

# 出租汽车 GPS 数据的空间分布特征及应用 ——以广州市为例

苏跃江, 吴德馨, 李晓玉

(广州市交通运输研究所, 广州市公共交通研究中心, 广东 广州 510635)

**摘要:** 出租汽车 GPS 数据记录了车辆的出行轨迹、时间、起讫点等信息, 利用相关算法可以挖掘车辆或乘客的运行特征、时空分布特征等, 应用于交通规划和运营管理。在总结出出租汽车 GPS 异常数据的处理方法以及时空分布特征主要算法的基础上, 对广州市约 2 万辆出租汽车的上下客流热点区域、时间分布、OD 分布等进行总结。以广州市为例进行实证研究, 从交通规划、运营管理、服务模式等方面选择典型案例, 剖析广佛出租汽车返程点布局规划、公共汽(电)车夜班线路优化, 以及如约巴士+出租汽车新服务模式的应用尝试。

**关键词:** 城市交通; 出租汽车; GPS 数据; 运行特征; 时空分布; 实践应用; 广州市

Characteristics of Taxi GPS Data Spatial Distribution: A Case Study in Guangzhou

Su Yuejiang, Wu Dexin, Li Xiaoyu

(Guangzhou Transport Research Institute, Guangzhou Guangdong 510635, China)

**Abstract:** The taxi GPS data, containing travel path, time, origin and destination (OD) and etc., can be used to reveal a passenger travel characteristics in spatial and temporal terms, which is valuable to the transportation planning and traffic operation management. This paper presents a case study that analyzed the GPS data from 20,000 taxies in Guangzhou, which clearly reveals the passenger hot spots, temporal distribution, OD distribution. This real world case study provides the empirical evidence not only to the Guangzhou transportation planning and operational management but also for optimizing night-time public transit service routes as well as new combined service of paratransit and taxi.

**Keywords:** urban transportation; taxi; GPS data; operational characteristics; temporal and spatial distribution; practical application; Guangzhou

收稿日期: 2018-02-21

作者简介: 苏跃江(1983—), 男, 贵州六盘水人, 硕士, 高级工程师, 主要研究方向: 公共交通、交通信息。E-mail: 250234329@qq.com

出租汽车作为城市公共交通的重要组成部分, 兼具公共交通和私人交通的双重属性, 在城市客运交通体系中发挥非常重要的作用。出租汽车 GPS 数据记录了车辆的出行轨迹、时间、起讫点以及乘客出行特征等信息, 对其数据挖掘可以分析出租汽车的运行特征和时空分布特征, 对城市规划、交通管理等方面意义重大。文献[1]总结出出租汽车上下客事件轨迹呈现的线状特征以及城市道路网络空间不同时段出现频率的动态分段特征; 文献[2]利用空间聚集模式探测方法, 挖掘出租汽车轨迹中城市热点区域; 文献[3]针对 GPS 数据在道路网分布的不均衡性, 分析路网分布的时变特征以及在不同区域、不同

通道上的分布特征; 文献[4]利用 GPS 数据分析不同时段出租汽车空驶率的变化情况和 OD 分布, 在此基础上规划出租汽车停靠站; 文献[5]通过 GPS 数据的分析和预处理, 提出了出租汽车运营管理指标计算模型; 文献[6]从出租汽车的行程分布、里程利用率、时间和空间分布等方面, 分析上海市出租汽车的运行特征并提出相关发展建议。然而, 既有文献较少关注和研究利用出租汽车 GPS 数据挖掘出行时空分布特征并应用到交通规划和运营管理中。因此, 本文以广州市为例, 重点研究出租汽车的出行起讫点位置识别算法、时空分布特征以及 GPS 数据在交通规划和管理方面的应用。

## 1 研究方法和数据基础

### 1.1 研究方法

本文采用文献法和数据挖掘法两种研究方法。文献法主要查阅了出租汽车GPS数据挖掘的相关理论、研究进展以及实践应用,归纳和整理有用素材;数据挖掘法主要利用出租汽车GPS数据对车辆的运行特征、载客特征以及时空分布特征等进行挖掘。

### 1.2 数据基础

出租汽车GPS数据主要包括车牌号、GPS时间、入库时间、经度、纬度、高度、速度、行驶方向、车辆运营状态以及数据有效性等共10项属性(见表1)。广州市约2万辆出租汽车基本都安装了GPS定位模块与计价器模块的车载终端,可动态采集出租汽车运

