离分别为4.44、4.59 km/次, MSD同指标分别为4.56、5.81 km/次, 手机信令数据识别的平均出行距离较长。分析不同出行距离的占比分布可知, HTS显示杨浦区出行距离1 km以内的出行占比为37.70%, 而MSD占比仅为21.82%, 偏差大小为15.88%; 松江区1 km以内的出行, MSD和HTS的占比分别为7.32%、25.59%, 偏差为18.27%(图2), 说明MSD中较多短距离的出行被忽略。这是由于识别过程中采用空间阈值使邻近基站间的出行无法识别, 且手机信令数据无法判断未跨越基站的出行而造成的。同时, 杨浦区基站间距在500~1500 m之间, 而松江区间距在1000~2000 m之间。较低的基站密度使得MSD在松江区的空间精度更差, 识别出的出行距离比实际情况更长。

进一步比较OD分布的差异, 发现两类数据在街道以上尺度的OD分布吻合度高。在跨区出行上, 可以看出杨浦区、松江区两类数据的出行始末点空间分布基本吻合(图3)。以区为单元对OD分

布进行相关性检验, 相关系数分别高达0.997、0.999, 相关性显著($P<0.001$)。以街道为单元再次进行检验, 相关性依旧显著($P<0.001$), 相关系数仍达0.8以上。但在居委会单元层级, 相关性出现明显降低, 相关系数未达0.2。对首次出行点的分布进行比较, 可以发现: 以街道为统计单元, MSD和HTS的首次出行起点分布相似度较高(图4), 杨浦区和松江区两类数据相关性均显著($P<0.001$), 相关系数分别为0.692、0.932。而以居委会为统计单元, 由于HTS在很多居委会无抽样样本, 导致首次出行起点空值较多, 使得两类数据首次出行点分布差异较大(图4), 这是出行OD分布在居委会层面产生较大差异的重要原因。

2.4 出行目的

居民出行调查数据将居民出行目的分为上班、上学、文化体育娱乐休闲、业务、接送人、回家、就医、日常生活8类, 而手机信令数据无法获取此类信息。因此, 本文主要将出行目的分为通勤出行(即上下班、上下学)和非通勤出行两类, 与出行调查数

