欧阳铭亮, 刘心雨, 张宇遨. 融合多源数据的出租车运营时空特征分析[J]. 智能计算机与应用,2026,16(8):61-67. DOI: 10.20169/j.issn.2095-2163.24121603

融合多源数据的出租车运营时空特征分析

欧阳铭亮[1], 刘心雨[1], 张宇遨[2]
(1 上海理工大学 管理学院, 上海 200093; 2 同济大学 交通学院, 上海 201804)

摘 要: 本文融合巡游出租车运营轨迹数据和网约车订单数据,旨在分析两类出租车的运营时空特征。首先提出了一套网约出租车和巡游出租车的数据标准化处理流程,并从供给、供需、盈利三个层面提出 28 个指标分析出租车时空特征。以上海市出租车数据为案例进行研究,时间特征从日变、时变、高峰三个时间区间进行分析,空间特征从市域、跨行政区和区内三个空间维度进行分析,结果发现:网约车的运营效率整体高于巡游车;在时间特征上,巡游车和网约车都呈现明显的早晚高峰,网约车相比较巡游车的价格竞争优势主要体现在平峰和夜间时段;在空间特征上,巡游车和网约车活跃热点区域均以上海市中心区域为主。

关键词: 巡游车; 网约车; 时空特征; 指标工具箱
中图分类号: U492.4 **文献标志码**: A **文章编号**: 2095-2163(2026)08-0061-07

Spatial and temporal characteristics of taxi operation based on multi-source data

OUYANG Mingliang[1], LIU Xinyu[1], ZHANG Yu'ao[2]
(1 Business School, University of Shanghai for Science and Technology, Shanghai 200093, China;
2 College of Transportation, Tongji University, Shanghai 201804, China)

Abstract: This paper integrates the operational trajectory data of conventional taxis and ride-hailing order data, aiming to analyze the operational spatiotemporal characteristics of the two types of taxis. First, a standardized data processing procedure for ride-hailing and conventional taxis is proposed. Additionally, 28 indicators are introduced from three perspectives—service provision, supply-demand, and operational profitability—to analyze the spatiotemporal characteristics of taxis. Using the taxi data from Shanghai as a case study, the temporal characteristics are analyzed daily, hourly, and peak hours. The spatial characteristics are analyzed from three spatial dimensions: citywide, inter-administrative districts, and district. The results show that the operational efficiency of ride-hailing cars is generally higher than that of conventional taxis. Regarding temporal characteristics, both conventional taxis and ride-hailing cars exhibit distinct peak hours in the morning and evening. The competitive pricing advantage of ride-hailing cars is most evident during off-peak and nighttime periods compared to conventional taxis. As for spatial characteristics, the hotspots of activities for both conventional taxis and ride-hailing cars are primarily concentrated in the central areas of Shanghai.
Key words: conventional taxis; ride-hailing cars; spatial and temporal characteristics; indicator toolbox

0 引 言

近年来,巡游出租车和网约出租车共同构成了城市公共交通的重要组成部分,其服务特征和运营效率对城市交通管理和居民出行具有重要影响。因此,构建一套能够分析巡游车与网约车时空特征的数据分析方法和指标工具箱,能够深入剖析两类车辆在不同时空维度下的服务特征差异,为城市交通规划和管理提供决策支持。

目前对巡游车和网约车的相关研究主要分为以下方面:张学东等[1]探讨了厦门市巡游车与网约车的运营效益和时空变化特征,发现巡游车运营强度高,而网约车服务范围广;杨燕等[2]比较了成都巡游车与网约车的载客空间特征,发现两者在热点区

作者简介: 欧阳铭亮(2002—),男,硕士研究生。主要研究方向:公共交通。
通信作者: 刘心雨(1995—),女,博士,讲师,硕士生导师。主要研究方向:公共交通管理与服务,交通数据挖掘,共享交通。Email:liuxinyu@usst.edu.cn。
收稿日期: 2024-12-16

