

基于多源数据的网络约租车与出租车运营特征分析

林鹏飞¹, 翁剑成¹, 刘文韬², 徐 硕², 尹宝才¹

(1. 北京工业大学 交通工程北京市重点实验室, 北京 100124; 2. 北京市运输管理技术支持中心, 北京 100073)

摘 要: 提出了以出租车计价器数据、出租车 GPS 数据与网络约租车订单数据为基础的运营强度、时空分布等运营特征指标分析模型, 构建了基于分区均衡系数的出租车需求空间分布评价指标。结果表明: 网络约租车的日均载客次数、载客里程以及载客时间明显低于出租车, 运营强度仅为传统出租车的 40%, 两者次均载客里程、载客时间无明显差异; 网络约租车订单主要集中在早晚高峰以及夜间时段, 约占总订单量的 50% 以上, 其服务热点载客区域与出租车服务对应的热点区域大部分吻合, 对城市外围区域出租车供给紧张区域补充作用显著, 高峰时段均衡系数最高可提升 9%。本文可对出租车行业的科学管理和提升服务提供有效支撑。

关键词: 网络约租车; 出租车; 计价器数据; GPS 定位数据; 运营特征

中图分类号: U 491

文献标志码: A

文章编号: 2096-3432(2020) 01-26-08

Operational Characteristics Analysis of Ride-hailing Services and Traditional Taxis based on Multi-source Data

LIN Pengfei¹, WENG Jiancheng¹, LIU Wentao², XU Shuo², YIN Baocai¹

1. Key Lab of Traffic Engineering, Beijing University of Technology, Beijing 100124, China

2. Beijing Transport Management Technical Support Center, Beijing 100073, China

Abstract: This paper is based on multi-source data such as taximeter data, taxi GPS data and ride-hailing service data, and proposes the operational characteristic index analysis model including operation intensity and spatial-temporal distribution, and establishes the spatial distribution index of taxi demand based on a regional balance coefficient. The research shows that the average daily operation mileage, operation time and transaction frequency are significantly lower for ride-hailing services, and their operation intensity is only 40% of that of traditional taxi. There is no obvious difference between the ride-hailing services and traditional taxi in the average carrying mileage and carrying time of each time. The orders of ride-hailing services are mainly concentrated in the morning, evening peaks and nighttime hours, accounting for more than 50% of the total order volume. The pick-up hotspots of ride-hailing services cover the most hotspots of the traditional taxi. The ride-hailing services have a significant complementary effect on the periphery region where the taxi is in short supply, and the balance coefficient of spatial distribution in these regions also could be highest increased by 9% in peak hours. This paper provides an important support for the scientific management and service quality improvement of the taxi industry.

Key words: ride-hailing services; taxi; taximeter data; taxi GPS data; operational characteristics

收稿日期: 2019-09-14.

基金项目: 国家自然科学基金国际合作项目资助项目 (No. 61420106005) 和北京市“科技新星”计划项目 (No. Z171100001117100).

作者简介: 林鹏飞(1993—), 男, 硕士, 研究方向为交通大数据智能处理技术, E-mail: pengfei_lin@qq.com.

0 引言

城市出租车客运系统是城市公共交通的重要组成部分,是城市轨道交通和常规公交的重要补充,提供了比传统公共交通运输服务更舒适、更个性化的居民出行服务。目前,北京市出租车约6.8万辆,出租车客运量占城市公共交通客运量的7%左右^[1]。Uber、滴滴专车等互联网约租车(以下简称网约车)作为“互联网+交通”一种创新融合,解决了乘客与出租车司机之间信息不对称的问题,有效缓解了乘客“打车难”的问题。截至2016年底,全国已有近400个城市提供网约车服务。