表1 出租汽车GPS数据表结构及示例

Tab.1 Structure and example of taxi GPS data

字段名称	注释	数据示例
LICENSEPLATENO	车牌号	粤AF7Z76
GPS_TIME	GPS时间	2016年10月31日0:03:26
IN_DATE	入库时间	2016年10月31日0:03:27
LONGITUDE	经度	+113.279 8
LATITUDE	纬度	+23.115 5
HEIGHT	高度	000
SPEED	速度	022
DIRECTION	行驶方向	110
CAR_STAT	车辆运营状态: 1为防盗, 2为防劫, 4为空车, 5为重车, 6为点火, 7为熄火	5
EFF	数据是否有效: 1为有效, 0为无效	1

行状态,通过无线通信装置以约每15~60 s不等的的时间间隔上传至后台控制中心存储,当日未运营车辆GPS数据待其开机后传输至控制中心。本次采集的数据包括广州市1个月和佛山市1星期的出租汽车GPS数据,平均每天约产生数据1亿条、20G容量。

### 1.3 异常数据处理

出租汽车GPS数据在从终端传输到控制中心过程中可能存在一些误差和错误,需要进行预处理、过滤和剔除,从而提高原始数据的可靠性和精确性。一般主要包括三个方面:1)GPS终端设备故障导致数据出现接收时间延迟和错报;2)出租汽车进入隧道、地下停车场以及高层建筑密集区等,GPS终端发射信号时被遮挡导致数据传输或信号中断、信息无法上传;3)驾驶人在短时间内重复打表产生无效或异常数据。

## 2 主要算法

### 2.1 地图匹配

地图匹配是一种常见的定位修正方法,基本思路是通过特定算法(本文采用最短距离算法),将GPS轨迹数据与实际数字地图进行匹配,实现车辆空间属性与GIS信息的对应,以确定车辆所处的路段与位置,保证车辆的运行轨迹或者位置真实地反映出来。首先,基于GIS道路网制作缓冲区,将出租汽车行驶轨迹数据匹配到缓冲区中,并剔除缓冲区外轨迹点(漂移点);其次,通过比较GPS目标点到与其相邻路段的距离及角度,将目标点投影到匹配度最高的弧段上,并依据校正后的GPS轨迹点分析车辆的运行特征、时空规律等(见图1)。

匹配度计算是否准确直接决定着匹配效果,需要考虑距离和方向两个因素。距离是在一定范围内有多条可能路段的前提下,需要考虑哪一条路段与GPS点位置最接近;角度主要考虑GPS数据中方向字段与道路的夹角,夹角越小匹配度越高。计算公式为

$$\Omega = \int(\overline{d}, \theta) = \epsilon_d \times \overline{d} + \epsilon_\theta \times \overline{\theta}$$

利用ArcGIS工具的近邻分析功能,计算GPS点与各候选路段的距离;以正北方向角为0°,顺时针旋转构建角度度量,计算GPS点的方向与各候选路段之间的夹角;最终计算得到匹配度。具体分析流程框架如图2所示。

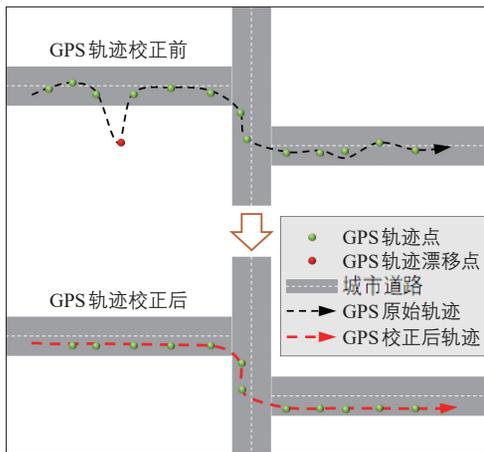


图1 出租汽车GPS轨迹点校正思路

Fig.1 Differential correction of taxi GPS trajectory data

## 2.2 出行起讫点识别

出租汽车起讫点的识别就是搜索车辆状态从4变为5和从5变为4的信息记录。采用C#语言与Oracle数据库结合，其中Oracle数据库用于GPS数据和车辆行程数据的存储，C#语言用于GPS数据中的逻辑判断及行程数据提取(见图3)。主要步骤如下：

- 1) 从数据库中提取EFF=1的有效数据，基于车牌号码和GPS时间对数据进行排序；
- 2) 读取出租汽车的所有GPS记录，从第一条记录开始进行循环；
- 3) 当出租汽车载客状态从4变为5时，认为一次载客行为开始，记为乘客的上车点，当出租汽车载客状态从5变为4时，认为该次载客行为结束，记为乘客的下车点；
- 4) 将上下车记录之间GPS轨迹记为该次载客行程中途驶过的路径；
- 5) 将出行行程起讫点和轨迹数据存储于数据库的OD表和重车轨迹表中。

