

国内图书分类号：U491
国际图书分类号：656

密级：公开

西南交通大学
研究生学位论文

基于功能区的共享单车出行预测
及调度研究

年 级2017 级
姓 名李玥阳
申请学位级别工程硕士
专 业交通运输工程
指 导 老 师帅斌

二零二零年 七 月

Classified Index: U491

U.D.C: 656

Southwest Jiaotong University

Master Degree Thesis

RESEARCH ON PREDICITON AND
SCHEDULING OF SHARED BICYCLE TRIPS
BASED ON FUNCTIONAL AREA

Grade: 2017

Candidate: Li Yueyang

Academic Degree Applied for: Master of Engineering

Speciality: Traffic and Transportation Engineering

Supervisor: Shuai Bin

July. 2020



2017210997

西南交通大学

学位论文版权使用授权书

本学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权西南交通大学可以将本论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复印手段保存和汇编本学位论文。

本学位论文属于

1. 保密□，在 年解密后适用本授权书；
2. 不保密☒，使用本授权书。

（请在以上方框内打“√”）

学位论文作者签名：

李明阳

指导老师签名：

mp文武

日期：2020.5.18

日期：2020.5.18



2017210997

西南交通大学硕士学位论文主要工作（贡献）声明

本人在学位论文中所做的主要工作或贡献如下：

（1）对影响出行者选择共享单车的因素从宏观及微观两个层面进行分析，并结合成都市共享单车大数据揭示了使用者的出行特征。其中除了包括最基本的出行时间、距离等要素外，还增加了与地铁接驳过程中的单车 OD 分布，及用数据可视化的方法展示了单车出行与城市功能区属性间存在的潜在联系。

（2）由于共享单车的时空分布特性具有一定的区域性，本文创新性的引入功能区的概念。并选择能较好提取共享单车时空关联性的长短记忆网络和能较好捕捉共享单车功能区关联性的卷积神经网络作为构建单车 OD 分布时空预测框架的基本模型。深度学习模型所展示的单车到达预测结果能为调度者提供一定的理论依据。

（3）探讨基于功能区及不同时间戳下的共享单车出行规律，建立基于功能区划分的共享单车时空调度模型，并在模型中对单车行为和调度行为进行相应约束。通过对时空调度的剖析，设计了改进型蚁群算法进行求解。

本人郑重声明：所呈交的学位论文，是在导师指导下独立进行研究工作所得的成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对本文的研究做出贡献的个人和集体，均已在文中作了明确说明。本人完全了解违反上述声明所引起的一切法律责任将由本人承担。

学位论文作者签名：

李明阳

日期：2020.5.18



摘 要

伴随着我国交通问题日益严峻,共享单车的出现很好地迎合了这一症结,成为城市出行链中必不可少的一环。这种灵活性和波动性较大的交通工具从诞生之初起就成为了当今的研究热点,尤其是在时空特征、时空行为、理论方法和技术手段等方面仍然值得深入研究。

本文深入分析了影响出行者选择共享单车的各种因素,再根据数据挖掘的结果用数据可视化的方式展示了共享单车用户的使用特征,并着重从功能区属性上分析了相应的时空特征与行为规划。此外,针对共享单车的动态时空分布特征,采用一种新的深度学习模型对单车的 OD 分布进行预测研究,并在此基础上建立了基于功能区划分的共享单车时空调度模型和相应的求解算法。主要工作与创新如下:

(1) 首先对成都市共享单车的大数据进行相应的数据处理,并从出行行为层面和空间功能属性两个层面对单车使用特征进行可视化分析。数据与特征挖掘将从工作日和非工作日两部分进行分析,涉及骑行的距离分布、距离的时长分布、频率分布、时间分布、地铁功能区以及其他功能区的潮汐行为等方面。通过数据可视化的方法,还发现了共享单车的出行行为以及与功能区属性的潜在联系。

(2) 基于共享单车的时空特征与功能区属性存在的潜在规律,创新性地将功能区属性这个概念引入 OD 预测模型中。并将长短时记忆网络(LSTM)和卷积神经网络进行并联构建了一种新的共享单车 OD 分布的时空预测框架。这种框架实现了功能区关联性与时间关联性的同步提取,将区域出行数量分布、单车密度分布、功能区属性、时间戳属性等空间和时间变量分别输入相应的网络结构中并通过最后的全连接层进行加权,实现了一个端到端的训练框架,精确地预测了后一个时段内的区域单车的到达分布情况。

(3) 在基于功能区的共享单车时空预测结果的基础上,根据功能区出行规律和不同时间戳下的出行特征建立了基于功能区划分的共享单车时空调度模型,并在模型中对单车行为和调度行为进行相应的约束。通过对时空调度模型的剖析,设计了改进型蚁群算法进行求解。通过对模型和算法进行算例仿真,得出了分时段的近似最优调度方案,证明了模型和算法的有效性。

关键词: 共享单车; 大数据分析; 功能区属性; OD 分布预测; 长短时记忆网络(LSTM); 卷积神经网络(CNN); 时空调度; 蚁群算法



Abstract

With the increasingly serious traffic problems in China, the emergence of shared bicycles has catered to this crux, and has become an indispensable link in the urban travel chain. This kind of vehicle with greater flexibility and volatility has become a research hotspot since its inception. Especially in terms of space-time characteristics, space-time behavior, theoretical methods and technical means, it is still worthy of further study.

This article analyzes in depth the various factors that affect travellers' choice of bike sharing, and then uses the data visualization to show the use characteristics of bike sharing users based on the results of data mining, and analyzes the corresponding space-time characteristics and behaviors from the attributes of the functional area. planning. In addition, based on the dynamic spatio-temporal distribution characteristics of shared bicycles, a new deep learning model is used to predict the OD distribution of the bicycles. Based on this, a shared bicycle air-conditioning model based on functional area division and a corresponding solution algorithm are established. . The main work and innovation are as follows:

(1) First of all, the corresponding data processing of the big data of bicycle sharing in Chengdu is carried out, and the characteristics of bicycle use are visually analyzed from the two levels of travel behavior and spatial function attributes. Data and feature mining will be analyzed from working days and non-working days, including the distance distribution of riding, the length distribution of distance, frequency distribution, time distribution, tidal behavior in subway functional areas and other functional areas. Through the data visualization method, the travel behavior of the shared bicycle and the potential connection with the attributes of the ribbon were also discovered.

