引文格式:汪晓龙,石岩,王达,等.基于自适应密度聚类的单车出行源汇时空区域探测方法[J].武汉大学学报(信息科学版),2025,50(11):2339-2351.DOI:10.13203/j.whugis20230362

Citation: WANG Xiaolong, SHI Yan, WANG Da, et al. Spatiotemporal Extent Detection Method of Bike Travel Source-Sink Based on Adaptive Density Clustering[J]. Geomatics and Information Science of Wuhan University, 2025, 50(11):2339-2351.DOI:10.13203/j.whugis20230362

基于自适应密度聚类的单车出行源汇时空区域探测方法

汪晓龙[1] 石 岩[1,2] 王 达[1] 汤仲安[2,3] 刘宝举[1,2]
陈炳蓉[1] 邓 敏[1,2]

1 中南大学地球科学与信息物理学院,湖南 长沙 410083
2 湖南省地理空间信息工程技术研究中心,湖南 长沙,410018
3 湖南省第三测绘院,湖南 长沙,410018

摘 要:无桩共享单车凭借灵活性强、成本低、全天供应等优势,已成为当前中国共享单车服务的主要形式。然而,过于灵活的无桩共享单车经常出现特定时空域内稀缺借不到或堆积还不进的供需失衡现象。准确、精细地探测共享单车出行OD(origin-destination)分布失衡时空范围与失衡程度是可靠认知单车供需关系、科学调度单车资源和发展绿色可持续公共交通的关键问题。当前研究大多利用预定义网格等方式划分空间单元,进而构建指标度量出行OD失衡程度,并设置失衡阈值提取源汇范围。基于人工先验硬性划分时空单元和设置失衡阈值的方式极易低估源汇时空范围与失衡程度,且当前缺乏出行源汇区域及时序演化的耦合分析,为此,提出一种基于时空密度的二元聚类方法自适应地提取出行源汇。首先,基于密度显著性统计判别视角适应性地提取OD核点;进而构建二元变量分布失衡统计判别提取源汇邻域;最后,利用时空邻接约束扩展提取出行源汇的精细时空范围。利用厦门岛无桩共享单车轨迹数据进行实验,发现在早高峰内共识别出11 351个源区域和9 453个汇区域,每小时源汇数量最高可达1 925个,其中源区域失衡指数集中于[0.5,1],汇区域失衡指数集中于[-1,-0.5],证明所提方法能够精细、可靠地提取表征车辆借还分布失衡的出行源汇空间区域、持续时间及其失衡程度。

关键词:出行源汇;分布失衡;密度聚类;二元聚类;时空分析;共享单车

中图分类号:P208 **文献标识码**:A **收稿日期**:2024-08-21

DOI:10.13203/j.whugis20230362 **文章编号**:1671-8860(2025)11-2339-13

Spatiotemporal Extent Detection Method of Bike Travel Source-Sink Based on Adaptive Density Clustering

WANG Xiaolong[1] *SHI Yan*[1,2] *WANG Da*[1] *TANG Zhong'an*[2,3] *LIU Baoju*[1,2]
CHEN Bingrong[1] *DENG Min*[1,2]

1 School of Geosciences and Info-Physics, Central South University, Changsha 410083, China
2 Hunan Geospatial Information Engineering and Technology Research Center, Changsha, 410018, China
3 Hunan Third Institute of Surveying and Mapping, Changsha, 410018, China

Abstract: **Objectives**: Due to its strong flexibility, low cost, and continuous availability, dockless shared bicycles have become the primary form of bike-sharing services in China. However, the excessive flexibility of dockless shared bicycles often leads to a supply-demand imbalance, manifested by scarcity within specific temporal and spatial domains, where bikes cannot be found or accumulate without being used. Accurately and precisely detecting the spatial and temporal extent, as well as the degree of imbalance in the origin-des-

基金项目:国家自然科学基金(42071452,42371477);湖南省自然科学基金(2022JJ20059);国家重点研发计划(2021YFB3900904);湖南省科技创新计划(2023RC3032);中南大学创新驱动研究计划(2023CXQD013)。

第一作者:汪晓龙,硕士生,主要从事时空聚类分析与异常探测。8211180227@csu.edu.cn

通信作者:王达,博士。215001023@csu.edu.cn

tination (OD) distribution of shared bicycle trips is a crucial issue for understanding the reliable cognition of bike supply and demand relationships, scientifically dispatching bike resources, and developing green and sustainable public transportation. Currently, most studies use predefined grids to divide spatial units and then construct metrics to measure the degree of OD imbalance, setting imbalance thresholds to extract source and sink regions. The method of manually predefining rigid spatial units and setting imbalance thresholds tends to underestimate the spatial and temporal extent of sources and sinks and the degree of imbalance. Additionally, there is a lack of coupled analysis of the temporal and spatial evolution of travel source and sink areas. **Methods**: To this end, this paper proposes a bivariate clustering method based on spatiotemporal density to adaptively extract trip sources and sinks. First, spatial kernel density estimation and information entropy theory are introduced to adaptively extract OD core points from the perspective of density significance test. Second, a bivariate distribution imbalance statistical method is constructed to identify source and sink core points of shared bike trips, and density-based clustering extension strategies are implemented for fine-grained identification of source-sink spatial regions. Finally, based on the detected source-sink spatial regions, OD point event intensity and imbalance significance tests are designed based on OD point temporal density distribution and event type probability distribution, to identify the sources and sink time periods of shared bike trips. **Results**: Experimental results using dockless shared bicycle trajectory data in Xiamen Island reveal that within the morning peak hours, this method identified 11 351 source areas and 9 453 sink areas. The number of sources and sinks can reach up to 1 925 per hour. The imbalance index for source areas is concentrated in the range [0.5, 1], while for sink areas, it is concentrated in the range [−1, −0.5]. This demonstrates that the proposed method can accurately and reliably extract spatial regions, durations, and degrees of imbalance that characterize the distribution imbalance of vehicle borrowing and returning. **Conclusions**: In order to overcome the problems that the spatial detection results of shared bicycle travel source and sink are heavily dependent on the rigid pre-division of spatiotemporal units, the predefined threshold of algorithm parameters and source and sink indicators, and the lack of spatiotemporal coupling mining analysis of source and sink, this paper proposes an adaptive spatial detection algorithm for traffic travel source and sink based on binary variable density clustering. A case study was carried out in the actual data of shared bicycles in Xiamen Island.

Key words: travel source-sink; distribution imbalance; density clustering; binary clustering; spatiotemporal analysis; bike sharing

随着中国城市化进程的不断推进,日益复杂的城市结构和快速增长的城市人口导致城市交通运输体量日益庞大,居高不下的拥堵程度和能源消费水平对中国交通绿色可持续发展和交通强国建设带来巨大挑战。作为服务居民出行“最后一公里”的主要出行方式,共享单车加强了公共节点之间的连接[1],有效缓解了出租车、私家车等机动车运输压力,对中国交通强国建设和绿色低碳出行具有重要意义。中国共享单车行业发展报告(http://www.caict.ac.cn/sytj/201803/t20180314_172823.htm)显示,共享单车的投入使用减少了约55%的私家车、网约车、出租车等小汽车出行。截至2017年,仅摩拜一家公司的单车出行已经累计减少了440万t碳排放量,节约的总体经济效益相当于1.94亿美元。然而,大量过于自由便利的无桩共享单车出行衍生出一系列管理和调度难题[2-3]。居住区、工作地、交通接驳口等重要出行区域经常出现单车稀缺借不到或过度堆积还不进等供需失衡的潮汐现象[4],已经成为阻碍共享单车服务效益最大化、交通系统可持续健康发展的现实难题[5-6]。因此,准确精细地探测共享单车源汇区域与持续时间是可靠度量单车供需失衡、科学调度单车存放资源的关键,对于实现城市居民绿色低碳出行、道路交通高效运行和城市空间优化建设具有重要指导意义。

出行源汇表示驶入、驶出流量大且显著不平衡的时空事件[7-9],如图1所示,流入比流出显著多的时空事件表示出行源模式,反之则为出行汇模式。预设时空单元及其特征阈值是探测人类异常出行模式的常用策略[10-11],例如,柯日宏等[6]利用共享单车在电子围栏内的借还需求数量,识别流量较大的热点区域,而基于时空单元特征预定义的出行源汇