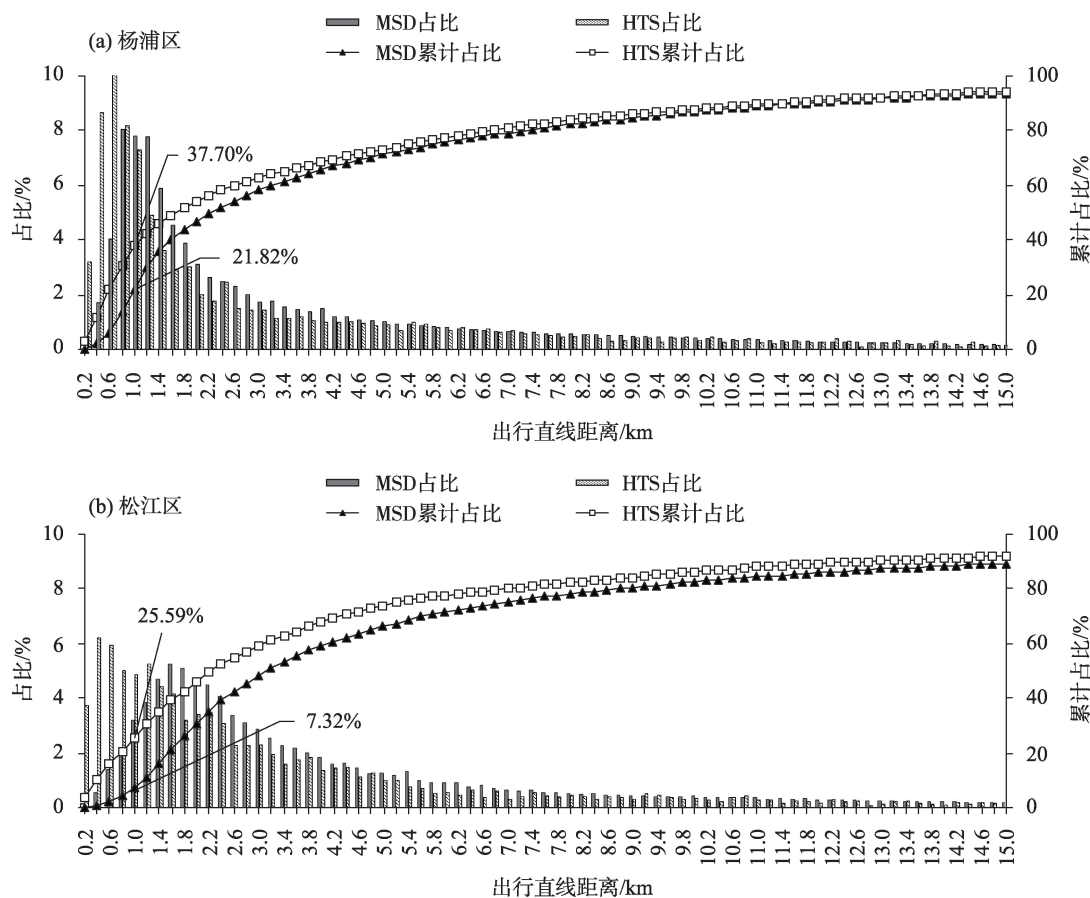


图2 出行直线距离分布比较

Fig.2 Comparison of travel straight-line distance

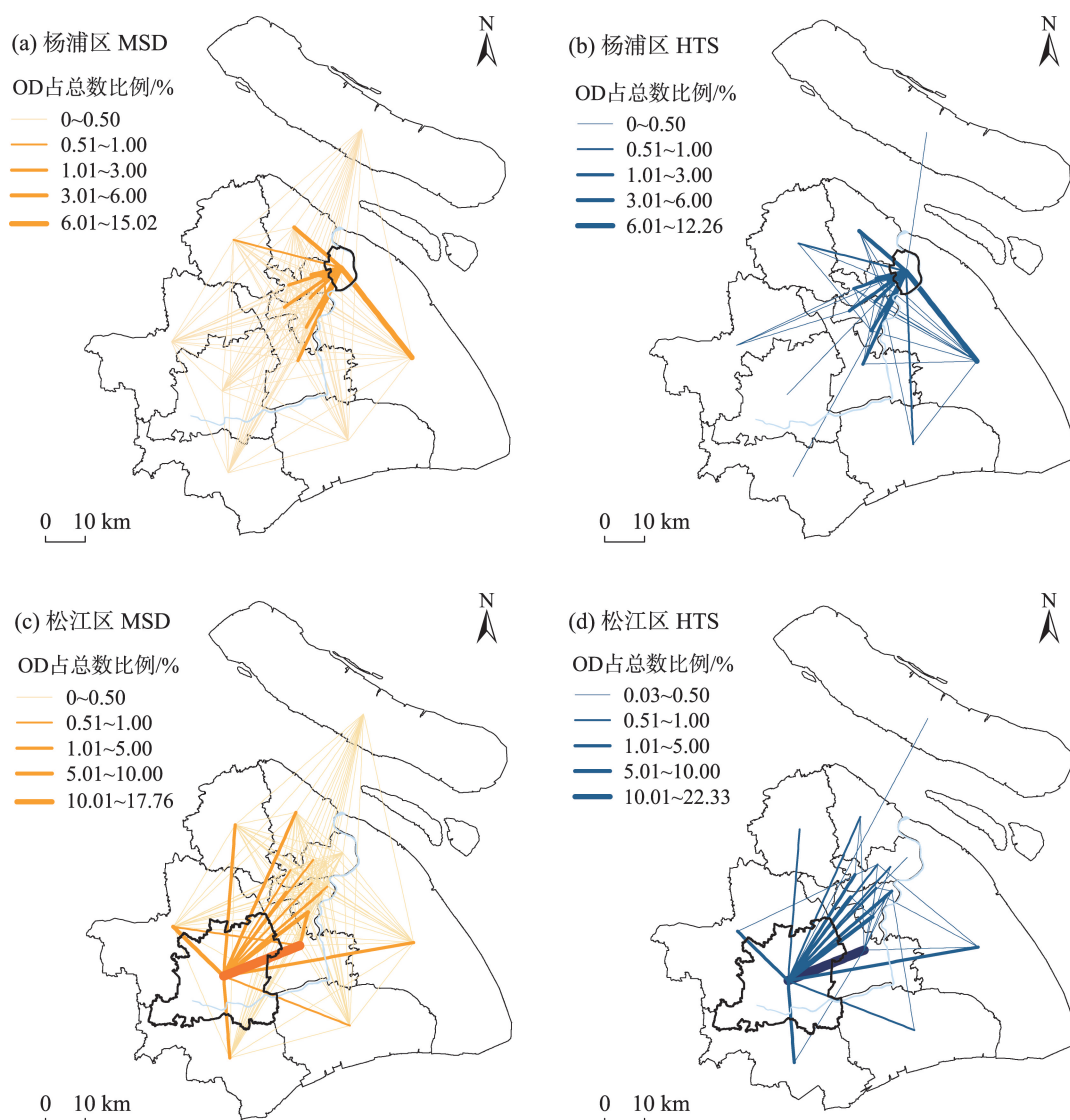


图3 跨区出行OD分布对比

Fig.3 Cross-district trip OD distribution

据进行比较分析。

手机信令数据记录的工作日日均水平上杨浦区、松江区通勤人口占出行总人口比例分别为31%、34%，出行调查同一指标分别为54%、69%，手机信令数据通勤人口比例显著低于出行调查数据。比较出行人口的通勤出行占比，HTS中杨浦区、松江区该指标分别为46%、55%，而MSD计算的通勤出行比重均仅为12%(图5)，显著低于HTS。具体比较HTS和MSD工作日通勤次数分布的特征(图6)，杨浦区、松江区每日通勤次数为1次的人数占比均高达68.9%，明显不符合实际情况，说明MSD中有大量的通勤出行并未被识别出来。

以全部出行人口为样本时，可能存在MSD通勤

人口识别比例低导致计算得到通勤出行比例偏低的情况。因此，为避免这一问题的影响，本文通过筛选，以杨浦区工作日的通勤人口为样本，重新测度出行目的构成。如表4所示，MSD识别的杨浦区通勤人口的狭义通勤出行比例、人均通勤次数分别为33%、0.85次，而HTS同指标高达90%、1.95次。MSD识别计算的通勤出行比例与通勤次数仍远远低于HTS。

