

轨迹数据驱动的出租车出行特征分析与 时空分布建模

王君悦,董春娇,王 菁,王明智

(北京交通大学 综合交通运输大数据应用技术交通运输行业重点实验室, 北京 100044)

摘 要: 基于北京市出租车全球定位系统(GPS)轨迹数据,结合 ArcGIS 地图匹配功能,选定合适阈值构建了出租车出行信息数据库。首先,构建了出租车出行网络,结合核密度和热点分析方法,分析了出行轨迹空间分布特征和时序特征,揭示了出租车出行密度在早、晚高峰的变化趋势。然后,考虑土地利用性质对出租车出行分布的影响,建立了基于土地利用性质的双约束人口权重机会模型。结果表明,本文模型提高了出租车跨区出行的预测精度,均方根误差由 4.54 辆次降为 4.41 辆次,预测结果与真实数据的相似度由 0.961 提高为 0.970,发生量平均相对误差由 0.48% 降为 0.19%。

关键词: 城市交通;出行特征;人口权重机会模型;土地利用性质;出租车

中图分类号: U491 **文献标志码:** A **文章编号:** 1671-5497(2026)08-2106-10

DOI: 10.13229/j.cnki.jdxbgxb.20250061

Trajectory data-driven taxi travel characteristics analysis and spatial-temporal distribution modeling

WANG Jun-yue, DONG Chun-jiao, WANG Jing, WANG Ming-zhi

(Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport, Beijing Jiaotong University, Beijing 100044, China)

Abstract: Based on the global positioning system (GPS) trajectory data of taxis in Beijing and the map matching function of ArcGIS, appropriate threshold values were selected, a taxi travel information database was constructed, and the travel characteristics of taxis were analyzed. Firstly, a taxi travel network was built, and spatial and temporal distribution features of travel trajectories were analyzed using kernel density and hot spot analysis methods, revealing the changing trend of taxi travel density during peak hours. Then, considering the impact of land use on taxi travel distribution, a dual-constraint population weight opportunity model based on land use was established to predict taxi travel distribution. The results show that the model improves the prediction accuracy of taxi inter-district travel, with the root mean square error reduced from 4.54 trips to 4.41 trips, and the similarity between the predicted results and

收稿日期: 2025-01-17.

基金项目: 中央高校基本科研业务费专项资金项目(2019RC027).

作者简介: 王君悦(1999-),女,博士研究生.研究方向:城市交通工程理论与技术.E-mail: 22110286@bjtu.edu.cn

通信作者: 董春娇(1982-),女,教授,博士生导师.研究方向:交通运输规划与管理,交通安全.

E-mail: cjdong@bjtu.edu.cn

actual data improved from 0.961 to 0.970. The average relative error in the number of trips decreased from 0.48% to 0.19%.

Key words: urban traffic; travel characteristics; population weighted opportunity model; land use characteristics; taxis

0 引言

社会的快速发展使得出行需求日益增加,然而私家车数量远远小于有驾照的人数。道路资源容纳量有限,“有本无车”的现象愈发明显。共享出行模式的普及可以降低车辆使用成本。出租车因其分布广泛性与多目的出行特性,虽不完全符合严格意义上的共享出行标准,但仍可作为共享出行体系的重要补充形式^[1]。因此,研究出租车的出行分布特征对于实现共享出行具有重要参考价值。

已有出行行为分析多关注出行特征指标分布规律的研究。Jiang等^[2]分析了瑞典50辆出租车6个月的全球定位系统(Global positioning system, GPS)轨迹数据,发现乘客的移动距离服从幂律分布。Bazzani等^[3]通过分析意大利佛罗伦萨市区的车载GPS数据,发现车辆停留时间和不同地点访问次数服从幂律分布。Liang等^[4]分析了北京市1万多辆出租车的2000万条GPS轨迹数据,发现城市出租车的出行位移服从指数分布,经过时间也可以用指数分布近似拟合。

此外,部分学者考虑不同因素构建了不同角度的出行分布预测模型。Simini等^[5]结合人类空间移动特性提出了辐射模型,显著提高了预测精度。Riccardo等^[6]通过分析意大利地区2%的私家车GPS数据,发现两次出行之间的时间分布存在规律,研究了平均访问频率的排名分布并提出简单的随机模型。Peng等^[7]对上海市的出租车GPS轨迹数据进行分析,也得到类似的结果。Ren等^[8]提出基于成本的人类移动性辐射模型,并用美国人口普查和公路交通数据证实了该模型。Zhao等^[9]选取3类典型大规模数据集构建随机游走模型,模型同时纳入3类影响因素:回访偏好、惯性效应和对新地点的探索。Yan等^[10]将记忆效应和人口数量相结合建立一个通用模型,并基于人口分布准确预测人类的流动性。Liu等^[11]提出一个普遍机会模型,考虑探索性倾向和谨慎性倾向,并且涵盖经典的辐射模型和机会优先选

择模型。基于城际游客客源地分布规律较为相近的特性,陶思然等^[12]将旅游偏好变量引入基本重力模型中构建城际客运出行分布预测模型。马健霄等^[1]研究发现,土地利用混合度在早晚高峰时段对网约车和出租车的出行需求影响显著。陈启香等^[13]将土地利用属性作为特征变量,研究了出租车与地铁的竞争合作关系。邢雪等^[14]根据兴趣点(Point of interest, POI)的类型划定载客热点区域的类型。

