

# 基于手机信令定位数据的居民出行时空分布特征提取方法

李祖芬<sup>1</sup>, 于 雷<sup>1,2</sup>, 高 永<sup>3</sup>, 吴亦政<sup>4</sup>, 龚大鹏<sup>1</sup>, 宋国华<sup>1</sup>

- (1. 北京交通大学 城市交通复杂系统理论与技术教育部重点实验室, 北京 100044;  
2. 德克萨斯南方大学 理工技术学院, 美国 休斯顿 77004; 3. 北京交通发展研究中心, 北京 100161;  
4. 加利福尼亚大学戴维斯分校 土木与环境工程系, 美国 戴维斯 95618)

**摘要:** 为了得到可靠的居民出行时空分布特征, 并为城市交通规划提供准确的出行现状数据, 基于手机信令定位数据设计了提取居民出行时空分布特征的方法。通过对重复冗余的手机数据进行处理、运用地理信息系统将手机数据映射至所研究的交通区域、划分交通小区、定义出行识别、建立OD矩阵及绘制出行期望线等出行数据挖掘, 得到了居民的出行时空分布特征。为了验证设计方法的可行性, 以北京市的手机信令定位数据为例, 提取出北京市居民的出行时空分布特征, 并将所得的结果与北京市第4次综合交通调查的数据进行对比得出: 两者的出行时间分布特征平均偏差为0.78%, 早晚高峰进出城方向比例的偏差为0.1, 全市的出行发生量与吸引力的平均偏差均小于3%。

**关键词:** 手机信令定位数据; 时空分布; 居民出行; OD矩阵; 期望线

中图分类号: U491.1

文献标识码: A

文章编号: 2095-9931(2016)01-0051-07

## Extraction Method of Temporal and Spatial Characteristics of Residents' Trips Based on Cellular Signaling Data

LI Zu-fen<sup>1</sup>, YU Lei<sup>1,2</sup>, GAO Yong<sup>3</sup>, WU Yi-zheng<sup>4</sup>, GONG Da-peng<sup>1</sup>, SONG Guo-hua<sup>1</sup>

- (1. MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, Beijing Jiaotong University, Beijing 100044, China; 2. School of Science and Technology, Texas Southern University, Houston 77004, USA; 3. Beijing Transportation Research Center, Beijing 100161, China;  
4. Department of Civil and Environmental Engineering, University of California, Davis, Davis 95618, USA)

**Abstract:** In order to extract valid temporal and spatial characteristics of residents' trips and provide data for urban traffic planning, a method based on cellular signaling data was proposed. By processing redundant data, mapping cellular signaling data on traffic analysis zones through using Geography Information System, defining the traffic analysis zones, defining trip identification, establishing OD matrix and expected lines, the temporal and spatial characteristics of residents' trips were extracted. In order to verify the feasibility of the proposed method, the cellular signaling data of Beijing was used as a case study to extract the temporal and spatial characteristics of residents' trips. The extracted results were com-

收稿日期: 2015-12-02

基金项目: 北京市科技计划项目(Z141100004014021)

作者简介: 李祖芬(1994—), 女, 江西宜春人, 硕士研究生, 研究方向为交通运输规划与管理。

E-mail: 11271003@bjtu.edu.cn。

pared with those of the 4th Comprehensive Transportation Survey in Beijing. The conclusion is drawn that the average deviation of the temporal characteristics of residents' trips was 0.78%, the deviation of the rates of trips for entering into the urban and leaving out of the urban during morning peak and evening peak is 0.1, the average deviation of trip generation and attraction is less than 3%.