探测方法通过预设基本时空单元或失衡指标度量出行源汇范围与源汇程度。高楹等[8]首先将研究区域划分为规则的空间网格,进而构建净流入率指标提取了共享单车的强源汇点,最后利用地理探测器分析了共享单车强源汇点的空间分布格局。FANG等[12]基于每小时的手机信号塔累积网络流量,利用净流量指标正负识别人群聚集模式和分散模式,进一步将一天划分为7个时间段,对人群聚集和分散过程的稳定性进行评估。高楹等[13]针对地铁站周围接驳的共享单车,通过映射到固定网格单元和净流量计算的方法确定源汇网格,进而识别源汇区域。基于时空单元与指标预定义方法的出行源汇探测方法直观易懂、可执行性强。然而,此类探测结果严重依赖于源汇失衡先验阈值的设定和基本时空分析单元的划分,算法参数敏感性相对较高。

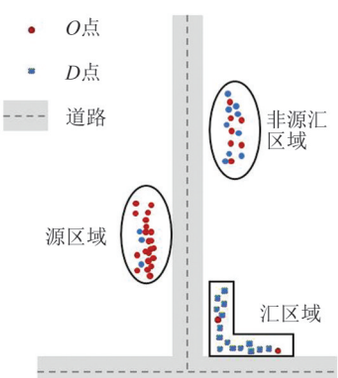


图1　源汇示意图

Fig. 1　Source-Sink Schematic Diagram

面对上述出行源汇探测研究中预定义时空单元的可塑面积单元难题[14-15],时空聚类可以直接利用出行轨迹点探测簇内紧密相似、簇间远离相异的聚集模式[16-17],已经被广泛用于出行热点提取[18]、交通拥堵探测[19]和潜在疫区发现[20]等人类出行模式研究。出行源汇模式可视为二元变量的时空聚类分析问题[7,9,21-22]。姜晓等[9]提出一种基于网络约束的时空图聚类算法,将具有相似时空特征的电子围栏聚到同一社区内,并在此基础上进行单车潮汐特征挖掘。LIU等[7]预先划分网格和时间单元,通过网格单元的上下车点时间序列聚类得到源汇区域。PEI等[21]提出一种面向起点-终点(origin-destination,OD)数据的密度聚类方法,通过分别对OD点事件设置高低密度阈值方式提取源汇核点,并基于密度可达策略生成OD分布失衡的空间簇,结合交通枢纽、居住区等地理要素分布分析源汇模式空间分布。LIU等[22]提出一种网络约束的双变量聚类方法探测城市交通出行中的“黑洞”和“火山”模式,即流入量明显大于或小于流出量的路段集合,反映了城市交通系统中的源汇区域。仝照民等[4]首先通过密度聚类得到流簇,然后利用簇内流的起终点生成空间分布热力图并提取高密度区域,最后基于区域聚散关系识别源汇区域。时空聚类是一种可以从轨迹点数据提取交通事件的非监督方法,能够有效规避时空单元划分难题。然而,现有基于聚类的单车出行源汇探测方法大多依赖先验知识预设聚类算法参数,难以适应源汇异质性分布特征。

综上所述,现有出行源汇探测研究仍然存在3个挑战。首先,先验时空单元硬性划分极易造成均质区域割裂或异质区域概化,造成源汇时空范围粗糙或源汇程度低估问题。同时,预定义源汇指标阈值或算法参数的方式难以自适应地提取源汇模式,探测结果严重依赖先验阈值设置。最后,当前单车出行源汇研究仅关注空间或时间的单一视角,缺乏出行源汇区域与其演化时序的耦合挖掘与可视化分析。基于此,本文基于城市无桩共享单车出行轨迹大数据,提出一种OD点密度显著性统计检验的双变量聚类方法自适应地提取出行源汇时空范围,以期助力城市规划管理者和共享单车企业准确认知共享单车出行失衡格局,科学调整电子围栏,提高借还可用性与调度效率。

1　研究方法

1.1　技术路线

如图2所示,本文方法包括数据预处理、算法描述和实验分析。其中,数据预处理部分包括数据源描述及数据预处理,算法描述部分详细介绍了基于OD点密度显著性统计检验的双变量聚类方法,实验分析展示了研究区域共享单车源汇时空分布特征。

1.2　数据源及数据预处理

本文研究区域为厦门岛(图3)。《2021年厦门市城市居民出行交通大调查》指出(http://xm.fjsen.com/2022-05/13/content_31033042.htm),在出行结构上,非机动车出行比例较高,全市及岛内非机动车出行比例分别为65.2%和76.5%,其中,自行车、电动车等慢行出行比例占比约为47.5%,是非机动车出行的主要方式。

厦门岛作为厦门市的经济、文化、金融中心[23],岛内居民通勤环境复杂,单车出行量十分庞大[24]。因此,本文选择厦门岛内城市居民共享单车出行轨迹大数据提取出行源汇模式并进行结果分析。由于早、晚高峰单车出行目的集中且

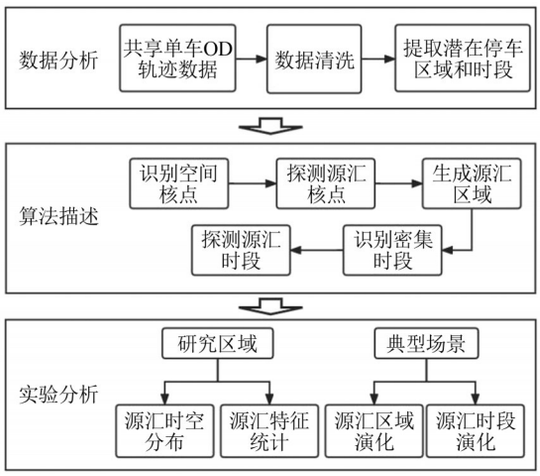


图2 技术流程

Fig. 2 Technical Process



图3 研究区域概况

Fig. 3 Overview of the Study Area

出行量大,因此,源汇模式主要出现于早晚高峰期间的车站、社区、工作场所等区域。实验数据清单如表1所示,包括:(1)2020年12月21日—25日早高峰期间52 934辆共享单车产生的约58万条的订单轨迹数据(https://data.xm.gov.cn/contest-series/digit-china-2021),已被广泛应用于共享单车失衡现象探索[6];(2)来源于OpenStreetMap(https://www.openstreetmap.org/)的研究区域道路网络数据。

数据预处理过程如下:(1)单车轨迹数据清洗与处理。首先,删除异常轨迹数据,如超出研究范围、落在水域或建筑内、时间戳格式错误等轨迹点。然后,对同一订单中位置和锁状态重复的记录进行筛选,仅保留时间戳最早的记录。进一步,提取同一订单的最早和最晚的轨迹点分别作为O点(出发地)和D点(目的地)。最后,删除订单出行时间间隔小于1 min或者大于30 min的OD数据。(2)共享单车潜在停取区域提取。因为O点和D点均表现了共享单车可停车的位置,因此基于所有OD点坐标以50 m半径生成空间缓冲区,将所有OD点生成的空间缓冲区叠置合并为共享单车的潜在停取区域,定义为ODA(OD area)。

表1 实验数据

Table 1 Experimental Data

数据名称	数据时间	数据规模	数据描述	
			字段名称	字段含义
厦门岛共享单车订单数据	2020年12月21日—25日AM 06:00—10:00	约58万条	ORDER_ID	加密后的单车订单ID
			LATITDUE	纬度/(°)
			LONGITUDE	经度/(°)
			LOCK_STATUS	锁状态(1:开锁;0:关锁)
			UPDATE_TIME	锁状态更新的时间
厦门岛路网数据	2020年	约8 000条道路	Length	道路长度/m
			name	道路名称

1.3 源汇区域识别

首先利用核密度估计OD轨迹点的空间分布密度,基于熵度量空间密度分布的异质性程度,选择密度分布异质性最强的核密度估计带宽作为密度聚类最优半径[25]解决先验设置聚类参数的问题,基于此构建OD点的邻域关系和空间密度域。然后,构建空间点密度统计假设检验方法,自适应识别出邻域密度显著稠密的空间核点,进而构建二元变量密度失衡统计检验模型,适应性地提取出行源汇邻域,解决预定义源汇识别阈值与时空单元的问题。最后,利用空间邻接约束扩展策略提取出行源汇的精细空间范围。