土地利用差异对出行分布的影响仍是一个复杂的研究领域,当前研究仍将重心放在土地利用与出行分布的相关关系,构造模型仍将其作为普通的影响因素,而未用模型解释二者存在的内在规律。本文结合GPS轨迹数据,在地图匹配的基础上,构建出租车出行信息数据库;应用核密度和热点分析方法研究出租车出行特征和轨迹时空分布;将跨区出行数据作为基础数据,构建基于土地利用性质的双约束人口权重机会模型,以揭示北京市出租车出行分布与土地利用性质的内在规律。本文研究结果能够为刻画未来城市共享出行需求的时空分布特征以及交通基础设施规划提供支持。

1 轨迹数据驱动的出租车出行特征分析

1.1 数据预处理与出行定义

出租车GPS轨迹数据包含大量出行特征的详细信息,然而存在无法区分出行起、终点的难题。如何进行数据预处理,并从大量轨迹数据中区分每段出行,进而提取出行活动指标是亟待解决的难题。本文对出行的定义如图1所示(彩图参加电子版,后文同)。

图1中,左下方红色点代表出行的起点,蓝色点为停留点。红色箭头表示实际的路线和方向。当在停留点的停留时间 T_{waiting} 大于阈值 θ 时,认为已经完成一次出行,由蓝色点到右上的红色点则视为第二次出行。用 r_s 表示出行实际距离, T_{travel} 为出行时间, r_d 为起点与终点之间的出行直线距离,其中 r_s 与 r_d 的比值为出行非直线系数。

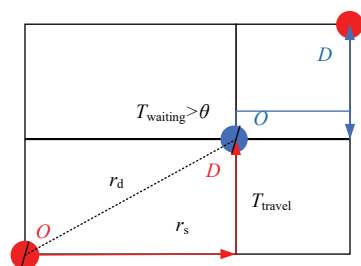


图 1 出租车出行示意图

Fig. 1 Schematic diagram of taxi travel

已有研究多基于私家车的出行特征指标(如出行距离、时间及停留时间),采用函数拟合分析其分布规律^[15]。由于出租车具有共享出行特征,是多人出行在时空上的优化衔接,因此本文将分析出租车的出行距离、直线距离、非直线系数、出行时间、停留时间、出行次数 6 个表征时间特征和空间特征的指标。其中,非直线系数反映交通工具在实际运行过程中相对于理想最短路径的绕行程度。出租车出行非直线系数对于提高交通规划效率、减少能源消耗和环境污染、提升乘客体验、降低经济成本以及科学研究与政策制定都具有重要意义。该指标为出租车出行实际距离与出行起讫点之间直线距离(欧氏距离)的比值。已有研究表明,私家车的出行距离、出行时间及停留时间等出行特征指标多服从指数分布或是幂律分布^[16]。本文结合出租车特性,利用韦布尔分布和有序分布对出行特征指标进行拟合。

有理函数是通过多项式的加、减、乘、除得到的函数,可以用分数形式表示。一个有理函数 h 可以写成 $h=f/q$, 其中 f, q 均为多项式函数。有理函数是特殊的亚纯函数,它的零点和极点个数有限。有理函数拟合对于处理复杂非线性关系有重要意义。韦布尔分布常用于描述产品寿命、可靠性工程、风速模型等,其概率密度 f 的计算公式为:

$$f(x; \lambda, l) = \begin{cases} \frac{l}{\lambda} \left(\frac{x}{\lambda}\right)^{l-1} e^{-(x/\lambda)^l}, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (1)$$

式中: $x \geq 0$; 形状参数 $l > 0$; 比例参数 $\lambda > 0$ 。

当形状参数 $l=1$ 时,韦布尔分布可简化为指数分布;当 $l=2$ 且适当选择参数时,可转化为瑞利分布。当一个随机变量服从韦布尔分布时,它的 l 次幂会服从指数分布,其中, l 为该随机变量的形状参数。同理,当一个随机变量服从指数分布

时,其 $1/l$ 次方会服从韦布尔分布。

1.2 出行特征分析

以北京市 2019 年 9 月 15 日~2019 年 10 月 27 日的 3 万余辆出租车 GPS 轨迹数据为基础进行研究,该数据为传统巡游出租车轨迹,未包含拼车订单。其中,数据采集间隔约为 30 s,日均数据量约为 4 300 万条,原始数据格式为 .txt 格式,共约 110 GB。6 列的字段依次为车辆 ID、经度、纬度、速度和数据接收时间。依据出租车数据处理流程,结合北京市实际情况和出行特征指标,以 5 min 为阈值构建出租车出行信息数据库,以此确保出行划分的合理性。提取出行信息数据库中出租车的出行距离、直线距离、非直线系数、出行时间、停留时间和出行次数 6 个表征时间特征和空间特征的指标进行分析,并与以往研究进行对比,出租车出行特征指标分布如图 2 所示。

由图 2(a)(b)可知:出租车出行距离近似服从比例参数为 50.613、形状参数为 0.764 的韦布尔分布,其拟合优度为 0.981;出租车出行的直线距离近似服从比例参数为 9.757、形状参数为 0.802 的韦布尔分布,拟合优度为 0.989。出行距离与出行直线距离拟合的韦布尔分布形状参数比较接近,二者的相关系数仅为 0.421,分析推测,该现象主要由路网复杂程度、城市社会经济属性以及出行回转特性等因素共同造成。

