

DOI:10.13718/j.cnki.xsxb.2018.10.013

基于大数据的南京市共享单车时空特征研究^①

周 超, 周亚男, 李振世, 杨泽航

河海大学 地球科学与工程学院, 南京 211100

摘要:共享单车的兴起极大地便利了人们的短距离出行,利用单车大数据来挖掘和分析短距离的出行规律已成为智慧城市和智慧交通的重要热点.利用南京市的摩拜单车时空数据,研究了单车的时空分布规律与差异,并重点挖掘了早晚高峰时刻以及区域间单车流动规律等,研究表明:①南京市整体早、晚高峰时间分别为 7:10—8:37 和 17:15—18:25;副中心早高峰比主城区早 35 min 开始,且晚高峰持续时间长;②早、晚高峰高强度骑行主要分布在以新街口、百家湖等为中心的区域;主城区和副中心区有着骑行差异,短距离骑行主城区相对副中心具有更高的比例,副中心则具有更多的长距离骑行;③单车骑行围绕地铁站而显现聚集趋势,并且集中于地铁 1,2,3 号线区域.

关键词:共享单车大数据;短距离出行;早晚高峰;区域流动

中图分类号: P208

文献标志码: A

文章编号: 1000-5471(2018)10-0066-08

随着经济的飞速发展与人心的急剧膨胀,城市建城面积不断扩张,相应地人们的生活与工作半径越来越大,城市内人口流动的强度与范围也越来越大^[1],但同时伴随的是交通拥挤日趋严重,而拥堵主要原因是城市交通设施的规划和建设落后于交通需求的增长^[2-3],因此在早、晚高峰期间选择地铁、公交等出行方式更为便捷,而共享单车有效地解决了城市轨道交通和每个人家门口之间的“最后一公里”问题.大规模的单车数据为早、晚高峰分析研究提供了可靠的数据源.基于轨迹大数据的分析^[4],能够深层次地挖掘诸如城市早、晚高峰时刻、道路网拥堵^[5-6]、车站停车场设置优劣^[7]等信息,进而为城市道路与站点规划^[8]、交通预测^[9]、工作与出行^[10]等提供决策支持.本文基于 2017 年 4 月上中旬南京市摩拜单车停放数据,利用时空统计与分析^[11-12]方法,分析南京市主城区单车骑行的日变化曲线,并进一步定义和分析南京市的早、晚高峰时刻;分析单车骑行热度的时空分布^[13],并对比南京市主城区与江宁副中心的出行特征;分析南京市城区早、晚高峰时刻单车的流动情况,进而结合地铁与公交站点的空间分布,挖掘单车流动与南京市轨道交通的衔接度^[14].

1 研究区与数据概况

南京市是我国东部地区重要的中心城市,同时也是摩拜、OFO、哈罗等共享单车首批及重点运营的城市.选择城市发展水平、人群职住分布、城市轨道交通等因素,本文以南京市城区为研究区(图 1),研究区包括鼓楼、秦淮、玄武、雨花台、江宁等行政区;其中秦淮、鼓楼、玄武和建邺是南京市主城区,人口密度大,单车聚集度高、骑行强度大;江宁区作为副中心区,其面积大,人口基数多,单车聚集度低,单车使用

① 收稿日期: 2018-05-04

基金项目: 国家自然科学基金项目(41501453);中央高校基本科研业务费项目(2016B11414).

作者简介: 周 超(1997-),男,本科生,主要从事地理信息科学研究.

通信作者: 周亚男,博士.

率高。因此南京市主城区与江宁副中心区的单车分布与流动能够大致反映出南京市城区短距离出行整体情况。

单车数据共包括 2017 年 4 月 5 日(周三)至 11 日(周二)和 13 日(周四)8 天的单车停放数据; 其中因为 6 日全天降雨, 因此在分析工作日单车数据时用 13 日数据代替。单车数据记录了研究区域内全天每隔 5 min 的单车停放信息, 包括单车的 ID 标识、单车类型和单车空间位置。TimeStamp 记录单车停放的时间, BikeID 记录单车的 ID, BikeType 记录单车的类型(本文不作研究), Lon 和 Lat 记录单车停放的经纬度, 它们分别反映单车的时间特性、属性特性、位置特性^[15]。



图 1 研究区域范围

2 数据处理与分析

图 2 为数据处理与分析的整体流程, 包括单车数据的静止状态到运动状态转换、单车整体骑行时间变化与早晚高峰定义、早晚高峰单车与骑行的空间分布、早晚高峰时段内主城区与副中心的单车骑行对比以及单车的流动与地铁衔接分析等。

