

DOI: 10.11918/j.issn.0367-6234.201708042

手机信令与出租车 GPS 数据融合车源定位方法

王 璞¹, 鲁恒宇¹, 谭 倩¹, 熊雨沙¹, 毛应萍², 李 琳²

(1. 中南大学 交通运输工程学院, 长沙 410000; 2. 深圳市城市交通规划设计研究中心, 广东 深圳 518021)

摘 要: 为揭示居民出行行为与城市交通拥堵的内在关联, 并为缓解城市交通拥堵提供技术支持, 利用高覆盖率、低精度的手机数据和低覆盖率、高精度的出租车 GPS 数据, 构建了数据驱动的车源定位方法. 利用手机数据获取出行需求信息, 利用出租车 GPS 数据获取交通状态信息; 提出基于数据融合的出行 OD 估计方法, 进行交通流分配, 对城市道路车流来源及城市拥堵源进行动态定位. 结果表明: 道路车流主要来自于少量车源小区, 且拥堵状态下这些小区更加集中; 同时受居民通勤行为的影响, 城市全局拥堵源在早晚高峰表现出不同的特征. 利用数据融合的车源定位可以用于揭示拥堵形成的内在机理及演化规律, 辅助制定有针对性的拥堵缓解策略.

关键词: 城市交通; 车源定位; 数据融合; 手机数据; 出租车 GPS

中图分类号: U491 **文献标志码:** A **文章编号:** 0367-6234(2018)09-0096-05

A data fusion approach for locating driver sources using mobile phone signaling data and taxi GPS data

WANG Pu¹, LU Hengyu¹, TAN Qian¹, XIONG Yusha¹, MAO Yingping², LI Lin²

(1. School of Traffic and Engineering, Central South University, Changsha 410000, China;

2. Shenzhen Urban Transport Planning Center, Shenzhen 518021, Guangdong, China)

Abstract: To investigate the relation between people's travel behavior and urban traffic congestion, and provide technical support to alleviate urban traffic congestion, a data-driven source localization approach was constructed based on high-coverage-rate low-precision mobile phone data and high-precision taxi GPS data. Travel demand and traffic condition information was obtained by personal mobile phone data and taxi GPS data, respectively. Mobile phone data and taxi GPS data were combined to estimate origin-destination matrix, conduct traffic assignment, and dynamically locate traffic sources of road segments and the congestion sources of the city. Result showed that a majority of traffic flow was generated by a few sources, and the sources became more concentrated when traffic jam occurred. Urban congested sources were affected by commuting behaviors, and exhibited different characteristics during morning and evening peak hours. Traffic source location by data fusion can uncover the internal mechanism and evolution law of traffic congestion, and help in making target policies dealing with traffic congestion.

Keywords: urban transportation; traffic source location; data fusion; mobile phone data; taxi GPS data

中国大城市的交通需求持续快速增长, 交通拥堵问题非常严重, 已对城市居民的工作、学习、生活造成极大不便. 尽管目前已有一些基于道路扩容的解决方案, 但这类方案具有实施周期长、建设成本大、缺乏灵活性等缺点, 难以适应当今快速变化的城市交通结构^[1]. 近年来学术界陆续提出很多新颖的交通拥堵缓解方法: 文献[2-6]提出了基于交通状态预测的拥堵缓解策略; 文献[7-8]从城市规划角度研究了改善城市交通的方法; 文献[9-10]利用复杂网络分析技术优化交通网络结构.

无论哪种解决拥堵问题的方法, 理解、掌握交通

拥堵的形成机制和源头信息都至关重要. 近年交通大数据的大量涌现为精准锁定交通拥堵成因提供了新的机遇. 文献[11-12]基于手机基站划分泰森多边形, 利用手机数据估计交通需求 OD 并研究了居民的出行规律. 文献[13-15]提出了基于 GPS 数据的交通状态检测与出行模式识别方法; 文献[16-18]采用手机数据和基于位置的社交数据分析了城市道路的使用模式. 其中, 文献[16]通过分析道路车流的来源, 发现造成道路拥堵的车流主要由少量小区贡献, 对这些小区实施交通限行能够更加有效地缓解交通拥堵. 然而, 文献[16]将手机用户的居住小区作为车源小区, 无法获取居民出行的起点, 提出的方法难以应用于动态交通管控. 本文通过对手机数据和 GPS 数据综合运用, 提出了一种能够动态定位道路车源和城市全局拥堵源的方法.