1.3.1 基于核密度估计的最优邻域设置

空间核密度估计认为地理事件在空间点稠密的区域发生概率大,在空间点稀疏的区域发生概率小[26]。度量空间稠密程度的空间核密度估计函数天然地映射了任意空间位置的局部聚集强度,局部聚集强度高的空间位置更容易成为聚类结构中的高密度簇[27]。然而,空间是无数位置的集合[28],无法穷尽某区域内所有点位置而判断

区域是否稠密。因此,本文基于核密度估计函数度量所有n个OD点位置的局部密度:

$$f(x_i)=\begin{cases}\dfrac{1}{nh}\sum_{j=1}^{n}K\left(\dfrac{\|x_i-x_j\|}{h}\right), \|x_i-x_j\|<h\\ 0, \|x_i-x_j\|\geqslant h\end{cases} \tag{1}$$

式中,h表示核密度估计函数带宽,也是空间点事件x的邻域半径。由于核函数的选择对密度分布估计的影响较小,因此本文利用信息熵评价密度分布异质性的不确定程度。熵越小,则密度分布的异质性越显著,可以用来发现空间点事件密度分布的最稳定点[29]:

$$H=-\sum_{i=1}^{n}\frac{f(x_i)}{\sum_{i=1}^{n}f(x_i)}\log\frac{f(x_i)}{\sum_{i=1}^{n}f(x_i)} \tag{2}$$

熵值最小时,所有OD点邻域的局部密度不对称性最高,即空间稠密区域显著性最强[30]。因此,本文给定带宽候选范围$h\in(0,100]$ m,自动选择熵最小值对应的带宽$h^{\min}$作为识别空间聚集模式的OD点最优邻域半径Eps。给定所有OD轨迹点中任一点x,本文以点x为圆心、Eps为半径生成空间邻域Eps_x,下一节将构建空间统计假设检验方法识别显著稠密的空间邻域作为空间聚类的子区域。

1.3.2　基于假设检验的空间核点识别

给定n个空间点和某目标区域$R\in$ODA,OD点是否存在于区域R内的概率符合二项分布$B(n,p_0)$。在不存在聚集结构的随机分布情况下,p_0仅与目标区域R的面积相对于总区域ODA的面积area_N有关[31]。SHIODE等[31]认为,当点事件数量n非常大而目标区域R面积很小时,区域R内存在OD点的概率可以近似地认为服从齐次泊松分布。因此,对于本文中任一点x_i的空间邻域Eps_x_i及其内实际观察到的点事件数n_x_i,在随机分布假设下,Eps_x_i中的点事件数量超过n_x_i的概率可以表示为:

$$P\left(\left|Eps_x_i\right|\geqslant n_x_i\right)=1-P\left(\left|Eps_x_i\right|<n_x_i\right)=1-\sum_{k=0}^{n_x_i-1}\mathrm{e}^{-\lambda'}\frac{\lambda'^{k}}{k!} \tag{3}$$

$$\lambda=n\times P \tag{4}$$

$$P=\frac{\mathrm{Area}\left(Eps_x_i\right)}{\mathrm{area}_N} \tag{5}$$

$$\mathrm{Area}\left(Eps_x_i\right)=\pi\times Eps^2 \tag{6}$$

式中,λ代表Eps_x_i中点事件的预期数量。

如图4所示,当$P\left(\left|\mathrm{Eps}_x_i\right|\geqslant n_x_i\right)$低于显著性水平$\alpha$(如0.05)时,Eps_$x_i$内的点事件分布通常足够稠密,显著区别于随机分布的空间邻域点密度,此时标记x_i为空间核点cx_i。

1.3.3　源汇核点识别及源汇区域探测

相对于聚集大量OD点事件的热点区域,源汇区域表现出O事件和D事件分布密度显著不均衡的特点。给定一个目标区域$R\in$ODA,在不存在O事件和D事件分布显著失衡的情况下,R内O点和D点的比例应该接近于1:1。因此,对于任一空间核点cx_i,其空间邻域Eps_cx_i内能够观察到O点的理论概率同样符合二项分布。给定$\left|Eps_cx_i\right|^{O}$和$n'$分别表示$Eps_cx_i$内理论分布下$O$点数量和实际观察下OD点的总数量,$n_cx_i^{O}$表示在$Eps_cx_i$中实际观察到的$O$点数量,OD点事件均衡分布假设下,$Eps_cx_i$中$O$点事件数量超过$n_cx_i^{O}$的概率可以表示为:

$$P\left(\left|Eps_cx_i\right|^{O}\geqslant n_cx_i^{o}\right)=1-P\left(\left|Eps_cx_i\right|^{O}<n_cx_i^{O}\right)=1-\sum_{k=0}^{n_cx_i^{O}-1}C_{n'}^{k}p^{k}(1-p)^{n'-k},\quad p=0.5 \tag{7}$$

如图4所示,当$P\left(\left|Eps_cx_i\right|^{O}\geqslant n_cx_i^{O}\right)$低于显著性水平$\alpha$(如0.05)时,认定其为源核点,记为$Ocx_i$;当$P$值高于$1-\alpha$(如0.95)时,认定其为汇核点,记为$Dcx_i$。在此基础上,源汇核点的空间邻域内与其密度相连的点都划入对应源汇区域,即源汇区域包含核点及其邻域内与其密度相连的点。具体来说,以源核点Ocx_i为例,首先将其空间邻域Eps_Ocx_i内所有点视为与其密度相连,然后依次判断密度相连点x_j是否也为源核点,如是则以x_j为中心继续连接密度相连点,直至无核点,最后合并所有密度相连点构成源区域。类似地,基于汇核点进行空间邻域扩展可以提取汇区域。

1.4　源汇时段识别

首先基于§1.3识别的源汇区域,构建OD点的时间邻域关系和时间密度域;然后,构建时间点密度统计假设检验方法,自适应识别出邻域密度显著稠密的密集出行时段;最后,构建二元变量密度失衡统计检验模型,可以自适应地提取源汇区域的源汇时段,解决缺乏出行源汇区域与其演化时序的耦合挖掘与分析问题。

1.4.1　密集出行时段识别

给定某目标时段t,任意一个OD点是否存在于时段t内的概率符合发生概率为p_0^t的伯努利试验。在不存在聚集结构的随机分布情况下,p_0^t仅与目标时段t相对于总时段长短有关[31]。因此,

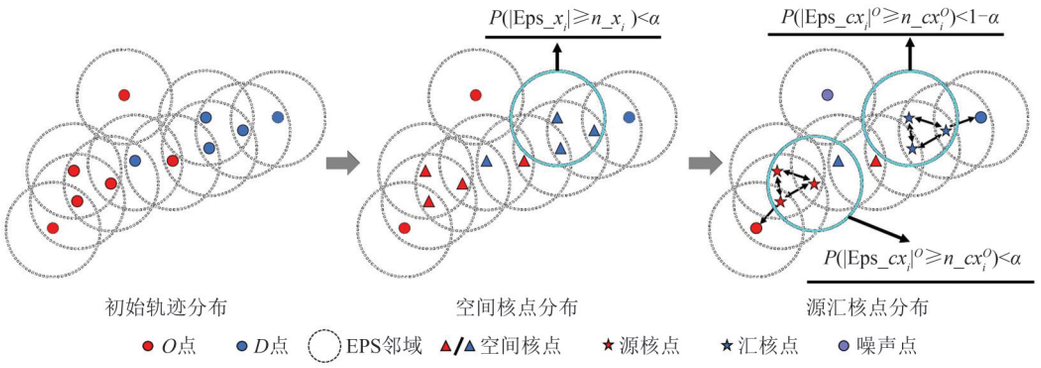


图4 源汇核点识别示意图

Fig. 4 Schematic Diagram of Source and Sink Core Points Detection

给定任意时段t_i，Length_N表示总时段的时长，理论上t_i内能够观察到超过实际存在的n_t_i个OD点事件的概率计算为：

$$P(|t_i|\geqslant n_t_i)=1-P(|t_i|<n_t_i)=1-\sum_{k'=0}^{n_t_i-1}C_{n'}^{k'}p_t^{k'}(1-p_t)^{n'-k'} \tag{8}$$

$$p_t=\frac{\mathrm{Length}(t_i)}{\mathrm{Length_}N} \tag{9}$$

出行量密集时段的点事件分布通常足够稠密，以至于几乎不可能被其他任意时段的点密度超过。因此，若$P(|t_i|\geqslant n_t_i)$小于给定足够小的显著性水平α（如0.05），则t_i可以被识别为一个密集出行时段，记作ct_i。