1) 数据预处理 原始数据去重, 将具有相同时间、单车 ID 且相近位置(位移小于 50 m)的记录删除, 保留首尾记录; 单车状态转换, 将在 2 个相邻时刻发生位移的单车标记为该时间内被骑行的状态, 获得单车骑行数据。

2) 统计单车骑行时空分布与变化 在骑行数据基础上, 以 500 m 建立空间格网, 统计整体和局部区域的单车数量、骑行次数、距离、空间分布、强度等; 依据时间变化曲线分析南京市(包括主城区与江宁副中心)早、晚高峰时刻。

3) 早、晚高峰单车与骑行的空间分布与差异对比 依据骑行强度空间分布, 划分南京市主城区与江宁副中心区并进行区域间的单车骑行情况(距离、复杂度等)与早晚高峰时刻对比。

4) 单车的流动分析 将研究区进行格网化处理, 分析格网间单车的流动情况与早晚高峰的骑行相关性, 并结合南京地铁站点分布分析单车与轨道交通的衔接情况。

5) 数据可视化 将分析成果叠加到高德地图上进行可视化分析^[16]。

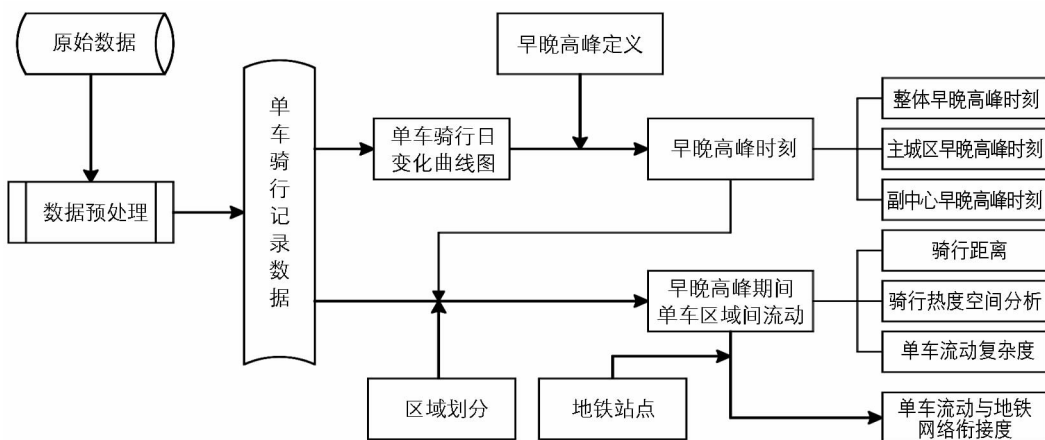


图 2 数据处理与分析流程

2.1 单车骑行与早、晚高峰定义

2.1.1 单车骑行数日变化曲线

本文以 10 min 为时间间隔, 统计研究区一天内单车的骑行次数; 并利用移动平均法对数据进行平滑, 得到单车骑行次数日变化曲线(图 3). 由图 3 可知, 6 日与 10 日曲线异常, 可能与天气情况有关(6 日全天阵雨, 10 日上午无雨、下午阵雨); 相对于周末日变化曲线(周末曲线没有峰值, 从 8:00 骑行次数开始增长, 直到 18:00 后下降), 工作日的单车骑行具有早、中、晚 3 个高峰.

相关分析是研究随机变量之间相关关系的一种统计方法, 通过相关分析可以快捷、高效地发现事物间内在关联^[17]. 进一步分析一周内单车骑行次数的相关性, 如表 1 所示为一周内骑行次数的相关系数. 工作日之间(除去周一曲线)、周末之间分别具有高度相似性, 如周二与周三、周四、周五的相关系数分别达到 0.975, 0.988, 0.946; 周六与周日的系数为 0.968; 但工作日和周末之间的相关性就较低, 如周日与周二、周三、周四、周五的相关系数分别为 0.859, 0.903, 0.845, 0.906, 远低于工作日之间、周末之间的相关性. 结果表明, 骑行次数在工作日每天相同时间段内具有高度的相似性, 说明人们每天的出行时间几乎保持着同步. 曲线间高相似性为下文将工作日曲线拟合成一条曲线提供了理论可行性.