(2) Based on the potential law of the spatio-temporal characteristics and functional area attributes of shared bicycles, the concept of functional area attributes is innovatively introduced into the OD prediction model. A long-term and short-term memory network (LSTM) and a convolutional neural network are connected in parallel to construct a new spatio-temporal prediction framework for shared bicycle OD distribution. This framework realizes the simultaneous extraction of functional area correlation and time correlation. Space and time variables such as regional trip number distribution, bicycle density distribution, functional area attributes, and time stamp attributes are entered into the corresponding network structure and passed through the final The fully connected layer is weighted to achieve an end-to-end training framework



that accurately predicts the arrival distribution of regional bicycles in the latter period.

(3) Based on the space-time prediction results of the shared bicycle based on the functional area, based on the travel rules of the functional area and the travel characteristics under different timestamps, a model of air-conditioning for shared bicycles based on the functional area division is established, and the behavior of the bicycles in the model Constraints with scheduling behavior. Based on the analysis of the time-conditioning model, an improved ant colony algorithm is designed to solve the problem. By simulating the model and algorithm with examples, an approximate optimal scheduling scheme for each period is obtained, which proves the effectiveness of the model and algorithm.

Keywords: bike sharing; big data analysis; functional area attributes; OD distribution prediction; long and short-term memory network (LSTM); convolutional neural network (CNN);



目 录

第 1 章 绪论.....	1
1.1 研究背景.....	1
1.2 研究意义.....	2
1.2.1 理论意义.....	2
1.2.2 实际意义.....	2
1.3 研究现状.....	3
1.3.1 共享单车出行行为的相关研究.....	3
1.3.2 基于深度学习的出行预测研究.....	5
1.3.3 共享单车的调度优化研究.....	7
1.4 研究主要内容及技术路线.....	9
第 2 章 基于大数据挖掘的共享单车出行行为分析.....	11
2.1 共享单车选择影响因素分析.....	11
2.1.1 宏观因素.....	11
2.1.2 微观因素.....	13
2.2 共享单车数据处理及分析.....	16
2.2.1 数据来源.....	16
2.2.2 数据处理.....	16
2.2.3 出行数据大数据分析.....	17
2.3 本章小结.....	26
第 3 章 基于城市功能区的共享单车出行时空预测.....	27
3.1 城市功能区的划分.....	27
3.1.1 POI 数据的获取及类型判定.....	28
3.1.2 区域划分.....	29
3.2 深度学习模型.....	30
3.2.1 卷积神经网络 (Convolution Neural Networks).....	30
3.2.2 长短时记忆网络 (Long Short Term Memory Network).....	32
3.3 城市功能区下的单车时空建模框架 (整体结构已调整).....	34
3.3.1 关联性分析.....	34
3.3.2 变量分析.....	34
3.4 基于 CNN-LSTM 的 OD 分布预测.....	36
3.5 实例分析.....	38
3.5.1 划分功能区.....	39
3.5.2 共享单车 OD 分布预测.....	40
3.6 本章小结.....	44



第 4 章 基于功能区划分的共享单车时空调度模型.....	45
4.1 基于功能区特征的共享单车调度问题分析.....	45
4.2 基于功能区的共享单车时空调度模型构建.....	46
4.2.1 模型的假设.....	46
4.2.2 模型的符号说明.....	46
4.2.3 模型的构建.....	47
4.3 本章小结.....	50
第 5 章 模型的求解及算例分析.....	51
5.1 共享单车调度问题的算法分析.....	51
5.2 算法原理及求解思路.....	51
5.2.1 蚁群算法原理.....	51
5.2.2 算法求解思路.....	53
5.3 算例分析.....	55
5.4 本章小结.....	58
结论与展望.....	59
致谢.....	61
参考文献.....	62

李阳

申文斌



第 1 章 绪论

1.1 研究背景

随着共享理念的先行者——网约车在交通领域取得空前的成功,受到该模式的启发,另一种新的共享交通模式——共享单车也随之出现^[1]。共享单车作为一种公共资源,与传统公共交通所用的共享模式有明显差异。共享单车是一种全民共享,是为了解决出行者最后 1 公里问题所提出的一种可行方案。这种随用随停的“无桩”理念与“有桩”公共自行车相比,无疑是扩大了交通服务的范围,为全体市民都提供了一定程度的便利。

为解决大城市拥堵问题,仅仅依靠公交系统及轨道交通系统是远远不够的,若想提高绿色出行的整体比例还需大力鼓励出行者选择步行及非机动车。在国外,纽约作为一个世界级大都市正为了打造“绿色之都”、“世界之都”的形象而大力发展自行车公共交通。纽约市长布隆伯不仅将公共自行车系统定义为新兴公共交通工具,更是呼吁纽约市市民及游客尽可能的抛开汽车及出租车出行,选择价廉物美的自行车上下班或旅游。积极拓展全市自行车道不仅可以减少交通拥堵,更能使得城市空气质量得到极大改善^[2]。

智能共享单车的发展元年为 2016 年,而中国也正是世界上第一辆智能无桩自行车的诞生地。共享单车的诞生是中国创新精神的集中体现,是互联网技术和中国制造中国创造相结合的产物,也是一次分享经济模式的成功试验^[3]。根据有关数据显示,截止 2017 年 6 月,我国已有 1.06 亿的共享单车用户,其比例约为我国网民数量规模的 14.1%。另外,共享单车业务已经逐渐从一二线城市渗透到了三四线城市,甚至还有企业将商业蓝图规划到了海外^[5]。

不仅如此,目前的中国共享单车行业在市场的引导下快速的完成了优胜劣汰及整合优化。现如今的共享单车市场已不再处于当初的爆发式发展阶段,而是过度到了一个相对精细化发展的阶段^[4]。具体来说,在 2017 年交通运输部等部门联合发布《关于鼓励和规范互联网租赁自行车发展的指导意见》后,很多小型共享单车企业逐渐退出市场,而作为单车巨头的 OFO 和摩拜在 2018 也遭遇了破产或被兼并的结局。也就是说,共享单车行业在政策及市场的双重指引下,其市场混乱、高额红利、无人监管的疯狂时代已然过去,共享单车必将走向市场化、规范化、秩序化的正确道路上。

但共享单车虽作为物联网技术取代了传统停车桩的新事物,为方便人们出行的同时也出现了一些问题。首先,相较于普通自行车来说,共享单车的车载设备明显增加,这也在一定程度上导致使用者的骑行体验降低。目前,共享单车整个行业仍处在探索的阶段,政府的监管,企业的经营管理都有待探讨。只有企业和政府都对现阶段存在的问题从整体上把握,才能引导这个新兴产业蓬勃发展。



其次,随取随用所带来的出行结构及出行行为上的革新也是一个不可忽视的问题。共享单车是由用户行为驱动的动态系统,因此,动态性和规律性将是共享单车研究的核心问题。如何从用户行为驱动共享单车时空分布的角度去解析共享单车的时空特征及规律,是研究共享单车的基本切入点。