**Key words:** cellular signaling data; temporal and spatial distribution; resident's trip; OD matrix; expected line

## 0 引言

居民的出行时空分布特征主要包括居民的出行时间分布、出行流量及方向,其能够反映居民的交通需求在时间、空间上的分布,不仅能为城市的交通规划提供出现状数据信息,也能为解决交通拥堵、提高道路通行效率等提供数据支持。其中,有效、实时的交通信息的获取是交通规划及决策的基础。而传统的交通信息采集方法(人工调查法和固定式交通信息采集技术)具有调查及维护成本高、传感器损坏率高、实时性较差及采集的交通数据精度较低等缺点<sup>[1]</sup>。随着手机终端的普及,采用手机定位等新技术采集交通信息逐渐受到重视。2015年中国手机用户已达到13.06亿,庞大的手机用户群为交通数据的采集提供了大量的数据源。另外,手机定位的交通信息采集技术还具有成本低、覆盖范围广等优点<sup>[2]</sup>。因此,手机数据可作为现有的交通数据采集技术的重要补充,其为居民出行时空分布特征的提取提供了很好的技术支持。

目前国内外已有一些针对手机定位数据的相关研究。Schlaich J.等<sup>[3]</sup>提出了利用手机定位数据获取出行轨迹。张健钦等<sup>[3]</sup>、赖见辉等<sup>[4]</sup>等提出了利用手机定位数据的出行路径识别方法。毛晓汶<sup>[5]</sup>对手机信令数据特征进行了分析,并以重庆为例得到出行OD分布、出行目的地分布和出行时间分布等出行特征,不过其研究虽然定义了手机信令定位数据中的重复冗余数据,却未详细阐述处理的方法。冉斌<sup>[6]</sup>根据手机信令定位数据和手机话单定位数据获取了出行时耗、出行距离等出行特征,但未详细阐述提取的方法。Caceres N.等<sup>[7]</sup>利用仿真系统模拟现实环境的手机网络,并采用单一路段利用位置区更新的数据来仿真获取交通出行OD和相应道路的流量数据。Gur Y. J.等<sup>[8]</sup>根据手机定位数据获取个人出行信息,并基于这些信息进行交通规划。Tettamanti T.等<sup>[9]</sup>运用手机定位

数据预测居民的出行路径选择。

从上述研究现状可以看出,国内外根据手机定位数据获取交通出行特征的研究还较少,并且在已有的研究中缺乏对获取交通出行特征方法的描述。因此,本文提出基于手机信令定位数据的出行时空分布特征提取方法,并运用所提出的方法对手机信令定位数据进行挖掘,以得到可靠的居民出行时空分布特征,可为城市的交通规划提供准确的出行现状数据。

## 1 手机信令定位数据的应用原理

### 1.1 手机信令定位数据获取的原理

本文采用的手机定位法是基于小区的定位方法。基于小区的定位方法是一种起源于蜂窝小区定位技术的简单手机定位方法,基于手机用户所在的基站小区ID来确定位置信息。实现的原理是:GSM具有“蜂窝”的网络结构<sup>[10]</sup>,手机用户如果进行通信业务就需要在所处的基站小区进行位置注册,通过提取注册的基站小区的ID号<sup>[11]</sup>就可以将用户定位到该基站小区信号覆盖的区域。相对其他的精确定位技术(如GPS),该定位方法在样本量、覆盖范围以及实施成本和周期上更具有优势。基站小区覆盖范围半径在市区大约为100~500m,郊区大约为400~1000m。该范围一般小于交通小区的覆盖范围,因此满足交通规划的应用要求。

当正在通话的手机用户从基站小区移动到另一个基站小区时,将发生位置区切换。为了保证用户通信的质量和连续性,基站平台将移动台从某一通话信道切换到另一通话信道的过程,即为基站小区切换<sup>[12]</sup>。

### 1.2 手机信令定位数据库设计

手机信令定位数据主要包括IMSI, Time Stamp, Longitude, Latitude和Event ID等字段。其中,IMSI

为手机用户的唯一标识码, 能唯一地标识用户; Time Stamp 表示手机信令的发生时间; Longitude 和 Latitude 共同描述手机用户的位置, 分别为用户的经度和纬度; Event ID 表示事件类型, 其字段含义如表 1 所示。

表 1 Event ID 字段及触发方式类型

触发方式类型	Event ID	事件类型
主动方式	0	主叫
	2	被叫
	4	短信
	5	位置区切换
被动方式	7	定时扫描

按照触发手机定位信息的方式对定位数据进行如下区分: 主动方式 (主叫、被叫, 短信、位置区切换) 和被动方式 (定时扫描)。其划分依据为该信令产生过程是否受到手机用户主观行为的影响<sup>[13]</sup>。由于主叫、被叫、短信和位置区切换属于手机用户的主观行为, 而定时扫描与手机用户本身的行为无关, 是 GSM 通信系统定时更新手机用户的位置信息。因此, 定时扫描属于被动方式触发的定位信息。