收稿日期: 2017-08-14

基金项目: 国家自然科学基金 (61473320, 51208520);

湖南省科技计划 (2015RS4011)

作者简介: 王 璞 (1983—), 男, 博士, 教授, 博士生导师

通信作者: 王 璞, wangpu@csu.edu.cn

1 数据描述

1.1 手机数据

本文所使用的手机数据为深圳市 2012 年某工作日的手机信令数据. 数据中共包含 1 630 万个手机用户的 6 亿条时间位置信息, 见表 1. 手机用户的位置由手机基站(共 5 952 个)记录, 密集的基站分布保证了用户位置信息具有较高的准确性. 从手机信令数据中可以获取居民的出行需求 OD、用户职住地点等信息.

表 1 手机信令数据的格式

Tab.1 Format of mobile phone signaling data		
记录时间	基站经度/(°)	基站纬度/(°)
2012-03-25T19:51:47	114.212 97	22.607 39
2012-03-25T15:27:26	114.212 97	22.607 39
2012-03-25T01:53:00	114.212 97	22.607 39

根据人们一般的生活规律, 将夜晚时段(00:00—6:00 时)内持续具有手机记录且基站位置保持不变的用户作为深圳市常住人口. 从手机数据中识别到的常住人口数为 879.94 万, 占深圳市实际常住人口(2013 年深圳人口普查数据为 1 062.89 万)的 80% 以上. 图 1 展示了手机用户记录量空间分布.

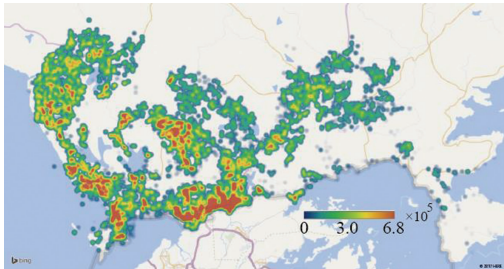


图 1 深圳手机用户记录热力图

Fig.1 Thermal map of mobile phone records in Shenzhen

1.2 出租车 GPS 数据

本文使用深圳市连续两周(2016 年 8 月 15 日至 2016 年 8 月 28 日, 其中 15 日至 19 日, 22 日至 26 日为工作日, 20、21、27、28 日为非工作日)的出租车 GPS 数据. 数据共包含 7.2 亿条坐标记录, GPS 数据格式见表 2, 每条记录包含出租车的经纬度坐标、时间戳与载客状态等信息, 其中 1 为载客状态, 0 为空乘状态. 其中记录对应出租车的 1 个轨迹点, 连续 2 个轨迹点的时间间隔约为 20 s.

表 2 出租车 GPS 数据的格式

Tab.2 Format of taxi GPS data			
轨迹点经度/(°) 轨迹点纬度/(°)		记录时间	载客状态
113.845 100	22.626 101	2016-08-15T9:24:00	1
113.846 603	22.622 801	2016-08-15T9:24:21	1
113.844 398	22.622 200	2016-08-15T9:24:41	0