域和空间相关性上存在相似性与差异;洪少枝等[3]对比了厦门市巡游车与网约车的运营特征,发现巡游车空驶率高,而网约车单位时间收入更高;胡松等[4]基于厦门市数据,研究了疫情对巡游车与网约车运营特征的影响,指出网约车在长距离出行中更受欢迎;路庆昌等[5]以上海市为例,分析了出租车空载率的时空特征及其影响因素;刘涵等[6]利用兰州市出租车 GPS 数据,分析了出租车上下客的时空分布规律;Li 等[7]分析了成都市网约车需求的时空变化,揭示了不同旅行距离下的需求特征及建成环境的影响;Zhang 等[8]提出了一种分析北京出租车时空特征的方法,为城市出租车运营管理提供了动态监测和决策支持;汪洪涛等[9]提出了一种考虑行驶路径的出租车 GPS 数据地图匹配算法;覃正桃等[10]基于轨迹数据对出租车交通运行特征进行了研究。

上述研究表明巡游车和网约车在运营效益和时空分布上存在显著差异,这些差异对城市交通规划和管理具有重要影响。因此,亟待形成一套围绕两类出租车运营特征分析的标准化流程、并在已有研究基础上进一步拓展对两类车辆时空分布规律和运营效益的理解,为城市交通政策制定提供更科学的依据[11-16]。

