

从大规模短期规则采样的手机定位数据 中识别居民职住地

许宁^{1,2} 尹凌² 胡金星²

1 中南大学地球科学与信息物理学院,湖南长沙,410000
2 中国科学院深圳先进技术研究院,广东深圳,518055

摘要:使用大规模手机定位数据获取居民职住地分布是大数据趋势下城市研究的新兴技术。然而,现有的研究主要使用了长期不规则稀疏采样的手机通话数据,对短期规则采样的手机定位数据缺乏尝试。基于大规模短期规则采样的手机定位数据,提出了一种居民职住地识别的方法。这是首次从大规模短期规则采样的手机定位数据中进行居民职住地识别的尝试,并对识别结果进行了较全面的验证。该研究成果为职住平衡等相关城市问题研究探讨了一种新型大规模数据源的可行性,在低成本大幅度提高相关研究的样本代表性和分析结果可靠性上具有重要意义。

关键词:手机定位数据;时空数据挖掘;职住平衡;通勤距离;深圳市

中图法分类号:P208

文献标志码:A

从微观个体的角度,准确掌握居民居住地与工作地的空间分布是众多城市研究与城市规划的基础。基于一定规模的居民职住地分布样本,可分析职住分离特征,揭示职住空间不匹配的影响机制,提出引导职住平衡的建议^[1-7];也可分析通勤流的时空特征,研究缓解交通拥堵的空间规划与引导政策^[8-11]。获取城市居民的职住地分布数据成为上述研究的关键环节。现有研究中,问卷调查是获取职住地分布数据的主要手段。然而,大规模问卷调查需耗费大量的人力和财力,调查执行时间长,被调查者负担重。另外,问卷调查的样本量大都集中在几百至几千个之间^[7]。相较于国内大中型城市的人口总量,大规模增加研究样本量对提高分析结果的广泛代表性和可靠性具有重要意义^[7,12]。

近年来,通过分析大规模手机定位数据得到具有广泛代表性的居民活动特征已经成为国际前沿的研究热点^[13-15]。由于手机定位数据的覆盖人群广,数据实时生成,采样周期长,能够大规模客观地长期记录城市居民活动的时空特征,并且被调查者零负担,因此学界认为其具有持续有效的支撑大规模居民活动模式分析的潜力^[16-17]。使用

大规模手机定位数据获取居民职住地分布,一方面可以支持大规模样本的职住平衡等相关实证研究,增加研究人群的完备性和多样性,减少因为样本量少和样本偏差而造成的认识偏差;另一方面,实时更新的数据也可以支持连续观察和快速更新的研究。

虽然基于手机定位数据分析居民活动模式的研究取得快速的进展,但基于手机定位数据的城市居民职住地识别研究仍然具有较大的完善空间。① 现有研究中缺乏对短期规则采样的手机定位数据的分析。目前广泛使用的手机通话位置数据是一种长期的不规则稀疏采样数据。研究表明,很多用户的手机使用事件(如打电话或者发短信)往往具有随机性和短期爆发性,导致其反映出的个体的时空特征不完备,甚至具有误导性^[18-19]。相比之下,规则连续采样数据,即手机跟踪定位数据(如每间隔1h记录一次手机用户的位置)则可以较完整地反映个体的时空特征,是研究居民活动和出行模式的理想数据源。但是这种数据源通常是短期的,如手机运营商通常只保存几天的跟踪定位数据。从理论上讲,这两类数据各有优缺点。如果能够补充对短期规则采样数据的分析,