2 基于手机数据的出行 OD 估计

基于手机数据的居民出行 OD 估计方法如下: 1) 停留基站的确定. 尽管有时用户没有发生位移, 其记录的位置也会在邻近基站间跳动, 这就是“乒乓效应”. 为了消除“乒乓效应”, 对出行 OD 估计的影响, 同时避免将长距离出行划分成多次的短距离出行, 定义用户记录中连续两次及其以上的基站为其停留基站. 2) 出行 OD 估计. 用户在一天内有多个停留基站记录, 当停留基站发生改变, 且两基站间直线距离超过 0.5 km 时, 认为用户发生了一次出行, 将前后两个停留基站的位置作为出行的起讫点. 通过对所有用户的记录进行上述统计, 获取基于基站位置的出行 OD. 3) 以交叉口为起讫点的出行 OD. 上述所获取的 OD 矩阵是以基站作为起讫点单位的出行 OD 矩阵, 难以直接用于交通流估计. 故将其转化为基于交叉口的出行 OD, 具体步骤如下: a) 利用 ArcGIS 划分手机基站的泰森多边形, 作为基站的服务区域, 如图 2 所示. b) 根据出租车的载客状态获取乘客上车点和下车点的位置, 计算上(下)车点与各个交叉口的距离, 选取距离最近的交叉口作为上(下)车实际位置, 计算服务区域内每个交叉口的上(下)车概率. c) 按照基站服务区域内的上(下)车概率选取交叉口作为出行起讫点, 建立以交叉口为单位的出行 OD 矩阵, 如图 3 所示.

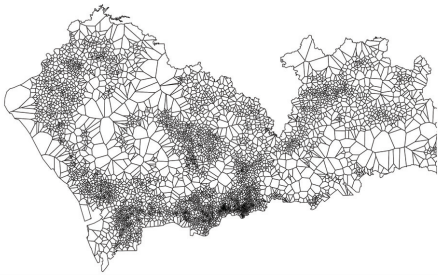


图 2 手机基站的服务区域

Fig.2 Service area of mobile phone towers

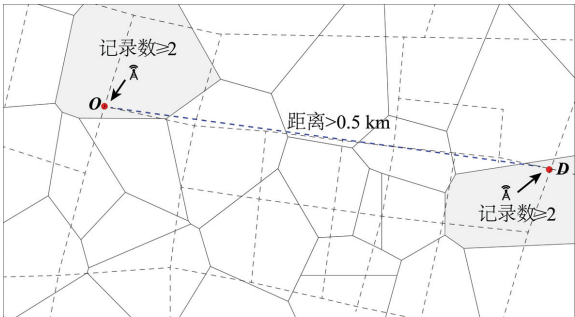


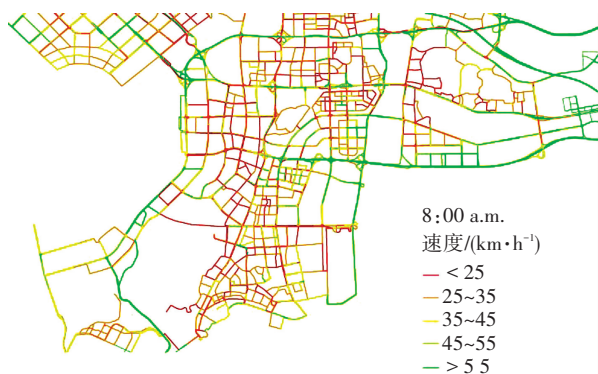
图 3 出行需求 OD 估计

Fig.3 Estimation of travel demand OD

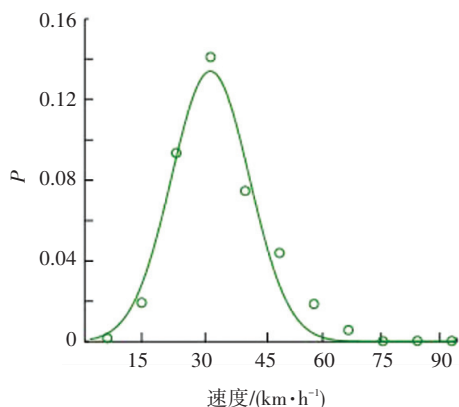
3 基于 GPS 数据的交通状态检测

本文采用 ST-Matching 算法^[15]对出租车 GPS 记录进行地图匹配. 所使用的路网覆盖范围包括了快速路、主干路、次干路和部分支路以及高速公路, 共有 21 115 条有向边及 85 389 个节点. 将 GPS 记录按照车牌号和 GPS 的返回时间排序, 得到出租车在一天中的轨迹, 选取载客状态为 1 的连续记录作为出租车的出行轨迹点. 利用 ST-Matching 算法将出行轨迹匹配到路网上, 获取对应的出行路径, 计算道路的车辆行驶速度和通行时间.