表 1 骑行次数相关系数

时间	周一	周二	周三	周四	周五	周六	周日
周一	1.000	0.523	0.513	0.600	0.357	0.426	0.328
周二	0.523	1.000	0.975	0.987	0.946	0.903	0.859
周三	0.513	0.975	1.000	0.961	0.921	0.927	0.903
周四	0.600	0.988	0.961	1.000	0.936	0.903	0.845
周五	0.357	0.946	0.921	0.936	1.000	0.938	0.906
周六	0.426	0.903	0.927	0.903	0.938	1.000	0.968
周日	0.328	0.859	0.903	0.845	0.906	0.968	1.000

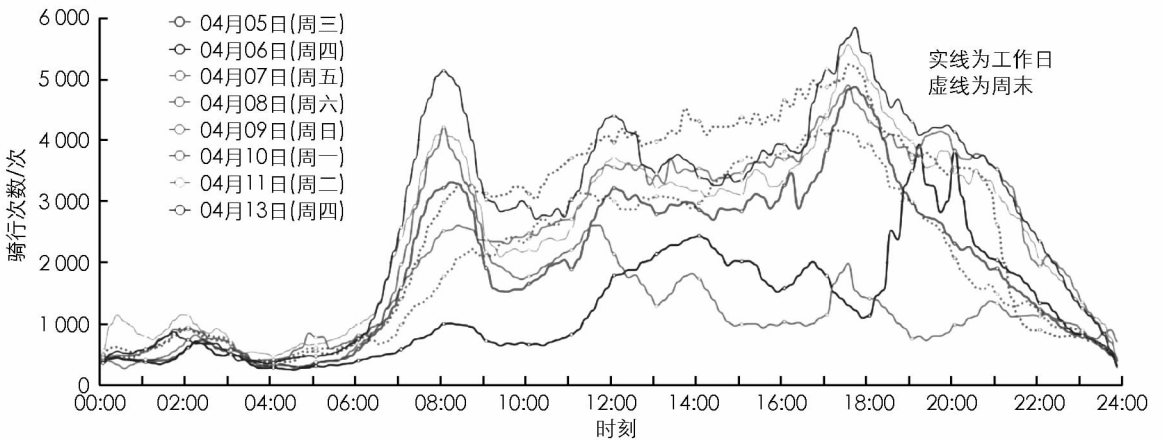


图 3 骑行次数日变化曲线图

2.1.2 早、晚高峰定义

对工作日骑行次数曲线拟合分析发现, 其高峰近似符合高斯曲线, 早高峰曲线形似于图 4a, 晚高峰曲线形似于图 4b; 为此定义并计算南京的早、晚高峰时刻. ①图 4a: T 点是峰值点, 对应时刻为 t_p . A, B 点分别是时段 $0-t_p, t_p-t_c$ 内曲线中切线斜率最大的点, 对应时间分别为 t_s, t_c , 则定义高峰时间段为 t_s-t_c . 由此结合图 3 确定南京市整体早高峰时间为 7:10—8:37. ②图 4b: T 点是峰值点, 对应时刻为 t_p , A 点是时间 $0-t_p$ 内曲线中切线斜率最大的点, 对应时间为 t_s ; B 点是 $y(B)=y(A)$ 的某点, 对应时间为 t_c ; 则定义高峰时间段为 t_s-t_c (反偏态曲线同理). 由此确定南京市整体晚高峰时间为 17:15—18:25.