1.4.2 源汇时段探测

源汇时段应该有非常显著的O事件与D事件分布密度差异。基于此，本文构建给定时段内O事件与D事件分布比例接近1∶1的零假设，进而基于OD事件实际分布的分布概率进行假设检验确定OD失衡显著性。给定密集出行时段ct_i及其时段内包含的n'个OD点事件，$|ct_i|^O$表示随机分布下ct_i内的O点数量。给定$n_ct_i^O$表示实际观察下ct_i内的O点事件数，则OD点事件均衡分布假设下ct_i中的O点事件数量超过$n_ct_i^O$概率可以表示为：

$$P\left(|ct_i|^O\geqslant n_ct_i^O\right)=1-P\left(|ct_i|^O<n_ct_i^O\right)=1-\sum_{k'=0}^{n_ct_i^O-1}C_{n'}^{k'}p^{k'}(1-p)^{n'-k'},p=0.5 \tag{10}$$

基于此，如果$P\left(|ct_i|^O\geqslant n_ct_i^O\right)$小于给定的显著性水平$\alpha$（如0.05），则$ct_i$将被识别为一个源时段，记为$Oct_i$；如果累积概率大于给定的显著性水平$1-\alpha$（如0.95），则$ct_i$将被识别为一个汇时段，记为$Dct_i$，如图5所示。

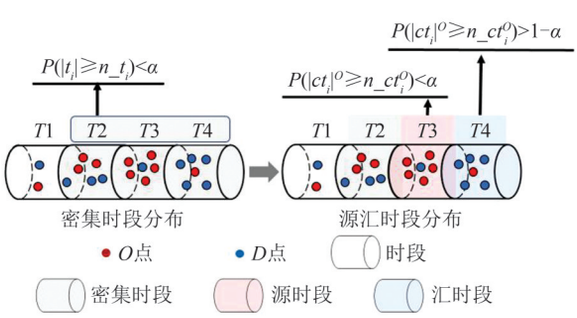


图5 源汇时段识别示意图

Fig. 5 Schematic Diagram of Source and Sink Periods Detection

2 实验与结果分析

2.1 单车出行源汇区域分布分析

图6展示了本文方法探测的源汇区域数量折线

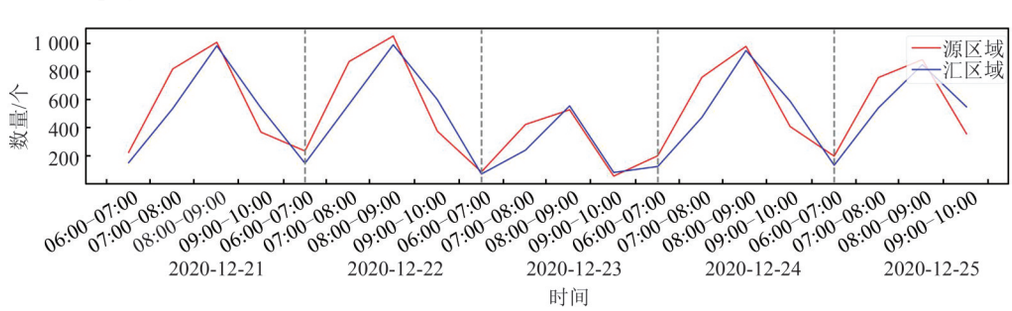


图6 本文方法探测的2020年12月21日—25日每小时源汇区域数量变化

Fig. 6 Hourly Changes in the Number of Source and Sink Areas from December 21 to 25, 2020 Detected by the Proposed Method

图，在研究区域内共识别到 11 351 个源区域，9 453 个汇区域，其中，每小时源汇数量最高可达 1 925 个。整体来看，研究时间内源汇区域数量变化模式较为稳定，均呈现先增加后减少的趋势：源汇区域数量在时段 06:00—09:00 内逐渐增加，时段 08:00—09:00 内达到最大值，在时段 09:00—10:00 逐渐降低。时段 06:00—09:00 内，源区域数量通常大于汇区域数量，直到早高峰最后 1 h，汇区域数量逐渐反超源区域数量。这可能是由于早高峰前期居民从居住地出发形成大量源区域，称之为居民出行的“第一公里”，而早高峰后期大量居民均接近工作地点，导致形成大量汇区域，对应形成了居民出行的“最后一公里”。研究时段内每天同一时刻源汇区域数量接近，不同的是，2020-12-23 共享单车出行源汇区域数量显著少于其他工作日，且源汇区域数量曲线差距较小。考虑当天厦门市为中雨天气，且有群体活动，为了排除天气等干扰因素，探索更普适的城市居民单车出行规律，本文选择 2020-12-25 时段 06:00—10:00 内的共享单车出行轨迹数据做进一步分析。

图 7 显示了本文方法探测的厦门岛共享单车出行源汇区域空间分布，其中红色表示源模式，蓝色表示汇模式。可以看出，源汇区域主要分布在厦门岛的西南部、中部、北部和东南部。其中，厦门岛北部中心区域存在较多的出行源模式，可能由于该区域集中了较多公寓群，因此数量庞大的职工群体大多在早高峰期间统一由此区域出发通勤前往周边交通枢纽换乘或工作场所上班。厦门岛西南和东南部中心位置的源汇区域分布较均匀，尤其是西南部分多种城市功能混合的老市区，居民没有形成密集且区分度高的出行需求。同时，厦门岛西南部、东南部和东北部的海滨区域存在较多的汇模式，这些区域集中了大量接驳码头、写字楼和商业中心等场所。例如，厦门岛西南部的海滨区域是厦门码头最集中的区域，有出航、换乘、游玩等需求的人群一般使用共享单车前往此区域接驳，导致了汇区域的密集分布。

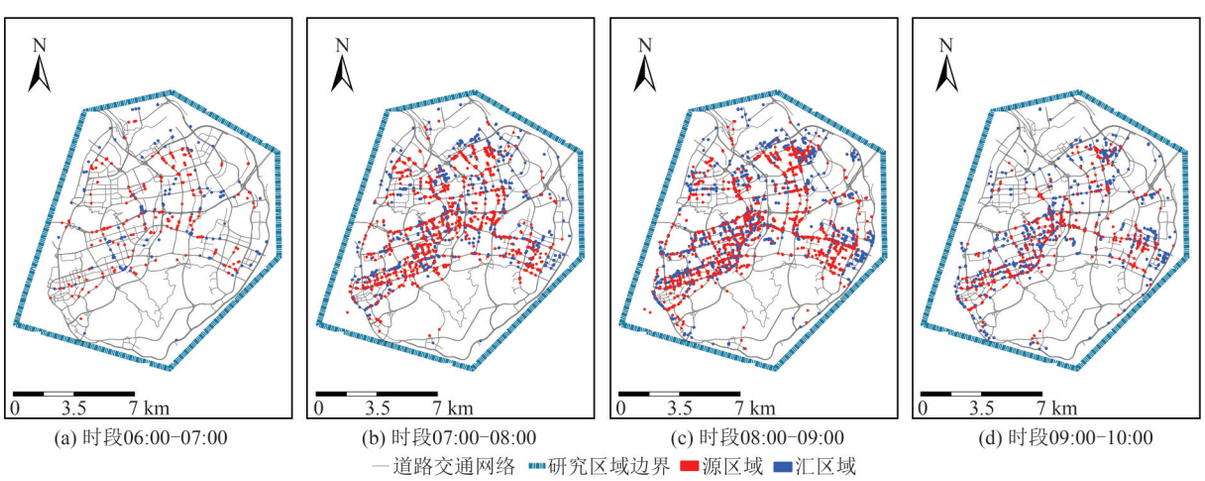


图 7　本文方法探测的 2020-12-25 每小时源汇空间分布

Fig. 7　Hourly Spatial Distribution of Source and Sink on December 25, 2020 Detected by the Proposed Method