图 2(c)中横坐标 k 为非直线系数,表示出租车每次出行实际距离与直线距离的比值。由图 2(c)可知,出租车出行的非直线系数近似满足参数为 0.104 和 -0.972 的有理函数关系(即满足拟合函数式 $P(x)=a/(x+c)$, 其中参数 $a=0.104$ 为决定曲线幅度的尺度参数;参数 $c=-0.972$ 为决定渐近线位置的平移常数),拟合优度为 0.962。由于非直线系数与出行距离、出行直线距离存在紧密联系,利用增补变量法求出出行距离与非直线系数的联合概率密度函数,并通过求边际分布得到非直线系数的分布(形如 $a_1/(b_1x+c_1x^{d_1})$ 的形式),为有理函数的扩展形式,基本可以证实非直线系数符合有理函数关系的合理性。

由图 2(d)(e)可知,出租车出行时间近似服从参数为 -1.043 的幂律分布,拟合优度为 0.974;出租车停留时间近似服从参数为 -1.358 的幂律分布,拟合优度为 0.991。由图 2(f)可知,出行次数的众数为 4 次,中位数为 10 次,平均出行

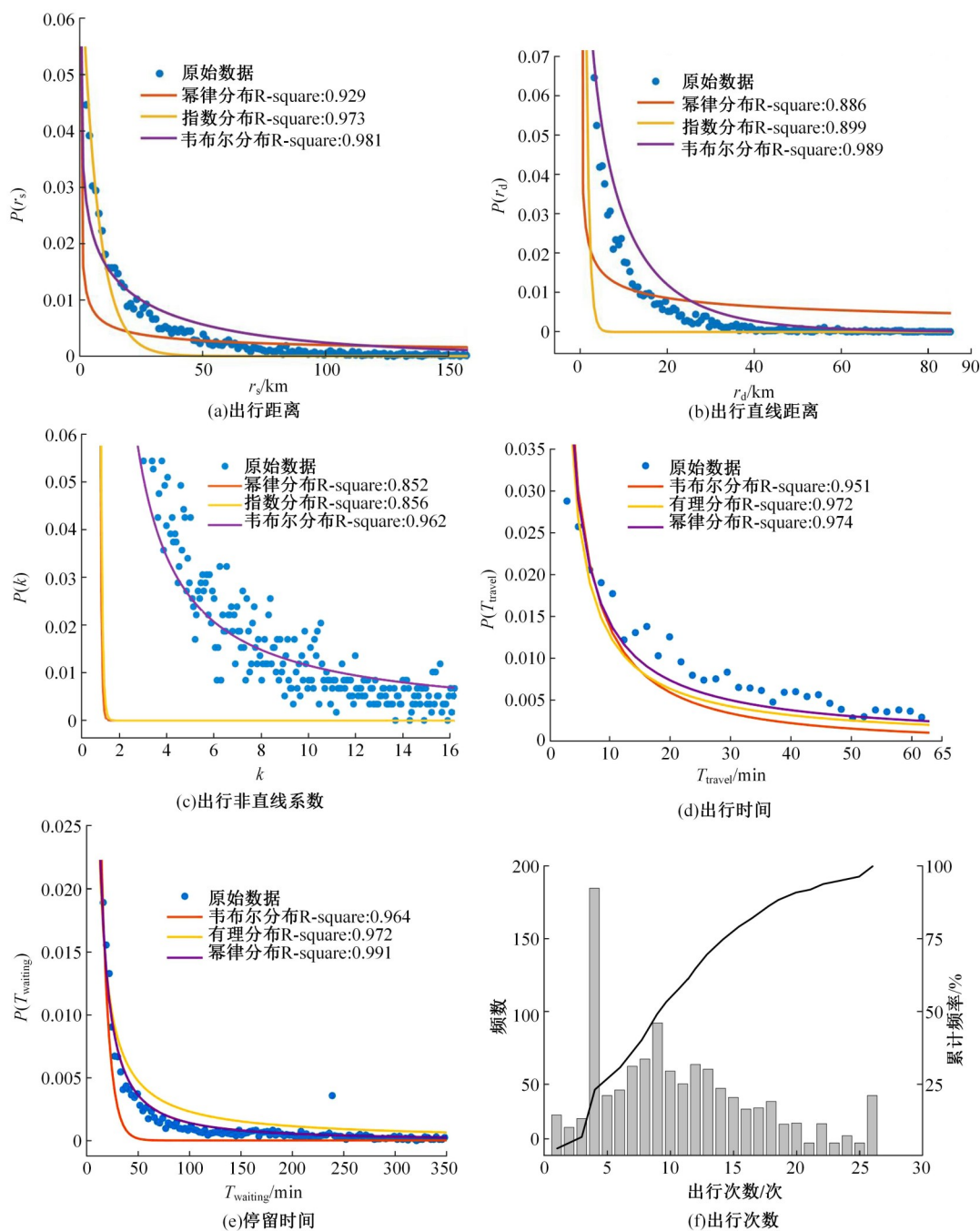


图 2 出租车出行特征指标分布特征

Fig. 2 Characteristics of distribution of taxi travel behavior indicators

次数为 10.94 次。出行次数基本可以看作是单峰分布。出行次数为 3~4 次时曲线上升速度最快,斜率最大;在 10 次之后,曲线整体上升趋势比较平缓。出租车出行以短距离的区内出行为主,且出行频次较高。

综上,在出行距离方面,出租车表现出与私家车不同的分布特征,可用韦布尔函数分布拟合。出行非直线系数符合有理函数分布。出行时间和停留时间特征与私家车相似,可用幂律函数拟合。

上述分析结果可为出租车出行以及共享出行的仿真模拟提供重要参考依据。

2 出租车出行分布时空模式

采用 ArcGIS 对轨迹与道路网进行匹配,绘制早高峰(7:00~9:00)和晚高峰(17:00~19:00)的出租车轨迹分布图。其中,早晚高峰的轨迹图都是截取主要影响的区域进行放大,如图 3 所示。