## 3 广州市出租汽车出行空间分布特征

### 3.1 上下客流热点区域

通过GPS数据分析可知，广州市出租汽车上下客流点主要分布在两个区域：1)城市主要干路，增槎路—东风路—黄埔大道、环市路—中路大道、内环路—燕岭路—天源路、东晓路—内环路、新港路、新滘路、广州大道、天河东路—猎德大桥—江海大道、科韵路、机场路等10条通道(见图4)；2)交通枢纽、重要商圈以及三甲以上的医院附近，例如广州火车站、广州东站、体育西路商圈、天河路商圈、珠江新城商务区、广东省人民医院等重点区域附近。以广东省人民医院客流来源地为例，出租汽车客流来源地不仅包括先烈路、中山路等周边地区的就医出行，而且包括广州站、广州东站、天河客运站、广园客运站等区域(见图5)。从出租汽车客流的来源地角度分析可知，广东省人民医院不仅服务于广州市本地市民，而且服务于其他城市市民的就医需求<sup>[7]</sup>。

### 3.2 出行OD分布特征

从行政区角度来看，广州市中心六区内部出行量为23.8万车次·d<sup>-1</sup>，占比57.6%，天河、白云、越秀、海珠4个区的出行量较高，日均出行总量为3.5万车次·d<sup>-1</sup>以上，占

比分别为18.3%，11.7%，10.9%，8.7%；并且越秀、天河地处城市核心，与各区交换量都在1.2万车次·d<sup>-1</sup>以上。从街(镇)角度来看，出租汽车日均出行量超过1000车次·d<sup>-1</sup>的出行集中分布于天河商圈、海珠区内部以及白云区各组团间，其中冼村街道内部出租汽车日均出行量最高(2441车次·d<sup>-1</sup>)，其次为赤岗街道2051车次·d<sup>-1</sup>，广州南站(石壁街)与万博商圈(南村镇)、白云机场与天河商圈均有较强交流。数据表明，交通枢纽与商