3 数据偏差原因分析及其校核

经过反复的比较分析，本文认为MSD在数据采集、数据处理和识别、统计分析等环节都会有误

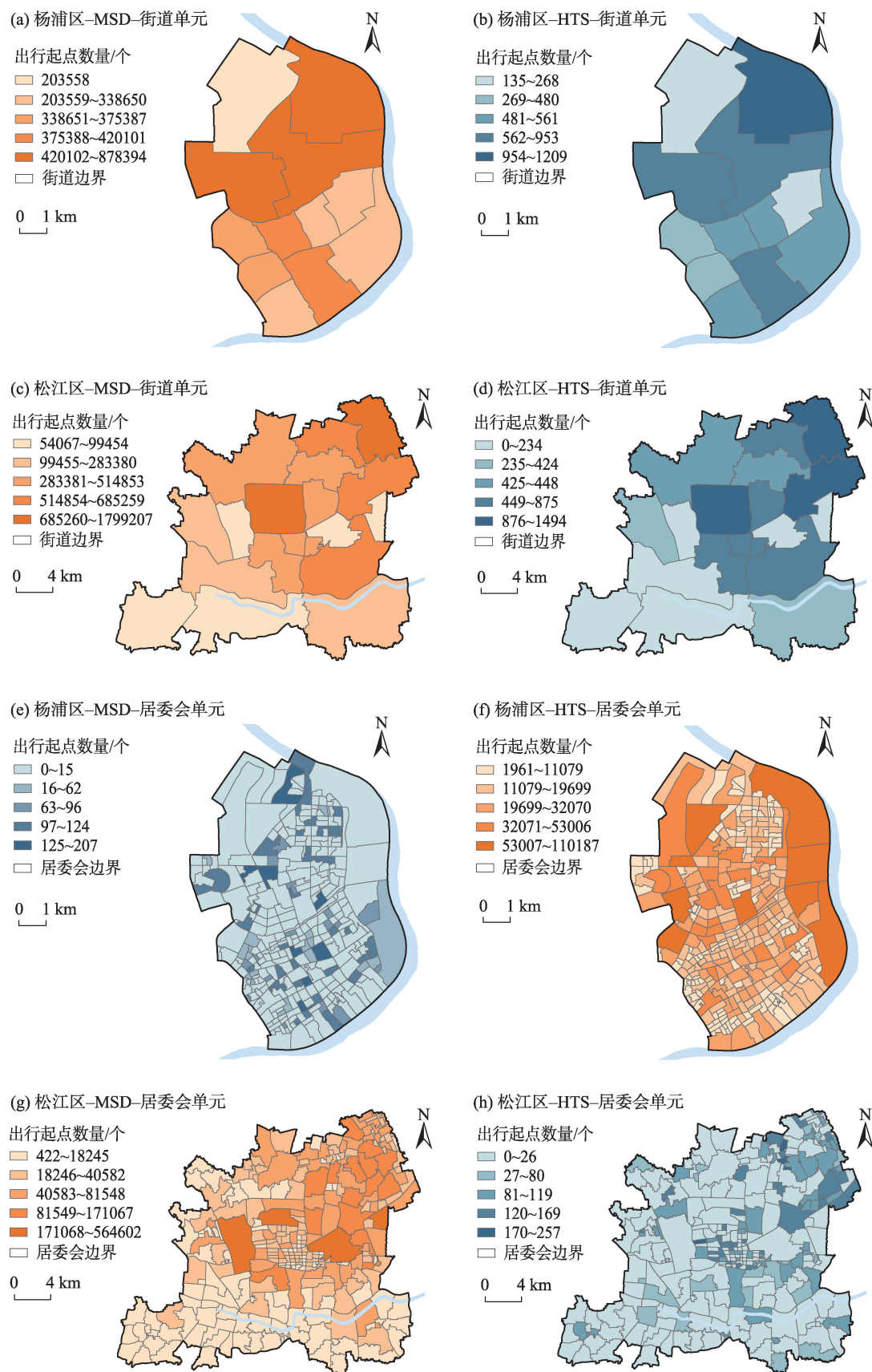


图4 首次出行空间O点分布对比

Fig.4 Comparison of the origin of the first travel

差产生。通过回溯MSD自身特征与数据识别方法,本文认为MSD产生偏差的原因,主要由信令数据质量、时间阈值、空间阈值选取和职住地识别偏差4个方面引起。

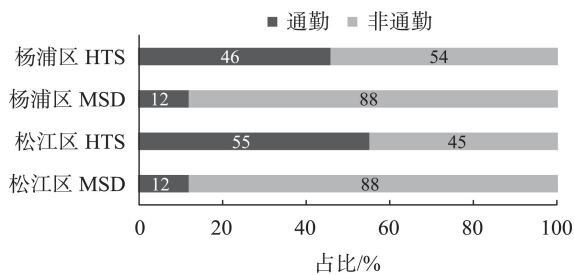


图5 杨浦区和松江区通勤出行对比
Fig.5 Comparison of travel purpose

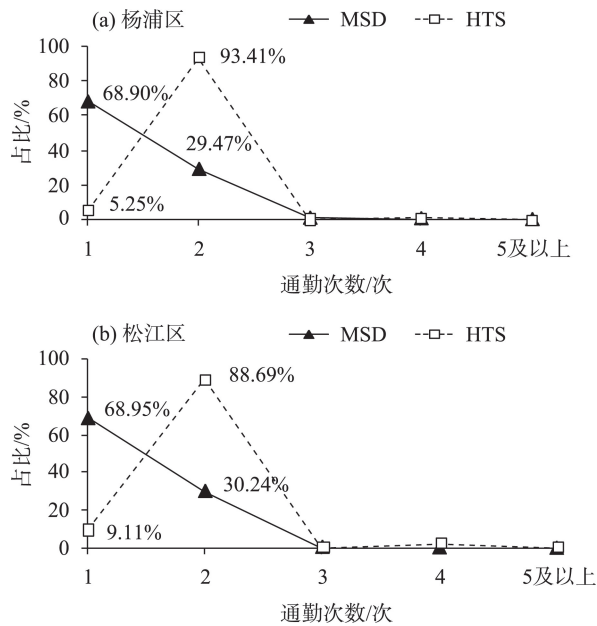


图6 杨浦区和松江区居民每日狭义通勤次数
Fig.6 Comparison of daily commutes in narrow sense

3.1 信令数据质量影响

在数据采集环节中,手机信令数据中存在部分低质量数据,导致用户的时空间行为难以被准确描述,进而导致出行特征的偏差。其影响可从记录时段的完整性、出行链完整性、位置记录时间间隔三个维度分析。

(1) 记录时段的完整性方面,大量存在的低质量信令数据的位置记录在时间上并非均匀连续,难以完整描述用户的时空间行为,从而导致出行特征的偏差。其中,原始信令数据记录时段不完全,对于出行次数的识别影响明显。以杨浦区工作日样本为例,对MSD样本日间活跃时间段6:00~24:00进行筛选,记录时段越多样本数量越少(表5)。由分析结果可知,日间记录时段越完整,MSD识别的人均出行次数越高(图7)。对MSD人均出行次数的校核修正可综合考虑信令数据记录质量和其样本数量的代表性,选择有12 h以上记录样本为高质量样本,MSD识别人均出行次数可由2.35次/(人·d)提高至2.58次/(人·d)(图7)。

(2) 出行链的完整性方面,常规对MSD的切割方法会破坏出行链的完整性。一般而言,为了提高MSD的计算效率,通常会采用将MSD在0:00进行切割的方法,但这种数据处理方式会导致出行的欠识别。当夜间在居住地位置记录数量较少时,数据的切割就使得在家驻留的情况难以清晰识别。以杨浦区通勤人口为例(表6),样本中全天末次出行的终点无法确定而以出行途中最后一个位置记录代替的样本量占比4%。但通常情况下,通勤人口末次出行的终点为家,即这部分样本的到家出行可能被误识别为其他类出行。此外,日出行终点为其他地点的情况占比28%,这部分样本存在以下可能:

表4 杨浦区通勤人口通勤出行比例比较

Tab.4 Percentage of commute of commuters in Yangpu District

数据类型	上班(上学)比例/%	下班(放学)比例/%	通勤出行比例/%	人均通勤次数/次	人均出行次数/次
MSD	21	12	33	0.85	2.61
HTS	46	45	90	1.95	2.16

表5 杨浦区高质量样本筛选

Tab.5 High-quality samples of Yangpu District

指标	样本1	样本2	样本3	样本4
日间活跃时段记录时段数	全样本	12 h及以上	14 h及以上	16 h及以上
样本数量/万人	642	498	445	370
占全样本比例/%	100.00	81.27	71.57	57.21