收稿日期:2014-01-27

项目来源:国家自然科学基金资助项目(41301440);深圳市战略新兴产业发展专项资金资助项目(JCYJ20130401170306842)。

第一作者:许宁,硕士生。现从事时空数据挖掘研究。E-mail: ning.xu@siat.ac.cn

通讯作者:尹凌,博士,副研究员。E-mail: yinling@siat.ac.cn

将会加深对手机定位数据这种新型数据源的系统认识,使其能够被科学推广。^②对基于手机定位数据的职住地识别方法缺乏较全面的验证。现有文献中,基于手机通话位置数据提出的职住地识别方法主要分为两类:第一类是分别在自定义的居住和工作时间段内,根据手机用户通话事件的频繁程度识别居住地和工作地点^[20-22]。第二类是根据手机通话的位置分布,通过空间聚类等分析识别个体的重要活动地点,并结合活动时间来判定居住地与工作地点^[23]。与问卷调查相比,手机定位数据只具有时空记录,而缺乏活动的属性。因此,验证活动地点的属性识别结果是这种新型数据源被广泛使用的关键步骤。虽然现有文献中提出了上述职住地识别方法,但是其对识别结果的验证都相对片面和薄弱,有的仅对比了少量样本的真实职住地信息^[23],有的仅比较了交通出行调查数据^[22]。为了证明基于海量手机定位数据的职住地识别结果的可靠性和适用性,必须进行更加全面深入的结果验证。

本文使用深圳市约 1 000 万手机用户在某一工作日规则采样的手机定位数据,结合土地利用类型,提出了一种识别居民职住地的方法,并且通过对比深圳市人口普查数据和交通出行调查数据,多方位讨论了识别出的职住地分布的准确度以及其对通勤特征研究的支持度,为相关城市问题和城市规划研究提供了一种新型大规模数据源的可行性探讨。

1 研究区域与数据源

1.1 研究区域概况

本研究以深圳市为研究区域。深圳市总面积为 1 953 km²,下辖六个行政区:福田区、罗湖区、南山区、盐田区、宝安区、龙岗区(2007 年以后新划分了 4 个功能区)。其中,福田、罗湖和南山区原属于特区,即“关内”地区,宝安和龙岗地区原属于非特区,即“关外”地区。2009 年以后,特区区域逐渐扩大至整个辖区。2012 年,深圳市常住人口数量达到 1 054.74 万人,其中非户籍人口 778.85 万人,人口密度在全国大中城市中最高,人口年均增长速度最快,年龄结构最年轻,适龄劳动人口的平均文化程度最低,城市流动人口总数在全国最多。“关外”的经济发展主要依赖劳动密集型企业、低档制造加工业;“关内”则以高技术产业为主。关内外的产业结构差异导致流动人口在关外密集,而深圳市户籍人口集中于“关内”区域。

1.2 数据源

本研究使用约 1 000 万手机用户规则采样的手机定位数据,时间为 2012 年某一工作日。原始记录包括用户的匿名 ID、时间戳以及所在手机基站的经纬度坐标。每位用户的采样时间间隔约为 1 h。

数据预处理步骤包括去除原始数据中信息不完整的数据(如缺少定位时间或坐标的数据),将时间戳转换为北京时间以及合并每个用户 ID 连续 24 h 的记录。最终将原始数据整理为每个手机用户约为 1 h 一次的基站位置记录。

2 居民职住地识别方法

本研究提出的从短期规则采样的手机定位数据中识别居住职住地的方法由三个主要步骤组成(见图 1):^①通过构建活动序列将个体约 1 h 一次的基站位置记录转换为个体活动序列;^②从个体活动序列中识别出居住地与工作地点所对应的基站位置;^③由于基站服务范围从几百 m²到几十 km²不等,如果以基站位置代表职住地位置,在进行不同空间单元统计和通勤距离计算时可能导致一定的偏差。因此,本研究结合土地利用类型,在基站服务范围内模拟职住地的位置。所述方法的具体步骤如下。

1) 构建个体活动序列。本研究使用“活动地点-活动开始时间-活动时长”数据模型构建个体活动序列。其中活动地点代表一个手机用户停留过的基站位置,活动时长代表在某活动地点的持续停留时间。

2) 识别职住地所对应的基站位置。基于每位用户的活动序列,筛选用户在工作时段累计停留时间超过阈值 Δt_w 的基站作为该用户的工作地所对应的基站;筛选用户在夜间睡眠时段内累计停留时间超过阈值 Δt_n 的基站作为该用户的居住地所对应的基站。其中,根据深圳市的企事业单位工作时段和深圳市居民生活习惯,本研究设定工作时段为 09:00~18:00,夜间睡眠时段为 00:00~06:00。对于 Δt_w 和 Δt_n 的选择,本研究参考了龙瀛等^[12]使用的北京居民平均工作时间以及深圳居民的作息规律,分别设置 Δt_w 和 Δt_n 为 6 h 和 4 h。