再者,仅从物联网技术对共享单车进行定位,管理者则很难限制和诱导自行车的运动,这也将对共享单车的调度及再分配问题增加难度。对出行者而言,调配不及时所导致的车辆密度不足,将很大程度上影响出行者的出行体验。但随着物联网和信息技术的不断发展,实时获取共享单车的出现轨迹将也成为可能,这也为共享单车的出行时空特征预测及调度规划提供了可能。

1.2 研究意义

作为科技与创新所结合的新事物——共享单车能有效解决出行者出行最后一公里问题,但共享单车如今的发展还处在不断摸索的阶段。作为城市的管理者和单车的运营者都应该设法保持共享单车的生命力,并将他和谐的融入整个城市。因此,如何满足使用者的日常骑行需求,如何更合理的对单车进行管理再调度,如何提高单车的使用效率及整个系统的服务水平都是值得本文探讨的问题。

1.2.1 理论意义

(1) 共享单车作为一种的新的公共交通工具,在运营初期缺乏相关的理论研究,无论是探索共享单车的出行行为规律还是研究共享单车的供需关系以及建立合理的调度路径优化方案都需要充实的理论基础加以指导,从而改善整体的交通情况。

(2) 共享单车的被使用次数多且使用群体覆盖面广,因此涉及的可研究数据量极大,这将导致传统的预测模型不能以较好的精度完成对共享单车出行方面的相关预测。而本文建立了新的适用于解决该类问题的预测模型,为研究共享单车领域相关问题提供了新思路。

1.2.2 实际意义

(1) 由出行者完全驱动的共享单车,其扩散规律更为复杂,且共享单车的分布更具有随机性及动态性,这也导致其时空复杂性大为提升。因此管理者需要动态的分析出行者的 OD 分布从而制定相适宜的调度策略。所以,研究共享单车的出行行为及时空特点,并揭示其运动和传播规律对构建新时代绿色交通具有重要的现实意义。

(2) 若想共享单车吸引更多的用户,满足更大范围的出行需求,就必须处理好共享单车精细化的投放及再调度问题。引导出行者把共享单车作为传统公共



交通系统的接驳方式,能有效弥补城市轨道交通所存在的缺陷,从而扩大公共交通的影响范围和吸引力^[6-8]。

1.3 研究现状

依据本论文研究侧重及目的,需对国内外学者在共享自行车出行行为及深度学习等相关方法所取得研究成果进行梳理。

1.3.1 共享单车出行行为的相关研究

1.3.1.1 国外研究现状

Jensen 等^[9] 根据法国里昂的有桩公共自行车站点及用户的使用情况,统计出该市的平均出行距离是 2.5km,而平均出行时间是 14.7min,在早高峰时期的平均时速是 15km/h。由此可见,在高峰时段的共享单车平均时速高于其机动车,这说明该时段公共自行车是比机动车更快捷的交通方式。

Buck 和 Buehler^[10] 分析了自行车道占比与自行车使用量间的关系时,加入了区域人口密度、站点生活服务水平及家庭机动车保有量等变量,得出了二者的相关性显著增加的结论。同时,作者还建议增加自行车道数量及周围的服务设施。

Harms^[11]调查了荷兰有桩自行车交通系统,发现市区的使用频率高于郊区;使用者层面来说,老年人最多,移民最少,但用户群体正在向年轻化方向爆发式增长。

Fishman^[12]在最近几年公共自行车及共享单车的研究成果基础上。分别从性别、目的、频率、收入等方面分析了二者的优缺点。并从行车安全及身体健康两方面给予了一定的建议。

Ji Hu^[13] 通过引导用户在不同网点重新平衡自行车,发现绝大多数自行车都在少数网点间循环。同时,通过对各个网点的活跃性进行定义,并优化站点间的循环路线,可提高低活性站点与高活性站点间单车的交流频率。

Pengtian Zhang^[14]根据对使用者特征及单车共享模式的探讨,分析其对共享单车需求产生的影响。提出了一种基于多阶段质量功能配置的混合模糊多准则决策的方法。该方法能较好的帮助企业提高单车的运行性能,并优化用户使用体验。

Yahya^[15]用设备综合分析法(OEE)测量了共享单车的被使用率。并通过收集真实的数据判定了自行车的自然生命周期,为公共交通系统提供了有价值的结论,同时也用辩证的思维论证了扩张与可持续发展战略。

Shen^[16]基于新加坡的有桩单车数据进行了实例分析,并在此基础上进行了时空模型的建立。模型考虑了共享单车的输入、周边环境探讨了单车规模、周围建筑环境、与公共交通的换乘与接驳及天气气候条件的影响等等,这些影响要素分析为城市运营商提供了技术支撑。



Wang^[17]利用纽约的花旗有桩自行车系统数据,建立负二项式模型估计某站台 5 个不同年龄人群的共享单车的小时交通量。结果表明天气相关变量、土地利用和建筑环境特征对总体自行车共享使用有显著影响,且不同年龄段人群的共享单车使用情况也有所差异。

Liu^[18]将基于共流(CF-Cluster)的聚类分析算法运用到有桩共享单车的站间关系,以有桩系统中所存在的“潮汐问题”,该问题主要涉及到车站布局、车辆调度策略、自行车流量分析等方面的研究。

综上,国外文献中的研究对象多为“有桩”自行车或私人自行车,而研究无桩共享单车的文章较少。虽然共享单车是以自行车为载体,但其流动性和不确定性以及随机性都与普通的非机动车有着很大的差别,因此运用传统的方法与思路是难以解决共享单车的一系列问题的。另外,较多文献是通过调查用户来推知共享单车的相关信息,而获取用户信息的方式多为问卷调查法,该法费时费力且准确性较低。用此法得到的共享单车信息可靠性不高,而在此基础上进行相关分析及预测更是不准确。共享单车本身就是 GPS、互联网+等前沿技术逐渐成熟的产物,因此应充分利用该点,从终端直接获取每年共享单车的实时数据,为今后的分析提供基本保障。

1.3.1.2 国内研究现状

薛强^[19]调研并分析了共享单车的用户组成,发现 4 成受访者的使用频率较高,可达每周五次以上;而约半数受访者都会使用共享单车作为出行最后一公里的交通工具;在影响出行者选择共享单车的因素中,被访者认为便捷程度是影响使用的关键。

刘立巧^[20]从交通安全的角度讨论了共享单车和传统公共交通系统的协调问题,并通过研究单车投放量分析如何更有效的提高单车的骑行安全;最后分析目前共享单车在管理中存在的不足,并提出建设性建议。