## 2 提取出行时空分布特征数据的方法

为了提取居民的出行时空分布特征, 本文首先对手机信令定位数据进行预处理, 然后对所研究的交通区域进行小区划分, 在此基础上提取居民的出行时空分布特征, 并将所得的结果进行扩样, 得到北京市总体人群的出行时空分布特征。

### 2.1 数据的预处理

#### 2.1.1 重复冗余数据的处理

本文中的重复冗余数据是指由主动方式触发所产生的冗余数据。主动方式触发所产生的冗余数据指的是手机用户在不进行位置切换的前提下, 因频繁通话或短信向 GSM 通信网络发送的位置信息<sup>[5]</sup>。对于这些数据, 本文的处理方法是将主动方式触发下的数据全部删除, 只保留被动方式触发下的数据。由于被动方式触发的周期约为 55min, 因此相比主动方式触发机制, 被动方式触发具有获取数据完整且稳定的优点, 且触发的周期能够满足居民出行时空分布特征分析的精度要求。因此在后续的研究中, 本文主要对被动方式触发的手机用户数据进

行分析处理。

#### 2.1.2 手机信令定位数据映射至交通区域

手机信令定位数据虽然显示了数据点的位置信息, 但无法直观地表示手机用户在交通区域的运动情况。因此本文运用地理信息系统将手机信令定位数据映射至所研究的交通区域, 根据手机信令定位数据和区域空间位置的关系确定每个数据点在区域中的具体位置, 便于计算和分析手机用户的交通出行。

### 2.2 交通小区的划分

本文为了直观地反映出行时间和空间分布特征, 综合考虑交通区域内每个子区域手机用户数和面积这两个因素对交通小区进行划分。本文没有严格按照交通规划中小区划分的规则进行小区划分, 因为按照划分规则进行的小区划分所得到的交通小区范围较小<sup>[14]</sup>。当基站分布数量较多时, 切换干扰多, 将难以保证手机信令定位数据位置信息的准确性, 导致获取的出行信息误差较大。当划分的交通小区范围较大时, 由于基站识别范围远远小于交通小区的范围, 切换干扰大大降低, 获取的手机信令定位数据准确度较高, 误差较小。

### 2.3 手机信令定位数据的出行时空分布特征提取

#### 2.3.1 出行时间分布特征的提取

出行时间分布特征指居民在不同时间段的出行量分布, 通常以 1h 为单位统计 1d 内发生的出行量, 分别计算各时段出行量占全日出行量的比例 (即各时段出行发生率)。

本文主要针对居民的出发时间分布特征进行分析, 定义 1h 内经过的不同小区数大于 1 的手机用户在这 1h 内产生出行, 相反则视为手机用户在这 1h 内未发生出行, 根据该定义统计 1d (24h) 的出行量, 并计算每小时出行量比例。

#### 2.3.2 出行空间分布特征的提取

出行空间分布通常使用出行量的期望线来表示, 它能表征出行的流量和方向。而出行量的期望线基于 OD 矩阵获得, 其中 OD 矩阵是指城市一定范围内所有出行起点 (O) 和终点 (D) 之间的出行交通量<sup>[15]</sup>。本文的出行空间分布特征数据主要包括早高峰、晚高峰及全天的 OD 矩阵及期望线。

由于本文以北京市的手机信令定位数据为基础, 因此根据北京市的出行时间分布特征定义出行



识别。根据北京市第4次交通综合调查的结果<sup>[16]</sup>,早高峰的出发时间分布集中在6:30—8:30,晚高峰的出发时间分布集中在16:30—18:00,且0:00—5:00和22:00—23:59这两个时间段的出行量比例均小于2%。因此对于早高峰的出行,本文将手机用户在1:00—5:00所在的小区标记为“O”,9:00—11:00所在的小区标记为“D”,根据出行的定义,将“O”和“D”不一致的手机用户视为发生了出行。同理,对于晚高峰的出行,将手机用户14:00—16:00所在的小区标记为“O”,22:00—23:59所在