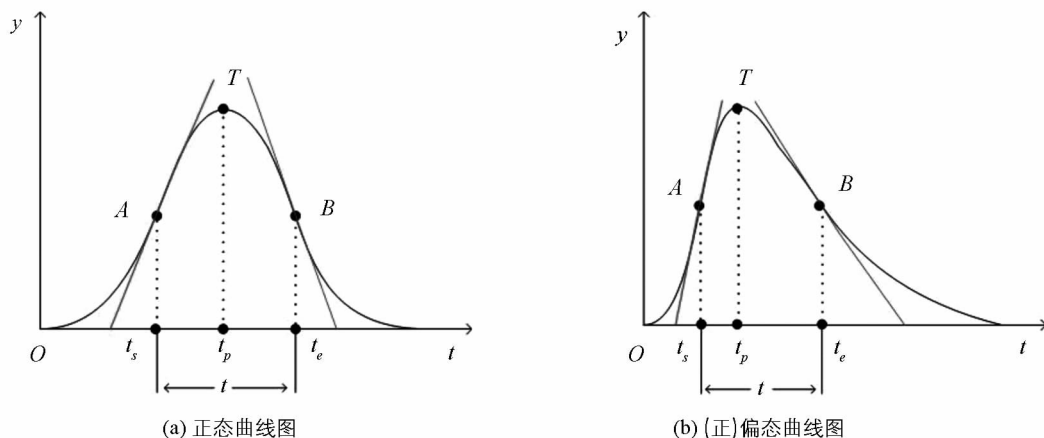


图 4 早晚高峰定义图

2.2 早晚高峰的单车总体分析

2.2.1 单车空间分布

如图 5 所示, 统计 7 天内零点时刻的单车停放空间分布情况, 由分布图可知: ①主城区单车聚集度很大, 郊区聚集度较小; ②单车聚集在地铁站、学校、社区等人流量较大的区域, 尤其是换乘站如油坊桥站、新街口站等更是单车高聚集度区域; ③以地铁站为中心分别作 200 m、500 m 的缓冲区, 平均有 7.06% 单车位于 200 m 缓冲区内、28.90% 的单车位于 500 m 缓冲区内. 定量研究地铁站缓冲区面积与单车数量间关系, 统计发现缓冲区面积增加 5.25 倍, 而单车数目只增加 3.09 倍; 结果表明随着与地铁站距离增大, 缓冲区面积大幅度增加, 而单车数目增加幅度较小, 即单车的分布基于南京市地铁轨道交通而显现聚集趋势.

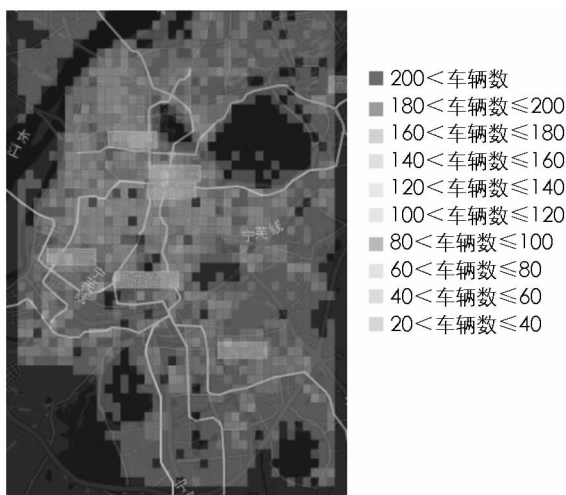


图 5 单车分布图

2.2.2 骑行率空间分布

骑行率为某一时段区域内骑行次数与总骑行次数的比值, 反映单车骑行的强度信息. 图 6 为早、晚高峰期间单车骑行强度的空间分布图, 图中围绕新街口、奥体中心、百家湖等形成了局部中心. ①早、晚高峰期间, 南京市主城区的新街口、副中心百家湖和奥体中心区域都是车辆骑行热度大、单车聚集度高的区域, 且主城区骑行强度远远大于郊区; ②早高峰期间有 41 971 次单车被骑行, 晚高峰期间有 40 534 次单车被骑行; 且早高峰持续 87 min, 晚高峰持续 70 min, 晚高峰单位时间内单车的骑行次数大于早高峰, 表明晚高峰人口流动密度大于早高峰; ③早、晚高峰骑行强度有着对应关系, 早高峰期间骑行强度大的区域晚高峰骑行强度依然很大, 早高峰期间骑行强度小的区域晚高峰骑行强度依然小; ④单车主要依靠城市轨道交通流动, 含有地铁站的区域车辆流动比未含有地铁站区域多.

2.3 核心区与副中心区早晚高峰对比分析

2.3.1 早、晚高峰时间对比

依据早、晚高峰的单车数量和骑行强度空间分布, 将南京市分为以新街口为中心的主城区、以奥体为中心的河西区和以百家湖为中心的江宁副中心区. 分析与对比主城区与副中心区的单车骑行情况能够进一步揭示城市中心与周边的早、晚高峰差异. 为此以主城区(图 1 中 A 区域)和副中心区(图 1 中 B 区域)为例, 对比分析单车骑行的区域差异.