① 日出行链终点为第二居所;② 其他地点至家的出行发生于当天,但由于记录时间间隔问题欠识别;③ 其他地点至家的出行发生于次日。针对出行

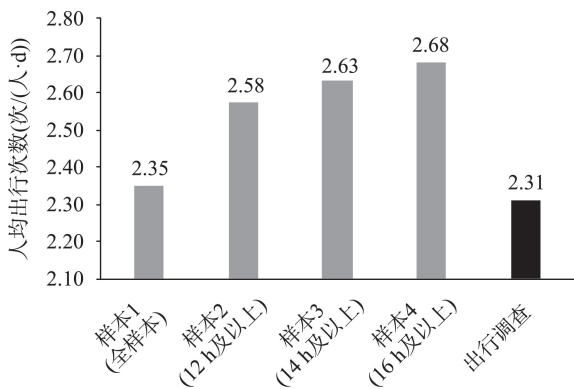


图7 杨浦区MSD各记录时段量出行强度及与HTS的比较
Fig.7 Comparison of travel intensity at various recording intervals between MSD and HTS in Yangpu District

表6 杨浦区通勤人口出行起讫点占比
Tab.6 Characteristics of starting points for commuting population in Yangpu District (%)

地点	日出行链起点	日出行链终点
家	74	55
工作地	18	13
其他地点	8	28
丢失出行起讫点	0	4

表7 出行链不完整情况修正方法
Tab.7 Correction methods for incomplete trip chains

	前日末点	后日首点	修正
可修正	其他地方O	家H	增加“O-H”出行
	家H	其他地方O	增加“H-O”出行
可信(第二住所)	其他地O	其他地O	确定可信
不确定	其他地M	其他地N	增加“M-N”出行

链不完整的情况,可通过前后两日出行链联合修复或者不对数据进行切割来校核,从而有效避免断点附近的出行欠识别的情况(表7)。通过对切割方式进行修正,本文发现由数据切割导致的欠识别出行比例达12%。

(3) 位置记录时间间隔长则是出行时间产生误差的主要原因。本文对MSD中的噪音数据进行清洗后,得到的人均记录量为23.5~25.1条/人,即平均打点间隔时间为1 h;其中周期性更新的信令事件占比23%,时间间隔为1 h或2 h。对杨浦区早通勤出发时刻进行分析,结果显示,平均出发时刻MSD相比HTS提前10 min,平均到达时刻MSD却延后57 min,平均通勤时间比HTS多67 min,偏差在1 h左右(图8)。由此可见,出行时耗的偏差较大程度是时间记录精度低引起的。但打点时间间隔长的偏差,是由MSD自身记录特征造成的,无法进行有效修正。

3.2 时间阈值选用造成的出行欠识别

在MSD处理识别环节中,出行链识别的时间阈值选用会造成两种形式的出行欠识别,从而造成出行量测算结果的偏差。

一是非往返出行途中的欠识别,即非往返出行途中驻留没有留下记录或出行中驻留有记录但记录时间间隔没有达到时间阈值,因而无法区分是驻留还是路过(图9a),导致实际的两段出行被误识别为一段出行,使得MSD出行量被低估。这类出行欠识别,可以以HTS为参照进行校核。根据HTS,杨浦区居民非往返出行中停留20 min以内的情况占出行总量的2.86%,其中无停留的情况占出行总量的1.22%。这类中途停留的出行主要以接送人为主,但MSD无法识别为两段出行。由此推算,本文中MSD采用20 min的时间阈值,会导致约3%以短

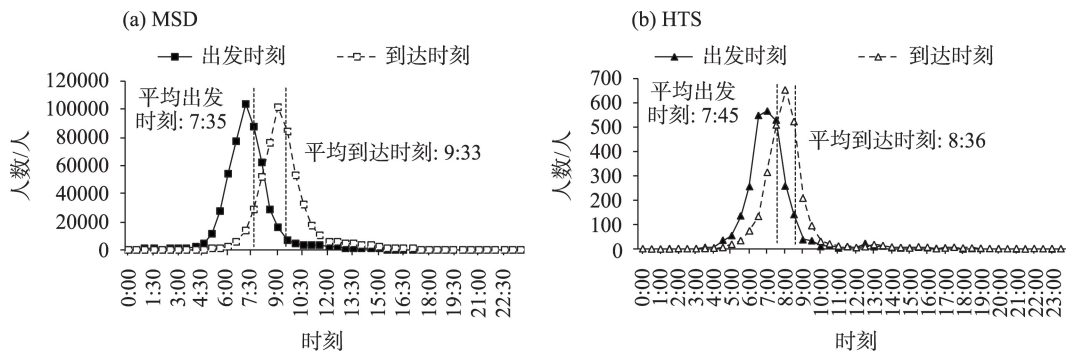


图8 两类数据早通勤出发及到达时刻比较
Fig.8 Departure and arrival times comparison of morning commutes

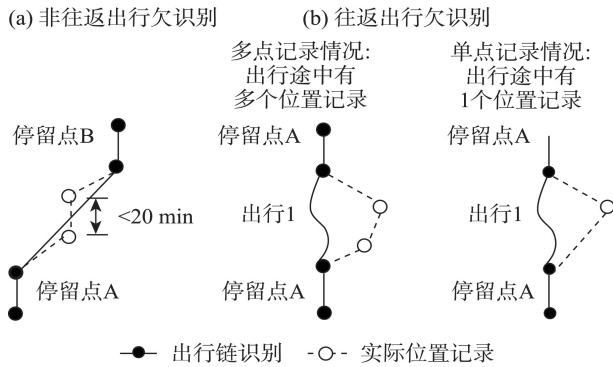


图9 出行欠识别示意图

Fig.9 Schematic diagram of trips not identified

时间停留为目的的出行被忽略,即人均出行次数比实际情况低0.07次/(人·d)。

二是起讫点相同的往返出行的欠识别,即出行途中有一或多次位置记录但由于停留时间短而无法被识别为驻留,如起讫点均为家且无长时间停留某点的散步行。按照HTS对出行的定义,上述情况均应记作2次出行,而在出行链识别中仅识别为1次出行(图9b),导致MSD出行量被低估。这类识别偏差可将MSD出行链中最远点作为目的地,对MSD出行量进行重新计算来校核。校核结果显示杨浦区居民首末点相同的出行占比5%,其中单点记录占比3%,多点记录占比2%,即由于起讫点相同的往返出行欠识别情况导致的人均出行次数缺失约0.13次/(人·d)。