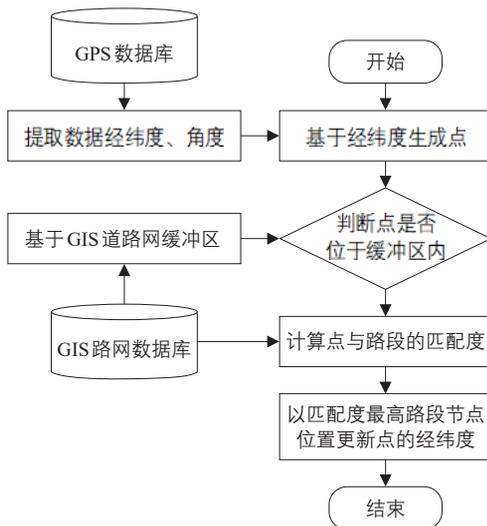


图2 GPS数据地图匹配流程

Fig.2 Process of GPS data map matching

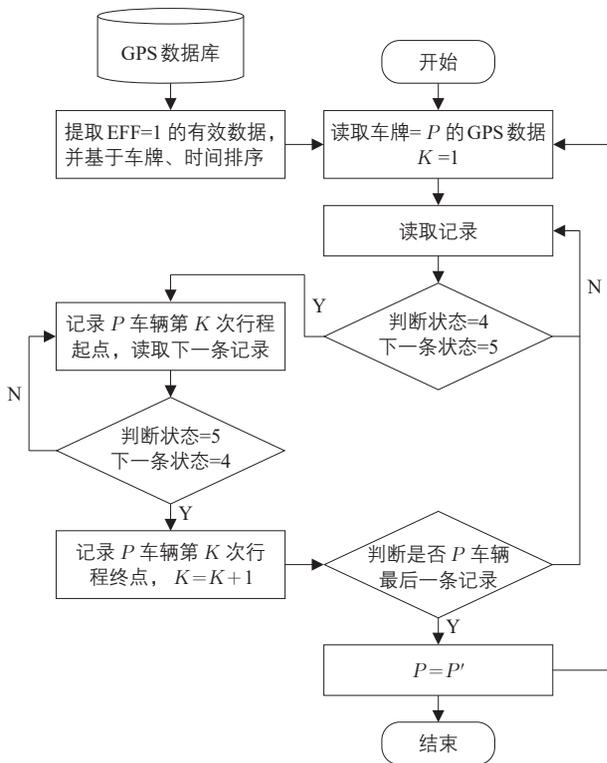


图3 出行起讫点的识别流程

Fig.3 Identification of travel origin and destination

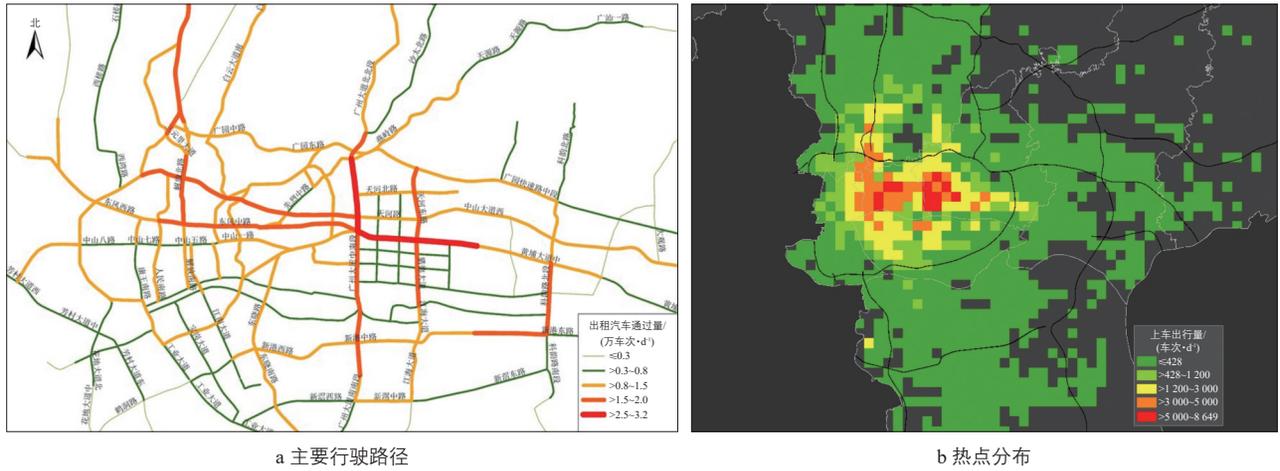


图4 广州市出租汽车出行主要行驶路径和热点分布

Fig.4 Major taxi driving routes and hot spots in Guangzhou

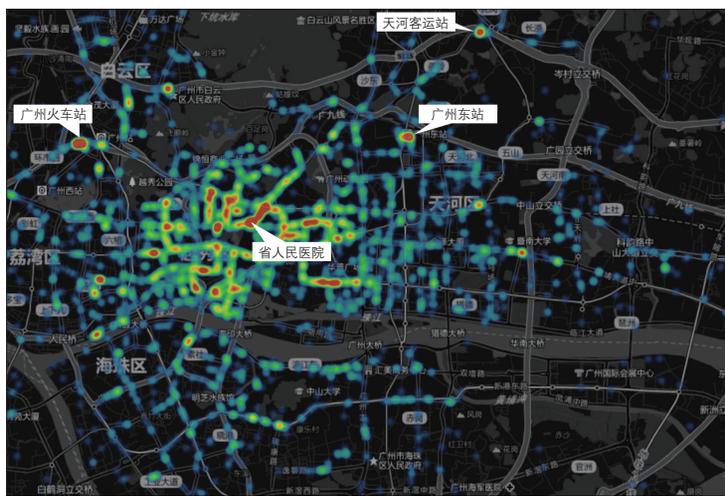


图5 广东省人民医院就医出租汽车客流来源热点分布

Fig.5 Distribution of taxi hot spots for passengers travel to Guangdong Provincial People's Hospital