3) 模拟职住地位置。Voronoi 多边形是表达基站服务范围的简单有效的方法^[13-14,24],因此,本研究建立了 Voronoi 多边形表达基站服务范围。在每个基站服务范围内,使用一种简单的随机点模拟法生成职住地的坐标^[21,24]。职住地的分布

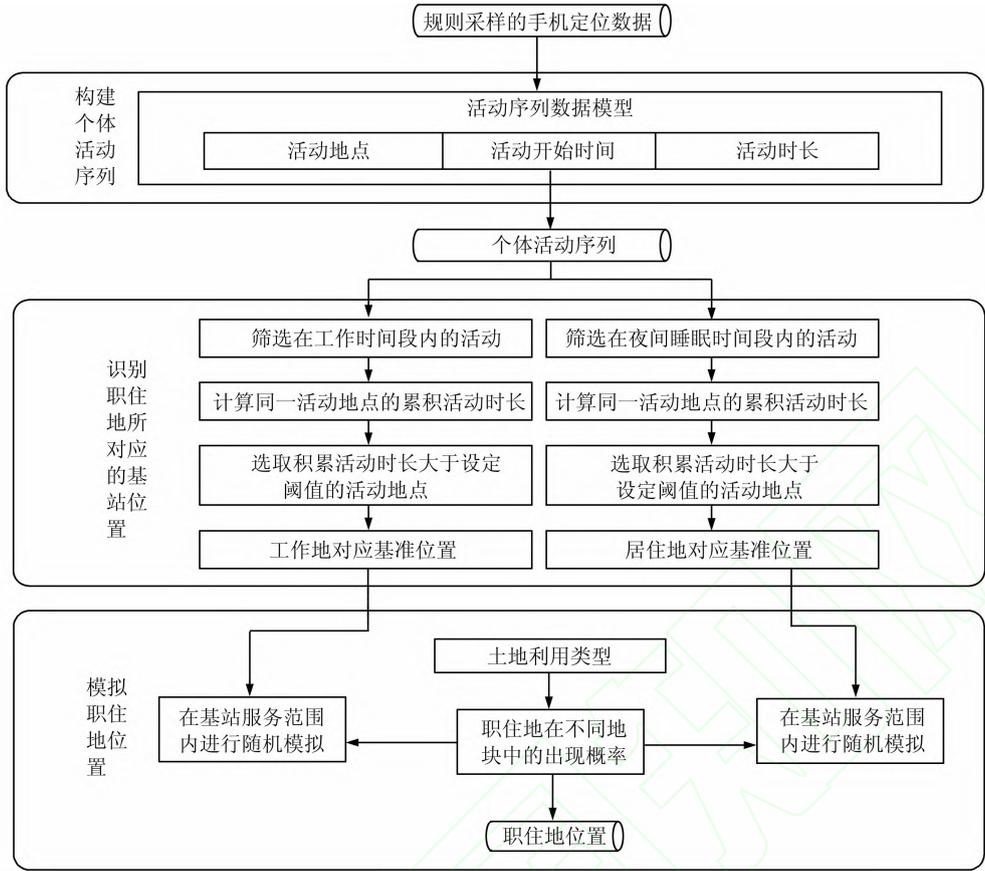


图1 从规则采样的手机定位数据中识别职住地的方法

Fig. 1 Identifying Home-work Locations from Regularly Sampled Mobile Phone Tracking Data

依赖于基站服务范围内的用地结构构成。不同的用地类型具有与之相应的居住地或工作地出现的概率。一个居住地或工作地在一个基站服务范围内某一地块中出现的概率 P_i 为:

$$P_i = (W_i \frac{A_i}{\sum A_i}) / (\sum W_i \frac{A_i}{\sum A_i}) \quad (1)$$

式中, i 为地块编号; W_i 为地块 i 的用地类型概率权重; A_i 为地块 i 的面积。参考威伟等^[25]提出的人口-昼夜-土地利用模型, 根据深圳市 2011 年土地利用数据, 将地块划分为三类: ① 包括滩涂、水