由图 3(a)(b)可知,北京市出租车早高峰出

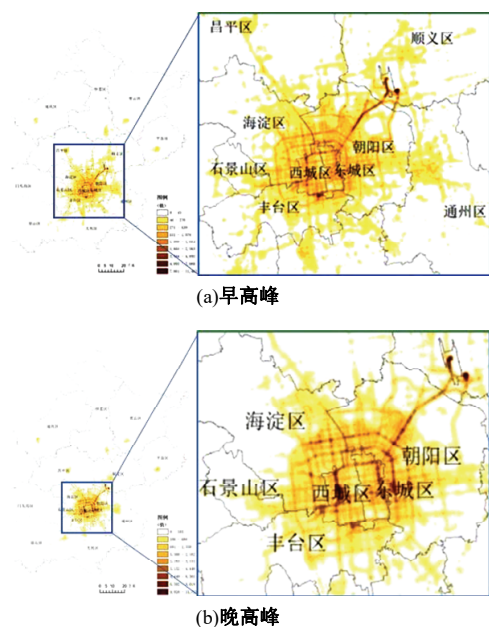


图 3 早晚高峰出租车轨迹分布

Fig. 3 Trajectory distribution of taxi during morning and evening rush hours

行区域主要集中于五环以内,轨迹的密集程度从中心向四周逐渐降低。从轨迹密集程度来看,快速路及主干路的轨迹密集程度明显高于次干道及支路,晚高峰与早高峰的轨迹特征基本一致。五环内 CBD、国贸等地区附近轨迹明显多于其他地区。早晚高峰的出租车出行轨迹都是中心城区轨迹更密集,越远离市中心出行轨迹越少。晚高峰市中心的轨迹密集程度较早高峰更高。

以匹配后的出租车出行起讫点为基础进行核密度分析。由于 GPS 轨迹数据点大多没有高程,位置信息由经纬度表达,故只考虑针对点要素的二维核密度。通过分析出租车出行信息数据库可以得到出租车的轨迹点,在理论上可以认为每个点上方都存在一个很小的平滑曲面。曲面在该点处的特征值最高,随着距离点的距离越远越小,在距离恰好等于搜索半径时,表面值为 0,且邻域为圆形邻域。每个输出栅格像元的密度均为叠加在栅格像元中心的所有核表面值之和。核函数以四次核函数为基础。预测点的密度 Den 为:

$$Den = \frac{1}{r^2} \sum_{i=1}^n \left[\frac{3}{\pi} \left(1 - \left(\frac{d_i}{r} \right)^2 \right)^2 \right] \quad (2)$$

式中: $i=1, 2, \dots, n$ 为输入点; d_i 为 i 点与预测点之间的距离,小于等于搜索半径; r 为预测点的搜索半径。

由于使用四次核函数,需要对每一个待估计密度的位置分别进行公式计算。

以出租车的 OD 信息为基础进行出租车出行热点区域的分析。热点分析会对数据集中的每一个要素计算 Getis-Ord G_i^* 统计,得到 z 得分和 p 值。这两个值用于帮助判断是否可以拒绝零假设(即要素的完全空间随机性)。正值表示热点,负值表示冷点。Getis-Ord G_i^* 局部统计可由式(3)(4)(5)确定。

$$G_i^* = \frac{\sum_{j=1}^n w_{ij} x_j - \bar{X} \sum_{j=1}^n w_{ij}}{S \sqrt{\frac{n \sum_{j=1}^n w_{ij}^2 - \left(\sum_{j=1}^n w_{ij} \right)^2}{n-1}}} \quad (3)$$

$$\bar{X} = \frac{\sum_{j=1}^n x_j}{n} \quad (4)$$

$$S = \sqrt{\frac{\sum_{j=1}^n x_j^2}{n} - \bar{X}^2} \quad (5)$$

式中: x_j 为点要素 j 的属性值; w_{ij} 为点要素 i 和 j 之间的空间权重; n 为点要素的总数; G_i^* 为 z 得分。

如果要对冷点和热点作出判断,则必须确定置信度。表 1 为不同置信度和临界 p 值、临界 z 得分之间的关系。

表 1 不同置信度临界 p 值和 z 得分的关系Table 1 Relationship between different confidence level critical p -values and z -scores

z 得分	p 值	置信度/%
<-1.65 或 $>+1.65$	<0.10	90
<-1.96 或 $>+1.96$	<0.05	95
<-2.58 或 $>+2.58$	<0.01	99

基于 ArcGIS 的核密度和热点分析,利用出行信息数据库提取的 OD 信息,得到北京市 9 月 25 日出租车出行核密度以及逐小时的出行热点。选择早晚高峰进行分析,结果如图 4 所示。其中,热点分布图中,黄色代表无明显统计特性,橙色和红色分别代表置信度为 90% 和 95% 的热点。

由图 4(a)(b)可知,早晚高峰出租车出行密度最大的是市中心区域,出行密度由中心向四周逐渐减少。早高峰存在明显的往中心城区聚集的特征,从而导致中心区域的出租车出行密度较大。从区域的土地利用性质可知,大型办公区域、商业服务业设施用地的出行密度普遍偏大。推测其原因是,北京目前仍存在比较明显的“职住分离”现象,大部分工作岗位仍集中在中心城区。早高峰