在出租车运营特征分析方面,李艳红等^[2]在出租车OD数据分析基础上,构建了出租车出行特征、出行时间分布及出行空间分布的指标体系和分析方法。翁剑成等^[3]通过对大量出租车的计价器与GPS数据的预处理和分析,完善面向出租车运营管理应用的广域浮动车数据处理方法体系,提出基于浮动车数据的出租车运营管理指标计算模型。庄立坚等^[4]依托大规模浮动车数据的分析和预处理,分析出租车GPS数据的特征,提出面向出租车运营管理的大规模浮动车数据处理框架,并综合出行需求、运营效益和运营效率,构建基于浮动车数据的出租车运营管理指标体系及其计算模型。扈中伟^[5]基于出租车运营信息数据分析和实际调查,研究了北京市出租车运营的时空分布特点,重点对城市出租车合理规模和驾驶员劳动强度2个专题进行了研究。Leng等^[6-8]分别以北京、厦门和义乌的出租车计价器数据为基础,在网约车不同发展阶段,利用出租车载客次数、里程利用率、司机收入等指标变化间接分析网约车对出租车行业的影响。

针对网约车的运营特征分析,国内外相关学者重点关注了运营管理、运营特征等方面的定性分析。汪光焘等^[9]总结国内外打车软件的主要区别,指出

中国打车软件的管理更为复杂、更容易引起社会矛盾,并建议在国家层面提出指导性意见。Fischer-Baum等^[10]通过研究2014—2015年4—6月曼哈顿地区Uber和出租车全部订单数据发现,Uber在曼哈顿并没有造成接单量的增加,相反拼车服务在市中心取代了出租车,同时对热点区域的乘车需求进行了补给。高永等^[11]通过将问卷调查数据扩样测算结果显示约租车的无序发展是导致城市小汽车出行量增加、局部区域(时段)拥堵加剧的重要原因之一。

目前基于实际交易数据的网约车运营特征的系统分析相对较少,并缺乏与传统出租车在运营属性、运营时空特征等方面的差异量化分析。本研究依托出租车计价器数据、出租车GPS数据与网约车订单数据等多源数据,提出了面向大规模运营数据处理的数据预处理流程以及运营特征指标模型,从运营强度、订单时空分布特征等角度剖析出租车、网约车2类个体机动车出行服务车辆的运营特征差异。

1 多源数据基础

1.1 数据基础

出租车计价器数据、GPS数据以及网约车的订单数据记录了车辆运营过程的时空分布信息,是出租车和网约车运营特性对比分析的重要数据源。统计分析2015年9月10个典型工作日数据,出租车数据为全样本数据,北京市6.7万辆出租车平均每天交易次数约90~110万次,每天回传的GPS数据约1亿5000万条;网约车数据为抽样数据,包含了1万名司机约20万条网约车订单数据。

出租车计价器数据包括了出租车司机的基本信息、每一次交易的乘客上下车时间、等待时间、行驶距离、空驶距离、金额等信息。计价器通常定期回传一次,数据具有滞后性,但运营数据记录更为准确。

目前北京全市出租车都搭载了GPS定位模块,

表1 出租车计价器数据属性

字段名	说明	字段名	说明
记录ID	2885451093	空驶里程	60
交易类型	9	司机标志	284892
交易金额	1300	上车天数	0
交易时间	2015-09-11 10:33:17	上车时间	102711
等候时间	0002	车牌号	BJXXXX
行驶里程	14	单价	200

车辆以 30 ~ 60 s 时间间隔回传 GPS 数据,包括出租车的时间、位置坐标和角度等基本信息,以及出租汽

车运营中的状态(如空载、满载、驻车、停运等)信息,能够详细记录出租汽车的实时运营状态。

表 2 出租车 GPS 数据属性

字段	说明	字段	说明
ID	527672720	纬度	40.029 713
终端号	133XXXXXX	速度	28
车牌号	BJXXXX	运营状态	1
GPS 时间	2015-09-11 0:01:02	方向	326
经度	116.341 637		