面、道路建设等用地的非居住和非工作用地, 在该区域内, 设定居住地和工作地的出现概率为 0。② 包括城镇住宅用地和农村住宅用地的主要居住用地和主要工作用地, 在该区域内, 设定居住地的出现概率为 90%, 工作地的出现概率为 50%。③ 包括其他用地类型的非主要居住用地和主要工作用地, 在该区域内, 设定居住地的出现概率为 10%, 工作地的出现概率为 50%。图 2 为一个居住地模拟的示例。

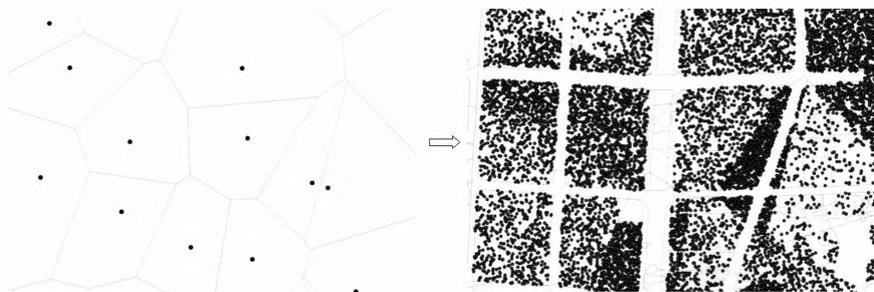


图2 根据土地利用类型在基站服务范围内模拟居住地位置

Fig. 2 An Example of the Home Location Simulation Within Cell Phone Tower Service Areas Based on Land Use Structures

3 结果分析与验证

为了评价上述方法识别出的居民职住地分布的准确性以及对通勤特征分析的支持度,本研究采取了从宏观整体特征到微观局部特征分层次评价的思路。首先将识别出的居住人口和就业人口分布与深圳市 2010 年第六次人口普查数据进行了街道级别的对比;然后计算了识别出的全市居民平均通勤距离,与深圳市 2010 年交通出行调查结果进行了交通分析小区(traffic analysis zone, TAZ)级别的对比;最后选取了典型的居住区和就业区,分析识别出的通勤特征与实际特征在微观小区级别的吻合情况。

3.1 职住人口分布的识别结果及验证

从约 1 000 万用户于 2012 年某一工作日规则采样的手机定位数据中,本研究识别出 9 269 561 名具有工作地的用户和 8 164 108 名具有居住地的用户。其中,2 793 411 名用户同时具有职住地,约为 2012 年深圳市常住人口的 26.48%。这 279 万用户中,职住同地(在本研究中,即为在同一基站服务范围内)的用户占 61%。

以深圳市 55 个街道为空间单元,计算得到本研究识别出的居住和就业人口分布与 2010 年第六次人口普查数据的 Spearman 相关系数分别为 0.95 和 0.90(街道为我国人口普查数据发布的基本统计单元。此外,此处使用 Spearman 相关系数是因为人口分布不是正态分布,不宜使用 Pearson 相关系数)。较高的相关系数说明识别出的居住和就业人口分布与人口普查结果在整体上有较好的一致性。

基于每个街道区域的人口比例,本研究进一步进行了相对误差分析。其中,

$$\text{相对误差} = 100\% \times \left| \frac{\text{识别人口比例} - \text{普查人口比例}}{\text{普查人口比例}} \right|$$

计算发现,居住人口比例和就业人口比例的平均相对误差分别为 22.83% 和 34.53%。经过对每一个街道区域的审查,总结出误差产生的原因如下:① 手机数据与人口普查数据在时间上间隔约 15 个月。在这期间,由于深圳市的快速发展,部分区域的实际人口分布产生了较大的变化。如坪地街道在这 15 个月内进行了大量的工业区扩张,使得该区域的居住和就业人口均大幅度增加。② 在客流量大的风景名胜区内,游客在工作时间段内的长时间停留会造成本方法对就业地的误判。如南澳街道和梅沙街道以旅游业为主,一天的游