的小区标记为“D”。

由于早晚高峰的出行以通勤出行为主,因此大部分居民1d内的平均出行次数为2次,故本文将手机用户早高峰OD矩阵和晚高峰OD矩阵相加得到全天的OD矩阵。

## 2.4 扩样

为了得到区域总体人群的出行时空分布特征数据,本文基于各区域的人口数,将所得的手机出行数据扩样为总体人群的出行数据,具体计算方法如下:

$$\text{区域总体人群出行发生量(吸引量)} = \frac{\text{出行发生量(吸引量)}}{\text{区域手机用户样本量}} \times \text{区域人口}$$

## 3 实例应用

本文基于北京市的手机信令定位数据,结合已提出的研究方法,对北京市的居民出行时空分布特征进行分析,并将所得的结果与北京市第4次综合交通调查的相应结果进行比较,从而验证方法的可行性。

表2 手机信令定位数据映射后的数据属性表

手机识别号	日期	时间	经度	纬度	区域
4600*****14000	2010-09-02	00:00:00	116.249 4	39.857 74	丰台区
4600*****28000	2010-09-02	00:00:00	116.301	39.881 48	丰台区
4600*****04000	2010-09-02	00:00:00	116.203 9	39.970 51	海淀区
4600*****00000	2010-09-02	00:00:00	116.355 8	39.879 87	西城区
4600*****57000	2010-09-02	00:00:00	116.281 3	39.844 55	丰台区
4600*****18000	2010-09-02	00:00:00	116.633 3	40.104 47	顺义区

## 3.2 小区划分

本文统计北京市各行政区的手机用户样本量,综合考虑北京市各行政区的面积和手机用户样本量进行小区划分,并参考北京市第4次交通综合调查的小区划分结果,最终将北京市的16个行政区划分成26个小区。

## 3.3 手机信令定位数据的出行时空分布特征提取

### 3.3.1 出行时间分布特征数据的提取

根据2.3节中对出行发生的定义,对北京市手机信令定位数据进行处理,得到每小时的出行量比例,将得到的结果与第4次交通综合调查的结果进行比较分析(见图1),得出基于手机定位数据提取的出发时间分布特征与北京市第4次综合交通调查获取的分布特征相似,都呈现出“双高峰”

特征,且两曲线的平均偏差为0.78%。

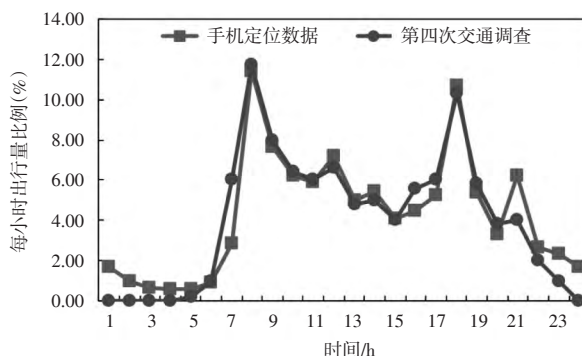


图1 出发时间分布对比图

### 3.3.2 出行空间分布特征数据的提取

根据2.3节中对出行发生的定义,对手机信令

定位数据进行OD标定,得到早高峰、晚高峰及全天的OD矩阵及期望线。

根据OD矩阵及期望线可得到许多出行空间分布特征数据,本文主要描述根据OD矩阵及期望线所得的早晚高峰进出城方向比例和全市出行发生与吸引量这两项出行空间分布特征数据。本文定义三环以内为城区,即进出三环的出行量为进出城的出行量。并将这两项数据与北京市第4次交通综合调查的结果进行对比,从而验证本文提出的出行时空分布特征提取方法的可行性。