网约车软件均基于移动定位技术,网约车运营数据包含司机、乘客的基本信息,还包括乘客下单、

司机抢单、乘客上下车时间、起终点位置、行驶距离、金额等信息。

表 3 网约车订单数据

字段	说明	字段	说明
司机编号(已加密)	C772DB20	终点经度	116.460 012
乘客编号(已加密)	A944EC43	终点纬度	39.955 664
订单创建时间	2015-09-18 09:48:31	起点地址	东城区安定门东大街小区
订单状态	1	终点地址	朝阳区霄云路天元港中心
城市名称	北京市	成功抢单时间	2015-09-18 09:48:50
当前经度	116.418 738	下车时间	2015-09-18 10:02:55
当前纬度	39.948 964	乘客支付费用	24.8
起点经度	116.418 396	订单距离	4.2
起点纬度	39.949 224	上车时间	2015-09-18 09:55:22

1.2 数据预处理与分析流程

出租车计价器数据、GPS 数据以及网约车的订单数据在采集过程中存在异常数据,为了提高数据质量要对多源数据进行预处理。出租车计价器数据存在的问题主要是载客距离、载客时间过长或过短等,研究采用阈值法剔除错误数据与异常数据。出租车 GPS 数据存在的问题主要有坐标重复、缺失、漂移等。本研究利用 GPS 生成时间与计价器中上、下车时间进行匹配,选取时间最接近的经纬度数据作为上下车的位置,允许时间匹配误差在阈值内,即认为匹配成功。网约车数据除了存在载客距离、载客时间异常的问题之外,还存在着少量的无效订单,即司机未抢单的情况,采用阈值法剔除错误数据与异常数据。经过预处理及数据融合后,分别提取出租车与网约车的运营特征,从运营强度和订单时空分布特征 2 个角度进行对比分析。多源数据处理与分析流程如图 1 所示。

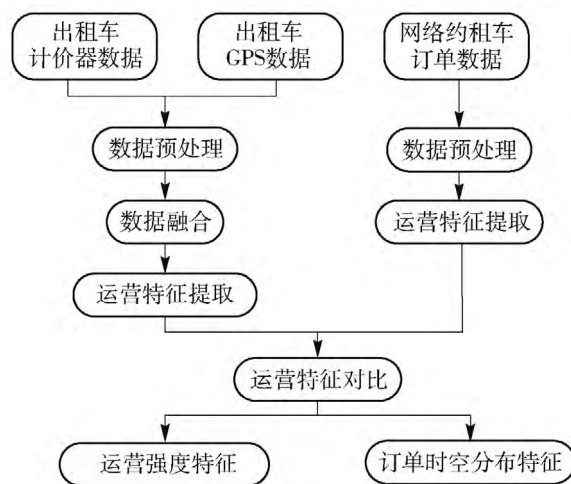


图 1 多源数据处理与分析流程

2 运营特征指标构建

结合多源数据基础,根据网约车与出租车 2 类

营运车辆特点,遵照系统性、可操作性、差异性的指标选取原则,研究构建了运营强度指标、订单时空分布指标2类运营特征指标。

2.1 运营强度指标

运营强度指标主要包含载客时间、载客里程、载客次数3项指标,这3项指标可以准确表征2类运营车辆运能运量水平及自身运行特点,同时能间接反应居民的出行选择行为是否存在差异。

载客时间表征出租车司机及网约车司机单日的平均劳动强度等,指标内容包括日均载客时间、次均载客时间如式(1)(2):

$$\text{日均载客时间: } T = \frac{\sum t_i}{m}; \quad (1)$$

$$\text{次均载客时间: } \bar{t} = \frac{\sum t_i}{\sum n_i}; \quad (2)$$

式中 t_i 为第 i 辆车单日总载客时间; n_i 为第 i 辆车单日载客次数; m 为单日所有运营车辆的数量, $i=0, 1, 2, \dots, m$ 。

载客里程表征两类营运车辆对道路资源的占用情况,指标内容包括日均载客里程、次均载客里程如式(3)式(4):

$$\text{日均载客里程: } L = \frac{\sum l_i}{m}; \quad (3)$$

$$\text{次均载客里程: } \bar{l} = \frac{\sum l_i}{\sum n_i}; \quad (4)$$

式中 l_i 为第 i 辆车单日总载客里程。

载客次数一方面可以表征司机的劳动强度,另一方面可以反应区域的交通出行需求,客次数越多,表明打车需求量越大。指标内容包括日均载客次数、车均载客次数等如式(5):

$$\text{日均载客次数: } C = \frac{\sum N_i^{\text{on}}}{m}; \quad (5)$$

式中 N_i^{on} 为第 i 辆车单日载客次数。

2.2 订单时空分布特征指标

订单在不同时段上的分布,客观上反映了城市居民的生活节奏以及乘客对出租车、网约车的交通需求在时间上的分布的差异性。以时间为横轴,纵向分析24h内不同时段统计分析订单分布,从而量化的掌握乘客对两类出租车需求强度的差异。