1 数据收集与处理

数据处理过程共包括 4 个主要步骤,分为数据清洗、时间序列位置提取、地图匹配、轨迹重构,数据处理过程如图 1 所示[17-20]。

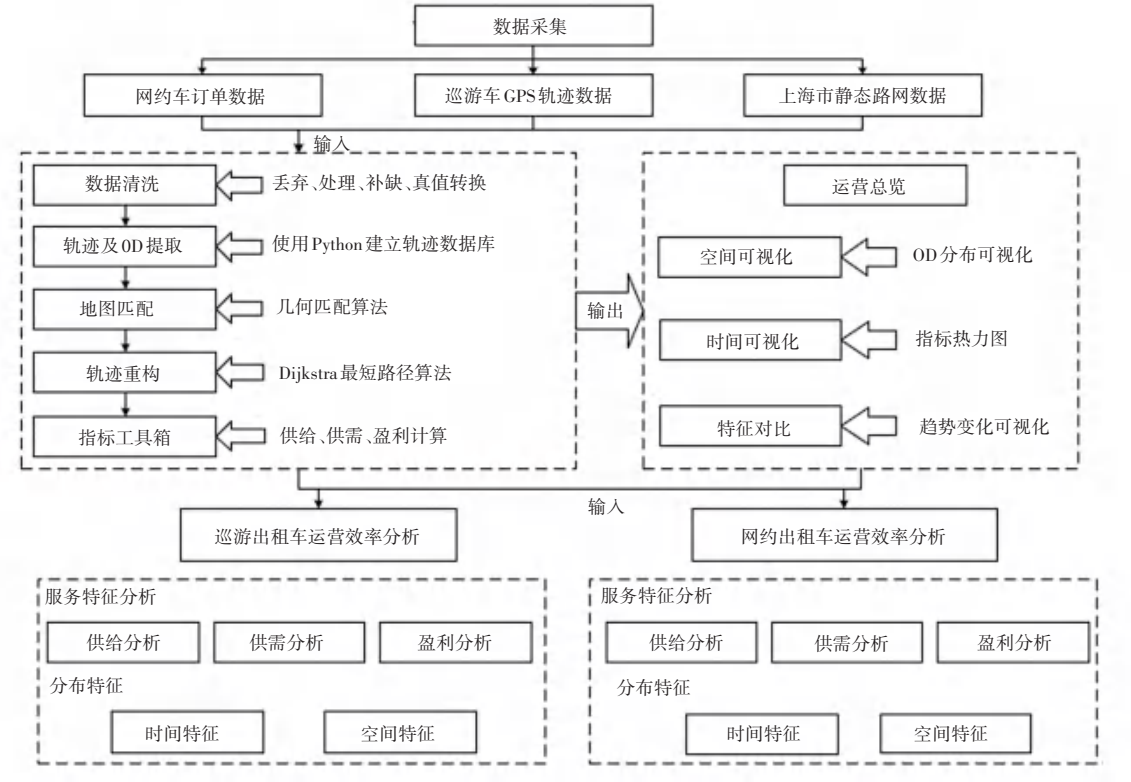


图 1 数据处理流程

Fig. 1 Data processing process

1)数据清洗

巡游出租车 GPS 数据具有覆盖面广、字段丰富及全天候等优点,但原始出租车 GPS 数据由于在传输过程中受到隧道、高层建筑物体等多方面的影响,存在一定比例的错误数据、缺失数据及无效数据。因此,有必要对数据进行预处理以方便下一步研究。在进行网约车运营效率和服务质量分析前,由于原始数据存在无效记录、部分字段信息无效、数据来源不同、格式不统一等问题,需要先对获取到的网约车历史订单数据进行预处理。

2)时间序列位置提取

出租车运营轨迹数据是分析上海市巡游出租车出行特性和时空分布模式及道路运行状况的数据基础,因此,需构建出租车运营轨迹数据库。网约车原始数据一般采用 YYYY-MM-DD HH:MM:SS 的时间格式记录时间,字母分别代表年、月、日、时、分、

秒。在应用过程中，为便于直观显示和统计分析，本研究将时间格式转换为时间戳（Unix timestamp，Unix）格式，该格式以格林威治时间的 1970 年 1 月 1 日 0 时 0 分 0 秒为起点，记录从起点到当前时刻的总秒数来表示时间。本研究利用 Python 编程对时间格式进行了统一，并结合后期研究需求按不同时间时间切片，对不同时间片和行政区域切片的起终点数量进行了统计。

3）地图匹配

出租车运营轨迹数据虽然具有时间、空间等多重属性，但并未与道路信息相关联，同时，由于系统误差，部分出租车运营轨迹数据的经纬度信息存在 20 m 左右的误差。因此，需要进行运营轨迹数据与路网的匹配分析以满足后续关键节点及关键路段的识别需求。地图匹配一般遵循两个步骤：一是确定与定位点对应的路段；二是确定定位点对应路段上的具体位置。运用几何分析法进行地图匹配。网约车订单数据也需要同巡游车运营轨迹数据一样，完成定位点到路网的匹配，以满足后续关键节点及关键路段的识别需求。

4）轨迹重构

本研究根据巡游出租车和网约车订单数据获取 OD（Origin-Destination，起点－终点）后，对每对 OD 计算最短路径，以便于后续与实际轨迹数据进行对比。其中最短路径是根据 OD 点生成的最优路径选择，轨迹匹配则对应实际出租车的轨迹路线。对于网约车轨迹的上下车点提取，需要将订单数据描述的出发地与目的地的地址分别赋予轨迹的上车点与下车点，再对两个坐标点之间的 OD 进行最短路径构建。