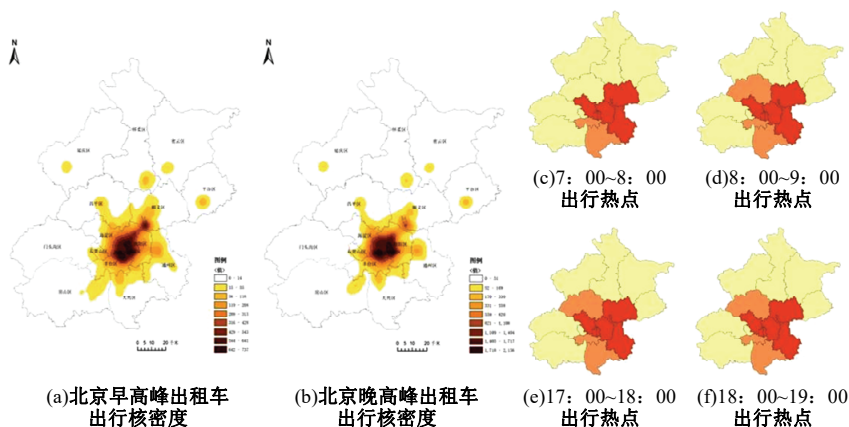


图4 2019年9月25日北京市出租车出行核密度及出行热点

Fig. 4 Taxi traffic density and hotspots in Beijing on September 25, 2019

影响的范围更大,晚高峰的影响范围相对较小。晚高峰市中心的出行密度较早高峰而言更大,推测是由于工作单位下班时间相对统一,而上班时

间会有一定间隔。由图4(c)~(f)可知,出行热点区域有大量居住用地,是出租车经过或者出行的起点。对早晚高峰的热点区域变化进行对比可知,早高峰前期的影响范围较小,但后期热点区域增加,影响范围逐渐增大;晚高峰前期的影响范围较大,后期热点区域无增加,影响范围保持不变。由上述分析可以看出,土地利用性质深刻影响出租车的出行分布。

3 考虑土地利用性质的双约束人口权重机会模型

进行出行分布预测时,人口权重机会模型^[17]具有较好效果,其优点是不需要估计可调参数,只需要输入各地点的人口分布数据就可以得到城市间的出行分布。该模型克服了辐射模型低估长距离出行、高估短距离出行的劣势。缺点是模型只带起点的出行约束,无法保证终点的出行约束。出租车的出行活动是多人联程出行组合,在时空层面融合多个目的地。相比于区域内的人口,出租车出行与土地利用性质更直接相关。本文保留人口权重模型的目的机会衰减机制,即出行者选择目的地的概率与目的地到起点之间人口数量成反比,体现竞争效应。目的地到起点的人口数量定义为以目的地为圆心,以起点之间的直线距离为半径做圆,圆内包含的总人口数如图5所示。此外,为精细刻画出租车的出行分布特征,本文以POI数据确定交通小区土地利用占比以及相互吸

引系数,结合佛尼斯法,在起点出行约束的基础上,保证终点出行约束,建立基于土地利用性质的双约束人口权重机会模型,预测出租车出行分布量。

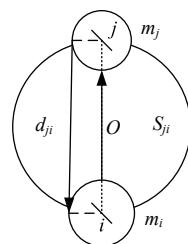


图5 目的地到起点的人口数量示意图

Fig. 5 Schematic diagram of population from destination to origin

基于土地利用性质的双约束人口权重机会模型(以下简称基于土地利用性质的出行分布模型)的基本假设是,出行者选择一个目的地的概率与该目的地的机会数成正比,与出行者选择的目的地到起点之间的人口总数成反比(由于空间影响,需要减去尺寸修正项)。由于地点机会数可以假设为正比于地点人口数与相互吸引系数,所以该模型的输入只有相互吸引系数和人口分布。由假设可知,目的地 j 的吸引量与该目的地的机会数成正比,与出行者目的地到起点之间的人口总数(减去尺寸修正项)成反比,如下所示:

$$A_j = o_j \left(\frac{1}{S_{ji}} - \frac{1}{M} \right) \quad (6)$$

式中: o_j 为地点 j 的总的机会数; S_{ji} 为以目的地 j 为圆心,以 j 到 i 的直线距离为半径的圆所包含的人口总和; M 为城市的总人口数,即尺寸修正项; A_j 为地点 j 的吸引量。

目的地与起点之间的人口总数(即 S_{ji})是模

型中至关重要的部分。假设每个区的人口均匀分布,即重叠面积占总面积的比例就是该区域占该区总人口的比例,即可得到 S_{ji} 。由于地点机会数可以假设正比于地点人口数与相互吸引系数,即得到目的地 j 的吸引量 A_j 为:

$$A_j = m_j \left(\frac{\alpha_{ji}}{S_{ji}} - \frac{\alpha_{ji}}{M} \right) \quad (7)$$

式中: α_{ji} 为目的地 j 对起点 i 的吸引系数; m_j 为目的地的 j 的人口数。

目的地被选择概率为:

$$P_{ij} = \frac{A_j}{\sum_{k \neq i} A_k} \quad (8)$$

式中: P_{ij} 为目的地 j 被选择的概率。

起点 i 到目的地 j 的出行量 T_{ij} 可以由式(9)得到:

$$T_{ij} = O_i P_{ij} = O_i \frac{m_j \left(\frac{\alpha_{ji}}{S_{ji}} - \frac{\alpha_{ji}}{M} \right)}{\sum_{k \neq i} m_k \left(\frac{\alpha_{ki}}{S_{ki}} - \frac{\alpha_{ki}}{M} \right)} \quad (9)$$