容量最高可达 6 万人。本方法识别出就业人口中有较大比例实则为游客,进而造成了该地区识别出的就业人口偏多。③ 深圳市“关外”以工业区为主,工业区内 24 h 轮班制度对于职住地的识别也造成一定的误差。如“关外”的松岗街道有大量的富士康工业区,其职工 24 h 轮班制的作息规律对职住地的识别造成干扰。④ 本研究使用的手机定位数据来自于单一运营商,该运营商用户在空间上可能存在不均匀分布。⑤ 居民在某一天的活动存在随机性。这种随机性也可能对职住地识别的结果产生影响。本研究去除掉明显由于上述①~③种原因造成误差的 17 个街道区域后,发现识别出的居住与就业人口的相对误差分别减小为 12.93% 和 17.11%。

3.2 居民通勤距离的识别结果及验证

本研究使用基于深圳市路网的最短路径距离,计算出深圳市职住分离用户(即职住地不在同一个基站服务范围之内)的平均通勤距离为 5.53 km。该分析结果与深圳市 2010 年交通出行调查结果(5.40 km)高度一致,说明本研究所提出的职住地识别方法在支撑通勤距离分析上具有较高的可靠性。

为了说明本研究识别出的通勤特征在空间分布上的可靠性,进一步在 TAZ 尺度上分析深圳市居民外出平均通勤距离分布。如图 3 所示,深圳市平均通勤距离由福田区与罗湖区的中心向外递增至宝安区和龙岗区中部后又递减。这种分布与深圳市的实际情况相吻合。深圳市的老城区(福田区与罗湖区)提供了充足的就业机会,因此,该区域的职住平衡性较强。而在宝安区和龙岗区有着大量的劳动密集型工业区和居住成本低廉的城中村,这两者的混合分布也满足了大量外来务工者就业和居住的就近选择需求。而位于这两块职住相对平衡的区域之间的部分则通勤距离相对较长。

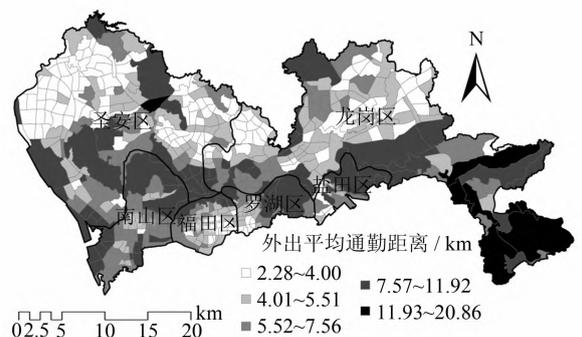


图 3 识别出的深圳市居民外出平均通勤距离
Fig. 3 Identified Average Commuting Distances of Shenzhen Residents in Each TAZ

3.3 典型居住区和就业区的通勤特征分析

本研究选取了深圳市位置相邻的三个典型居住区(皇岗村、梅林村和香蜜湖)和三个典型就业区(富士康工业园区、华强北商贸中心和高新科技园)进行通勤特征分析,以验证职住地的识别结果在微观小区级别的可靠性。一方面,分析了通勤流分布(见图4),另一方面,选取了外出通勤比例、外入通勤比例、外出平均通勤距离和外入平均通勤距离四个特征值描述小区的通勤特点(见表1)。其中,为了方便直观对比,通勤距离采用直线距离。外出和外入通勤比例的计算方法如下:

$$\text{外出通勤比例} = \frac{\text{居住样本} - \text{居住且在本区域就业样本}}{\text{居住样本}} \quad (2)$$

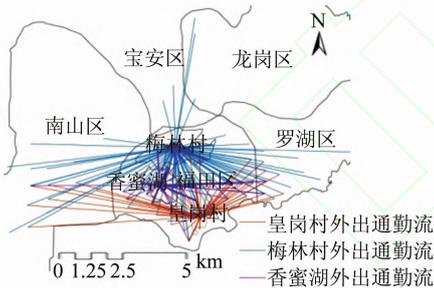
$$\text{外入通勤比例} = \frac{\text{就业样本} - \text{居住且在本区域就业样本}}{\text{就业样本}} \quad (3)$$

在居住区方面,三个区域的通勤特征均符合其居住人口特征。梅林村为典型的公务员和白领聚集居住区,其居民工作地分布较广,外出通勤距离较长,外出通勤比例较高。皇岗村为典型的城中村,外来务工人口密集,个体经营者比例也较高,工作地较为集中地分布在皇岗村周围,通勤距离短。香蜜湖既是高档住宅小区,又是度假村,外出、外入比例都较高,通勤距离较长,其居民的工作地离散分布于中心城区。