#### (1) OD矩阵及期望线

表3为所得的部分OD矩阵示例。由于本文划分的小区数较多,得到的全北京市的期望线数量也较多,因此为了清晰地展示得到的期望线图,本文

只选取了北京市城六区(海淀区、西城区、东城区、朝阳区、丰台区和石景山区)绘制全天出行期望线图,如图2所示。

表3 OD矩阵示例

O \ D	11	12	13	21	22	...	162
0	11	12	13	21	22	...	162
11	381	36	367	46	123	...	0
12	43	135	220	19	77	...	0
13	391	220	4028	439	1160	...	4
21	66	44	579	521	485	...	0
22	160	100	1482	453	1945	...	3
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
162	1	0	2	0	1	...	2

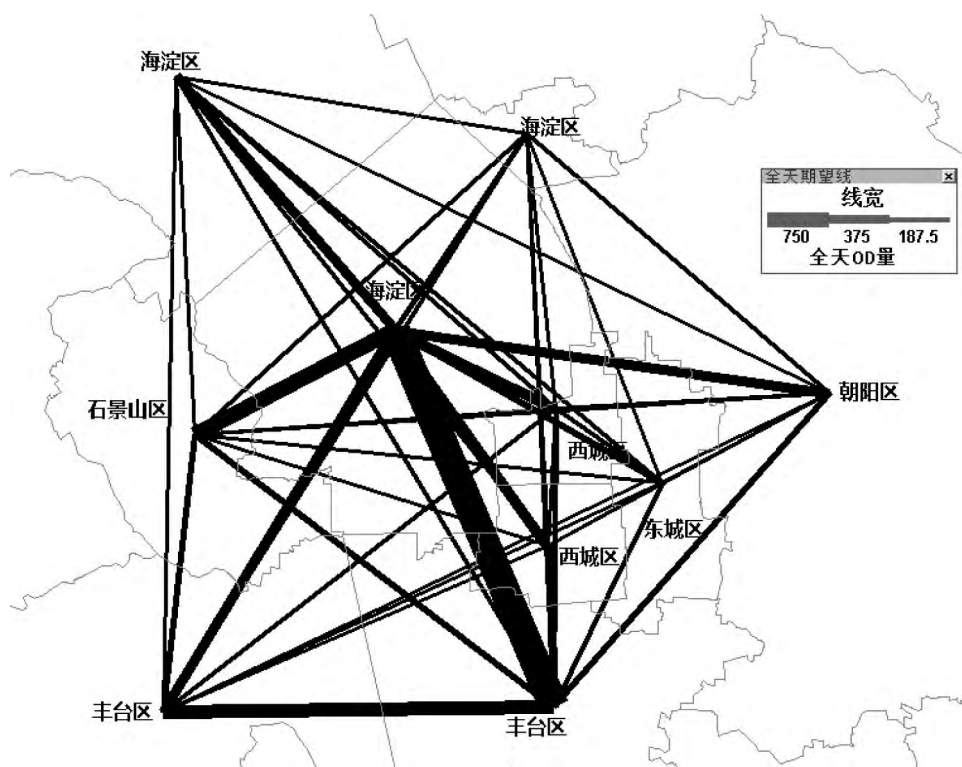


图2 全天出行期望线

(2) 早晚高峰时段进出城方向的不均衡性分析  
为了描述早、晚高峰时段进出城方向的不均衡性,本文基于OD矩阵分别计算早高峰与晚高峰进出城的出行量比,并与第4次交通综合调查中相应的结果进行比较分析后发现,基于手机定位数据得到的早高峰进出城出行量比与第4次交通综合调查

的结果偏差仅为0.1,晚高峰的偏差为0(见图3)。

#### (3) 全市出行发生量与吸引量

将获取的手机用户OD矩阵进行扩样,得到北京市各行政区的出行发生量与吸引量(见图4)。将所得结果与北京市第4次交通综合调查结果(见图5)进行比较,得到出行发生量的平均偏差为