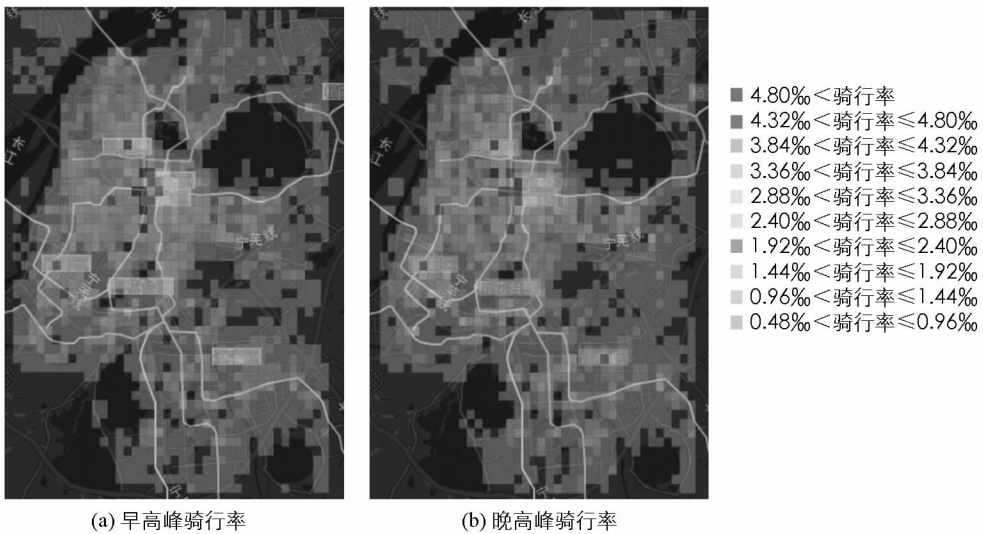


图 6 早晚高峰骑行率分布图

图 7 所示为主城区和副中心区的单车骑行次数日变化图(实线主城区, 虚线副中心区), 对比发现: ①主城区的早、晚高峰并没有副中心区的明显; ②主城区的早、晚高峰时段分别为 7:35—8:50 和 17:25—17:53, 副中心早、晚高峰时段分别为 7:00—8:35 和 16:55—18:05, 且副中心区表现出了中午高峰特征(时段为 11:35—12:30); 副中心区的早高峰相对于主城区要提前 35 min, 且晚高峰持续时间长; ③主城区早晚高峰的单车骑行次数约为副中心的 3 倍, 且早、晚高峰时段之外也有较多的骑行次数, 而副中心只在中午有个明显的小高峰, 其他时间段骑行次数较少; ④利用率为某一时段区域内骑行次数与单车数量的比值, 反映单车的使用率. 进一步对比单车利用率发现, 主城区单车使用率为 87.139%, 副中心为 99.556%, 虽然副中心车辆少, 骑行次数少, 但是副中心单车使用率高于主城区.

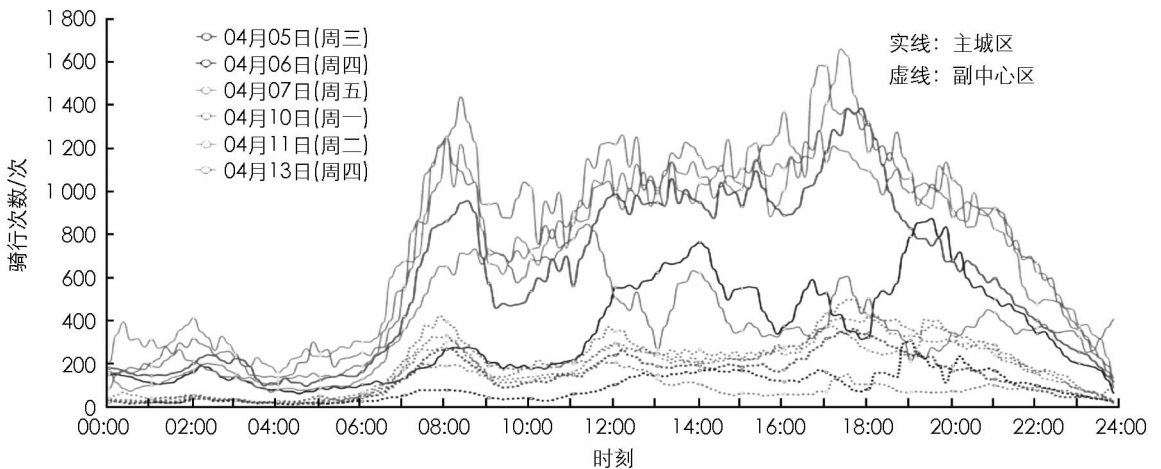


图 7 骑行次数日变化曲线图

2.3.2 骑行距离对比

基于高德地图骑行路径规划 API 对早、晚高峰期间主城区和副中心区的骑行距离(以高德地图给出的最短骑行路径距离作为真实的单车骑行距离)进行统计分析(考虑到距离大于 7 000 m 的骑行很少, 故只取小于 7 000 m 的骑行进行分析), 结果如图 8 所示. 早、晚高峰主城区和副中心区都以短距离骑行为主, 骑行距离在 1 500 m 以内的主城区占据 58.612%, 副中心占据 50.798%, 晚高峰期间, 主城区占据 50.960%, 副中心占据 49.302%. 晚高峰骑行距离相对于早高峰骑行距离要更远(可能是早高峰为上班时

段, 时效性要求较高); 1 000 m 距离内的短距离骑行, 主城区相对副中心具有更高的比例(主城区 34.42%, 副中心 29.05%), 副中心则具有更多的长距离骑行。