2　时空特征分析指标

2.1　指标分类

本文中所涉及的统计指标（统计指标工具箱）如图 2 所示，主要分为三类：供给指标、供需指标和盈利指标[21]。

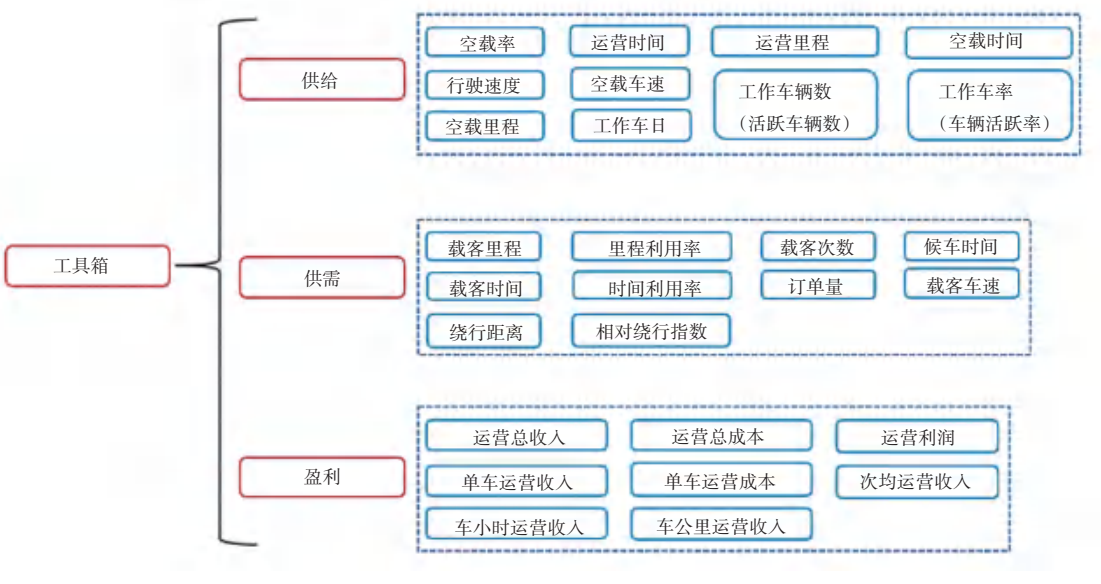


图 2　统计指标工具箱

Fig. 2　Statistical indicator toolbox

2.2　指标计算

三类指标的计算公式见表 1。

3　案例分析

本文以上海市的巡游出租车和网约车数据作为案例分析，并进行工作日和非工作日划分，从时间特征和空间特征两个角度进行。

3.1　运营效率对比

将巡游出租车和网约车的供需指标进行对比，结果如图 3 所示。巡游出租车在车均载客里程、载客时间、订单量、载客速度以及载客次数方面均领先于网约车；但网约车的车均里程利用率和时间利用率显著高于巡游出租车。造成这种差异的原因主要有两点：一方面，巡游出租车主要通过随机巡游的方式揽客，这使得其行驶里程和时间相对较长，且空驶情况较多，也因此载客里程较长；另一方面，网约车借助平台实现高效匹配乘客与司机，极大地减少了空驶里程，进而提高了车均里程利用率和时间利用率。