张雨宁等^[21]从共享单车产业现状及整个市场现状分析了共享单车行业在从无到有过程中的优劣势。最后,对共享单车的未来发展给予可行性建议并对其前景提出合理展望。

刘硕秋等^[22]分别从时间层面及空间层面建立了分时段的时空模型来分析共享单车的分布问题。并得出了共享单车在时间层面上具有波动性及不均衡性,而在空间层面上相对稳定和均衡。

赵曼^[23]建立了共享单车运行网络,并对网络特征值及聚类子群进行了分析,并结合分析结果建立了寻求最有调度策略的调度模型。



张俊英^[24]在现有理论基础上,采用问卷调查法分析了单车用户的使用意愿及实际出行行为。其结果表明,通过获取出行者的使用意愿能较好预测其真实的使用情况,且使用意愿与骑行体验关系密切。

吴淑婷^[25]通过对城市公共自行车管理模式的研究提高城市共享单车的使用率,利用大数据找到共享单车目前存在的问题及存在的原因,并提出解决办法。

王宇洋^[26]在大数据环境下,采用 Python 语言编写网络爬虫程序,实现了对共享单车位置信息进行实时获取,针对海量的共享单车实时位置信息,进行了数据采集、处理、分析和可视化展示,得出了不同地段、不同时段共享单车移动趋势和骑行偏好的研究结论,及时掌控城市中共享单车的数量、位置和移动规律,智能预测未来的用车高峰地点与时段。

综上所述,近几年来国内对共享单车出行行为的研究成果较多,通过问卷调查的方式获得了出行者选择共享单车的原因、偏好、出行和换乘特征以及影响使用的因素,为共享单车的发展方向、改进措施、调度协调等方面提供一些意见。目前,基于大数据对共享单车时空分布揭示以及衍生出来的调度、发展等问题的探讨性分析较多。但是,揭示的出行规律多为骑行时间、距离等基本特征,更深层的挖掘还有待研究。且基于数据做更进一步的预测的文章则多使用传统模型,预测效果欠佳。另外,也有较多文献分析了共享单车与其他公共交通方式的衔接与协作及存在的问题,但基本是采取文字探讨的方式,由于没有数据的支持,文章的说服力较弱。本文先从出行者的角度分析其选择共享单车的内部因素,再基于已有的大数据下,验证分析是否于情于理。同时,在数据的支持下,分析共享单车的存在的问题也更有说服力。

1.3.2 基于深度学习的出行预测研究

1.3.2.1 国外研究现状

共享单车出行行为预测能较好的为政府及其运营商提供高效可行的管理办法。共享单车的时空分布主要受使用者的出行行为影响,而单车运营者也必须根据新的供需关系对单车进行再分配。因此,共享单车分布的预测具有多个维度的依赖性,需使用更智能的深度学习模型对其时空特征进行提取。深度学习又称无监督特征学习,可直接从大量数据中提取有用的特性信息,并在分类及预测问题中使用广泛。

Álvaro Lozano^[27]提出了一种系统通用的多智能体系统,并用数据可视化的方式展示了不同类型的共享单车系统。该系统包括了数据的收集及清理,能根据每个自行车站点的实际情况建立需求预测模型。



Chen^[28]通过一种递归神经预测了每个站点的实际租金及需求，并在线上提供了一种有效的平衡策略。模型的输入变量是高维的时间序列及天气信息，预测结果是下一时间段的单车使用及回收数量。

Zhou^[29]发现预测车站的换乘需求对自行车站点的重新布局及数量优化有较好效果。由于传统交通需求预测模型存在局限性，不能适应换乘需求的预测，现提出一种基于马尔科夫链的预测模型。结果表明，预测结果精度高且泛化能力好。

Feng^[30]通过一个与时间相关的 PCTMC 矩量预测了未来某自行车站点的可用车辆数。该模型的原始方程由系统中所有站的当前状态得出，并通过有向图及贡献传播理论调整模型参数，最终得到一个对单车前往目标有重大贡献的站点。最后通过最大熵原理得出自行车的分布。

Liu^[31]针对共享单车的分布预测问题。通过分析影响单车分布的因素并将之融入到卷积神经网络的建模分析中。通过数据的验证和分析，证明了模型有很好的预测性能。

Xu^[32]在文中首先进行了时空分析，研究了无站位共享单车的移动模式。结果发现了共享单车出行的时空需求不平衡性。再利用长短时记忆神经网络(LSTM NNs)预测不同时间间隔下共享单车的出行量和吸引力。通过研究发现，与传统统计模型相比，先进的机器学习方法在不同的时间间隔下具有更好的预测精度。

Ai^[33]运用一种比传统长短时记忆网络更好的获取时空信息的模型——卷积长短时记忆网络(conv-LSTM)来处理共享单车的空间依赖和时间依赖。在此基础上，完成模型对共享单车时空分布的预测，并用可视化的方法更直观的展示结果。

Li^[34]运用一种基于时空强化学习的自行车重定位模型实现对有桩单车长时间有效地重新定位，从而到达最大限度地减少客户流失。该模型使用了一种相互独立的内平衡聚类算法，将单车的站群进行聚类。再针对每个站群增加一个时空强化学习模型以获得最优策略，最后通过 Citi bike 的实际数据集进行实验，验证了模型的有效性。

从上述研究成果可以发现，外国大多文献都是预测传统租赁车某个站点的单车数量及调度等衍生问题。近年来，对于共享单车的研究有明显增加，已有文献将预测模型建立在大数据挖掘之上，且涉及较为智能的深度学习模型，但多为比较基础的模型，且预测的内容也多为单车的分布问题。由于获取的单车数据类型多为单车的分布情况，而单车的 OD 数据较难获得，所以对单车的 OD 出行的预测非常少，因此选择一个合适的深度学习模型解决去更深层次问题将颇为有意义。



1.3.2.2 国内研究现状

段文欣^[35]通过建立科学的调度指标体系,针对不同区域单车的时空特征展开了需求预测。作者基于时间、人口密度、地域属性等多个维度的变量体系,利用 BP 神经网络进行了需求预测,模型展示了一定的预测性能。

陶启萌^[36]针对共享单车的骑行数据进行了时空特征的分析,通过统计不同区域的借还车频次以及使用时间的频次,揭示了共享单车在不同区域的峰值时段。最后通过聚类方法对这些区域属性进行了聚类分析。

王忱^[37]将共享单车的需求量进行了时空维度的解析,得出了共享单车在不同时间段的多项式函数;在此基础上,作者还基于不同城市的数据,利用 LM 神经网络对相关性比较强的指标进行了训练,得出了共享单车在这些指标影响下的需求量变化情况。