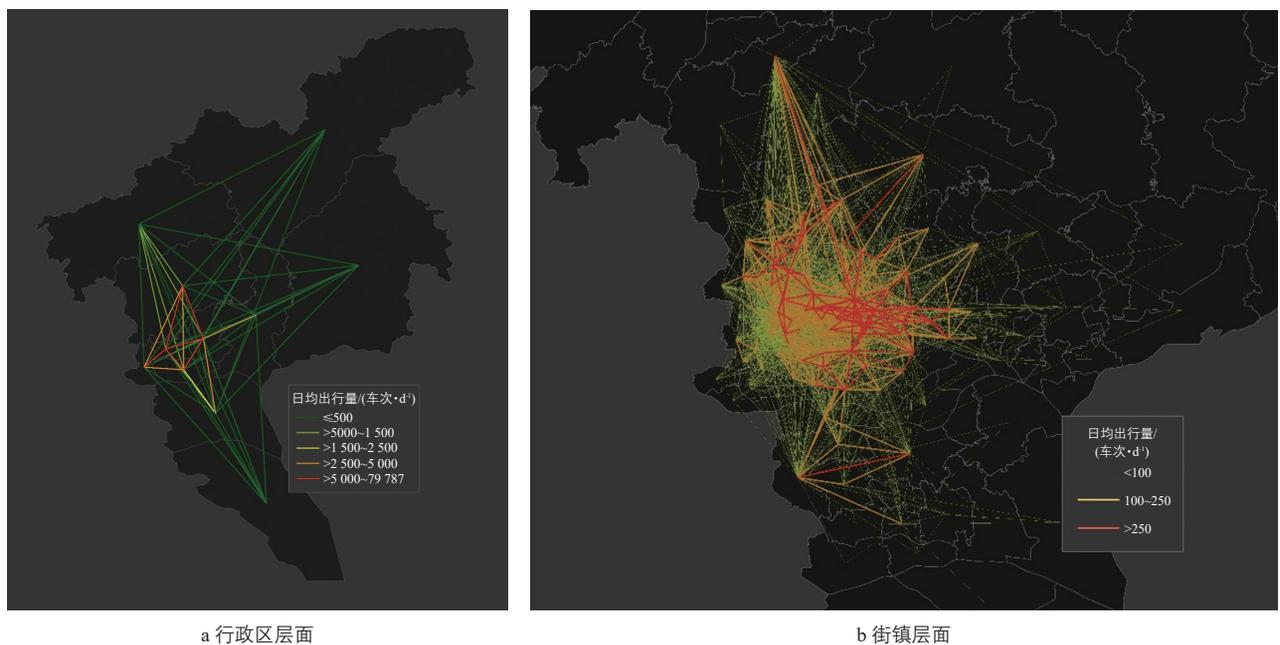
圈之间有较强的联系强度(见图6)。

#### 4 出租汽车GPS数据在交通规划和运营管理中的应用

##### 4.1 规划研究：广佛出租汽车返程点布局规划

由于广州、佛山两市出租汽车实行分区运营，出租汽车仅提供起点或终点位于运营管理行政区域内的客运服务<sup>①</sup>，即两市出租汽车载客到达对方城市可前往回程候客点配客回程，但由于回程配客点设置数量少，并且两市收费标准不统一，难以满足广佛出租汽车现状出行需求。