式中: O_i 为起点 i 的发生量。

相互吸引系数 α_{ji} 基于土地利用性质得出。本文根据《城市用地分类与规划建设用地标准》,将城市建设用地分为8大类、35中类、44小类。将研究区域划分为多个交通小区,既可以查找规划文件确定各个区的土地利用比例,也可以通过提取POI数据判断各个区的各类土地利用的比例。利用多元回归求得各种用地对出行的影响参数。将参数带回各个区得到各自的吸引系数,最后的相互吸引系数则是用目的地的吸引系数除以起点的吸引系数,得到相互吸引系数矩阵 $[\alpha_{ij}]$ 。该矩阵当 $i=j$ 时, $\alpha_{ij}=1$;若 $i \neq j$ 时, $\alpha_{ij}=1/\alpha_{ji}$ 。该矩阵是对数意义下的反对称矩阵。

上述模型只满足发生约束,所以需要对结果进行吸引约束。由于佛尼斯法的计算相对简单,迭代速度较快,本文采用佛尼斯法进行迭代,模型公式如下所示:

$$f_{FN}^1(F_{Oi}^m, F_{Dj}^m) = F_{Oi}^m \quad (10)$$

$$f_{FN}^2(F_{Oi}^m, F_{Dj}^m) = F_{Dj}^m \quad (11)$$

式中: F_{Oi}^m 为 i 小区的第 m 次计算发生增长系数, F_{Dj}^m 为 j 小区的第 m 次计算吸引增长系数。

先令吸引增长系数为1,求满足条件的发生增长系数,接着用调整后的矩阵重新求满足条件

的吸引增长系数,完成一个循环迭代过程;然后重新计算发生增长系数,再用调整后的矩阵求吸引增长系数,经过循环,直到发生和吸引交通量增长系数满足设定的收敛标准为止。

为验证基于土地利用性质的出行分布模型的有效性合理性,将模型的结果和原数据进行相似度检验和索伦森相似指数(Sørensen similarity index, SSI)检验。相似度检验是检验两个矩阵的相关性,越接近1,则相关性越强,相似度越高,其计算公式为:

$$S(B, C) = \frac{B \cdot C}{\|B\| \|C\|} \quad (12)$$

式中: B 为原数据的分布矩阵; C 为模型得到的分布矩阵。

索伦森相似指数常用于判断生物多样性以及DNA碱基对的相似程度。本文采用索伦森相似指数验证模型与实际数据的吻合程度,1表示完全吻合,0表示偏离极大,其计算公式为:

$$SSI = \frac{1}{N(N-1)} \sum_i \sum_{j \neq i} \frac{2 \min(q'_{ij}, q_{ij})}{q'_{ij} + q_{ij}} \quad (13)$$

式中: q_{ij} 为实际分布数据; q'_{ij} 为模型得到的交通分布数据; N 为小区的个数。

此外,本文还采用均方误差(Mean square error, MSE)、均方根误差(Root mean square error, RMSE)和平均绝对误差(Mean absolute error, MAE)评价基于土地利用性质的出行分布模型对出租车OD分布的预测效果。

4 实例验证

针对北京16个行政区,通过提取POI数据判断各个区的各类土地利用的比例。本文提取10 000个百度POI,利用Python编程通过字段判断各个行政区的各类用地的比例。为简化计算,本文的发生量数据是随机抽取的1 000辆出租车的出行信息汇总得到的。按照模型构造方法依次获得相互吸引系数、目的地与起点之间的人口总数,最后得到人口权重机会模型与基于土地利用性质的出行分布模型运算结果。两个模型检验及评价如图6所示。

由图6(a)可知,两个模型的矩阵相似度都比较高,改进前后分别为0.961和0.970。改进前后的索伦森相似指数分别为0.43和0.59,具有一定程度的提高,即添加相互吸引系数对提高模型精

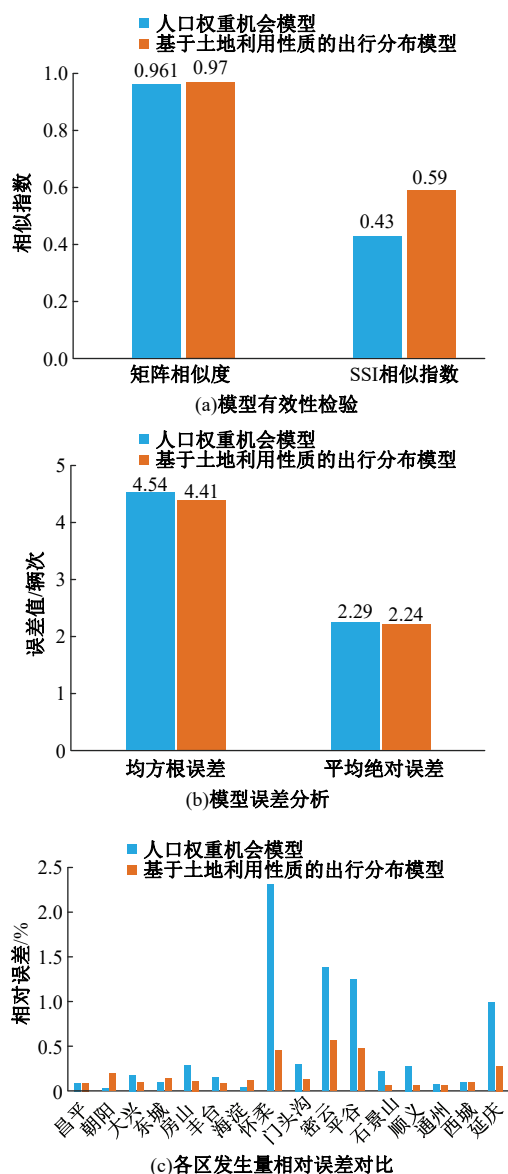


图6 基于土地利用性质的双约束人口权重机会模型有效性检验及误差分析

Fig. 6 Validity test and error analysis of opportunity model of population weight based on land use characteristics with double constraints