2.2　单车出行源汇时段分布分析

本节利用§1.4 提出的源汇时段探测方法识别研究区域内的源汇时序变化（图 8），其中，折线分别表示净流量和 OD 轨迹点数量，淡红色和蓝紫色分别表示本文方法探测出的源汇时段。净流量为区域内 O 点轨迹数减去区域内 D 点轨迹数。如图 8 所示，除了星期三受天气和集体活动影响外，其他工作日早高峰期间城市居民出行均呈现比较稳定的源→汇→源→汇→源→汇模式。其中，源时段经常出现的 07:30、08:05 和 08:35 这 3 个时刻通常是城市居民骑乘单车出发的集中时段，同时也是上班或上学的高峰时段；而汇时段经常出现的 07:50、08:20 和 08:45 这 3 个时刻通常是城市居民到达交通站点、学校或工作场所等目的地较为集中的时间段。有趣的是，源汇交替出现的时刻相差大约 15 min，这与骑乘单车通勤的时间长度相符[32]。

2.3　单车出行源汇特征统计分析

净流量率（net flow ratio，NFR）一般用来衡量区域的源汇失衡强度[32]，计算公式为：

$$\mathrm{NFR}_i=\frac{O_i-D_i}{O_i+D_i} \tag{11}$$

可以看出，NFR 越大，O 点事件数量越多于 D 点事件，表现出源模式特征；NFR 越小，则 O 点事件数量越少于 D 点事件，表现出汇模式特征。

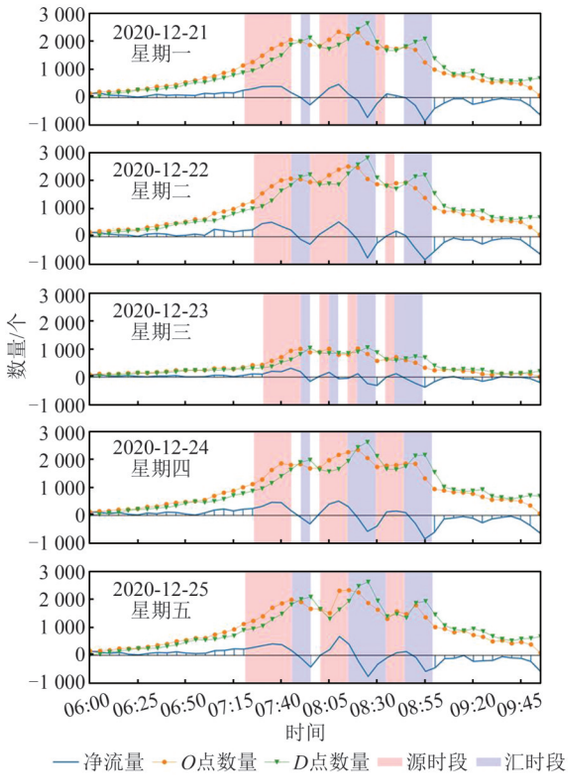


图8 本文方法探测的研究区域源汇时段变化

Fig. 8 Source and Sink Time Period Changes Detected by the Proposed Method in the Study Area

从图9可以看出,本文探测的源区域NFR指数均大于0,且最大值超过0.8,显著集中于[0.5,1]区间。时段08:00—09:00尤为明显,说明该时段内大量人群从出发地开始骑乘共享单车。与源区域不同,本文探测的汇区域NFR指数分布于[−1,−0.5]区间,大多集中于−1,说明大多汇区域内只包含D点事件,这与图7源汇区域空间分布显示的汇区域较集中现象相符。由此可以推断,早高峰骑乘共享单车出行的出发地比较分散,出行场景以居民区为主,同时存在大量骑行出发和少量到达行为,而目的地在空间上较为集中,出行场景以工作区为主,几乎不存在骑行出发行为。

2.4 典型单车出行源汇场景时空演化分析

本节选取典型单车出行源汇场景,进一步分析本文方法探测出的源汇模式时空分布演化态势。厦门岛北部湖里创新园附近作为2000年后新兴建成区,集中了大量商业公寓和写字楼,形成了显著的职住通勤和交通换乘相关的源汇模式。如图10所示,区域A和区域B分别表示住宅属性的围里社区和工作属性的湖里创新园,面积较大,通常需要借助共享单车通行;区域C作为位于区域A和区域B附近的县后公交站,每天承担着大量通勤于围里社区或湖里创新园的慢行交通量。具体地,城市居民从围里社区出发骑乘共享单车前往区域C附近的天桥、地铁站或公交站等交通节点。同时,工作出行目的的人群到达区域B附近后,骑乘共享单车前往湖里创新园的具体楼栋。

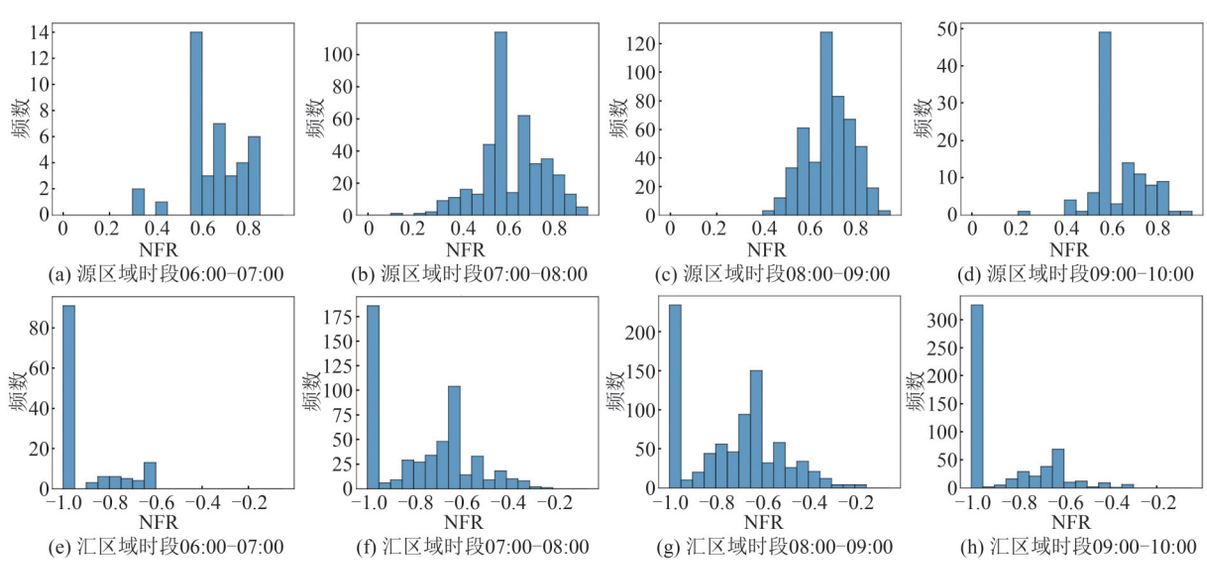
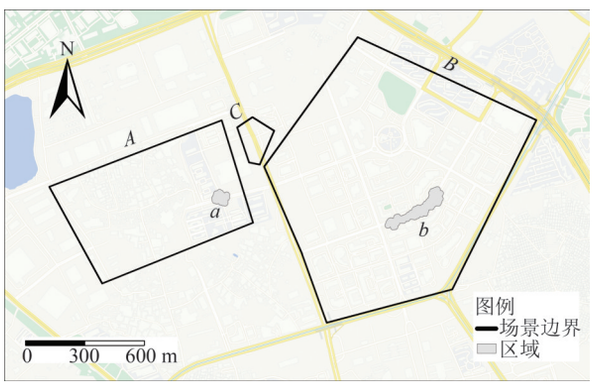


图9 2020-12-25时段06:00—10:00每小时源汇区域内NFR值频数分布直方图

Fig. 9 Histogram of Frequency Distribution of NFR Values in the Source and Sink Area Every Hour from 06:00 to 10:00 on December 25, 2020

从图11可以看出,区域A围里社区上午聚集了较多的源区域,同时几乎没有出现汇区域,这与围里社区以大量商业公寓为主有关。时段06:00—07:00相对少量人群从围里社区出发前往附近道路上的交通节点,如区域C内的县后公交站。时段07:00—09:00内围里社区的源区域数量最多、面积最大,此