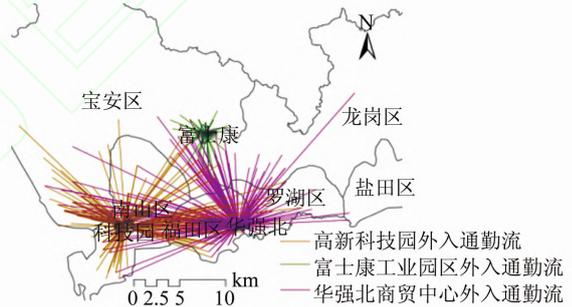
表1 三个典型居住区和就业区的通勤特点

Tab.1 Commuting Characteristics of Three Typical Residence Communities and Employment Zones

	特点	外出通勤距离/km	外入通勤距离/km	外入通勤比例	外出通勤比例
居住区	梅林村 公务员和白领居住集聚地	4.19	3.70	0.18	0.30
	香蜜湖 高档住宅小区和度假村	4.24	5.81	0.51	0.33
	皇岗村 流动人口集聚的城中村	2.70	2.59	0.14	0.23
就业区	富士康工业园 劳动密集型工业园	1.46	1.74	0.72	0.66
	华强北商贸中心 深圳最大的商业中心	2.00	5.06	0.72	0.14
	高新科技园 高新技术产业园	3.32	5.79	0.73	0.36



(a) 三个典型居住区的外出通勤流分布



(b) 三个典型就业区的外入通勤流分布

图4 深圳市典型居住区与就业区的通勤流分布

Fig. 4 Commuting Trips of Three Typical Residence Communities and Employment Zones

三个典型就业区的通勤特征也均符合其各自的特征。三个区域的外出通勤距离比外入通勤距离小,外出通勤比例比外入通勤比例小,这符合就业区的特点。富士康作为劳动密集型工业园区的代表,其职工通常居住在园区内或者附近的城中村,通勤距离十分短。华强北作为深圳市最大的商业区,就业辐射范围广泛,外入通勤距离较长。高新科技园作为深圳市最大的高新技术产业区,就业辐射范围基本覆盖了以科教产业为主的南山区,外入通勤距离较长。

以上典型案例分析表明,在微观小区级别上,

本研究的职住地识别结果对通勤特征分析具有较好的支持度。

4 结语

本研究首次从大规模短期规则采样的手机定位数据中进行居民职住地识别的尝试,并对识别结果进行了较全面的验证。分析结果说明,本研究提出的职住地识别方法对通勤特征分析具有较好的支持度,为相关城市问题和城市规划研究探讨了一种新型大规模数据源的可行性,在低成本、

大幅度提高相关研究的样本代表性和分析结果可靠性上具有重要意义。

本研究为初步探索,未来还有大量的研究工作有待完成:① 根据已识别出的产生误差的主要因素,对职住地识别方法进行改进,进一步提高准确性;② 比较短期规则采样和长期不规则采样的手机定位数据在特定分析结果上的不同,增加对手机定位数据这种新型数据源的更加全面系统的认识;③ 利用手机定位数据覆盖人群广泛的优势,关注大城市中特定群体(如城中村居民)的职住分离等活动特征,研究不同群体的生活状态;④ 研究如何保护手机定位数据中的用户隐私,使得该数据源能够被更加广泛的使用;⑤ 研究如何结合缺乏社会属性的手机定位数据和社会属性丰富的问卷调查等其他数据源,使其相互补充,在城市研究中发挥更大的作用。