根据广佛两市出租汽车GPS数据分析，



a 行政区层面

b 街镇层面

图6 出租汽车出行OD分布

Fig.6 Taxi OD distribution

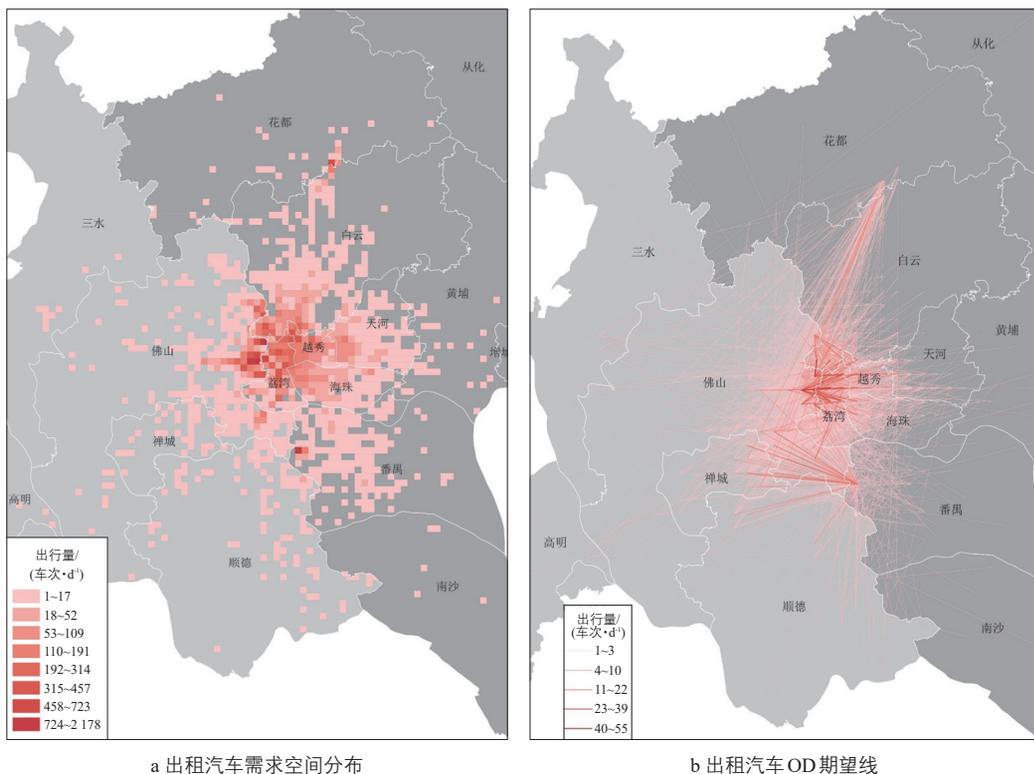


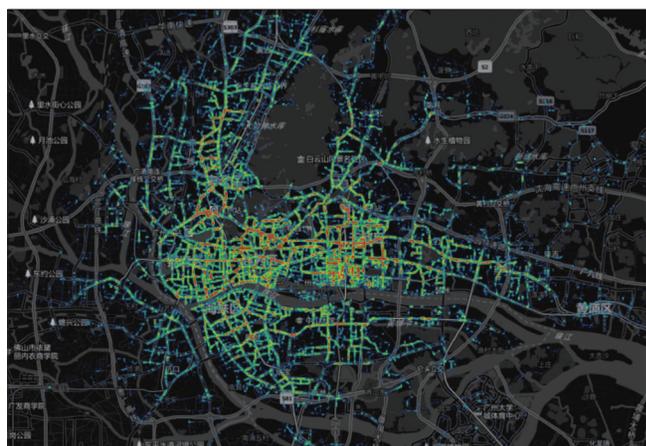
图7 广佛两市出租汽车需求空间分布和OD期望线

Fig.7 Travel demand distribution and OD desire line by taxi between Guangzhou and Foshan

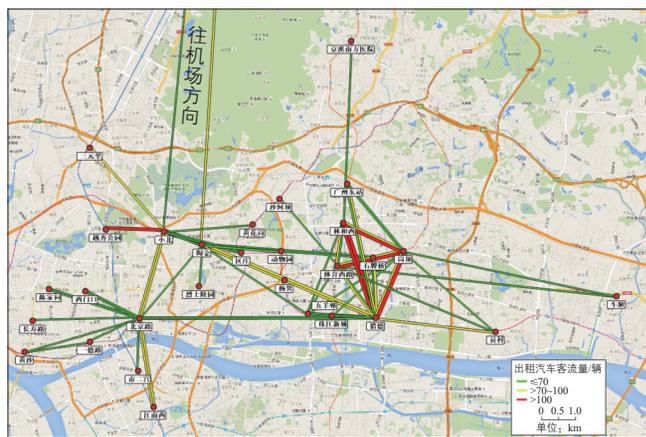
从客流方向分布来看，每天从佛山至广州方向的车辆数为1 203 车次·d<sup>-1</sup>，是返回佛山方向的3倍(日均407 车次·d<sup>-1</sup>)；每天从广州至佛山方向的车辆数为2 612 车次·d<sup>-1</sup>，是返回广州方向的1.5倍(日均1 724 车次·d<sup>-1</sup>)。从空间分布来看：1)需求集中在广佛两市毗邻(接壤地段)地区；2)需求集中于白云国际机场、火车站(广州南站、广州站、广州东站)、芳村客运站等交通枢纽以及地铁站(浔峰岗、黄沙、西朗等)；3)需求呈点、线状分布，点状为主要交通枢纽，线状集中在花地大道、广佛路、龙溪大道等路段。根据出租汽车需求和空间分布，结合现状出租汽车返程点设置，规划共计6个出租汽车回程候客点，包含佛山2个、广州市4个，覆盖佛山市和广州市重要综合交通枢纽，基本满足广佛两市出租汽车停靠和市民的城际出行需求(见图7)。同时，建议广佛两市建设统一的电召或信息平台，实现两市往返叫车信息互通且平台统一派车<sup>[8]</sup>。

#### 4.2 运营管理：夜班公共汽(电)车线路优化调整

通过对广州市出租汽车GPS数据挖掘分析，梳理夜间(22:00 至次日 2:00)居民出行集聚区域、客流热点、OD 期望线及客流走廊，在此基础上结合现状线路规划夜班公共



a 上下客热力分布



b OD期望线

图8 夜间出租汽车的上下客热力分布和OD期望线

Fig.8 Hot spots distribution of taxi loading and unloading and OD desire line at night