居民出行的空间分布特征具有一定的规律性。工作日居民出行多以通勤行为为主,周末节假日居民出行以休闲娱乐目的为主。订单的空间分布特征即将某个统计分析区域分为若干个单位网格,分别

统计每个单位网格中上车和下车次数。本研究提出分区均衡系数用于表征出租车和网约车订单在空间分布的差异性。分区均衡系数越接近于1,说明订单的空间分布越均衡。为保证函数有意义,在上车次数为0的单位网格利用0.0001代替0。均衡系数如式(6):

$$E = \frac{\sum_{j=1}^n x_j \log_2 x_j}{\log_2 \left(\frac{1}{n} \right)} \quad (6)$$

式中 x_i 为某区域内第 i 个单位网格内上车次数占该区域上车次数的百分比; n 为某区域内单位网格的个数。

3 运营特征分析

对出租车计价器数据与网约车的订单数据进行频率分布统计,95%的订单载客里程、载客时间分别在50 km和144 min以内,将这2个值作为载客里程、载客时间的阈值,将超出阈值的出租车计价器与网约车记录剔除。本研究中,研究范围为北京六环覆盖的区域,将超出范围的出租车GPS数据与网约车订单数据删除,同时剔除出租车GPS数据中的漂移点、重复点。GPS数据与计价器数据时间关联匹配的阈值为 ± 30 s。将统一坐标系后的订单起终点位置匹配到对应的分析区域中。

3.1 运营强度特征分析

出租车与网约车的日载客次数的频率统计结果显示,出租车的日均载客次数约为20次,网约车的日均载客次数约为8次,前者约为后者2.5倍。网约车的日载客次数频率最高的范围在0~5次,运营强度较低。北京市部分出租车实行“双班”的运营模式,因此载客次数高于网约车。出租车和网约车的日载客次数分布情况如图2所示:

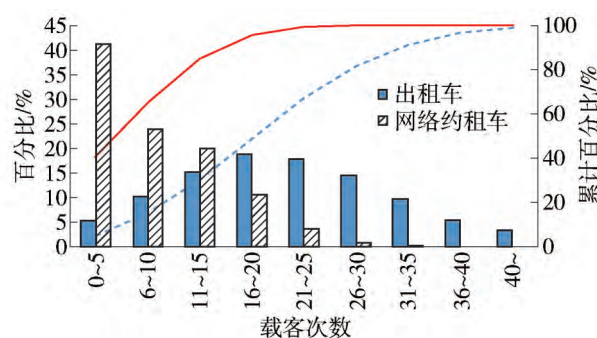


图2 日载客次数频率分布

图3 为出租车和网约车的日载客时间分布情

况,网约车的日均载客时间明显低于出租车,反映网约车驾驶员的工作劳动强度相对较小。出租车的日均载客时间为 6.7 h,而网约车的日均载客时间分别为 2.9 h,不足出租车日均载客时间的一半,约占 50% 的网约车的日均载客时间主要集中在 2 h 以内。

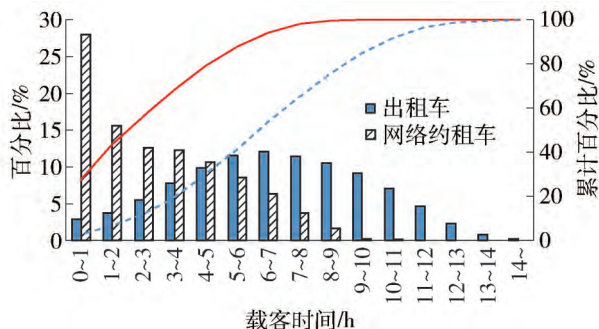


图3 日载客时间频率分布

图4为出租车和网约车的日载客里程分布情况,出租车的载客里程分布的偏度为 0.2,近似为正态分布,而网约车的分布的偏度分别为 1.1,属于偏态分布,说明大部分网约车的日载客里程较短。出租车日均运营里程为 178.9 km,网约车的日均载客里程为 71.9 km,80% 的网约车的每天的载客里程在 120 km 以内。