3.3 基站密度与空间阈值选用造成的偏差

手机信令数据的空间精度依赖于基站密度分布,当用户处于基站服务范围重叠区内,即使实际位置没有发生改变,也有可能被周边若干个邻近基站所记录,又称“跳基站”现象。该现象由于记录时间间隔特征与正常数据没有差别,无法通过类似清洗“乒乓数据”的方式去除,从而造成出行量识别的偏差。因此,需要采用空间阈值作为“空间容差”,来防止邻近基站的“跳基站”现象被误识别为出行。但由于城市尺度下基站分布不均匀,中心城区基站分布密集,郊区基站分布稀疏,所有基站都使用固定空间阈值进行出行识别时,会造成中心城区部分短距离出行未被识别,而郊区“跳基站现象”被识别为出行的情况(图10)。因此,更理想的方法是每个基站的空间阈值都会根据周边基站密度与距离来动态地调整阈值大小,即采用“动态空间阈值”算法。本文中杨浦区基站密度较高,动态空间阈值

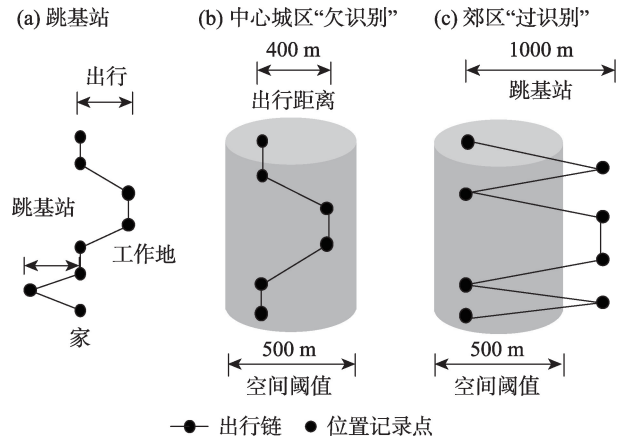


图10 固定空间阈值引起的偏差

Fig.10 Deviation caused by fixed spatial threshold

的选用集中在400~900 m之间;松江区基站密度较低,阈值集中在700~1800 m之间。

但即使采用动态空间阈值的方法进行改进,出行量的识别仍然会有误差,主要体现在以下两个方面:① 基站本身有一定信号覆盖范围,若用户位置在同一基站服务范围内移动,则无法被识别为出行。这部分偏差与手机信令数据的基站密度有关,难以直接进行校核修正。② 空间阈值可以很大程度防止识别过程中邻近基站间的“跳基站”现象而误识别为出行的情况,但却也造成了邻近基站间的短距离出行被忽略。由空间阈值设置造成的短距离出行量的偏差可以以HTS为依据进行校核。比较MSD与HTS短距离出行的人均出行次数差异(图11),可知MSD中杨浦区居民1 km以内短距离出行与松江区居民1.4 km以内短距离出行均存在明显欠识别的情况。

3.4 职住地识别偏差

在手机信令数据处理识别环节中,职住地识别偏差主要会导致出行目的中通勤或非通勤出行判别的误差,主要有两种形式:一是就业人口判别误差,二是职住地位置判别误差。具体来说,就业人口识别偏差表现为手机信令数据存在短距离通勤人口过识别现象。以杨浦区为例,MSD通勤距离1 km以内的通勤人口比例高于出行调查13.62个百分点(表8)。但这部分通勤人口被识别出通勤的天数大多数少于3 d,明显低于其他组别(图12)。选择部分样本观察其多日时空行为轨迹,发现MSD识别中可能将近家惯常活动点被误识别为就业地。这部分就业人口判别偏差可通过剔除短距离通勤样本进行校核。校核后,杨浦区手机信令数据全部通勤

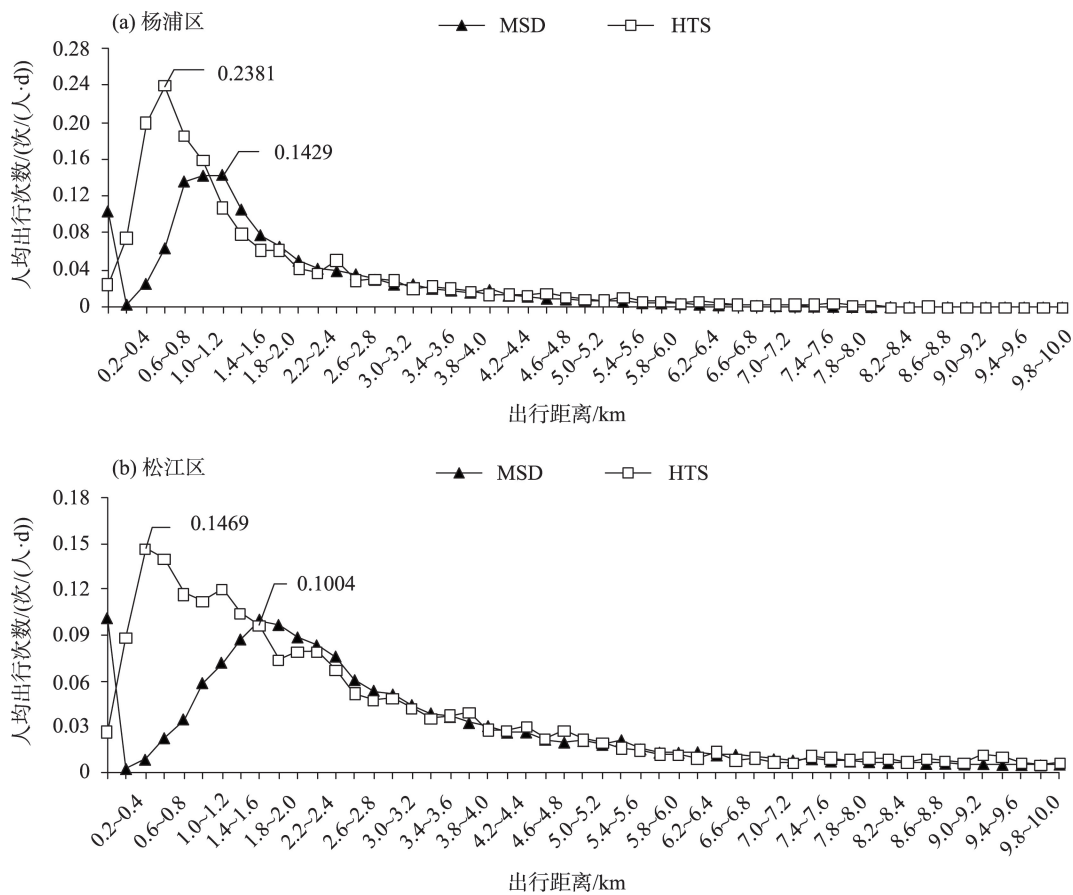


图 11 行政区区内出行距离(直线距离)分布比较
Fig.11 Comparison of travel straight-line distance within the districts