汽(电)车线路。具体步骤为: 1)对夜间出租汽车出行客流热点和聚集区域进行归集(见图8), 共有153个出租汽车客流聚集点, 其中客流生成量较大(超过400人次·h<sup>-1</sup>)的出行点共计13个; 2)以13个较强的客流生成点为基点, 形成联系强的OD(超过100人次·h<sup>-1</sup>)共计66对, 客流出行路径主要分布在环市西路、三元里大道、广园路、花城大道、中山大道、新港路等路段; 3)依托夜间客流走廊, 结合出租汽车客流出行强度较高的OD以及现状夜班公共汽(电)车线路, 共计新增6条夜班公共汽(电)车线路(见图9)。

#### 4.3 创新服务: 如约巴士+出租汽车等新模式应用尝试

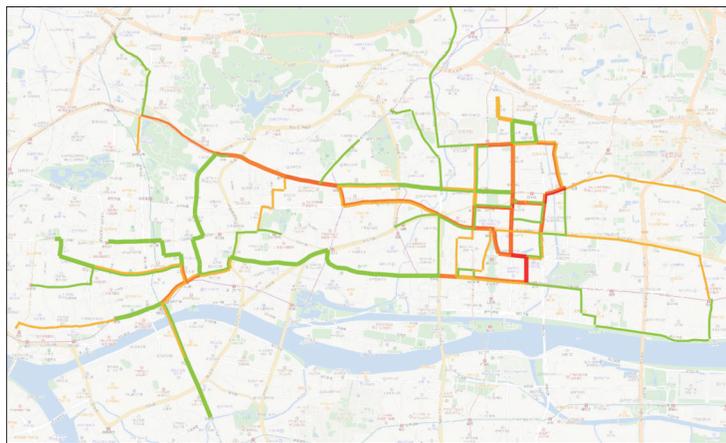
为打造安全春运、有序春运、温馨春运, 保障广州南站节后返程高峰期间交通组织有序, 为市民夜间出行提供多种交通方式选择, 及时疏运夜间到达旅客, 2017年春运期间, 广州市交通部门采取如约巴士+出租

汽车接驳新模式。主要思路为: 利用出租汽车GPS数据分析春运期间地铁停运后, 从广州南站(区位上处于郊区位置)出发前往中心区的客流通道和目的地分布(见图10), 在此基础上开行如约巴士<sup>®</sup>线路。从位置上看, 出租汽车乘客的目的地主要集聚在学院、天朗明居、天河软件园等附近大型居住区; 从路径上看, 主要有3条路径: 1)体育东路—林和东路—天寿路—广园快速路; 2)体育东路—天河路—中山大道; 3)体育东路—天河路—环市东路—机场路—广州大道北。2017年1月31日(正月初四)—2月13日(正月十七)夜间(地铁停运后), 开行广州南站至天河体育中心东门、广州火车站、广州火车站3条直达如约巴士, 并在体育中心东门如约巴士下客点设置出租汽车候客点, 让乘客下车即能换乘出租汽车, 进一步解决乘客夜间出行需求。经统计, 3条线路开行期间运输旅客共计4 004人次, 日均286人次·d<sup>-1</sup>, 最高峰在2017年2月2日(正月初六), 当晚运输旅客共计548人。

2018年2月20日至3月5日夜间共14天(正月初五至正月十八), 为返程旅客提供高铁(机场)+如约巴士(地铁)+出租汽车的快速出行服务。主要包括三个方面: 1)广州南站如约巴士+出租汽车夜间运输服务, 即开行广州南站至广州东站、南岗、大沙地、东平横岗村、广云路、中山八路、广州火车站、花都客运站、南海车站等9条如约巴士线路, 并在体育中心东门、骏景花园、中山八路、广州火车站等地设置出租汽车接驳点; 2)广州南站地铁二号线+出租汽车夜间运输服务, 即每天夜间22:30起, 在地铁昌岗站、公园前站、广州火车站设置3个出租汽车接驳点; 3)白云国际机场机场巴士+出租汽车夜间运输服务, 即每天夜间23:00至航班结束开通白云国际机场至体育中心东门、昌岗站、广州火车站等3条夜间直达巴士线路, 并在体育中心东门、昌岗站、广州火车站设置出租汽车接驳点<sup>[9]</sup>。

## 5 结语

本文在总结出出租汽车GPS数据预处理和主要算法的基础上, 重点分析广州市出租汽车上下客流热点分布以及出行空间分布等特征, 从规划研究、运营管理、创新服务三个方面, 总结同城化背景下广佛出租汽车返程



a 客流走廊



b 规划线路

图9 夜间出租汽车的客流走廊和规划夜班公共汽车线路

Fig.9 Taxi service corridor and night bus lines planning

点布局规划、广州市公共汽(电)车夜班线路优化,以及针对春运返程旅客夜间出行方式单一问题创新性提出如约巴士+出租汽车等新模式应用尝试,希望对其他城市利用出租汽车GPS数据进行数据挖掘及其在交通规划和运营管理中的应用提供启示。在未来研究中,可以考虑从不同时间(如极端天气、重大节假日)、不同空间(如出租汽车休息场地及停车场布局、不同路段在不同时间的打车概率)以及不同维度交叉融合分析(如出租汽车GPS与公共交通IC卡多源数据融合挖掘)角度,对出租汽车的运行特征和空间分布特征进行挖掘,从而为后续城市规划、交通规划以及运行管理方面提供强有力的数据支撑。