时段是城市居民出发通勤的高峰时段。时段09:00—10:00内围里社区的源区域基本消失,可以认为此时城市居民稀疏自由的日常通勤没有形成单车借还显著失衡的源汇区域。围里社区周边源区域主要分布在公寓群周边300 m范围内,存在明显向公交站点靠拢的趋势,说明居民区源模式与公共交通节点密切相关。总而言之,围里社区作为源区域聚集的典型场景,源区域呈现先增多后减少的演化趋势,与图6显示的变化趋势一致。



注:区域A表示围里社区,区域B表示湖里创新园,区域C表示县后公交站;区域a表示围里社区中的阳光公寓,区域b表示湖里创新园的安岭路。

图10　典型单车出行源汇场景

Fig. 10　Typical Source and Sink Scenario of Bicycle Travel

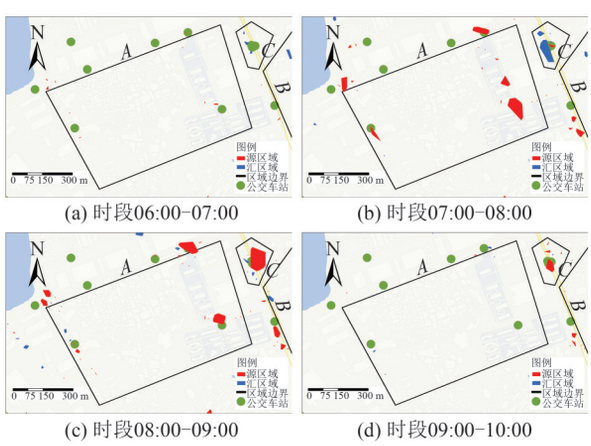


图11　本文方法探测的围里社区源汇区域分布变化

Fig. 11　Distribution Variation of Source and Sink Areas in Weili Community Detected by the Proposed Method

区域B湖里创新园作为占地约1.8 km^{2}的高新技术产业园区,白天集中了大量的工作人群。从图12可以看出,区域B聚集了大量汇区域。时段06:00—08:00内,湖里创新园仅形成了少量汇区域;时段08:00—09:00内,则形成了大量且密集的汇区域,这个时间段是工作人群到达工作地点的主要时间。时段09:00—10:00,湖里创新园内汇区域数量逐渐减少,但仍然较为密集,可能是业务来往人群骑乘前往产业园区。图12表现了不同人群到达工作场所的通勤特点。湖里创新园的汇模式主要分布在办公楼300 m范围内的沿街停放点,说明显著的工作场所驱动形成。虽然区域A围里社区的源模式和区域B湖里创新园的汇模式同样呈现先增加后减少的趋势,但是区域A围里社区的出行源模式高峰集中于时段07:00—08:00,体现了早高峰前期出发地人群集中出门的“第一公里”特点,区域B湖里创新园的出行汇模式高峰集中于时段08:00—09:00,表现出早高峰后期人群到达工作场所的“最后一公里”特点。可见,工作日居民通勤出行模式对源汇区域形成具有重要的影响。

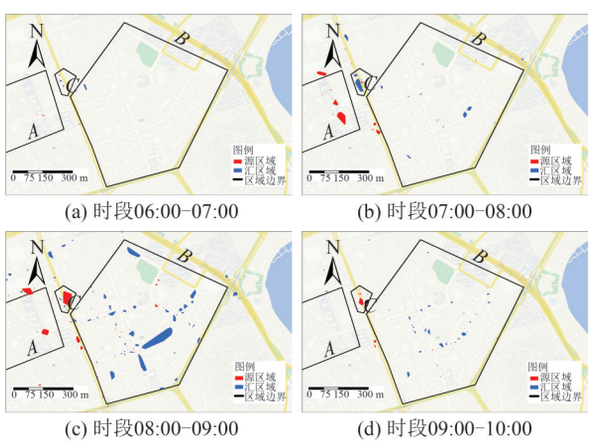


图12　本文方法探测的湖里创新园源汇区域分布变化

Fig. 12　Distribution Variation of Source and Sink Areas in Huli Innovation Park Detected by the Proposed Method

区域C县后公交站位于围里社区和湖里创新园中间,早高峰前期承担着大量从围里社区出发的慢行交通出行量,早高峰后期接驳了陆续前往湖里创新园的工作人群。图13展示了区域C县后公交

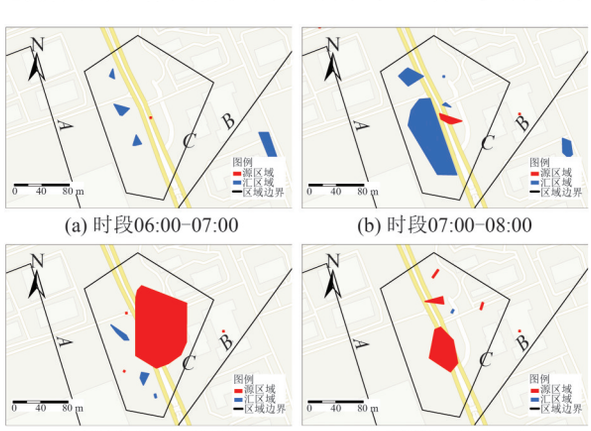
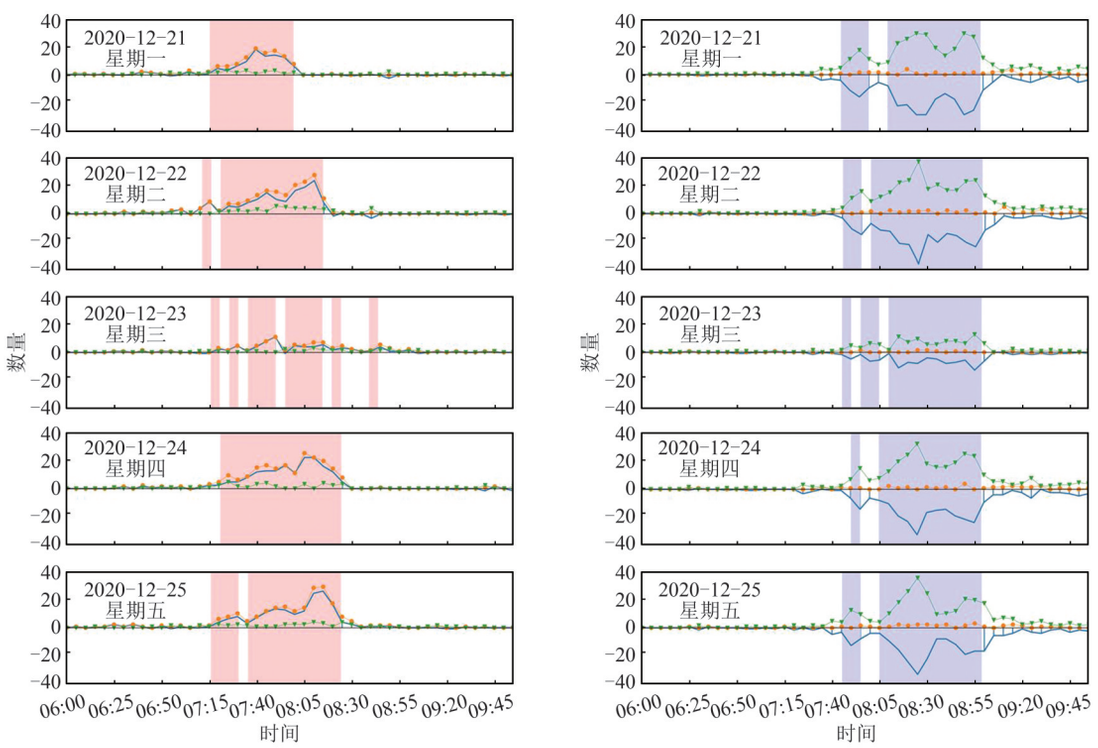


图13　本文方法探测的县后公交站源汇区域分布变化

Fig. 13　Distribution Variation of Source and Sink Areas in Xianhou Bus Station Detected by the Proposed Method