由于出租车轨迹覆盖面有限, 一些道路在某些时间窗的存在通行时间缺失. 对于这种情况, 将相邻道路上车辆行驶速度的平均值作为该道路的车辆行驶速度, 计算道路通行时间. 图 4 为深圳市南山区路网在早 8:00—8:30 时的车辆行驶速度分布. 大部分道路速度分布在 30 km/h 附近, 大部分二级道路均为拥堵状态, 交通状况较好 (绿色) 的道路多为城市快速路.



(a) 地图匹配结果



(b) 车辆行驶速度分布

图 4 地图匹配结果与道路车辆行驶速度分布

Fig.4 ST-matching result and distribution of vehicle speed

4 车源定位

4.1 交通流估计

利用通过手机数据获取的居民出行 OD 和通过

出租车 GPS 数据估计的道路通行时间计算深圳市路网交通流分布, 具体步骤如下: 1) 对于某时间窗的出行 OD, 将其前一个时间窗的道路通行时间作为路阻, 使用全有全无法进行交通流分配, 获取各次出行的路径. 2) 通过出租车 GPS 数据获取出行时段内的道路车辆行驶速度, 计算车辆到达各个交叉口的时间, 对车辆在路网上的行驶进行仿真.

4.2 道路车流溯源方法

定义 D 为全体交通小区的集合, (Z, S) 为从小区 Z 到小区 S 之间的所有出行 OD 的集合, q_{zs} 为从交叉口 $z (z \in Z)$ 到交叉口 $s (s \in S)$ 的出行量, p_{zs} 为 q_{zs} 经过的路径, $q_{zs}^l \in p_{zs}$ 为选取经过道路 l 的出行路径集合, Q_l 为路段 l 的交通流量, 统计道路在所选时段内各个小区对其流量的贡献率为

$$C_l(Z) = \left(\sum_{s=1}^D \sum_{q_{zs}} q_{zs} \right) / Q_l. \quad (1)$$

对各个小区的车流贡献率 $C_l(Z)$ 排序, 从贡献量最大的小区开始对贡献率依次累加, 当累加贡献率达到 θ 时, 定义参与累加的小区为道路主要车源, 拥堵状态下的主要车源定义为拥堵车源.

本文使用广泛应用的“二八定律”计量车源与交通流之间的关系, 选用 80% 作为主要车源的判别标准.

4.3 城市全局拥堵源定位方法

道路溯源定位了道路在各个时段的车辆来源, 有助于交通管理者针对某条道路进行交通管理. 同时城市管理者也需要掌握整个城市或某个片区的交通拥堵来源, 从全局规划城市交通.

车源定位方法流程如图 5 所示, 通过计算各个交通小区的总额外出行时间, 可在城市或区域层面定位城市全局拥堵源. 按 30 min 划分时间窗, 选取 OD 起点所在时间窗为出行时段 M , 根据出行出租车 GPS 数据获取路段 l 在时段 M 内的通行时间 t_l^a . 使用 23:00—24:00 时段的路段通行时间为自由通行时间 t_l^f , t_l^a 减去 t_l^f 即为路段额外行驶时间 t_l^e . 统计从某小区 Z 出发的所有出行经历的额外行驶时间 $T^e(Z)$ 为

$$T^e(Z) = \sum_{s=1}^S \sum_{p \in P} t_p \times q_p, \quad (2),$$

$$t_p = \sum_{l \in L} t_l^e. \quad (3)$$

其中: (z, s) 为小区 Z, S 间的 OD 对, P 为 (z, s) 的出行路径, L 为路径 p 通过的路段, t_p, q_p 分别为路径的额外行驶时间与流量.