表 1 指标计算表

Table 1 Index calculation table

指标类别	指标名称	计算公式	符号说明
供给指标	空载率	$\eta_m = \frac{M_e}{O_d}$	η_m 是空载率，M_e 是空载里程，O_d 是运营里程
	运营时间	$O_t = T_p + T_e$	O_t 是运营时间，T_p 是载客时间，T_e 是空载时间
	运营里程	$O_d = M_p + M_e$	O_d 是运营里程，M_p 是载客里程，M_e 是空载里程
	空载时间	$T_e = \sum_{i=1}^{n} T_{ei}$	T_e 是空载时间，T_{ei} 是车辆 i 的空载时间，n 是车辆数
	车辆活跃率	$\mathrm{ACI} = \frac{\mathrm{TACar}}{\mathrm{TCar}}$	ACI 是车辆活跃率，TACar 是总活跃车辆数，TCar 总车辆数
	行驶速度	$V = \frac{O_d}{O_t} \times 60$	V 是行驶速度，O_d 是运营里程，O_t 是运营时间
	空载里程	$M_e = \sum_{i=1}^{n} M_{ei}$	M_e 是空载里程，M_{ei} 是车辆 i 的空载里程，n 是车辆数
	空载车速	$V_e = \frac{M_e}{T_e} \times 60$	V_e 是空载车速，M_e 是空载里程，T_e 是空载时间
	工作车日	$\mathrm{WVD} = \sum_{i=1}^{n} \mathrm{WVD}_i$	WVD 是总工作车日，WVD_i 是第 i 天的工作车日数，n 是统计周期
	活跃车辆数	$\mathrm{TACar} = \sum_{i=1}^{n} \mathrm{ACar}_i$	TACar 是总活跃车辆数，ACar_i 是第 i 天活跃车辆数，n 是统计周期
供需指标	载客里程	$M_p = \sum_{i=1}^{n} M_{pi}$	M_p 是载客里程，M_{pi} 是车辆 i 的载客里程，n 是车辆数
	里程利用率	$F^s = \frac{M_p}{O_d}$	F^s 是里程利用率，M_p 是载客里程，O_d 是运营里程
	载客次数	$\mathrm{TM} = \sum_{i=1}^{n} M_i$	TM 是总载客次数，M_i 是第 i 天的载客次数，n 是统计周期
	载客时间	$T_p = \sum_{i=1}^{n} T_{pi}$	T_p 是载客时间，T_{pi} 是车辆 i 的载客时间，n 是车辆数
	时间利用率	$F^h = \frac{T_p}{O_t}$	F^h 是时间利用率，T_p 是载客时间，O_t 是运营时间
	候车时间	$W_t = \sum_{i=1}^{n} W_{ti}$	W_t 是候车时间，W_{ti} 是第 i 个订单的候车时间，n 是订单数
	绕行距离	$D_d = M_p - D_s$	D_d 是绕行距离，M_p 是载客里程，D_s 是最短可能行驶距离
	订单量	$N = \sum_{i=1}^{n} o_i$	N 是总订单量，o_i 是第 i 天的订单数，n 是统计周期
	载客车速	$V_p = \frac{M_p}{T_p} \times 60$	V_p 是载客车速，M_p 是载客里程，T_p 是载客时间
	相对绕行指数	$\mathrm{RDI} = \frac{M_p}{D_s}$	RDI 是相对绕行指数，M_p 是载客里程，D_s 是最短可能行驶距离
盈利指标	运营总收入	$\mathrm{OI} = \sum_{i}^{n} \mathrm{OI}_i$	OI 是运营总收入，OI_i 是订单 i 的运营收入
	运营总成本	$\mathrm{OC} = \sum_{i=1}^{n} \mathrm{CC}_i + \mathrm{LC}_i + \mathrm{MC}_i$	OC 是运营成本，CC_i 车辆 i 的车辆成本，LC_i 是车辆 i 的人工成本，MC_i 是车辆 i 的管理成本，n 是车辆数
	运营利润	OP=OI−OC	OP 是运营利润，OI 是运营总收入，OC 是运营成本，单位是元
	车小时运营收入	$\mathrm{RVH} = \frac{\mathrm{OI}}{H \times \mathrm{TCAR}}$	RVH 是车小时运营收入，OI 是运营总收入，H 是活跃车辆的载客总小时数，TCAR 是活跃车辆数
	车公里运营收入	$\mathrm{RVK} = \frac{\mathrm{OI}}{K \times \mathrm{TCAR}}$	RVK 是车公里运营收入，OI 是运营总收入，K 是活跃车辆的载客总公里数，TCAR 是活跃车辆数
	单车运营收入	OI_i	OI_i 是第 i 辆车的运营收入
	单车运营成本	OC_i	OC_i 是第 i 辆车的运营成本
	次均运营收入	$\mathrm{ARPT} = \frac{\mathrm{OI}}{\mathrm{TM}}$	ARPT 是次均运营收入，OI 是运营总收入，TM 是载客次数