对出租车、网约车的次均载客里程、载客时间进行频率分布分析,采用箱型图描述出租车与网约车次均载客里程、载客时间分布,如图5所示。出租车和网约车的次均载客里程分别为 8.9 km 和 9.5 km,均以中短途距离为主,四分位距分别为 8.7 km、8.9 km。就距离影响因素而言,乘客对差异化服务模式的选择倾向并不明显。两者的次均载客时间分布差异较大,次均载客时间分别为 21 min 和

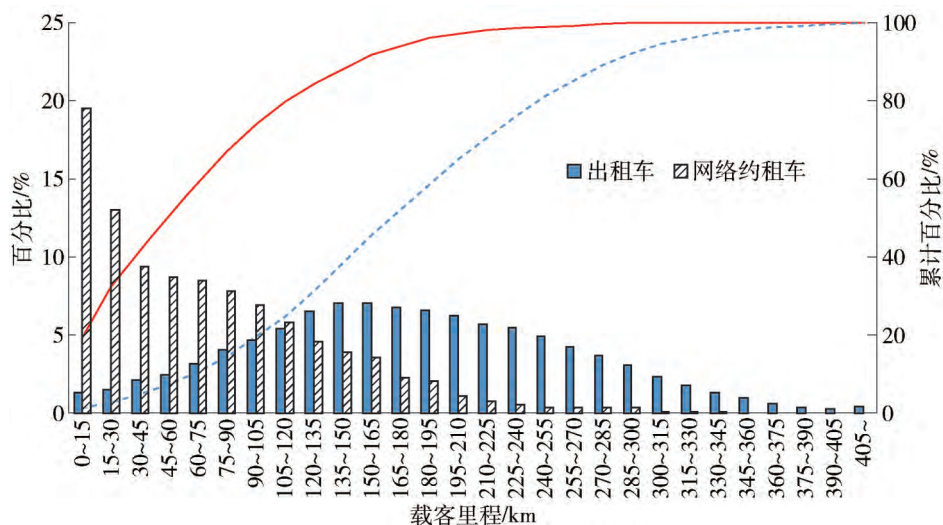


图4 日载客里程频率分布

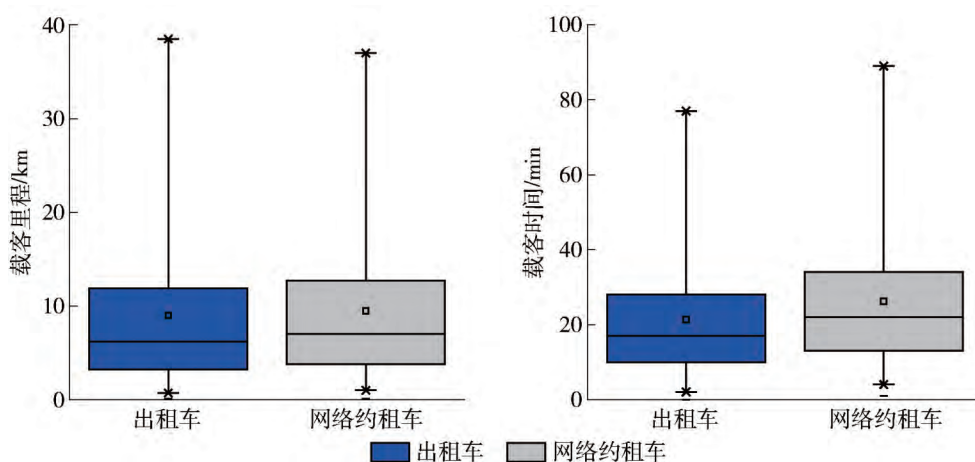


图5 次均载客里程、时间分布

26 min,这主要与订单的时空分布、载客区域的实时路况有关。

3.2 订单的时空分布特征分析

3.2.1 订单时变特征

订单时变特征分析即是以时间为横轴,时间区

间为0.5 h,分析订单数随时间的变化。因抽样后网约车与出租车的订单分布绝对数量差异较大,为了便于分析,以百分比为纵坐标,实现两类运营车辆的归一化处理。

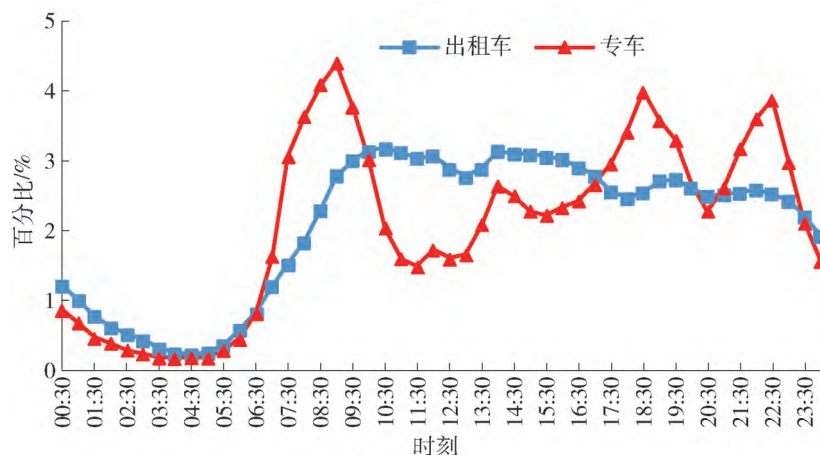


图6 订单时变特征

出租车订单分布相对平稳,高峰特征不明显。出租车的需求低谷期出现到上午07:00以前,在08:00—10:00上班早高峰期间,出租车需求量明显上升并达到需求峰值,在10:00以后出行需求有所回落,在随后的下午和晚上出行需求量基本维持在一个较高水平,10:00—23:00订单量占全天的72%。