对额外出行时间进行排序, 从额外出行时间最大的小区开始对贡献量依次累加, 当累加贡献量达

到总出行量的 20%时,定义参与累加的小区为城市拥堵源。

体车源的地理分布均各自相似,但主要车源的分布随时段的不同呈现动态变化的特性,如图 7 所示。

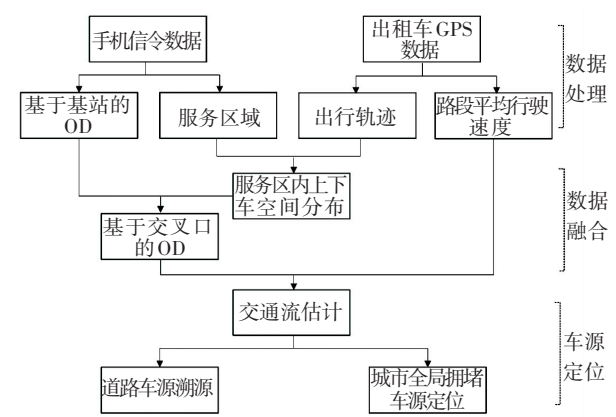


图 5 车源定位方法流程图

Fig.5 Flow chart of traffic source locating method

5 案例分析

5.1 拥堵道路车流溯源

选取深圳市交通拥堵状况较为严峻的梅观路南段与香蜜湖路进行道路车流溯源实例分析,并以 30 min为时间窗观测两条道路的车辆行驶速度和流量随时间的变化,如图 6 所示。图中阴影区域(梅观路的 7:00—9:00、香蜜湖路的 17:00—19:00)出现车辆行驶速度明显突降和流量明显突增,说明这些时段车辆缓行,车流量很大。因此,梅观路最拥堵时段为早高峰、香蜜湖路最拥堵时段为晚高峰。

选取 7:00—9:00(早), 11:00—13:00(午), 17:00—19:00(晚)3 个时段对梅观路和香蜜湖路进行车流溯源。图 7 为梅观路与香蜜湖路分别在早、午、晚 3 个高峰时段的车流溯源结果。对结果进行分析发现:梅观路和香蜜湖路在3个时间段内的全

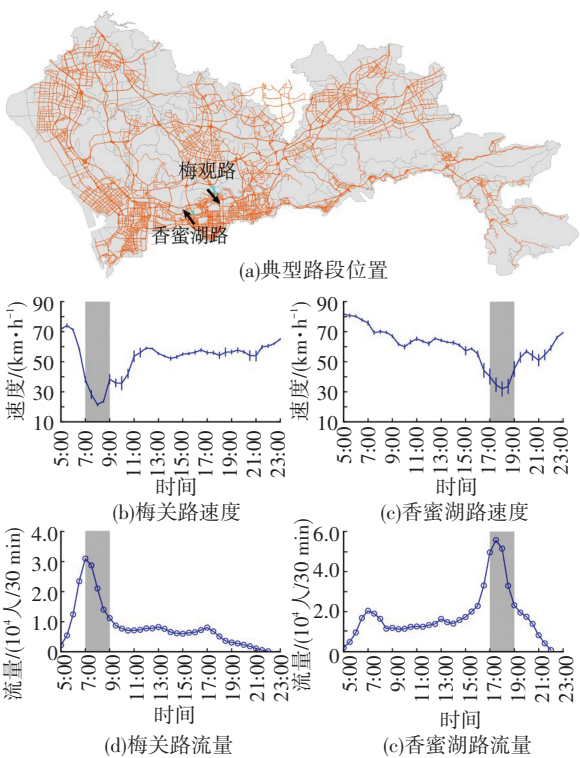


图 6 典型路段位置及车辆行驶速度与流量变化

Fig.6 Vehicle speed and traffic flow on typical road segments

定义 R_i 为路段 i 的拥堵车源个数占全体车源的比例,梅观路在早、午、晚 3 个时段的 R_i 值分别为 6.1%、17.8%、21.0%;香蜜湖路在早、午、晚 3 个时段的 R_i 分别为 12.3%、15.0%、9.9%。据前文分析已知梅观路的最拥堵时段为早高峰,香蜜湖路的最拥堵时段为晚高峰,这表明:1) 两条道路最拥堵时段的 R_i 均为当天的最低值;2) 拥堵状态下这些小区更加集中。