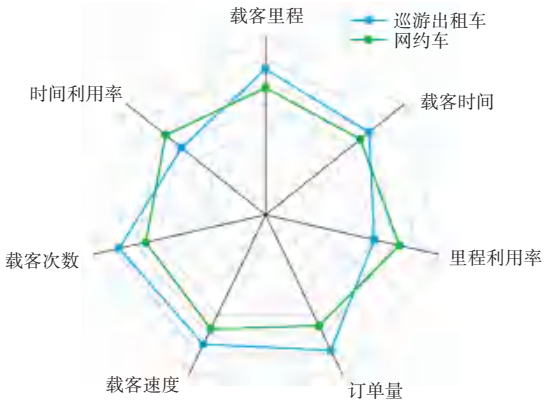


图 3　巡游出租车和网约车供需指标雷达图

Fig. 3　Radar chart of supply and demand indicators between cruising taxis and ride-hailing cars

3.2　时间特征

本文分别对巡游出租车和网约车在工作日和非工作日的车均载客里程进行分析，结果如图 4 所示。对于巡游出租车，工作日三天的载客里程逐渐递增，其中周五的载客里程最长。对于网约车，工作日与非工作日网约车车均载客里程区别比较小，非工作日中周六的载客里程均大于周日。将巡游车与网约车进行对比，可发现巡游车的载客里程大于网约车的载客里程，从整体上看，两类车型工作日的载客里程都大于非工作日的载客里程。

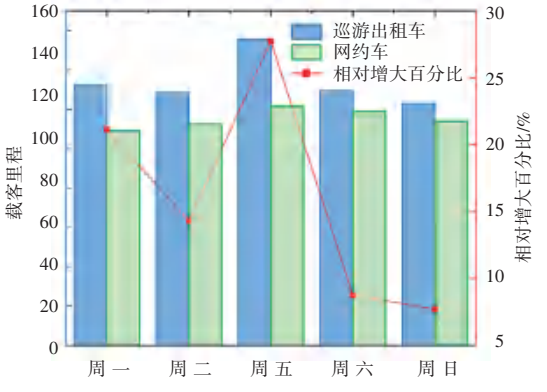


图 4　工作日与非工作日巡游出租车和网约车车均载客里程

Fig. 4　Average per-vehicle passenger mileage of cruising taxis and online car-hailing vehicles on weekdays and weekends

每小时巡游车和网约车车均载客里程变化见图 5，结果显示，两类车型里程的变化都有明显的高峰与低谷，且整体上看两者趋势大致相同。早高峰出现在 8：00–9：00，晚高峰出现在 17：00–18：00，夜间小高峰出现在 20：00–21：00，整体上网约车的载客里程高于巡游车。网约车工作日在 7：00–8：00 有明显的低谷，远小于非工作日。巡游车工作日和非工作日比较接近。

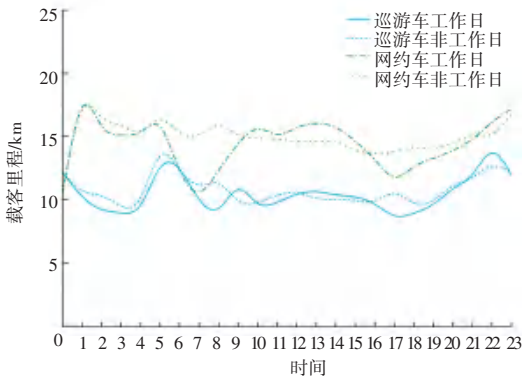


图 5　每小时巡游车和网约车车均载客里程变化

Fig. 5　Variation of hourly average per-vehicle passenger mileage of cruising taxis and online car-hailing vehicles on weekdays and weekends

3.3　空间特征

图 6 为工作日早高峰（7：00–9：00）巡游出租车活跃车辆数热力图，图 6 所示，市中心区域，尤其是黄浦区，活跃车辆数量最多，长宁区、静安区和徐汇区等其他市中心区域也处于较高水平。与此同时，宝山区、嘉定区等效区的活跃车辆较少，浦东新区的活跃车辆分布均匀。图 7 为工作日晚高峰（17：00–19：00）巡游出租车活跃车辆数热力图，可见市区的活跃车辆数量普遍下降，黄浦区仍为活跃车辆数量最多的区域。但黄浦区及其周边区域的活跃车辆数量也相对较高。