表 8 杨浦区各组别职住距离占比分布
Tab.8 Distribution of commuting distance by groups in Yangpu District (%)

数据类型	0~500 m	500~1000 m	1000~1500 m	1500~2000 m	2000 m 以上
MSD	13.01	17.30	8.03	5.27	56.39
HTS	7.84	8.85	7.48	5.15	70.68

样本的通勤出行占比由 33%提高至 35%。

此外,通过单一特定的基站点判断驻留点,会造成职住地位置判别误差。在 MSD 实际记录情况中,在居住地或工作地的驻留很可能被邻近的多个基站记录。因此,对职住地判别的误差,可采用动态空间阈值进行校核。以杨浦区为例,将全样本中通

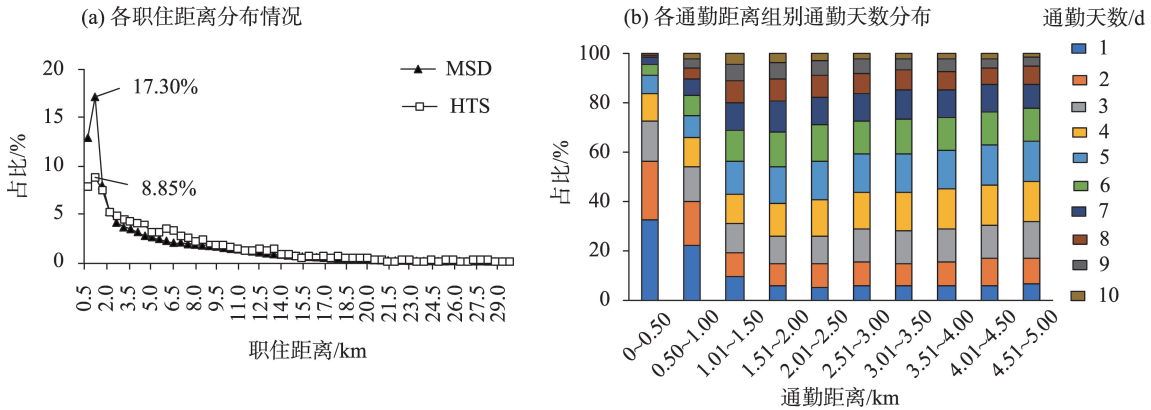


图 12 杨浦区居民职住距离分布及各通勤距离组别通勤天数
Fig.12 Commuting days of each commuting distance group in Yangpu District

勤距离1 km以下以及职住地距离小于动态空间阈值的样本进行剔除,校核后得到的通勤出行比例从35%提升为44%,平均通勤次数提高0.24次/(人·d)。

4 结论与讨论

4.1 结论

本文通过同一年份的手机信令数据与居民出行调查数据的比较分析,系统地探讨了手机信令数据测度居民出行特征的偏差大小,并提出了修正方法。比较结果显示:①在出行强度方面,手机信令数据出行人口识别上存在缺失,识别到的静止人口明显偏多,但手机信令数据出行人口的人均出行次数识别较为精确。②在出行时间方面,手机信令数据对出行谷峰识别能力较差,体现在手机信令数据的平均单次出行时耗较出行调查数据出行时耗高近1 h,出行量在一天中的分布更为均匀。③在出行空间特征,两类数据出行距离分布整体趋势一致,但手机信令数据在1 km以下的短距离出行识别比例偏低;两类数据出行OD分布的相关性随尺度减小而逐渐递减,在区县和街道空间尺度手机信令数据的可信度较强。④在出行目的特征的识别上,手机信令数据识别的通勤出行占总出行的比例显著低于出行调查数据,对短距离通勤人口的识别可信度较低。

本文认为手机信令数据偏差主要由4个方面因素造成:一是手机信令记录缺失与打点时间间隔影响。信令记录时段严重缺失的样本与打点稀疏问题会导致手机信令数据人均出行量偏低、出行时间偏长。因此,需要剔除手机信令数据中记录不完整的低质量样本,以及前后两日出行链联合来修正数据切割时间断面附近的出行。二是采用时间阈值对出行中的停留进行限定,必然会一定程度导致出行过程中的短时间驻留无法被识别,使得手机信令数据的出行量被低估。三是空间阈值选用以及基站密度的影响,会造成邻近基站间的短距离出行无法被识别。四是手机信令数据对就业人口以及居住就业地判断的不精准,会导致通勤出行目的比例偏低。

4.2 讨论

数据质量分析和可靠性验证是手机信令数据能够应用于居民出行特征研究的前提,本文有效补充了手机信令数据在居民出行研究方面的适用范

围。在研究内容上,本文发现手机信令数据能够较为精确地描述出行人口的人均出行量与长距离出行趋势特征、反映街道以上汇总尺度的出行空间分布格局,但在判定平均出行总时耗、通勤人口比例上存在较大偏差。同时,出行调查数据自身具有局限性^[29-30],如样本覆盖率有限^[11-12,31]及存在时空信息主观偏差^[7,32]等,也会导致两类数据对出行特征的识别产生偏差,研究中也发现出行调查也存在短距离出行漏报、午间工作地周边出行漏报等情况。总的来说,本文认为手机信令数据更适宜进行宏观尺度的定性描述与判断,而定量测算得到的数值结果准确性值得商榷。

在研究方法上,识别方法的运用对于偏差结果必定也会造成一定程度影响。本文已通过结合既有取值经验选择多个时间阈值进行测试,并采用动态空间阈值对识别方法进行改进,以尽可能减小识别方法可能引起的偏差。当然,本文仍存在进一步完善的空间:在偏差大小的定量分析准确性方面,本文未对部分误差进行更精确的测度。例如,手机信令数据打点记录时间间隔是引起出行时间误差的主要原因,但其引起的偏差大小难以准确衡量;短距离出行欠识别和短时间驻留欠识别引起的偏差存在交叠,难以进行测度,只能参照出行调查数据进行估算。在手机信令数据适用性方面,本文采用的2014年信令数据,部分定量结论无法直接套用于4G、5G的手机信令数据上,例如出行时间特征的分析准确度与基站记录的时间精度有关,随着4G基站记录时间间隔的缩短,出行时间识别的准确度必然也会得到提升。在出行链识别方法的多样性方面,本文未进行多类型识别方法的偏差比较,后续研究可对不同识别方法的偏差进行探索。