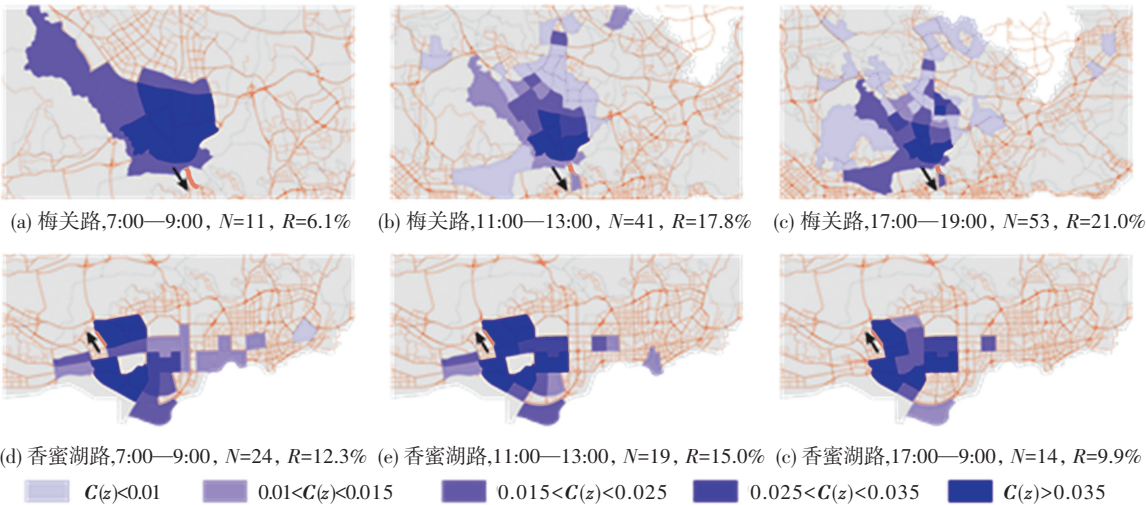


图 7 梅观路与香蜜湖路的主要车源分布

Fig.7 Major traffic sources of Meiguan Road and Xiangmihu Road

分别选取梅观路和香蜜湖路的最拥堵时段,分析发现梅观路在早高峰的拥堵车源(11个,占总车源的6.1%)都分布在上游地区,且该区域主要为住宅用地,从这些小区出发的早高峰通勤出行贡献了梅观路80%的高峰流量,引发了道路拥堵;香蜜湖路在晚高峰时段的道路拥堵车源(14个,占总车源的9.9%)集中分布福田区的中南部地区,该区域主要为商业用地,表明香蜜湖路在晚高峰时段的拥堵现象是由上游距离较近的商业用地产生的大量下班出行造成的. 统计结果见表3.

表 3 结果统计

Tab.3 Statistical results

典型路段	时段	速度/(km·h ⁻¹)	拥堵车源数	全体车源数	R _i /%
梅观路	早	27.9	11	180	6.1
	午	56.9	43	230	17.8
	晚	55.6	53	252	21.0
香蜜湖路	早	71.2	24	195	12.3
	午	64.1	19	127	15.0
	晚	35.1	14	141	9.9

5.2 深圳城市全局拥堵源定位

早高峰(7:00—9:00)深圳的总出行量为270万,晚高峰(17:00—19:00)出行量为231万,分别对早晚高峰进行城市全局拥堵溯源. 结果如图8所示,早高峰时段城市全局拥堵源为34个,占全体小区的R=7.2%,大多分布在火车站、汽车站周边及居住用地;晚高峰时段城市全局拥堵源个数为20,占全体小区的R=4.3%,主要集中在深圳市区,用地性质多为商业用地. 早高峰城市拥堵源的总平均额外行驶时间为3.16×10⁶s,晚高峰的拥堵车源的总平均额外出行时间为6.86×10⁶s.