焦志伦^[38]基于共享单车项目的多维度大样本数据,通过随机森林等机器学习模型,探讨了对共享单车需求预测产生影响的相关因素,并分析了多个模型的预测效果。结果发现,影响共享单车小时需求的主要因素包括特定的位置因素、时间因素以及天气条件因素。

从国内深度学习模型的研究成果中发现,深度学习算法还较少应用于共享单车领域,对共享单车的研究主要是使用机器学习模型,神经网络等非智能模型。但是共享单车明显的时间属性,一般模型是非常难准确掌握其时间特性的。因此,探索适合共享单车特性的深度学习可以更好地抓住其时空特征,并提供精确度更高的分布预测。

1.3.3 共享单车的调度优化研究

1.3.3.1 国外研究现状

Chemla^[39]通过建立总调度距离最短模型解决了调度车数量唯一情况下的调度问题,并采用了分支切割算法对模型进行求解。

Raviv 和 Kolka 等^[40]研究了自行车共享系统的调度问题,包括调度路线的确定、迁移或放置自行车的数量,并将目标函数定义为调度总距离及被租赁点拒绝的使用者数量。最后,文章通过实例中 104 个共享单车租赁站点完成了调度模型的基本验证。

Forma^[41]针对共享单车在系统中的定位问题,采用聚类分析法,对单车的各个租赁点进行聚类,找到不同租赁点之间的线路分布。

Contardo^[42]等将研究方向聚焦于公共自行车系统中的动态平衡问题,通过建立相应的系统动力学模型,可以得出不同状态下系统的最佳调度方案。



Caggiani^[43] 提出了基于调度成本最小化的单车二次配送的动态模型。模型考虑了需求的动态性和微模拟信号在时间和空间上的定位误差, 得出了客户满意度最高时的最优调度路径。

Alvarez-Valdes^[44] 等建立了共享单车系统的单随机模型, 通过研究了用户的随机性行为对站点及其分配数量的影响, 并在此基础上对公共自行车系统进行了相应的量化研究。

Flicker^[45]等提出了一个单一的共享单车系统的随机模型, 研究了用户选择的随机性对存在分配问题站点的数量的影响, 对公共自行车的调配进行了量化研究。

Pal^[46] 等针对共享单车的调度问题建立了混合整数规划模型。该模型可以对不同站点不同车辆进行统一规划, 对大规模公共自行车系统的调度和再分配问题提供了一种新的思维。

从国外学者研究成果可以发现, 国外学者对调度问题的探究更多的是基于有桩自行车的站点进行聚类分析, 涉及无桩自行车方面研究较少。且在调度问题研究过程中涉及的方面较为单一, 比如只从共享单车的骑行特征分析、单车需求量预测、调度模型中的某一方面着手研究共享单车调度问题, 又或针对某个区域的骑行数据进行时空维度的解析, 进一步提出共享单车调度的相关建议, 研究的层面较为单一。

1.3.3.2 国内研究现状

秦茜^[47]解决了需求不断变化下的公共单车调度问题。针对动态化的需求量设置滚动的时间区间, 并在模型中加入了运营成本和服务质量两个因数, 最终使用遗传算法和禁忌搜索法完成求解。

张辉^[48]研究表明共享单车的供需关系在时空上都存在的不均衡。基于此, 建立了时间窗模糊的单车调度路径优化模型, 并使用蚁群算法完成求解, 模型的求解精度提升了 26.96%。

于文超^[49]建立了 GM (1,1) 预测模型和调度模型反复迭代后反馈的调度方案, 模型的目标函数为最小调度成本和站点服务能力最大, 并采用 Pareto 占优排序的方法求解模型。

陈昕昀^[56]通过归纳分析借还车数据, 建立 BP 神经网络预测借还车的分布情况, 并通过调度时间窗内站点饱和度平均的原则分析出最佳的调度量。

刘路美^[50]在将整个城市的自行车站点分成多个调度区域, 并通过聚类的方法确定衔接站点属于的调度区, 并对调度区内需要调度的站点进行建模, 目标函数设置为满足需求的同时调度时间最短。



刘兆仁^[52]分析了共享单车目前存在的调度问题,并通过运用遗传算法,建立了解决 VRP 问题的模型。最后利用相关实例验证了模型对解决单车调度问题具有可行性、准确性和高效性。

王璐等^[53]对北京摩拜单车数据进行再处理,结合聚类的方法分析了居民使用共享单车的出行规律,并对高频使用区域进行了时空特征分析,为单车的投放数量及调度管理提供有效建议。

从国内学者的相关研究可以看出,大多数研究成果都是基于静态单车的空间分布进行调度,而时间轴等相关信息的引入及动态需求分析涉及较少。但是对于单车调度而言,单车的分布是随着供需关系变化而变化的,这既有空间上的不同,也可以是时间上的递进,而调度行为也是随着单车分布的时空变化而产生的有超前性的实时调整,这些供需行为和调度行为都需要结合时间序列进行研究。除此之外,也少有学者已经通过对空间属性进行分类及标注来帮助运营商对供需关系进行更加精确的判断,因此本文也将用新的方法对空间属性进行分类以提高预测及调度的准确性。

1.4 研究主要内容及技术路线

论文基于对共享单车的出行行为分析及到达预测,再以功能区为依托完成对共享单车的调度分析。在充分讨论共享单车发展情况的基础上,通过查找资料及已有的学术成果分析影响出行者选择共享单车的因素及原因,揭示共享单车 OD 时空分布特点。同时在已有的历史大数据的基础上,以功能区为载体运用模型对共享单车的出行进行预测。最后基于实际出行数据及预测结果,以调度总成本最低为目标完成最终的调度的过程,具体研究章节安排如下:

第一章为绪论。主要讲述共享单车出行行为、基于深度学习模型的预测问题的相关研究背景、意义及现状。同时也充分借鉴并改进其他文献中的方法和提出的问题。

第二章为共享单车大数据的挖掘与分析。首先对数据来源、对象及处理过程进行说明。再将影响出行者选择共享单车的因素从两个层面进行简要阐述。最后利用海量数据,分析共享单车出行行为特征,如不同时间段的单车分布、OD 分布、骑行距离、骑行时间等。

第三章以功能区为载体并基于深度学习模型完成对共享单车到达预测。由于不同功能区的潮汐现象存在较大差异,因此通过调取 POI 完成对城市的功能区划分。对深度学习基础的模型进行介绍后,建立 CNN 与 LSTM 并行的新模型——CNN-LSTM。该模型能较好的使 CNN 抓取共享单车空间特性的同时,使用 LSTM 抓取共享单车的时间特性,两个模型同时运行能较好的完成对共享单车出行发生的预测。同时,用一个实例来验证该模型的准确性。