网约车订单的时间分布则较为集中,尤其在早晚高峰时段形成较为明显的增长,基本与路网车流量早晚高峰时段重合,2个时段的总订单量接近全日订单的40%左右。另外,夜间时段(21:00至23:00)专车订单比例相对较高,夜间时段与晚高峰形成明显的“双高峰”的时间分布特征,与同一时期出租车的特征形成显著差异。网约车全天高峰时段的订单量约占总订单量的50%以上。

3.2.2 订单空间分布特征

为了研究出租车与网约车订单分布的时空差异,用ArcGIS软件生成 0.01×0.01 的经纬度网格覆盖北京市六环以内的区域,共计3484个单元网格。将出租车、网约车订单数据中的上、下车位置与生成的网格进行空间连接分析,分别统计每个单位网格中的上车和下车数量,由于三者订单绝对数量差异较大,对统计结果采用Jenks自然间断点法分类,进行归一化处理,上、下车热点区域空间分布如图7所示。

出租车的上车区域与下车区域的空间分布基本保持一致,热点载客区域主要集中于市中心,热度由

内向外递减。较密集的地区主要包括北京西站、北京南站、北京站以及首都机场等对外交通枢纽,订单量约占总量的10%。网约车服务热点载客区域与出租车服务对应的热点区域大量重合,在呼家楼—国贸区域、中关村、国贸CBD、东单附近、西单商圈等商业办公区,其服务人群空间分布集中度显著高于出租车,在一定程度上缓解了这些热点区域的出行压力。除此之外网约车出现了一些新的热点区域,如来广营、望京以及西二旗附近、通州北苑附近区域存在的上车热点,网约车在空间上弥补了部分出租车运力较为紧张的区域。为了进一步分析网约车对出租车的补充作用,分别计算在不同时段、不同空间范围网约车引入前后的均衡系数,结果如表4所示:

在市中心区域两类运营车辆载客区域大部分吻合。各时段引入网约车前后,二环以内区域均衡系数变化不明显,表明网约车的载客区域与出租车的载客区域重合度较高。

市区外围区域均衡系数的改善效果明显,网约车与出租车存在补充关系。从全天的均衡系数来看,出租车的订单空间分布由内到外不均衡程度逐渐加重,五、六环之间的区域订单分布最不均衡,均衡系数仅为0.8089,而引入网约车后,除二三环之间各个区域中的均衡指数均有明显增加。早高峰时段,均衡系数改善明显的区域在二三环之间、四六环之间;晚高峰和夜间时段均衡系数改善明显区域则主要分布在四六环之间,与图7中订单空间分布特

征吻合,这些区域的均衡指数分别在早、晚高峰和夜间时段最高提升了 9.0% 5.7% 和 5.8%,网约车对

城市外围区域(四环外)出租车供给紧张区域的补充作用最为明显,有效缓解了“打车难”问题。

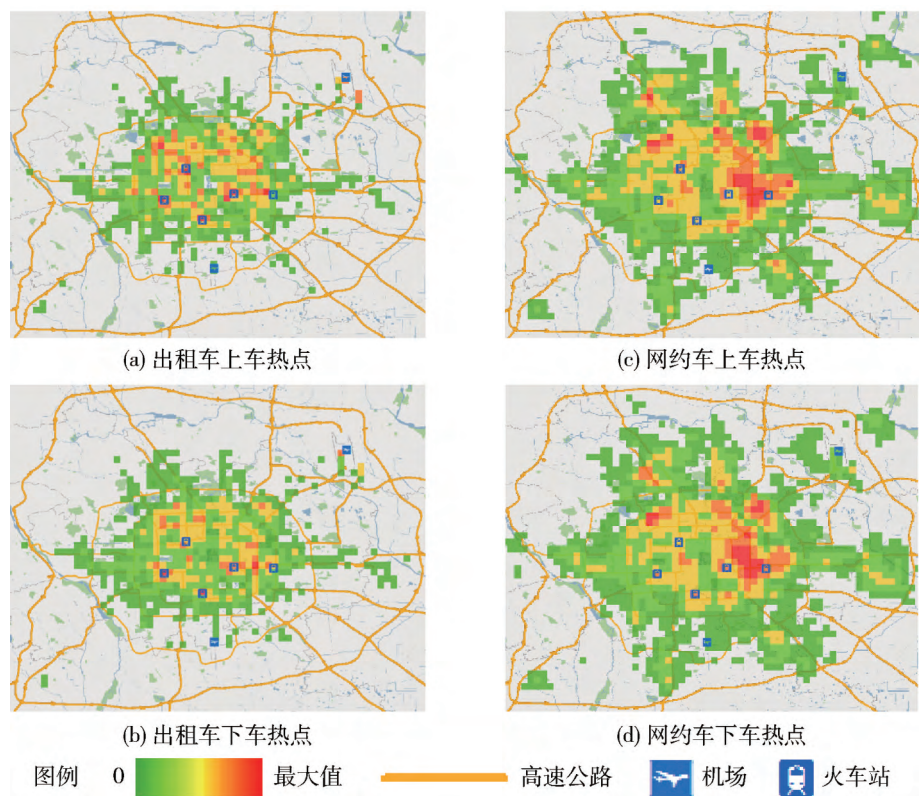


图 7 订单空间分布特征

表 4 网约车引入前后订单空间分布均衡系数变化

空间范围	早高峰		晚高峰		夜间时段		全天	
	引入前	引入后	引入前	引入后	引入前	引入后	引入前	引入后
二环以内	0.964 6	0.974 7	0.961 9	0.955 5	0.950 7	0.936 5	0.966 2	0.966 5
二三环之间	0.969 4	0.972 6	0.969 3	0.966 5	0.958 8	0.915 2	0.969 3	0.966 5
三四环之间	0.871 8	0.864 2	0.876 2	0.873 7	0.870 1	0.857 4	0.878 6	0.880 4
四五环之间	0.735 0	0.801 4	0.705 3	0.734 1	0.686 9	0.701 0	0.848 5	0.852 1
五六环之间	0.831 7	0.896 7	0.782 0	0.826 3	0.760 4	0.804 5	0.808 9	0.823 5
最大提升比例/%	9.0		5.7		5.8		1.8	