注释:

Notes:

- ① 广州、佛山两市出租汽车分别实行分区运营:广州市中心六区、番禺区、花都区、南沙区、增城区、从化区各区域分区运营;佛山市的禅城区、顺德区、南海区、高明区、三水区各区域分区运营。由于各区计费标准不一,两地出租汽车在运送乘客时收取的费用有差别。两市共设置19个出租汽车回程候客点,其中广州市境内3个,分别位于滘口客运站、坑口地铁站(近芳村客运站)和广州南站;佛山市境内16个,与广州接壤的南海区最多(有8个)、顺德区3个、禅城区高明区各2个,三水区1个;16个候客点大都位于广佛主干路的交叉口、各个大型商场以及客运站场附近。
- ② 如约巴士本质上是一种定制公交或定制巴士,实行一人一座制。2018年2月20日至3月5日(正月初五至正月十八)夜间,广州市如约巴士的票价为50 km以内5元·人<sup>-1</sup>·次<sup>-1</sup>、50 km以上10元·人<sup>-1</sup>·次<sup>-1</sup>,服务时间至末班高铁到达后20 min。

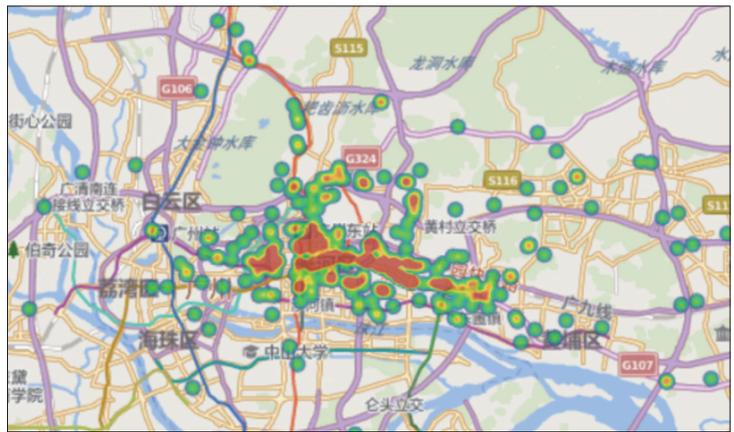
参考文献:

References:

- [1] 唐炉亮,郑文斌,王志强,等.城市出租汽车上下客的GPS轨迹时空分布探测方法[J].地理信息科学学报,2015,17(10):1179-1186.  
Tang Luliang, Zheng Wenbin, Wang Zhiqiang, et al. Space Time Analysis on the Pick-Up and Drop-Off of Taxi Passengers Based on GPS Big Data[J]. Journal of Geo-information Science, 2015, 17(10): 1179-1186.



a 客流通道



b 目的地分布

图10 春运期间地铁停运后广州南站至中心区的出租汽车客流通道和目的地分布  
Fig.10 Taxi service corridor and destination distribution from Guangzhou South Railway Station to city center after subway discontinuing service during the Spring Festival

- [2] 周勃,秦昆,陈一祥,等.基于数据场的出租汽车轨迹热点区域探测方法[J].地理信息科学学报,2016,32(6):51-56.  
Zhou Qing, Qin Kun, Chen Yixiang, et al. Hot-spots Detection from Taxi Trajectory Data Based on Data Field Clustering[J]. Journal of Geo-information Science, 2016, 32(6): 51-56.
- [3] 辛飞飞,陈小鸿,林航飞.浮动车数据路网时空分布特征研究[J].中国公路学报,2008,21(4):105-110.  
Xin Feifei, Chen Xiaohong, Lin Hangfei. Research on Time Space Distribution Characteristics of Floating Car Data in Road Network [J]. China Journal of Highway and Transport, 2008, 21(4): 105-110.
- [4] Chen Min. The Support of Floating Car Data for the Policy Decision and Planning on Taxi Industry[C]//The Proceedings of the 14th World Congress on ITS. Beijing: Chinese International Broadcasting and Press, 2007: 57-62.

(下转第8页)