第四章在功能区划分的基础上，完成对共享单车时空调度模型的建立。以问题导向为切入点，分析不同功能区存在明显的潮汐差异。结合第三章的到达预测结果和单车分布图及单车出行图，以综合运输成本最低为目标函数，调度阈值及调度车辆的容量、数量及服务时间等为约束条件完成最终模型的建立。

第五章在分析现有的启发式算法的基础上，确定适合求解本文模型的蚁群算法。在阐述蚁群算法求解原理及思路后，运用相同的实例完成对模型的可行性验证。

具体技术路线图如下图 1-1 所示。

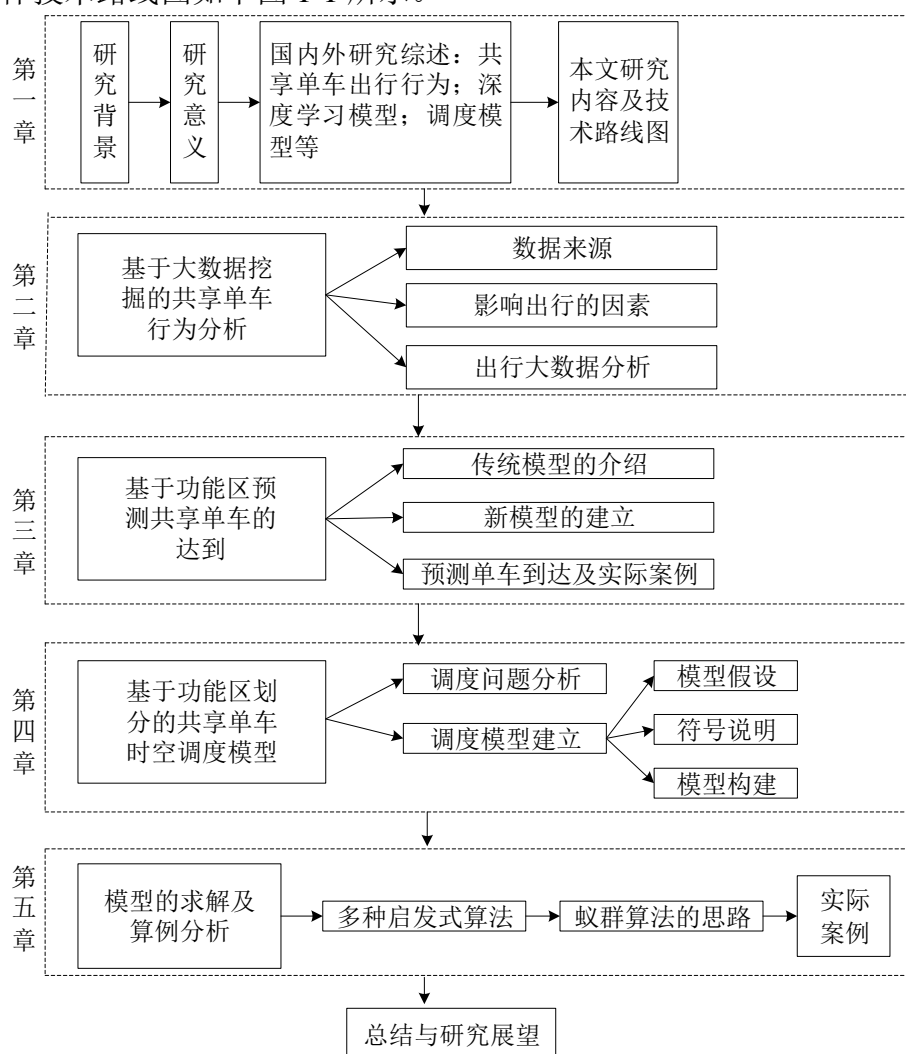


图 1-1 技术路线图



第 2 章 基于大数据挖掘的共享单车出行行为分析

2.1 共享单车选择影响因素分析

影响出行者选择共享单车的因素有很多。比如,用户的使用特点和出行特征;另外,整个城市的经济、结构、环境、运营基础设施和管理模式都将影响居民的选择^[54]。

2.1.1 宏观因素

共享单车是一个半开放性的系统,除了自身的系统构建与管理,城市的空间构造、自然条件、发展策略及基础设施等外部宏观因素都对共享单车的发展有至关重要的影响^[55]。

2.1.1.1 城市空间构造

(1) 城市规模

随着城市的迅速发展和人民的生活质量不断提高,导致城市内部的平均出行距离都有所增加。然而,不同的运输方式在运载能力、交通需求、出行分布密度等多个方面都存在很大差异^[56]。为尽可能的节约出行时间,步行、自行车等非机动车出行方式的比例却在不断降低,更多的出行者倾向于选择城市公共交通、私家车等机动化交通方式。但当城市扩张到一定程度时,虽然常规公交和轨道交通已足够发达,但依然具有一定的局限性。此时便需要共享单车来解决换乘或出行最后一公里问题。

(2) 城市空间布局

单中心、多中心组团、多中心带状、带状轴向和星形形态等是常见的城市空间布局方式。而在我国城市发展中最为普遍的模式是单中心连片密集式,该模式的特点是城市用地较为集中,从而城市居民的平均出行时间也相对较短。这种出行模式将加大城市中心的交通负荷。而共享单车的主要功能就是解决短距离出行需求,能很好的缓解此种类型的交通问题。

(3) 城市用地功能布局

城市土地利用性质一般有以下五种:居住、商业、办公、工业生产、旅馆用地等。而城市地块活动的特征与城市用地性质有非常紧密的关系。根据调查,居民总体出行中,有 80%为通勤^[57],因此城市用地的功能布局将对城市出行的交通结构产生间接影响,同时也对共享单车的使用率及投放区造成一定影响。

2.1.1.2 城市自然条件

由于共享单车为纯人力出行,因此城市的共享单车的使用及分布除了依靠平台的日常管理,更主要的还是凭借纯人力的非机动化出行。因此,城市的地形、气候等自然条件会对共享单车的发展产生较大的影响^[58]。



（1）地形条件

我国主要有平原、丘陵及山地等多种地形。成都是具有代表性的平原城市，其地貌特征主要表现为地面起伏不明显、坡面少坡度缓，是一种较利于共享单车骑行的地形。而丘陵和山地地形的地貌多为起伏较大的坡面，这对共享单车的使用产生了一定的负面影响。基于地形条件的限制，山地和丘陵地形的城市，自行车使用率都相对偏低。