站显著的源汇交替演化过程。时段06:00—08:00内,围里社区及其周边大量城市居民骑乘共享单车前往该公交站转往目的地,导致此时区域C内左侧汇区域较为集中,对应着单车作为交通出行的“第一公里”;同时,道路另一侧开始出现了小范围的单车出行源模式,可能是前往湖里创新园或附近其他区域。时段8:00—10:00内,大量其他区域的城市居民通过公交车等公共交通方法到达区域C,再从此处出发骑乘共享单车前往湖里创新园等工作场所,造成此时区域C源区域较为集中,对应了单车作为交通出行的“最后一公里”。

图14展示了区域A围里社区中某源区域a和区域B湖里创新园某汇区域b的源汇时序演化。由图14可知,作为工作日早高峰期间典型的源区域,区域a源时段分布连续且稳定集中在时段07:00—08:00。作为工作日早高峰期间典型的汇区域,区域b的汇时段主要集中在时段07:50—08:00和08:15—09:00,第一个时段可能是居住在附近公寓群的城市居民到达时间较早,第二个时段则主要是通过公共出行交通接驳后再骑乘共享单车的方式,整体到达时间较晚。不同的是,在2020-12-23星期三,由于下雨和群体活动对城市居民单车出行的影响,源汇模式较不连续,且分布时段发生偏移,但在这类典型职住场景下,整体规律仍然保持稳定。



(a) 围里社区区域a源汇时段演化特征 (b) 湖里创新园区域b源汇时段演化特征

图14 围里社区区域a和湖里创新园区域b源汇时段演化特征

Fig. 14 Evolution Characteristics of Area a of Weili Community and Area b of Huli Innovation Park

3 讨 论

通过对厦门市共享单车出行轨迹的源汇时空特征分析,可以得出以下几点发现:

首先,除个别受中雨天气影响的特殊日期外,工作日早高峰期间城市居民共享单车出行的变化规律较为稳定,每天相同时刻的源汇区域数量接近,数量变化曲线均呈现先增加后减少的趋势。单独观察源或汇区域的数量变化发现,早高峰前期时段06:00—09:00可能对应居民出行的“第一公里”,以大量的骑行出发行为为主,导致源区域数量通常大于汇区域数量;早高峰后期时段09:00—10:00可能对应居民出行的“最后一公里”,包含大量的骑行到达行为,汇区域数量逐渐反超。

其次,从源汇区域分布来看,源汇区域分布与城市职住场景布局关系密切。本文探测出的源汇区域NFR指数表现出强烈的OD失衡特征,其中,源区域NFR指数集中在[0.5,1]区间,汇区域NFR指数集中在[−1,−0.5]区间。有趣的

是,汇区域NFR指数大多集中于－1。观察可知,早高峰期间的汇区域位置较为集中,汇区域内几乎不包含骑行出发行为;相比之下,源区域对应位置较为分散,如各居民区,同时包含大量的骑行出发和少量到达行为。再者,从整个研究区域的源汇时段分布来看,工作日早高峰期间城市居民出行源-汇-源-汇交替出现,存在3个源-汇循环的子高峰。第一个源时段出现时间逐渐推迟后恢复,可能是因为工作日中居民心理逐渐懈怠导致通勤时间出现波动。最后,从典型出行场景的源汇演化来看,早高峰期间居民区聚集了较多的源区域且几乎没有汇区域,可能是受到居民工作出发

需求的影响,源区域数量呈现先增多后减少的演化趋势,源时段分布连续且集中在时段07:15—08:10;而工作区则形成了大量且密集的汇区域,受到类似的影响,汇区域数量同样呈现先增多后减少的演化趋势,而汇时段相对源时段存在延迟,主要集中在时段7:50—08:00和08:05—08:55。

本文提出的出行源汇时空探测方法为深入理解城市慢行出行时空特征提供了可靠的识别与定量分析途径,分析结果与日常认知相符,侧面印证了研究方法提取源汇时空分布的有效性。本文探测的源汇模式时空演化特征有助于为共享资源动态均衡配置优化提供决策信息。特别地,针对共享单车运营商车辆调度的需求,可以准确识别单车失衡区域,并支持面向特定场景的源汇时序演化特征进行及时停放疏导,例如,将地铁站周围积压的共享单车转移至附近需求量大的居民区,避免供过于求或供不应求的情况发生,有效提高单车利用效率,提高城市居民出行便利程度。

本文研究的数据源主要覆盖工作日早高峰时段,因此仅聚焦于早高峰期间城市居民单车通勤的源汇特征。若能获取更丰富的共享单车轨迹数据,将在早晚高峰和节假日等不同视角下对出行源汇模式进行更丰富的对比实验分析。未来工作将基于因果推断方法探究共享单车出行源汇模式的形成机制,为城市高质量发展与智慧交通建设提供科学可信的决策知识服务。

4　结　语

为克服共享单车出行源汇时空域探测结果严重依赖时空单元硬性预划分、算法参数与源汇指标阈值预定义,以及缺乏源汇时空耦合挖掘分析的问题,本文提出了一种基于二元变量密度聚类的交通出行源汇时空域自适应探测方法,并在厦门市共享单车实际数据中进行了实例验证。相较以往研究,本文的主要贡献如下:

1)自适应的源汇区域探测算法。针对预定义时空单元或依赖先验聚类参数的问题,本文顾及共享单车OD点事件空间分布密度,引入空间核密度估计与信息熵理论;同时针对源汇状态阈值主观设置的问题,设计二元空间点密度显著性检验方法识别共享单车出行源汇核点,基于密度聚类扩展策略实现源汇空间区域的精细识别。

2)协同挖掘出行源汇空间区域与持续时段。针对缺乏时空耦合视角挖掘分析源汇模式的现状,本文在源汇空间区域探测基础上,基于OD点事件的时序点密度分布与事件类型概率分布,设计OD点事件密集与失衡的显著性检验方法识别出行源汇时段,分析了单车出行源汇过程的稳定性和趋势性,同时验证了源汇区域探测结果的合理性与可靠性。

3)利用厦门市真实单车OD轨迹数据可视化出行源汇时空分布特征,拓展了不同出行场景与职住通勤结构下居民出行源汇流动性认知。

参　考　文　献

[1] SHAHEEN S A, ZHANG H, MARTIN E, et al. China's Hangzhou Public Bicycle[J]. *Transportation Research Record: Journal of the Transportation Research Board*, 2011, 2247(1): 33-41.

[2] AI Y, LI Z P, GAN M, et al. Retraction Note: A Deep Learning Approach on Short-Term Spatiotemporal Distribution Forecasting of Dockless Bike-Sharing System[J]. *Neural Computing and Applications*, 2024, 36(25): 15927.

[3] LIU Z Y, SHEN Y Y, ZHU Y M. Inferring Dockless Shared Bike Distribution in New Cities[C]//The 11th ACM International Conference on Web Search and Data Mining, Marina Del Rey, CA USA, 2018.

[4] 仝照民, 刘耀林, 张紫怡, 等. 骑行流密度聚类下的共享单车源汇空间识别[J]. 武汉大学学报(信息科学版), 2025, 50(1): 184-196.
TONG Zhaomin, LIU Yaolin, ZHANG Ziyi, et al. Bike-Sharing Source-Sink Space Recognition Based on Riding Flow Density Clustering Method[J]. *Geomatics and Information Science of Wuhan University*, 2025, 50(1): 184-196.

[5] CHEN L B, ZHANG D Q, PAN G, et al. Bike Sharing Station Placement Leveraging Heterogeneous Urban Open Data[C]// ACM International Joint Conference on Pervasive and Ubiquitous Computing, Osaka, Japan, 2015.

[6] 柯日宏,吴升,柯玮文.一种识别共享单车潮汐点的时空模型和基于KNN-LightGBM的租还需求预测方法[J].地球信息科学学报,2023,25(4):741-753.
KE Rihong, WU Sheng, KE Weiwen. A Spatial-Temporal Model for Identifying Tidal Shared-Bicycle Stops and Bicycle Sharing Demand Prediction Based on KNN-LightGBM[J]. *Journal of Geo-Information Science*, 2023, 25(4): 741-753.

[7] LIU Y, WANG F H, XIAO Y, et al. Urban Land Uses and Traffic 'Source-Sink Areas': Evidence from GPS-Enabled Taxi Data in Shanghai[J]. *Landscape and Urban Planning*, 2012, 106(1): 73-87.