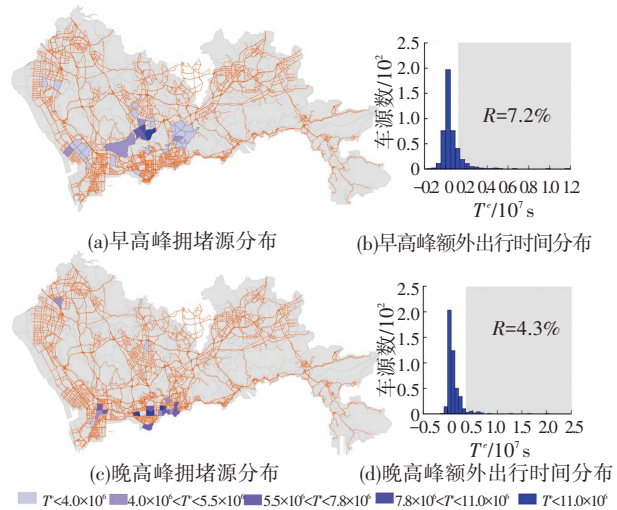


图 8 早晚高峰的城市拥堵车源和额外出行时间分布

Fig.8 Urban traffic congestion sources and distribution of extra travel time in morning and evening peak hours

通过分析可以发现,受居民通勤行为影响,早高峰的拥堵车源分布在城市二线关外住宅用地区域,晚高峰的拥堵车源集中分布城市内部工作区域,与现实情况相符. 同时,晚高峰城市拥堵源占比明显小于早高峰,而平均额外出行时间是后者的2.21倍,说明尽管早高峰的出行量较大,但拥堵源较为分散;而晚高峰由于车源较为集中,造成拥堵更为严重.

6 结 论

1) 通过融合手机数据和出租车GPS数据构建了道路车源定位和城市拥堵源定位方法,发现道路车流主要来自于少量车源小区,且拥堵状态下这些小区更加集中;同时受居民通勤行为的影响,城市全局拥堵源在早晚高峰表现出不同的特征.

2) 所揭示道路主要车源和城市拥堵源的分布特征及其动态变化规律可应用于动态交通管控,协助交通管理者从源头制定交通拥堵管理策略.

3) 文中获取的手机数据与GPS数据的时间分别为2012年、2016年,车源定位的结果可能由于时间迁移发生一定的变动,下一步可依本文提供的方法采用当前时间的手机数据和GPS数据,进行实时、精准车源定位,并接入城市交通管理控制平台,充分发挥大数据的潜力.

参考文献

[1] 高自友, 龙建成, 李新刚. 城市交通拥堵传播规律与消散控制策略研究[J]. 上海理工大学学报, 2011, 33(6): 701
GAO Ziyou, LONG Jiancheng, LI Xingang. Congestion propagation law and dissipation control strategies for urban traffic[J]. University of Shanghai for Science and Technology, 2011, 33(6): 701

[2] HERRERA J C, AMIN S, BAYEN A, et al. Dynamic estimation of OD matrices for freeways and arterials[M]. Berkeley: UC Berkeley, 2007

[3] LAURA W, WEI S. Real-time traffic prediction using GPS data with low sampling rates: A hybrid approach[C]// In 91st Transportation Research Board Annual Meeting, Washington DC: Transportation Research Board, 2012

[4] WANG Jingyuan, MAO Yu, LI Jing, et al. Predictability of road traffic and congestion in urban areas[J]. PloS One, 2015, 10(4): e0121825. DOI: 10.1371/journal.pone.0121825

[5] 吴奇兵, 陈峰, 黄焱, 等. 北京市机动车拥堵成本测算与分析[J]. 交通运输系统工程与信息, 2011, 11(1): 168
WU Qibing, CHEN Feng, HUANG Yao, et al. Calculation and analysis of traffic congestion cost in Beijing[J]. Journal of Transportation Systems Engineering and Information Technology, 2011, 11(1): 168

[6] 赵颀, 陆建, 张文珺. Logit 模型下的道路拥堵收费策略改善效果分析, 哈尔滨工业大学学报[J]. 2017, 49(3): 80
ZHAO Yi, LU Jian, ZHANG Wenjun. Improvement effect analysis of congestion pricing using Logit model[J]. Journal of Harbin Institute of Technology, 2017, 49(3): 80