度有一定作用。

由图6(b)可知,人口权重机会模型的均方误差和平均绝对误差分别为4.54辆次和2.40辆次,本文模型的均方误差和平均绝对误差分别为4.41辆次和2.29辆次,说明改进后的出行分布模型能够提高预测精度。

由图6(c)可知,各个区的发生量改进前、后的相对误差都小于2.5%。通过计算可知,人口权重机会模型的平均相对误差为0.48%,而本文模型的平均相对误差则为0.19%,明显优于人口权重机会模型。

预测结果中,共计12个区发生量的相对误差有所降低;1个区发生量的相对误差基本保持不变;3个区的相对误差有所提高。根据北京市2019年GDP数据,相对误差降低的12个区均处于GDP排名的末位,保持不变以及相对误差提高的区域则是经济发展迅速、GDP稳定在前4位的城区。这4个区域商业服务业的设施用地比例是16个区中的前4位,即增加商业服务业设施用地会很大程度上提高该区域对其他区域的相互吸引系数,进而导致对其出行分布有一定程度的高估,这也是基于土地利用性质出行分布模型的缺点,即对于商业服务业设施用地占比较大的区域的预测误差会比改进前略有提高。

相对误差有所降低的12个区中,怀柔、延庆、密云和平谷4个区域的发生量相对误差下降程度十分显著。基于土地利用的相互吸引系数可以很好地描述这些区域更容易被拥有更多商业与服务设施用地的区域所吸引,其出行分布会比改进前的模型更接近实际数据。所以,该模型的重要性和实际意义在于可以更好地预测经济发展欠发达地区的出租车的出行分布。同时,推测产生这一现象的原因是北京市目前还存在比较明显的职住分离现象,土地利用占比的差距较大,所以不同行政区之间的相互吸引现象比较明显。

经过多次测试可知,基于土地利用性质的出行分布模型迭代次数明显小于人口权重机会模型,迭代速度加快。将基于土地利用性质分布模型得到的交通分布进行统计,得到各个区出租车的发生量和吸引量如图7所示。

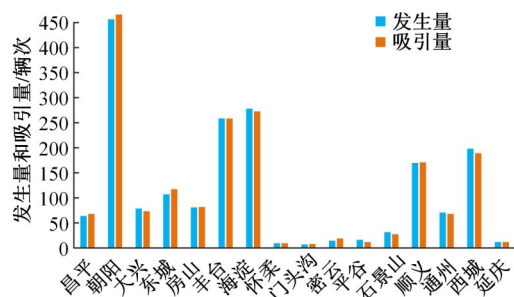


图7 北京市各区的出租车的发生量和吸引量

Fig. 7 Generation and attraction of taxis in each administrative region of Beijing

由图7可知,朝阳区的出租车发生量和吸引量最多;出租车出行的发生和吸引与土地利用性质和人口密度联系紧密,从侧面验证了模型的合理性。尽管由模型得到的结果显示,北京出租车

出行仍以中心城为主,但是也可以看出,通州区、大兴区的发生量和吸引力也相对较高。推测其原因是由于经济技术开发区和市政府搬到通州造成通州区和大兴区的用地性质改变,从而使发生量和吸引力与其他郊区相比要高一些。从整体北京市各区的发生量和吸引力可以看出,每个行政区出租车的发生量与吸引力差距都很小。跨区出行的出租车普遍呈现返回起点所在区域的周期性特征。在数据处理阶段,相当一部分出租车会在一天结束之前返回到一天开始时的地点,这也侧面反映本文推测具有一定合理性。本文认为造成这种情况的两个关键因素是出租车司机的习惯和出租车的交班机制。北京的出租车分为单班和双班,单班是一个司机开一辆车;双班则是一辆车由两个司机开。单班司机驾驶的出租车存在明显的回转周期特性。计算可知,出租车在不同区的出行分布量差距很大,中心城区之间的出行量要远远大于郊区之间的出行量。这是起讫点之间的人口以及小区之间的相互吸引系数造成的。比如:从朝阳到海淀的出行量大约为朝阳到大兴的4.5倍。利用ArcGIS的区域交集分析得到的海淀区到朝阳区之间的人口为990.18万,到大兴区之间的人口为871.88万,前者大约是后者的1.14倍;海淀区的人口则是大兴区的1.96倍;海淀区对朝阳区的吸引系数为0.57,大兴区对朝阳区的吸引系数为0.63。在满足发生和吸引约束的情况下,得到二者的倍数接近4.5。

综上所述,北京市各个小区之间的交通分布差距很大,但是各小区各自的发生量和吸引力差距较小。

5 结束语

本文对北京市3万余辆出租车GPS数据进行预处理,构建出租车出行信息数据库,分析了出租车出行特征指标。结合核密度分析法和热点分析法对出租车出行密度、空间时序特征进行研究。最后,结合土地利用性质对人口权重机会模型进行改进,建立了基于土地利用性质的双约束人口权重机会模型,结合佛尼斯法进行求解,并评价了模型的有效性和精度。结果表明:①出租车日出行距离和日出行直线距离都近似服从韦布尔分布;非直线系数近似满足有理函数关系;出租车日出行时间和停留时间近似服从幂律分布;日平均