[8] 高楹,宋辞,舒华,等.北京市摩拜共享单车源汇时空特征分析及空间调度[J].地球信息科学学报,2018,20(8):1123-1138.
GAO Ying, SONG Ci, SHU Hua, et al. Spatial-Temporal Characteristics of Source and Sink Points of Mobikes in Beijing and Its Scheduling Strategy[J]. *Journal of Geo-Information Science*, 2018, 20(8): 1123-1138.

[9] 姜晓,白璐斌,楼夏寅,等.基于多尺度时空聚类的共享单车潮汐特征挖掘与需求预测研究[J].地球信息科学学报,2022,24(6):1047-1060.
JIANG Xiao, BAI Lubin, LOU Xiayin, et al. Usage Patterns Identification and Flow Prediction of Bike-Sharing System Based on Multiscale Spatiotemporal Clustering[J]. *Journal of Geo-Information Science*, 2022, 24(6): 1047-1060.

[10] LI S Y, ZHUANG C G, GUO E T, et al. Research on Bike-Sharing Travel Behavior: A Review of the Chinese Language Literature[J]. *Transactions in Urban Data, Science, and Technology*, 2023, 2(2-3): 100-122.

[11] 盛宇裕,毕硕本,范京津,等.运用交通运行状况指标分析交通热点时空模式[J].武汉大学学报(信息科学版),2021,46(5):746-754.
SHENG Yuyu, BI Shuoben, FAN Jingjin, et al. Analyzing Spatiotemporal Patterns of Traffic Hotspots Using Traffic Operation Indicators[J]. *Geomatics and Information Science of Wuhan University*, 2021, 46(5): 746-754.

[12] FANG Z X, YANG X P, XU Y, et al. Spatiotemporal Model for Assessing the Stability of Urban Human Convergence and Divergence Patterns[J]. *International Journal of Geographical Information Science*, 2017, 31(11): 2119-2141.

[13] 高楹,宋辞,郭思慧,等.接驳地铁站的共享单车源汇时空特征及其影响因素[J].地球信息科学学报,2021,23(1):155-170.
GAO Ying, SONG Ci, GUO Sihui, et al. Spatial-Temporal Characteristics and Influencing Factors of Source and Sink of Dockless Sharing Bicycles Connected to Subway Stations[J]. *Journal of Geo-Information Science*, 2021, 23(1): 155-170.

[14] GAO F, LI S Y, TAN Z Z, et al. Understanding the Modifiable Areal Unit Problem in Dockless Bike Sharing Usage and Exploring the Interactive Effects of Built Environment Factors[J]. *International Journal of Geographical Information Science*, 2021, 35(9): 1905-1925.

[15] WONG D W S. The Modifiable Areal Unit Problem (MAUP)[M]// Janelle D G, Warf B, Hansen K,eds. WorldMinds: Geographical Perspectives on 100 Problems. Dordrecht :Springer, 2004:571-575.

[16] 唐建波,刘启亮,刘博,等.一种基于重排检验的时空聚类方法[J].武汉大学学报(信息科学版),2017,42(4):503-511.
TANG Jianbo, LIU Qiliang, LIU Bo, et al. A Spatiotemporal Clustering Method Based on Permutation Test[J]. *Geomatics and Information Science of Wuhan University*, 2017, 42(4): 503-511.

[17] 陆川伟,孙群,季晓林,等.一种基于核距离的车辆轨迹点聚类方法[J].武汉大学学报(信息科学版),2020,45(7):1082-1088.
LU Chuanwei, SUN Qun, JI Xiaolin, et al. A Method of Vehicle Trajectory Points Clustering Based on Kernel Distance[J]. *Geomatics and Information Science of Wuhan University*, 2020, 45(7): 1082-1088.

[18] 刘耀林,方飞国,王一恒.基于手机数据的城市内部就业人口流动特征及形成机制分析:以武汉市为例[J].武汉大学学报(信息科学版),2018,43(12):2212-2224.
LIU Yaolin, FANG Feiguo, WANG Yiheng. Characteristics and Formation Mechanism of Intra-Urban Employment Flows Based on Mobile Phone Data—Taking Wuhan City as an Example[J]. *Geomatics and Information Science of Wuhan University*, 2018, 43(12): 2212-2224.

[19] 李昊明,蔡忠亮.土地利用紧凑性与城市交通拥堵关联研究[J].测绘地理信息,2021,46(S1):20-24.
LI Haoming, CAI Zhongliang. Correlation Between Land Use Compactness and Urban Traffic Congestion[J]. *Journal of Geomatics*, 2021, 46(S1): 20-24.

[20] 石岩,王达,陈袁芳,等.流空间邻近关系约束下的流行病分布空间异常探测方法[J].测绘学报,2021,50(6):777-788.
SHI Yan, WANG Da, CHEN Yuanfang, et al. An Anomaly Detection Approach from Spatio Distributions of Epidemic Based on Adjacency Constraints in Flow Space[J]. *Acta Geodaetica et Cartographica*

Sinica, 2021, 50(6): 777-788.

[21] PEI T, WANG W Y, ZHANG H C, et al. Density-Based Clustering for Data Containing Two Types of Points[J]. *International Journal of Geographical Information Science*, 2015, 29(2): 175-193.

[22] LIU Q L, WU Z H, DENG M, et al. Network-Constrained Bivariate Clustering Method for Detecting Urban Black Holes and Volcanoes[J]. *International Journal of Geographical Information Science*, 2020, 34(10): 1903-1929.

[23] 曹昕婷, 刘昭. 厦门城市空间演化及发展探析[J]. 中南林业科技大学学报, 2009,29(3):181-184.
CAO Xinting, LIU Zhao. Analysis on Urban Spatial Evolution and Development of Xiamen City[J]. *Journal of Central South University of Forestry and Technology*, 2009,29(3):181-184.

[24] 李峰清, 赵民, 吴梦笛, 等. 论大城市"多中心"空间结构的"空间绩效"机理: 基于厦门LBS画像数据和常规普查数据的研究[J]. 城市规划学刊, 2017(5): 21-32.
LI Fengqing, ZHAO Min, WU Mengdi, et al. Polycentric Mage-City and Its Mechanism of Spatial Performance: Findings from Xiamen Based on LBS and Census Data[J]. *Urban Planning Forum*, 2017(5): 21-32.

[25] DENG M, YANG X X, SHI Y, et al. A Density-Based Approach for Detecting Network-Constrained Clusters in Spatial Point Events[J]. *International Journal of Geographical Information Science*, 2019, 33(3): 466-488.

[26] BORRUSO G. Network Density Estimation: Analysis of Point Patterns over a Network[C]//Computational Science and Its Applications — ICCSA 2005. Berlin, Heidelberg: Springer, 2005: 126-132.

[27] 吕峥, 孙群, 赵国成, 等. 顾及方向关系的农村居民地聚类方法[J]. 武汉大学学报(信息科学版), 2023, 48(4): 631-638.
LÜ Zheng, SUN Qun, ZHAO Guocheng, et al. A Clustering Method of Rural Settlement Considering Direction Relation[J]. *Geomatics and Information Science of Wuhan University*, 2023, 48(4): 631-638.

[28] 刘瑜, 汪珂丽, 邢潇月, 等. 地理分析中的空间效应[J]. 地理学报, 2023, 78(3): 517-531.
LIU Yu, WANG Keli, XING Xiaoyue, et al. On Spatial Effects in Geographical Analysis[J]. *Acta Geographica Sinica*, 2023, 78(3): 517-531.

[29] SHANNON C E. A Mathematical Theory of Communication[J]. *Bell System Technical Journal*, 1948, 27(3): 379-423.

[30] ZHAO P X, QIN K, YE X Y, et al. A Trajectory Clustering Approach Based on Decision Graph and Data Field for Detecting Hotspots[J]. *International Journal of Geographical Information Science*, 2017, 31(6): 1101-1127.

[31] SHIODE S. Street-Level Spatial Scan Statistic and STAC for Analysing Street Crime Concentrations[J]. *Transactions in GIS*, 2011, 15(3): 365-383.

[32] LI W W, WANG S H, ZHANG X Y, et al. Understanding Intra-Urban Human Mobility Through an Exploratory Spatiotemporal Analysis of Bike-Sharing Trajectories[J]. *International Journal of Geographical Information Science*, 2020, 34(12): 2451-2474.