（2）气候条件

天气条件是影响共享单车使用的重要影响因素。在我国北部许多城市，冬季多雨雪等天气，这将促使出行者放弃共享单车出行而选择公共交通等机动化交通工具。而在南部地区，夏天的高温天气使得出行者无法承受露天骑行，而其他相对舒适的出行方式将更加被青睐。通过以上分析，气候条件将引起交通出行的结构性变化。

2.1.1.3 共享单车交通基础设施

共享单车交通基础设施主要有非机动车道路、共享单车停车规划区域或电子围栏、共享单车本身等。基础设施的完善与否也对出行者的选择行为产生一定影响。为保证共享单车的实施顺畅，首先要充分考虑非机动车的行驶路权，设置合理规范的非机动车道。

同时，共享单车的停车区规划和电子围栏的布局也将直接影响出行者对共享单车的选择。共享单车采用的“无桩”模式，给出行者带来便利的同时也给管理者带来了很多挑战。整治共享单车的乱停乱放，合理规划停车区域和违停区域，加强共享单车在结构和功能上的设计，都能在一定程度上提高共享单车的使用率和使用年限^[59]。

2.1.1.4 城市政策

城市政策对共享单车的发展起着引领的作用，不仅决定共享单车的定位还直接决定其发展方向。城市政策具有明显的多样性，因此将众多影响政策分为直接（交通政策）和间接（经济政策）两个方面。

（1）交通政策

政府的基本职能就是对城市进行宏观调控，在公共交通方面则主要是参与其运营、组织和市场等多个方面。当新兴产业共享单车逐渐成为一种新的公共交通工具时，各地政府的相关部门也开始制定相关的政策保证共享单车的发展有序推进。2017 年 3 月，成都市政府为鼓励共享单车的发展出台了《成都市关于鼓励共享单车发展的试行意见》；同年 8 月，由交通运输部等部门联合发布的《关于鼓励和规范互联网租赁自行车发展的指导意见》中明确指出，相关部门应加快落实设置停车点或电子围栏等措施，并要求共享单车线上线下同发展，规范行业秩



序,不断提升服务质量。2018 年 1 月,成都相关部门负责人及共享单车企业就共享单车管理工作召开专题会议,对共享单车投放数量及后期运营调度展开讨论。

(2) 经济政策

经济发展政策主要是指政府为刺激或限制一定的社会经济内容而制定的法律、法规等。政策的对象主要有国家经济发展、土地开发建设、城市生态环境和可持续发展。依照出台的相关经济政策,城市交通有关部门要积极响应迅速调整,制定切实有效的发展战略,早日实现共享单车的社会效益、经济效益和环境效益的协调发展。

2.1.2 微观因素

根据共享单车的特性分析,将影响共享单车选择行为的微观因素分为两个方面:个人基本特征和使用者出行特征。个人基本特征包括很多方面,如年龄、性别、收入等。而使用者用车特征则包括:出行目的、使用频率、出行费用等。

2.1.2.1 共享单车使用者个人特性分析

(1) 性别

根据数据研究中心 iiMedia Research 数据^[60]显示,从性别分布来看,男性用户占总用户的 52.2%,而女性占比为 47.8%,女性比例略低于男性。共享单车属于身体活动较强的非机动车交通工具,而数据表明男性的接受程度较高。

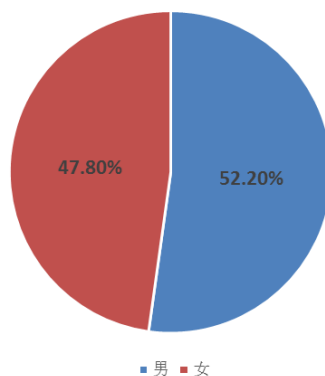


图 2-1 用户男女比例

(2) 年龄

共享单车以 APP 为平台,自行车为主体的自主性交通工具,对年龄较大或较小的出行者有一定的限制性。根据数据研究中心 iiMedia Research 数据^[60]显示,26 岁~40 岁的上班人群是共享单车用户的主要群体。随着市场运营的规范,未成年人使用共享单车的情况得以改善,因此 24 岁以下的学生群体占比有所减少。同时,随着年龄的增长,对需要消耗体力的交通工具青睐度明显降低。

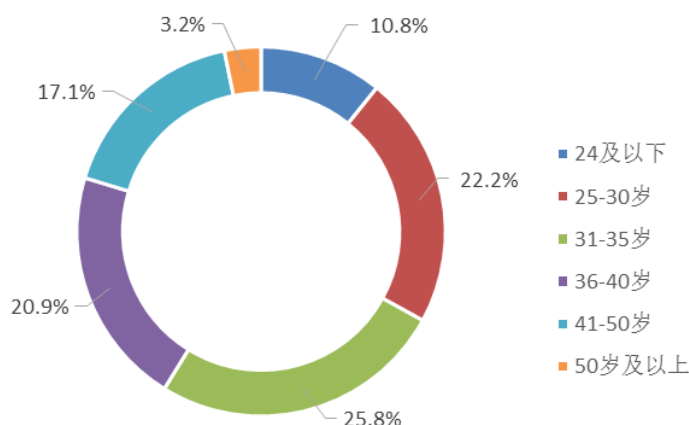


图 2-2 用户年龄比例

（3）受教育程度

根据数据显示^[61]，共享单车用户中，本科以上学历占比 76.4%，其中本科学历用户以占比 51.9%；硕士学历用户占比 18.3%；博士及以上学历占比 6.2%。共享单车的使用需要注册、扫码等相关简单操作，这要求骑行用户具备一定的互联网知识和技能。

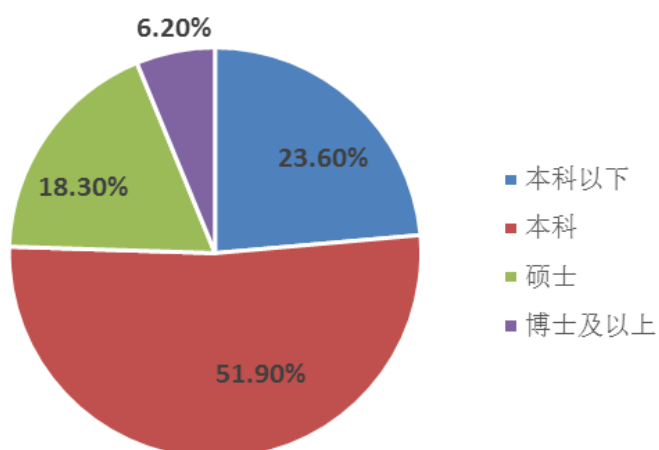


图 2-3 用户受教育程度比例

（4）收入

共享单车的注册及使用都相对简单，且使用价格较低，多为 1-2 元/小时。收入高低将导致出行者对交通工具的票价费用有不一样的理解力和承受能力。一般来说，不同收入的出行者有着不同的消费习惯，并且可支配收入水平较低的乘客对票价的敏感度更高，在做决策时更容易受到价格的影响，而高收入人群，更看重出行过程中的舒适度、时效性和安全性。根据艾媒数据^[60]显示，39.3%的用户收入在 3001 -10000 元，占收入比例最大比重，这说明中等收入人群对共享单车较为青睐。