参考文献(References)

- [1] 柴彦威,申悦,肖作鹏,等. 时空间行为研究动态及其实践应用前景[J]. 地理科学进展, 2012, 31(6): 667-675. [Chai Yanwei, Shen Yue, Xiao Zuopeng, et al. Review for space-time behavior research: Theory frontiers and application in the future. Progress in Geography, 2012, 31(6): 667-675.]
- [2] 王德,谢栋灿,王灿,等. 个体时空行为的规律性与可预测性研究: 以上海市居民工作日活动为例[J]. 地理科学进展, 2021, 40(3): 433-440. [Wang De, Xie Dongcan, Wang Can, et al. Regularity and predictability of individual spatiotemporal behavior: A case study of residents

- working day activities in Shanghai. *Progress in Geography*, 2021, 40(3): 433-440.]
- [3] 钮心毅, 岳雨峰, 李凯克. 长三角城市群中心城市与周边城市的城际出行特征研究 [J]. *上海城市规划*, 2020(4): 1-8. [Niu Xinyi, Yue Yufeng, Li Kaike. Inter-city travel characteristics between central and surrounding cities in the Yangtze River Delta urban agglomerations. *Shanghai Urban Planning Review*, 2020(4): 1-8.]
- [4] 丁亮, 钮心毅, 宋小冬. 利用手机数据识别上海中心城的通勤区 [J]. *城市规划*, 2015, 39(9): 100-106. [Ding Liang, Niu Xinyi, Song Xiaodong. Identifying the commuting area of shanghai central city using mobile phone data. *City Planning Review*, 2015, 39(9): 100-106.]
- [5] Chen C, Ma J T, Susilo Y, et al. The promises of big data and small data for travel behavior (aka human mobility) analysis [J]. *Transportation Research Part C: Emerging Technologies*, 2016, 68: 285-299.
- [6] Bonnetain L, Furno A, El Faouzi N-E, et al. TRANSIT: Fine-grained human mobility trajectory inference at scale with mobile network signaling data [J]. *Transportation Research Part C: Emerging Technologies*, 2021, 130: 103257. doi: 10.1016/j.trc.2021.103257.
- [7] Arentze T, Timmermans H, Hofman F. Creating synthetic household populations: Problems and approach [J]. *Transportation Research Record: Journal of the Transportation Research Board*, 2007, 2014(1): 85-91.
- [8] Stopher P R, Greaves S P. Household travel surveys: where are we going? [J]. *Transportation Research Part A: Policy and Practice*, 2007, 41(5): 367-381.
- [9] Calabrese F, Diao M, Di Lorenzo G, et al. Understanding individual mobility patterns from urban sensing data: A mobile phone trace example [J]. *Transportation Research Part C: Emerging Technologies*, 2013, 26: 301-313.
- [10] Schneider C M, Belik V, Couronné T, et al. Unravelling daily human mobility motifs [J]. *Journal of the Royal Society Interface*, 2013, 10(84): 20130246. doi: 10.1098/rsif.2013.0246.
- [11] 郭莉, 周军, 梁宇豪. 手机信令数据与传统居民出行调查的对比及扩样方法研究: 以深圳为例 [C]// 中国城市规划学会城市交通规划学术委员会. 交通治理与空间重塑: 2020年中国城市交通规划年会论文集. 北京, 2020: 1890-1897. [Guo Li, Zhou Jun, et al. Study on sample expansion methods of mobile signaling data compared with traditional resident travel survey: Based on Shenzhen // Academic Committee of Urban Transportation Planning of China Urban Planning Society. 2020 China annual meeting of urban transportation planning. Beijing, China, 2020: 1890-1897.]
- [12] 陈小鸿, 陈先龙, 李彩霞, 等. 基于手机信令数据的居民出行调查扩样模型 [J]. *同济大学学报(自然科学版)*, 2021, 49(1): 86-96. [Chen Xiaohong, Chen Xianlong, Li Caixia, et al. Sample expansion model of household travel survey using cellphone data. *Journal of Tongji University (Natural Science)*, 2021, 49(1): 86-96.]
- [13] López-Ospina H, Cortés C E, Pérez J, et al. A maximum entropy optimization model for origin-destination trip matrix estimation with fuzzy entropic parameters [J]. *Transportmetrica A: Transport Science*, 2022, 18(3): 963-1000.
- [14] 李明高, 杨冠华, 刘剑锋, 等. 手机数据在城市居民出行特征分析中的应用实践 [C]// 中国城市规划学会城市交通规划学术委员会. 创新驱动与智慧发展: 2018年中国城市交通规划年会论文集. 青岛, 2018: 1084-1091. [Li Minggao, Yang Guanhua, Liu Jianfeng, et al. Application of mobile data in the analysis of urban residents' travel characteristics // Academic Committee of Urban Transportation Planning of China Urban Planning Society. 2018 China annual meeting of urban transportation planning. Qingdao, China, 2018: 1084-1091.]
- [15] Alexander L, Jiang S, Murga M, et al. Origin - destination trips by purpose and time of day inferred from mobile phone data [J]. *Transportation Research Part C: Emerging Technologies*, 2015, 58: 240-250.
- [16] 周洋, 杨超. 基于时空聚类算法的轨迹停驻点识别研究 [J]. *交通运输系统工程与信息*, 2018, 18(4): 88-95. [Zhou Yang, Yang Chao. Anchors identification in trajectory based on temporospatial clustering algorithm. *Journal of Transportation Systems Engineering and Information Technology*, 2018, 18(4): 88-95.]
- [17] 杨飞, 姜海航, 姚振兴, 等. 基于手机信令数据的出行端点识别效果评估 [J]. *西南交通大学学报*, 2021, 56(5): 928-936. [Yang Fei, Jiang Haihang, Yao Zhenxing, et al. Evaluation of activity location recognition using cellular signaling data. *Journal of Southwest Jiaotong University*, 2021, 56(5): 928-936.]
- [18] 上海市城乡建设和交通发展研究院. 上海市第五次综合交通调查主要成果 [J]. *交通与运输*, 2015, 31(6): 15-18. [Shanghai urban and rural construction and Transportation Development Research Institute. The main results of the fifth comprehensive traffic survey in Shanghai. *Traffic & Transportation*, 2015, 31(6): 15-18.]
- [19] 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): 35-42, 50. [Li Na, Dong Zhiguo, Xue Meigen, et al. The practice of new survey technology and methodology in the 5th Shanghai comprehensive transportation survey. *Urban*