出行次数为10.94次。②出租车出行轨迹主要集中在北京市中心区域,密集程度从中心向四周逐渐降低。早晚高峰的出租车出行密度呈现中心密度大,远离市中心出行密度小的特征;早高峰前期影响范围较小,但后期热点区域增加;晚高峰前期的影响范围较大,但后期热点区域无增加。③本文基于土地利用性质的双约束人口权重机会模型,与经典的人口权重机会模型相比,对发生量预测的平均相对误差降低0.29%,整体预测精度提升。北京市各区之间的交通分布量差距很大,中心城区之间的分布量明显高于郊区之间的分布量,各区各自的发生量和吸引力差距较小。土地利用差异以及起讫点之间的人口分布使得各区之间的出行量差距较大,而出租车司机的出行习惯和交班机制导致各区的总发生量和总吸引力差距较小。本文研究结果可为优化出租车共享出行调度策略、完善城市交通规划提供理论依据,尤其对高峰时段车辆配额分配与跨区调度具有指导意义。同时,可为规划出租车临时停靠点及出租车交班时间调整提供理论依据。此外,该研究成果中的出租车出行时空特征可为差异化定价(如高峰溢价)提供数据支持,平衡供需时空分布,进而有效缓解城市交通压力。

参考文献:

- [1] 马健霄,赵飞燕,尹超英,等.建成环境和出租车需求对网约车出行需求影响的时空分异模式[J]. 交通运输系统工程与信息, 2023, 23(5): 136-145.
Ma Jian-xiao, Zhao Fei-yan, Yin Chao-ying, et al. Spatial-temporal heterogeneity effects of built environment and taxi demand on ride-hailing demand[J]. Journal of Transportation Systems Engineering and Information Technology, 2023, 23(5): 136-145.
- [2] Jiang B, Yin J, Zhao S. Characterizing the human mobility pattern in a large street network[J]. Physical Review E: Statistical, Nonlinear, and Soft Matter Physics, 2009, 80(2): No. 021136.
- [3] Bazzani A, Giorgini B, Rambaldi S, et al. Statistical laws in urban mobility from microscopic GPS data in the area of Florence[J]. Journal of Statistical Mechanics: Theory and Experiment, 2010, 2010(5): No. P05001.
- [4] Liang X, Zheng X, Lv W, et al. The scaling of human mobility by taxis is exponential[J]. Physica A: Statistical Mechanics and its Applications, 2012,

- 391(5): 2135–2144.
- [5] Simini F, González M C, Maritan A, et al. A universal model for mobility and migration patterns[J]. *Nature*, 2012, 484(7392): 96–100.
- [6] Riccardo G, Armando B, Sandro R. Towards a statistical physics of human mobility[J]. *International Journal of Modern Physics C*, 2012, 23(9): No. 1250061.
- [7] Peng C, Jin X, Wong K, et al. Collective human mobility pattern from taxi trips in urban area[J]. *PLoS One*, 2012, 7(4): No. e34487.
- [8] Ren Y, Ercsey-Ravasz M, Wang P, et al. Predicting commuter flows in spatial networks using a radiation model based on temporal ranges[J]. *Nature Communications*, 2014, 5(1): No. 5347.
- [9] Zhao Z, Yang Z, Zhang Z, et al. Emergence of scaling in human-interest dynamics[J]. *Scientific Reports*, 2013, 3(1): No. 3472.
- [10] Yan X, Wang W, Gao Z, et al. Universal model of individual and population mobility on diverse spatial scales[J]. *Nature Communications*, 2017, 8(1): No. 1639.
- [11] Liu E, Yan X. A universal opportunity model for human mobility[J]. *Scientific Reports*, 2020, 10(1): 4657.
- [12] 陶思然, 叶霞飞. 引入旅游偏好的城际客运出行分布预测模型[J]. *交通运输系统工程与信息*, 2021, 21(4): 140–147.
- Tao Si-ran, Ye Xia-fei. A forecasting model of inter-city trip distribution with tourism preference[J]. *Journal of Transportation Systems Engineering and Information Technology*, 2021, 21(4): 140–147.
- [13] 陈启香, 吕斌, 陈喜群, 等. 空间异质性建成环境对出租车与地铁竞合关系的影响[J]. *交通运输系统工程与信息*, 2022, 22(3): 25–35.
- Chen Qi-xiang, Lv Bin, Chen Xi-qun, et al. Impacts of built environment on competition and cooperation relationship between taxi and subway considering spatial heterogeneity[J]. *Journal of Transportation Systems Engineering and Information Technology*, 2022, 22(3): 25–35.
- [14] 邢雪, 王菲, 李佳楠. 结合载客热点和POI的出租车停车位划定方法[J]. *吉林大学学报: 信息科学版*, 2024, 42(1): 93–99.
- Xing Xue, Wang Fei, Li Jia-nan. Cab fixed parking area delineation method combining passenger hotspot and POI data[J]. *Journal of Jilin University (Information Science Edition)*, 2024, 42(1): 93–99.
- [15] Xiao Z, Wu L, Jiang H, et al. Exploring intercity mobility in urban agglomeration: Evidence from private car trajectory data[J]. *IEEE Transactions on Computational Social Systems*, 2024, 11(2): 2940–2954.
- [16] Deng J, Cui Y, Chen X, et al. Who are on the road? a study on vehicle usage characteristics based on one-week vehicle trajectory data[J]. *International Journal of Digital Earth*, 2023, 16(1): 1962–1984.
- [17] 闫小勇. 超越引力定律: 空间交互和出行分布预测理论与方法[M]. 北京: 科学出版社, 2019.