参 考 文 献

- [1] Buliung R N, Kanaroglou P S. Commute Minimization in the Greater Toronto Area: Applying a Modified Excess Commute[J]. *Journal of Transport Geography*, 2002, 10(3): 177-186
- [2] Zhou Suhong, Yan Xiaopei. Characteristics of Jobs-housing and Organization in Guangzhou[J]. *Scientia Geographica Sinica*, 2005, 25(6): 664-670(周素红,闫小培. 城市居住-就业空间特征和组织模式——以广州市为例[J]. *地理科学*, 2005, 25(6): 664-670)
- [3] Ma K R, Banister D. Excess Commuting: A Critical Review[J]. *Transport Reviews*, 2006, 26(6): 749-767
- [4] Liu Zhilin, Zhang Yan, Chai Yanwei. Home-Work Separation in the Context of Institutional and Spatial Transformation in Urban China: Evidence from Beijing Household Survey Data [J]. *Urban Studies*, 2009, 16(9): 110-117(刘志林,张艳,柴彦威. 中国大城市职住分离现象及其特征:以北京市为例[J]. *城市发展研究*, 2009, 16(9): 110-117)
- [5] Meng Bin. The Spatial Organization of the Separation Between Jobs and Residential Locations in Beijing[J]. *Acta Geographica Sinica*, 2009, 64(12): 1 457-1 466(孟斌. 北京城市居民职住分离的空间组织特征[J]. *地理学报*, 2009, 64(12): 1 457-1 466)
- [6] Du Hao, Zhen Feng. Medium Sized City Employment and Residential Space Characters and Expansion Model: Qingyuan Example [J]. *Planners*, 2011, 27(5):99-105(杜浩,甄峰. 中等城市就业与居住空间组织特征及拓展模式——以广东省清远市为例[J]. *规划师*, 2011, 27(5): 99-105)
- [7] Zhou Jiangping, Chen Xiaojian, Huang Wei, et al. Jobs-Housing Balance and Commute Efficiency in Cities of Central and Western China: A Case Study of Xi'an[J]. *Acta Geographica Sinica*, 2013, 68(10): 1 316-1 330(周江评,陈晓键,黄伟,等. 中国中西部大城市的职住平衡与通勤效率——以西安为例[J]. *地理学报*, 2013, 68(10): 1 316-1 330)
- [8] Giuliano G. Is Jobs-Housing Balance a Transportation Issue? [J]. *Transportation Record*, 1991, 13(5): 305-312
- [9] Zhou Suhong, Yang Lijun. Study on the Spatial Characteristic of Commuting in Guangzhou[J]. *Urban Transport of China*, 2005, 3(1): 62-67(周素红,杨利军. 广州城市居民通勤空间特征研究[J]. *城市交通*, 2005, 3(1): 62-67)
- [10] Sun Bindong, Pan Xin, Ning Yuemin. Analysis on Influence of Job-Housing Balance on Commute Travel in Shanghai [J]. *Urban Planning Forum*, 2008(1): 77-82(孙斌栋,潘鑫,宁越敏. 上海市就业与居住空间均衡对交通出行的影响分析[J]. *城市规划学刊*, 2008(1): 77-82)
- [11] Lu Huapu, Ding Yu, Zhang Yongbo. Urban Job-Housing Balance in China: Empirical Analysis and Counter Measures[J]. *Urban Transport of China*, 2013, 11(3): 1-6(陆化普,丁宇,张永波. 中国城市职住均衡实证分析与关键对策[J]. *城市交通*, 2013, 11(3): 1-6)
- [12] Long Ying, Zhang Yu, Cui Chengyin. Identifying Commuting Pattern of Beijing Using Bus Smart Card Data[J]. *Acta Geographica Sinica*, 2012, 67(10): 1 339-1 352(龙瀛,张宇,崔承印. 利用公交刷卡数据分析北京职住关系和通勤出行[J]. *地理学报*, 2012, 67(10): 1 339-1 352)
- [13] Gonzalez M C, Hidalgo C A, Barabasi A L. Understanding Individual Human Mobility Patterns [J]. *Nature*, 2008, 453(7 196): 779-782
- [14] Song C, Qu Z, Blumm N, et al. Limits of Predictability in Human Mobility [J]. *Science*, 2010, 327(5 968): 1 018-1 021
- [15] Schneider C M, Belik V, Couronné T, et al. Unravelling Daily Human Mobility Motifs[J]. *Journal of the Royal Society Interface*, 2013, 10(84): 246-253
- [16] Chai Yanwei, Zhao Ying, Ma Xiujun, et al. Mobile Positioning Method for Spatial Temporal Behavioral Data Collection and Its Geographical Applications [J]. *Areal Research and Development*, 2010, 29(6):1-7(柴彦威,赵莹,马修军,等. 基于移动定位的行为数据采集与地理应用研究[J]. *地域研究与开发*, 2010, 29(6): 1-7)