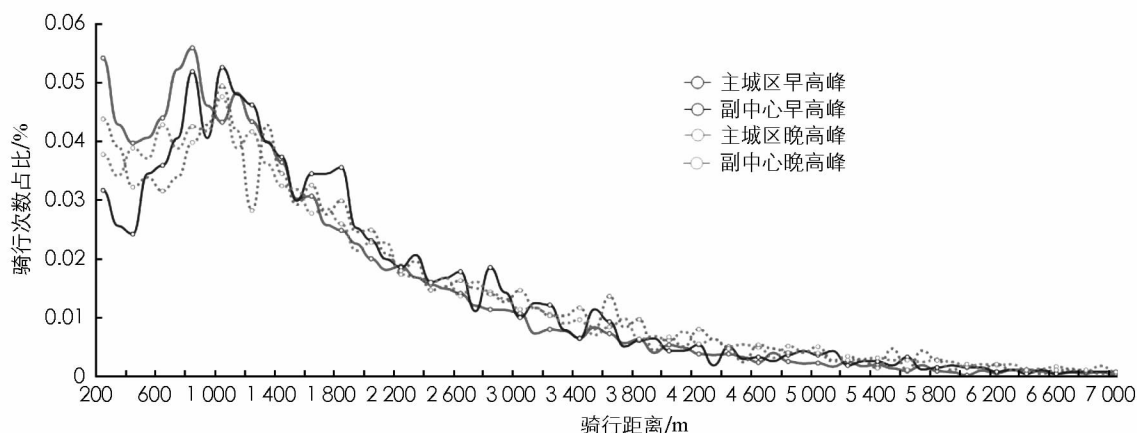


图 8 骑行距离占比图

2.3.3 早高峰与晚高峰关系对比

将研究区进行大小为 1 000 m 的格网化处理(左下角编号为 0, 向上编号依次增加, 后续列同理), 分析早、晚高峰时段区域间的单车流动情况, 结果如图 9 所示. 其中纵坐标为骑行次数, 横坐标为 $[A, B]$ (表示车辆从区域 A 移动到 B); 同一横坐标下, 早高峰曲线表示 A 区域到 B 区域间流动次数, 而晚高峰曲线表示 B 到 A 的次数(即晚高峰表示早高峰流动的逆向); 且考虑到远距离骑行的数量较少和偶然因素造成单车移动状况发生, 故舍掉骑行距离大于 7 000 m 和骑行次数小于 10 次的区域间流动. 主城区和副中心的单车骑行多数(主城区 41.90%, 副中心 55.04%)发生在自己的区域内, 即反映出单车的短距离行驶特性; 主城区的横坐标刻度远远多于副中心, 而横坐标表示发生单车流动的区域, 可见主城区单车流动网络更为复杂; 主城区在早高峰和晚高峰间区域单车流动量有着同步关系, 即早高峰在 $[A, B]$ 内单车骑行次数多, 晚高峰在 $[A, B]$ 内依然多; 早高峰在 $[A, B]$ 内单车骑行次数少, 晚高峰在 $[A, B]$ 内依然少, 而江宁区不存在同步关系, 这可能是由于晚高峰按照早高峰的路线原路返回, 主城区车辆多可以提供这个条件, 而副中心因为车辆少的原因, 无法为晚高峰期间提供充足骑行。

2.4 单车与轨道交通衔接分析

从上述分析中得知, 单车主要依靠城市轨道交通流动, 含有地铁站的区域车辆流动比未含有地铁站区域多. 现以南京市早、晚高峰单车流动为基础研究早、晚高峰期间流入每条地铁线内地铁站 200 m 缓冲区和 500 m 缓冲区的单车次数, 发现早高峰骑行到 200 m 缓冲区有 6 749 次, 500 m 缓冲区有 14 374 次; 晚高峰 200 m 缓冲区有 5 204 次, 500 m 缓冲区 12 638 次; 进入每条地铁线次数结果如表 2 所示(斜线前为 200 m 缓冲区次数, 斜线后为 500 m). 统计进入地铁站 200 m 缓冲区各个线的占比发现, 地铁 1, 2, 3 号线早高峰总占比为 88.34%, 晚高峰总占比为 81.80%, 由此可知早、晚高峰人口流动主要集中地铁 1, 2, 3 号线. 定量研究缓冲区距离与单车数目及骑行到缓冲区次数之间规律发现, 地铁站缓冲区从 200 m 变为 500 m, 缓冲区面积增加 5.25 倍, 单车数目增加 3.09 倍, 而早、晚高峰骑行到缓冲区次数平均增加 1.28 倍. 通过单位面积作为单车骑行目的地的骑行次数这一特征值反映单车的流向规律发现, 与地铁站相距 0~200 m 区域特征值为 200~500 m 区域的 2.41 倍. 由此可知单车骑行围绕地铁站而显现聚集趋势, 即多数的单车骑行目的地为地铁站并与地铁轨道交通相衔接。