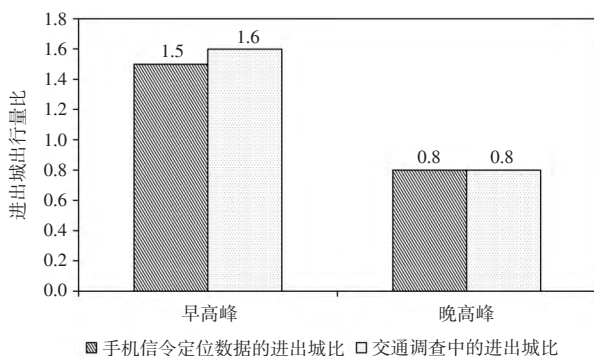


图3 早晚高峰进出城方向的不均衡性比较

2.04%，出行吸引力的平均偏差为1.35%。因此得出扩样的结果较好，能够反映北京市总体人群的出行空间分布特征。



图4 出行发生量与吸引力（手机信令定位数据）

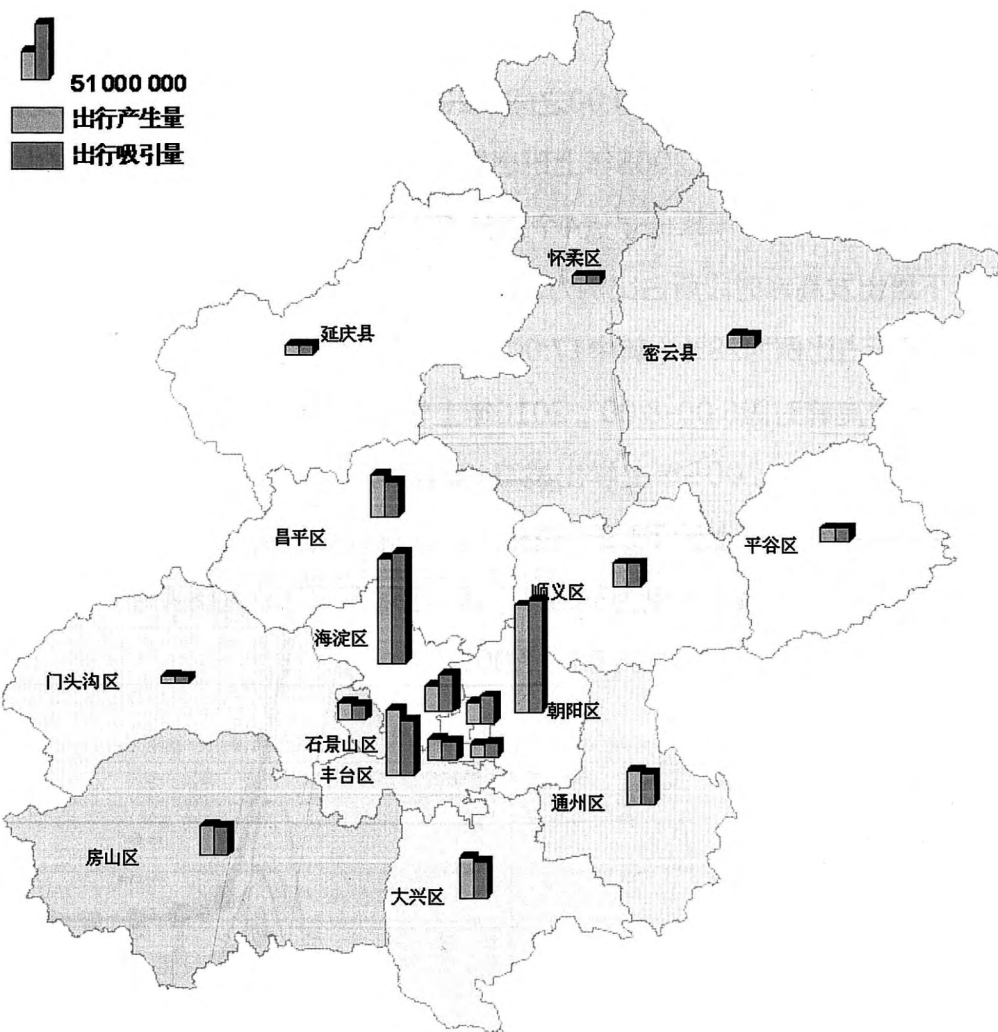


图5 出行发生量与吸引力（交通调查）<sup>①⑥</sup>

①为所有区域的出行发生量与吸引力中最大值的1/2。

## 4 结语

本文基于手机信令定位数据设计了提取居民出行时空分布特征的方法, 并对方法进行了详细地阐述, 弥补了已有研究的不足。另外, 本文还应用北京市手机信令定位数据, 将通过该方法所得的结果与北京市第4次交通综合调查结果进行比较, 得出两者的出行时间分布特征和空间分布特征的偏差均较小, 说明本文提出的方法能够得到较可靠的居民出行时空分布特征, 该方法具备可行性。