4 结束语

本文利用北京市部分工作日的出租车计价器、GPS 数据和网约车订单数据,从运营强度指标、订单时空分布 2 个方面构建运营特征指标模型。两者的运营特征差异可总结为如下 4 点:

1) 网约车个体的日运营强度明显低于出租车。从日均运营里程、日均载客时间和日均载客次数的角度来看,工作日网约车的日运营强度明显低于出租车,约为出租车的 40% 左右。

2) 出租车和网约车的次均载客特征差异性并不显著。两者次均载客里程为 8.9 km 和 9.5 km,均以中短途距离为主。

3) 网约车订单集中在早晚高峰及夜间时段。从订单的时变特征来看,出租车在全日的订单分布相对平稳,网约车订单在早晚高峰及夜间时段形成较为明显的增长,订单量占全天订单的 50% 以上,有效缓解了“打车难”的问题。

4) 网约车订单的空间分布更集中,成为出租车的有益补充。引入分区均衡系数衡量两者订单的空

间分布均衡性,引入网约车后市中心区域均衡系数变化不明显,表明网约车订单与出租车服务对应的热点区域大量重合,在市区外围“打车难”的区域,高峰时段均衡系数最高可提升9%,网约车形成对出租车有效补充。

本研究可以帮助政府和行业管理部门更好地了解网约车和出租车的运营特征,尤其是在网约车合法化的背景下,为网约车监管、出租车行业改革提供有效的数据支撑。

致谢

感谢滴滴政策研究院和北京市交通运行监测调度中心对本研究提供的数据支持。

参考文献:

- [1] 北京交通发展研究院. 2016 北京市交通发展年度报告. 北京, 2016.
- [2] 李艳红, 袁振洲, 谢海红, 等. 基于出租车 OD 数据的出租车出行特征分析[J]. 交通运输系统工程与信息, 2007, 7(5): 85-89.
- [3] 翁剑成, 刘文韬, 陈智宏, 等. 基于浮动车数据的出租车运营管理研究[J]. 北京工业大学学报, 2010(6): 779-784.
- [4] 庄立坚, 韦清波, 何兆成, 等. 基于浮动车数据的出租车运营管理特性建模与分析[J]. 重庆交通大学学报(自然科学版), 2014, 33(4): 122-127.
- [5] 扈中伟. 基于浮动车技术的北京市出租车运营管理研究[D]. 北京: 北京工业大学, 2007.
- [6] Leng B, Du H, Wang J, et al. Analysis of taxi drivers' behaviors within a battle between two taxi apps[J]. IEEE Transactions on Intelligent Transportation Systems, 2015, 17(1): 1-5.
- [7] Hai D, Yang Y, Li W, et al. How do chauffeured cars affect the taxi market? a case study in Xiamen, China[C] // Transportation Research Board 95th Annual Meeting, 2016 (16-4784).
- [8] Sun Z, Yu M, Zeng J, et al. Assessment of the impacts of app-based ride service on taxi industry: evidence from Yiwu City in China[C] // Transportation Research Board 95th Annual Meeting, 2017(17-06426).
- [9] 汪光焘, 齐彤岩, 陈小鸿, 等. 关于加强打车软件综合管理的建议[J]. 城市交通, 2014, 12(5): 1-3.
- [10] Reuben Fischer-Baum and Carl Bialik. Uber Is Taking Millions of Manhattan Rides Away From Taxis [EB]. [2015-12-09]. <https://fivethirtyeight.com/features/uber-is-taking-millions-of-manhattan-rides-away-from-taxis/>
- [11] 高永, 安健, 全宇翔. 网络约租车对出行方式选择及交通运行的影响[J]. 城市交通, 2016, 14(5): 1-8.