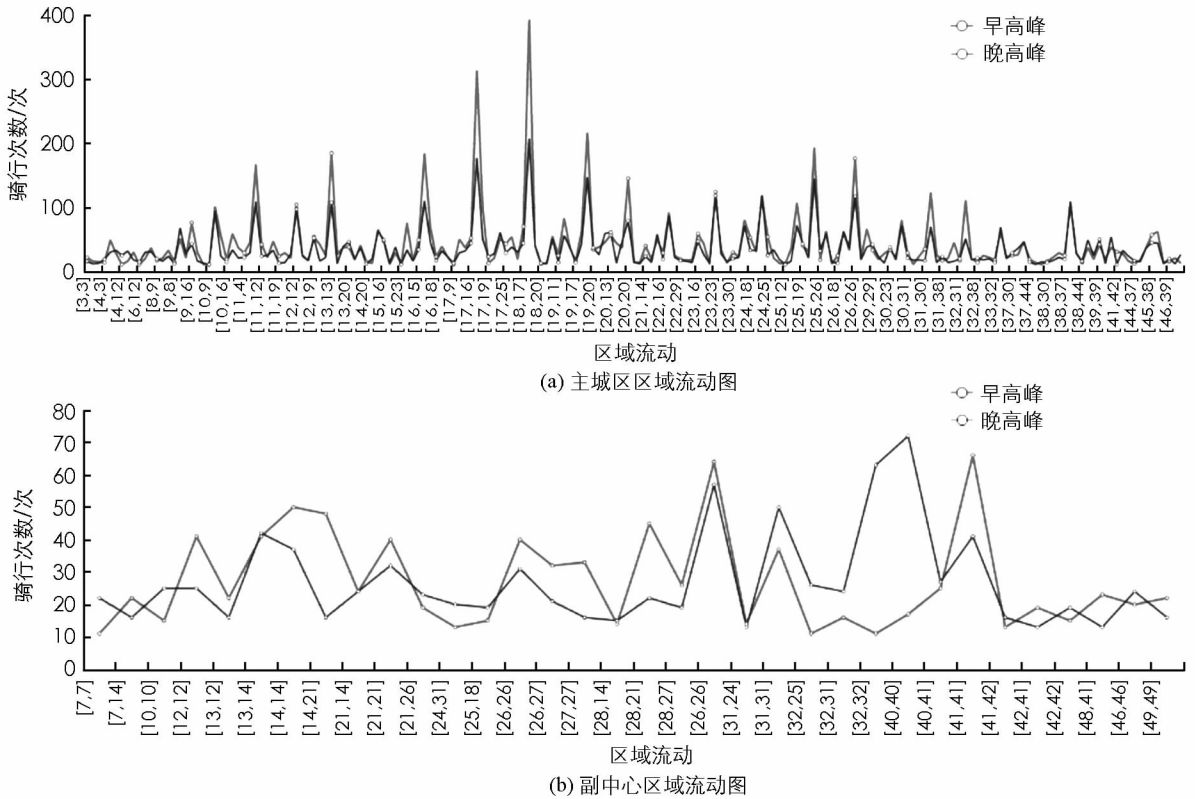


图 9 车辆区域间流动数值图

表 2 早晚高峰期间单车进入各个地铁线情况

	1 号线	2 号线	3 号线	4 号线	10 号线	其他线
(早)进入地铁线/次	2 684/5 901	1 844/3 680	1 434/2 972	418/955	199/511	170/315
(晚)进入地铁线/次	1 557/4 166	1 537/3 322	1 163/2 829	359/855	353/555	235/911

3 结 论

本文基于 2017 年 4 月上中旬南京市城区的摩拜单车时空位置数据,研究南京市单车的时空分布规律与差异,并重点挖掘早、晚高峰时刻、单车骑行空间分布规律.结果表明:①南京市整体早、晚高峰时间分别为 7:10—8:37 和 17:15—18:25;副中心早高峰比主城区早 35 min 开始,且晚高峰持续时间长;②早、晚高峰高强度骑行主要分布在以新街口、奥体中心和百家湖等为中心的区域;并且主城区和副中心区有着骑行差异,短距离骑行内主城区相对副中心具有更高的比例,副中心则具有更多的长距离骑行;③约三成单车骑行基于轨道交通,并且集中于地铁 1,2,3 号线内地铁站区域.本文只从单车方面研究规律,缺少单车使用者的相关信息,例如姓名、年龄、职业、性别等,如若将这些因素一起研究,即可进入更高层次单车出行目的地预测研究.