在未来的研究中可从以下三方面展开深入研究: ①运用手机数据进行其他出行量化指标的分析, 如出行目的、出行方式、出行次数等量化指标; ②提出居民出行链识别的算法, 更加准确地获取居民出行特征数据; ③比较分析主动方式和被动方式的差异, 建立OD修正模型, 从而获取更加准确的OD矩阵。

## 参考文献

- [1] YAO Z F, YANG F, CHENG Y, et al. Multimode Trip Information Derivation from Personal Trajectory Data[C]// Transportation Research Board. TRB 94th Annual Meeting Compendium of Papers. Washington D.C.: Transportation Research Board, 2015: 1-14.
- [2] SCHLAICH J, OTTERSTÄTTER T, FRIEDRICH M. Generating Trajectories from Mobile Phone Data[C]// Transportation Research Board. TRB 89th Annual Meeting Compendium of Papers. Washington D.C.: Transportation Research Board, 2010: 1-17.
- [3] 张健钦, 仇培元, 徐志浩, 等. 一种基于手机定位数据的出行行程识别方法[J]. 武汉理工大学学报 (交通科学与工程版), 2013, 37(5): 934-938.
- [4] 赖见辉, 陈艳艳, 钟园, 等. 基于手机定位信息的地铁乘客出行路径辨识方法[J]. 计算机应用, 2013, 33(2): 583-586.
- [5] 毛晓汶. 基于手机信令技术的区域交通出行特征研究[D]. 重庆: 重庆交通大学, 2014.
- [6] 冉斌. 手机数据在交通调查和交通规划中的应用[J]. 城市交通, 2013, 11(1): 72-81, 32.
- [7] CACERES N, WIDEBERG J P, BENITEZ F G. Deriving Origin-Destination Data from a Mobile Phone Network[J]. IET Intelligent Transportation System, 2007, 1(1): 15-26.
- [8] GUR Y J, BEKHOR S, SOLOMON C, et al. Inter-City Person Trip Tables for Nationwide Transportation Planning in Israel Obtained from Massive Cell Phone Data[J]. Transportation Research Record: Journal of the Transportation Research Board, 2009(2121): 145-151.
- [9] TETTAMANTI T, DEMETER H, VARGA I. Route Choice Estimation Based on Cellular Signaling Data[J]. Acta Polytechnica Hungarica, 2012, 9(4): 207-220.
- [10] CHENG P, QIU Z J, RAN B. Traffic Estimation Based on Particle Filtering with Stochastic State Reconstruction Using Mobile Network Data[C]// Transportation Research Board. TRB 85th Annual Meeting Compendium of Papers CD-ROM. Washington D.C.: Transportation Research Board, 2006: 1-15.
- [11] 秦艳珊, 宁彬, 徐凯, 等. 蜂窝网络单基站定位技术的研究与实现[J]. 计算机时代, 2015(7): 20-22.
- [12] CACERES N, WIDEBERG J P, BENITEZ F G. Review of Traffic Data Estimations Extracted from Cellular Networks[J]. IET Intelligent Transport Systems, 2008, 2(3): 179-192.
- [13] 吴亦政. 基于手机定位信息和出行调查的动态OD获取方法[D]. 北京: 北京交通大学, 2014.
- [14] 李佳. 基于手机定位数据的动态OD矩阵获取方法研究[D]. 昆明: 昆明理工大学, 2013.
- [15] 芦方强, 陈学武, 胡晓健. 基于公交OD数据的居民公交出行特征研究[J]. 交通运输工程与信息学报, 2010(2): 31-36.
- [16] 北京市交通委员会. 北京市第四次交通综合调查简要报告[R]. 北京: 北京市交通发展研究中心, 2010.