- Transport of China, 2016, 14(2): 35-42, 50.]
- [20] 张艳, 柴彦威. 基于居住区比较的北京城市通勤研究 [J]. 地理研究, 2009, 28(5): 1327-1340. [Zhang Yan, Chai Yanwei. Characteristics of commuting pattern in Beijing: Based on the comparison of different urban residential areas. Geographical Research, 2009, 28(5): 1327-1340.]
- [21] 张文佳, 柴彦威. 基于家庭的居民出行需求理论与验证模型 [J]. 地理学报, 2008, 63(12): 1246-1256. [Zhang Wenjia, Chai Yanwei. Theories and confirmed model of urban resident's travel demand: Considering intra-household interaction. Acta Geographica Sinica, 2008, 63(12): 1246-1256.]
- [22] 周素红, 闫小培. 基于居民通勤行为分析的城市空间解读: 以广州市典型街区为案例 [J]. 地理学报, 2006, 61(2): 179-189. [Zhou Suhong, Yan Xiaopei. The impact of commuters' travel pattern on urban structure: A case study in some typical communities in Guangzhou. Acta Geographica Sinica, 2006, 61(2): 179-189.]
- [23] 方家, 王德, 谢栋灿, 等. 上海顾村公园樱花节大客流特征及预警研究: 基于手机信令数据的探索 [J]. 城市规划, 2016, 40(6): 43-51. [Fang Jia, Wang De, Xie Dongcan, et al. Research on dynamic change and early warning of large tourist flow based on mobile signal data analysis: A case study of Gucun park sakura festival in Shanghai. City Planning Review, 2016, 40(6): 43-51.]
- [24] 肖志权, 张子民, 毛曦, 等. 基于手机信令数据居民出行链提取算法 [J]. 北京测绘, 2019, 33(10): 1192-1195. [Xiao Zhiquan, Zhang Zimin, Mao Xi, et al. Resident travel chain extraction algorithm based on mobile signaling data. Beijing Surveying and Mapping, 2019, 33(10): 1192-1195.]
- [25] 邹戴晓, 过秀成, 樊钧, 等. 基于手机信令数据的苏州市域职住空间和通勤出行分析 [J]. 现代城市研究, 2018(12): 16-21. [Zou Daixiao, Guo Xiucheng, Fan Jun, et al. Analysis of living space distribution and commute patterns in Suzhou based on cellular signaling data. Modern Urban Research, 2018(12): 16-21.]
- [26] 张维. 基于手机定位数据的城市居民出行特征提取方法研究 [D]. 南京: 东南大学, 2015. [Zhang Wei. Study on methods of extracting resident trip characteristics based on cellphone location data. Nanjing, China: Southeast University, 2015.]
- [27] 王蓓, 王良, 刘艳华, 等. 基于手机信令数据的北京市职住空间分布格局及匹配特征 [J]. 地理科学进展, 2020, 39(12): 2028-2042. [Wang Bei, Wang Liang, Liu Yanhua, et al. Characteristics of jobs-housing spatial distribution in Beijing based on mobile phone signaling data. Progress in Geography, 2020, 39(12): 2028-2042.]
- [28] 谢栋灿, 王德, 钟炜菁, 等. 上海市建成环境的评价与分析: 基于手机信令数据的探索 [J]. 城市规划, 2018, 42(10): 97-108, 120. [Xie Dongcan, Wang De, Zhong Weijing, et al. Assessment and analysis on built environment of Shanghai: An exploration based on mobile phone signaling data. City Planning Review, 2018, 42(10): 97-108, 120.]
- [29] Wolf J L. Using GPS data loggers to replace travel diaries in the collection of travel data [D]. Georgia, USA: Georgia Institute of Technology, 2000.
- [30] 王德, 李丹, 傅英姿. 基于手机信令数据的上海市不同住宅区居民就业空间研究 [J]. 地理学报, 2020, 75(8): 1585-1602. [Wang De, Li Dan, Fu Yingzi. Employment space of residential quarters in Shanghai: An exploration based on mobile signaling data. Acta Geographica Sinica, 2020, 75(8): 1585-1602.]
- [31] 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. Xi'an, China: Chang'an University, 2017.]
- [32] 杨超, 朱荣荣, 涂然. 基于智能手机调查数据的居民出行活动特征分析 [J]. 交通信息与安全, 2015, 33(6): 25-32. [Yang Chao, Zhu Rongrong, Tu Ran. Analysis of the travel characteristics of residents in Shanghai using the itinerary data collected from smartphones. Journal of Transport Information and Safety, 2015, 33(6): 25-32.]

Accuracy analysis of mobile signaling data in measuring travel indices: Based on the comparison with household travel survey

WANG De¹, HAN Binli², ZHANG Tianran³, YOU Zhimin¹, LI Xintian¹

(1. College of Architecture and Urban Planning, Tongji University, Shanghai 200092, China;

2. Social Work Department of the CPC Chengdu City Qingbaijiang District Committee, Chengdu 610300, China;

3. Shanghai Urban Planning & Design Research Institute, Shanghai 200040, China)

Abstract: Traditional household travel survey provides the basic data for urban traffic planning and the most authoritative record of urban residents' travel behaviors. With the development of information technology, position data represented by mobile signaling data have shown great potential in the field of temporal-spatial behavior research. However, whether the mobile signaling data can replace the travel survey data is controversial because the travel data accuracy of mobile signaling data is difficult to assess. How credible the identification results are and whether the travel characteristics of residents can be accurately measured remains to be answered. In this study, the travel survey data of Shanghai residents and mobile signaling data with similar dates were used to compare the differences in intensity, time, spatial distribution, purpose, and other characteristics of travels. This study also analyzed the degree of deviation and summarized the causes of errors. The research revealed that the mobile signaling data are more accurate in describing the per capita travel times of the travel population and the travel origin-destination (OD) distribution at the scale above the street, but less accurate in describing the travel time distribution and travel purposes. Our research suggests that the main reasons for the deviation in mobile phone data are the impact of incomplete location records on data quality, the impact of retention time thresholds, spatial thresholds, and home and workplace identification. Due to the above systematic errors, mobile signaling data cannot replace household travel survey data. Nevertheless, the accuracy of mobile signaling data analysis can be improved by sample screening, trip chain data recovery, and improvement of identification methods.

Keywords: travel indices; trip chain; trip distance; deviation; correction; Shanghai