参考文献:

- [1] 王振坡,李 锦,程浩岩.我国大城市居民出行特征演变与交通发展研究[J].城市,2016(8):73—79.
- [2] 赵鹏军,万海荣.我国大城市交通拥堵特征与国际治理经验借鉴探讨[J].世界地理研究,2016,25(5):48—57.
- [3] 宗会明,戴技才,乔 宏,等.渝中区综合交通发展格局及拥堵原因解析[J].西南大学学报(自然科学版),2014,36(2):143—149.
- [4] 刘经南,方 媛,郭 迟,等.位置大数据的分析处理研究进展[J].武汉大学学报(信息科学版),2014,39(4):379—385.
- [5] 张俊涛,李志勇,张 浩,等.利用出租车轨迹数据估计城市道路拥堵状况[J].测绘工程,2016,25(9):68—72,76.

- [6] 付子圣, 李秋萍, 柳 林, 等. 利用 GPS 轨迹二次聚类方法进行道路拥堵精细化识别 [J]. 武汉大学学报(信息科学版), 2017, 42(9): 1264—1270.
- [7] 李 军, 秦其明, 游 林, 等. 利用浮动车数据提取停车场位置 [J]. 武汉大学学报(信息科学版), 2013, 38(5): 599—603.
- [8] 杨 伟, 艾廷华. 基于车辆轨迹大数据的道路网更新方法研究 [J]. 计算机研究与发展, 2016, 53(12): 2681—2693.
- [9] NIU Xiao-guang, ZHU Ying, CAO Qing-qing, et al. An Online-Traffic-Prediction Based Route Finding Mechanism for Smart City [J]. International Journal of Distributed Sensor Networks, 2015, 2015.
- [10] LI Qing-quan, ZENG Zhe, ZHANG Tong, et al. Path-Finding Through Flexible Hierarchical Road Networks: An Experiential Approach Using Taxi Trajectory Data [J]. International Journal of Applied Earth Observations and Geoinformation, 2011, 13(1): 110—119.
- [11] 吉根林, 赵 斌. 时空轨迹大数据模式挖掘研究进展 [J]. 数据采集与处理, 2015, 30(1): 47—58.
- [12] 夏 英, 梁中军, 王国胤. 基于时空分析的短时交通流量预测模型 [J]. 南京大学学报(自然科学版), 2010, 46(5): 552—560.
- [13] 赵鹏祥. 基于轨迹聚类的城市热点区域提取与分析方法研究 [D]. 武汉: 武汉大学, 2015.
- [14] 姜晓睿, 田 亚, 蒋 莉, 等. 城市道路交通数据可视分析综述 [J]. 中国图象图形学报, 2015, 20(4): 454—467.
- [15] 沈敬伟, 周廷刚, 朱晓波. 基于 GPS 浮动车数据的交通信息时空分布研究 [J]. 西南大学学报(自然科学版), 2015, 37(8): 157—162.
- [16] 王祖超, 袁晓如. 轨迹数据可视分析研究 [J]. 计算机辅助设计与图形学学报, 2015, 27(1): 9—25.
- [17] 梁吉业, 冯晨娇, 宋 鹏. 大数据相关分析综述 [J]. 计算机学报, 2016, 39(1): 1—18.

On Space-Time Characteristics of Shared Bikes in Nanjing Based on Big Data

ZHOU Chao, ZHOU Ya-nan, LI Zhen-shi, YANG Ze-hang

School of Earth Sciences and Engineering, Hohai University, Nanjing 211100, China

Abstract: The rise of the current shared bikes greatly facilitates people's travel of short distances, and with the big data of shared bikes, studies have been done to mine and analyze the traveling pattern of people in the field of Smart City and Smart Transportation which has become a important focus. With the space-time data of the MOBIKE located in Nanjing studies the space-time pattern regularity and differences of the shared bikes and focus on mining the rush hour and regional flow of the shared bikes, etc.. The results show that, firstly, the overall morning rush time of Nanjing is about 7:10-8:37, and the evening rush time is nearly 17:15-18:25; the morning rush hours of sub-center of Jiangning District began 35 minutes earlier than the main urban area, and the evening rush last for a longer time. Secondly, the high intensity riding in the morning and evening rush is mainly distributed in the main urban area centered around Xinjiekou and the sub-center of Baijia Lake. Thirdly, there are some riding difference laying between the main urban area and the Baijia Lake, the deputy central district. The main urban area has a greater proportion in short distance riding, while the deputy central district has more long-distance riding. Eventually, cycling shows a tendency to gather around subway stations, and is concentrated in metro line 1, 2, and 3.

Key words: the big data of shared bikes; short distance travel; morning and evening rush; regional core