图 6　工作日 7：00–9：00 活跃车辆数

Fig. 6　Number of active vehicles from 7：00–9：00 on weekdays



图 7　工作日 17：00–19：00 活跃车辆数

Fig. 7　Number of active vehicles from 17：00–19：00 on weekdays

图 8 为非工作日早高峰(7：00-9：00)巡游出租车活跃车辆数热力图,活跃车辆主要集中在静安区、虹口区、杨浦区和黄浦区,与工作日的分布特征有所区别。图 9 为非工作日晚高峰(17：00-19：00)巡游出租车活跃车辆数的热力图,可见上海市巡游出租车的活跃车辆分布与工作日特征相似,黄浦区、徐汇区等市区的活跃车辆数量始终处于较高水平。



图 8 非工作日 7：00-9：00 活跃车辆数

Fig. 8 Number of active vehicles from 7：00-9：00 on non-working days



图 9 非工作日 17：00-19：00 活跃车辆数

Fig. 9 Number of active vehicles from 17：00-19：00 on non-working days

本文采用弦图来表示上海市各行政区间出租车 OD 流量的相互关系。在弦图中,圆环被划分为若干扇区,每个扇区代表上海市的一个行政区,每个扇区的宽度表示该行政区往其他行政区出行的出租车 OD 流量大小。图 10 为工作日巡游出租车 7：00-9：00 的弦图,可见该时段从长宁区、松江区、嘉定区、青浦区以及徐汇区等区域开往闵行区的巡游出租车数量较多,推测这与闵行区的商业活动或居民区分布有关。同时,由黄浦区、徐汇区、静安区、杨浦区等市中心区域开往浦东新区的巡游出租车也显示出较高的流量,这可能与浦东新区就业岗位较多有关。

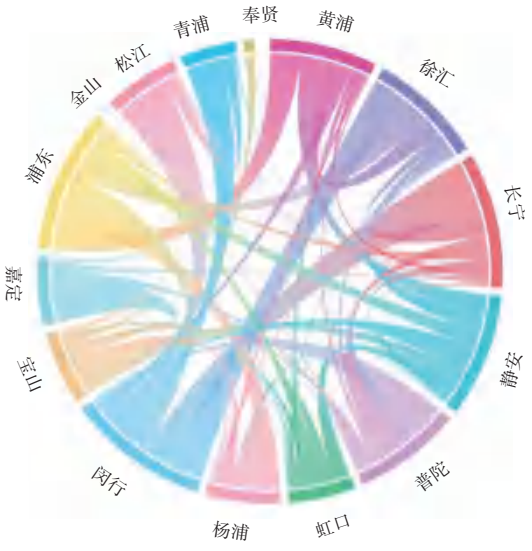


图 10 工作日 7：00-9：00 巡游车 OD 流量

Fig. 10 OD flow of cruising taxis from 7：00-9：00 on weekdays

4 结束语

本文提出了融合多源数据的出租车运营时空特征分析的框架,构建了包含供给指标、供需指标和盈利指标三类共 26 个指标工具箱。通过从时间特征和空间特征两个角度进行分析,得到以下结论：

(1)巡游出租车的空载率要高于网约车,而里程利用率、时间利用率都要低于网约车,即网约车的运营效率整体要高于巡游出租车。网约车的订单质量相比巡游出租车更高一些,巡游车服务的订单中短距离起步价所占比例更高。

(2)时间特征层面,巡游出租车和网约车都表现出明显的早晚高峰现象,巡游出租车和网约车在空载里程、空载时间、载客时间、载客里程、载客次数指标都呈现出明显的高峰和低谷。总体上使用网约车的花费比巡游出租车相对低一些,网约车相比较巡游车的价格竞争优势主要体现在平峰与夜间时段。

(3)空间特征层面,巡游出租车与网约车活跃热点区域均主要集中在上海市中心区域,并在工作日早、晚高峰时段呈现出差异化分布特征。

参考文献

[1] 张学东, 贾礼朋, 邓少存, 等. Gps 轨迹数据和 poi 数据约束的厦门市出租车与网约车运营特征研究[J]. 北京建筑大学学报, 2021, 37(4)：60-68.