- [17] Liu Yu, Xiao Yu, Gao Song, et al. A Review of Human Mobility Research Based on Location Aware Devices[J]. *Geography and Geo-Information Science*, 2011, 27(4): 8-13(刘瑜, 肖昱, 高松, 等. 基于位置感知设备的人类移动性研究综述[J]. 地理与地理信息科学, 2011, 27(4): 8-13)
- [18] Kang C, Liu Y, Ma X, et al. Towards Estimating Urban Population Distributions from Mobile Call Data[J]. *Journal of Urban Technology*, 2012, 19(4): 3-21
- [19] Ranjan G, Zang H, Zhang Z L, et al. Are Call Detail Records Biased for Sampling Human Mobility? [J]. *ACM SIGMOBILE Mobile Computing and Communications Review*, 2012, 16(3): 33-44
- [20] Zang H, Bolot J. Anonymization of Location Data Does not Work: A Large-scale Measurement Study [C]. The 17th Annual International Conference on Mobile Computing and Networking, Las Vegas, Nevada, USA, 2011
- [21] Yuan Y, Raubal M, Liu Y. Correlating Mobile Phone Usage and Travel Behavior—A Case Study of Harbin, China [J]. *Computers, Environment and Urban Systems*, 2012, 36(2): 118-130
- [22] Frias-Martinez V, Soguero C, Frias-Martinez E. Estimation of Urban Commuting Patterns Using Cellphone Network Data [C]. The ACM SIGKDD International Workshop on Urban Computing, Beijing, China, 2012
- [23] Isaacman S, Becker R, Cáceres R, et al. Identifying Important Places in People's Lives from Cellular Network Data [M]. Berlin Heidelberg: Springer, 2011: 133-151
- [24] Wu Jiansheng, Huang Li, Liu Yu, et al. Traffic Flow Simulation Based on Call Detail Records [J]. *Acta Geographica Sinica*, 2012, 67(12): 1 657-1 665(吴健生, 黄力, 刘瑜, 等. 基于手机基站数据的城市交通流量模拟 [J]. 地理学报, 2012, 67(12): 1 657-1 665)
- [25] Qi Wei, Li Ying, Liu Shenghe, et al. Estimation of Urban Population at Daytime and Nighttime and Analyses of Their Spatial Pattern: A Case Study of Haidian District, Beijing [J]. *Acta Geographica Sinica*, 68(10): 1 344-1 356(戚伟, 李颖, 刘盛和, 等. 城市昼夜人口空间分布的估算及其特征——以北京市海淀区为例 [J]. 地理学报, 68(10): 1 344-1 356)

Identifying Home-Work Locations from Short-term, Large-scale, and Regularly Sampled Mobile Phone Tracking Data

XU Ning^{1,2} YIN Ling² HU Jinxing²

¹ School of Geosciences and Info-Physics, Central South University, Changsha 410000, China

² Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China

Abstract: In urban studies acquisition of individual home-work locations from large-scale mobile phone tracking data is an emerging technology using big data. Long-term irregularly as well as sparsely sampled mobile phone call data are widely used in existing studies, but short-term regularly sampled mobile phone tracking data are less widely used. This study proposes a home-work location identification method based on short-term, large-scale, and regularly sampled mobile phone tracking data. To the authors' knowledge, this study is the first effort to identify home-work locations for urban residents from short-term, large-scale, and regularly sampled mobile phone tracking data. The findings of this study evaluate the feasibility of using this new type of large-scale data source for research on urban issues such as the job-housing balance, and is of great significance when improving the representativeness of samples and the reliability of analysis results in home-work location related research effectively in terms of low financial and labor costs.

Key words: mobile phone tracking data; spatio-temporal data mining; job-housing balance; commuting distance; Shenzhen City

First author: XU Ning, postgraduate, specializes in spatial-temporal data mining. E-mail: ning.xu@siat.ac.cn

Corresponding author: YIN Ling, PhD, associate researcher. E-mail: yinling@siat.ac.cn

Foundation support: The National Natural Science Foundation of China, No. 41301440; the Strategically New Industry Development Special Foundation of Shenzhen, No. JCYJ20130401170306842.