[2] 杨燕, 胥川. 出租车和网约车载客空间特征分析[J]. 交通运输工程与信息学报, 2020, 18(1)：68-76.

[3] 洪少枝，刘新田. 巡游与网约车运营特征对比分析:以厦门市为例[J]. 综合运输，2022，44(6)：145-153.
[4] 胡松，林鹏飞，孙宇星，等. 多场景模式下网约车与巡游车运营特征差异及调度策略研究[J]. 交通工程，2021，21(4)：29-38.
[5] 路庆昌，王琴，徐标，等. 出租车空载率时空特征及影响因素分析:以上海市为例[J]. 地理研究，2024，43(1)：149-163.
[6] 刘涵，刘聪. 基于 gps 数据兰州市出租车上下客时空特征分析[J]. 黑龙江交通科技，2023，46(6)：124-126.
[7] Li X，Xu J，Du M，等. Understanding the spatiotemporal variation of ride-hailing orders under different travel distances[J]. Travel Behaviour and Society，2023，32：100581.
[8] Zhang J，Qiu P，Duan Y. A space-time visualization analysis method for taxi operation in beijing[J]. Journal of Visual Languages and Computing，2015，31：1-8.
[9] 汪洪涛，赵靖，冯文秀. 考虑行驶路径的出租车 GPS 数据地图匹配算法[J]. 物流工程与管理，2017，39(5)：82-85.
[10] 覃正桃，赵靖，王家儒，等. 基于 GPS 数据的出租车交通运行特征研究[J]. 中国水运(下半月)，2017，17(6)：70-72.
[11] 章程，赵靖，杨晓光，等. 城市快速路入口匝道控制研究综述[J]. 上海理工大学学报，2023，45(4)：332-344.
[12] 赵靖，葛庆红，韩印. 考虑需求起讫点及需求等级的响应型社区公交行车调度优化[J]. 公路交通科技，2020，37(6)：128-137.
[13] 赵靖，杨晓光，章程. 交通设计技术发展与对策建议[J]. 前瞻科技，2023，2(3)：45-57.
[14] 刘佳俊，唐阳山. 基于 MATLAB 的交通拥堵状态预测研究[J]. 智能计算机与应用，2024，14(9)：199-204.
[15] 吴振阳，邢杰. 基于多元回归和 GM 的新能源汽车保有量预测[J]. 智能计算机与应用，2024，14(10)：214-217.
[16] 董洁霜，毛伟航. 考虑时空异质性的新能源汽车保有量增长态势研究[J]. 智能计算机与应用，2024，14(11)：9-15.
[17] 姚婷，章莹莹，杨莉. 基于出租车轨迹数据的居民出行碳排放测算模型研究与应用[J]. 智能计算机与应用，2024，14(10)：70-78.
[18] 郑智泉，罗忠琛，常国艳，等. 数据填补算法的参数优化问题[J]. 智能计算机与应用，2024，14(10)：1-11.
[19] 胡天伟，余靖，刘国华，等. 关系型大数据查询的层次结构[J]. 智能计算机与应用，2023，13(12)：56-61.
[20] 滕硕. 大数据在智慧城市应急管理中的应用[J]. 智能计算机与应用，2021，11(4)：179-181.
[21] 全国城市客运标准化技术委员会. 城市客运经济技术指标计算方法：JT/T 1373.3-2021[S]. 2021.