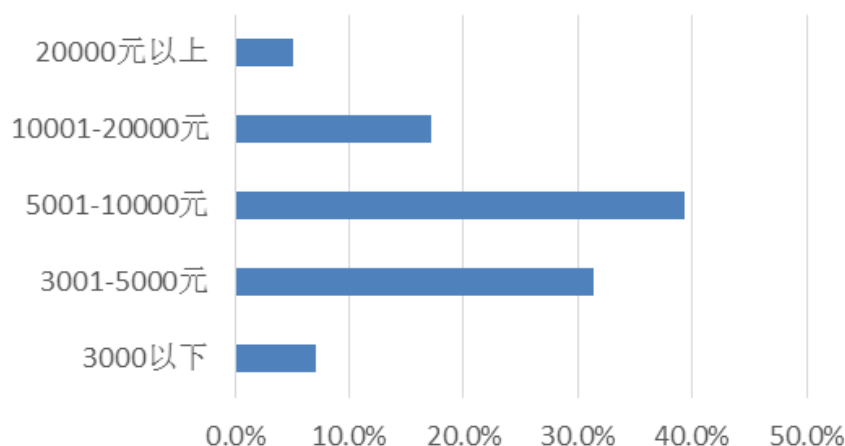


图 2-4 用户收入比例

2.1.2.2 共享单车使用者出行特征分析

(1) 出行目的

出行方式的选择很大程度上依赖于出行者的出行目的。目前出行者使用共享单车的目的大致分为以下几类：通勤、上下学、休闲娱乐、居家生活等。不同出行目的出行者出行特征有所不同，其中以通勤、上下学为目的的出行者占比较大，且使用的稳定性较高。其原因是以通勤、上下学为目的的出行者更看重时效性，在考虑出行方式时一般会考虑便捷性和准时性；而以生活娱乐为出行目的的出行者将更看重舒适性和安全性，而这类出行者对共享单车的使用随机性更大。

(2) 使用频率

不同人群使用各种交通方式的频率会存在明显差异，这也是研究供需问题的关键所在。对出行频率的准确把握能提高针对不同时段、人群、交通方式下出行意愿的预测精度，可以更好的整合社会资源配置并提供适当的出行诱导。根据艾媒数据^[60]显示，45.1%的用户平均每周使用共享单车的频率是 5 次以下；54.9%的用户平均使用在 5 次以上。这说明共享单车主要还是解决的是短距离出行需求，5 次及以下的低频次是属于正常范围。

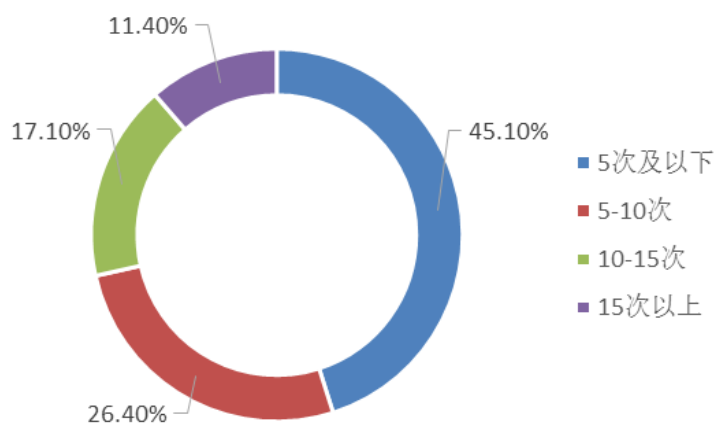


图 2-5 用户使用频率比例



(3) 出行费用

出行费用是指出行者完成一次出行所消耗的经济成本。出行费用同样会影响出行者的选择行为。尤其是,某一交通方式的出行费用较高,且有其他更经济的备选出行方案时,出行费用将是出行者的重点考虑因素。共享单车的使用成本较低,除了品牌收费标准不同以外,只与使用时间长短有关。而共享单车主要是解决短距离出行需求,所以出行费用对共享单车的影响较小。

2.2 共享单车数据处理及分析

2.2.1 数据来源

本文将 2017 年成都市摩拜单车的 GPS 数据作为研究数据来源。调查时间为: 2017 年 8 月 14 日至 9 月 4 日。调查对象为: 成都市的摩拜单车 APP 用户。调查方式为: 对调查区域内的所有共享单车进行不间断的坐标扫描。在此过程中,只有静止车辆的坐标能够被记录,即: 如果车辆正在被使用,则坐标无法被记录。由于单车数量庞大且扫描过程中网速不稳定等因素,导致完成一次扫描间隔大约为 5-8 分钟。具体调查内容主要为: 每个带编号的共享单车在扫描过程中所产生的时空 GPS 数据,即: 带编号的单车被扫描的时间和其位置坐标,如图 2-6 所示。调查数据来源为大数据公司的数据调研,有一定的保密性。

Time	ID	WEI	JING
2017-08-17 22:23:09+080286139975#		30.778708	103.938018
2017-08-17 22:23:09+080286022645#		30.777971	103.938748
2017-08-17 22:23:09+080286088958#		30.779682	103.939049
2017-08-17 22:23:09+080286110477#		30.780012	103.938828
2017-08-17 22:23:09+080286108869#		30.780237	103.938901
2017-08-17 22:23:09+080286062614#		30.780345	103.936046
2017-08-17 22:23:09+080286036309#		30.780475	103.939588
2017-08-17 22:23:09+080286113795#		30.779309	103.940183
2017-08-17 22:23:09+080286139007#		30.781046	103.936552
2017-08-17 22:23:09+080286013750#		30.780978	103.936058
2017-08-17 22:23:09+080286022420#		30.781292	103.936849
2017-08-17 22:23:09+080286040829#		30.781368	103.936724
2017-08-17 22:23:09+080286115227#		30.780178	103.94123
2017-08-17 22:23:09+080286730507#		30.778562	103.938459
2017-08-17 22:23:09+080286542729#		30.778791	103.939789
2017-08-17 22:23:09+080286732389#		30.779962	103.939374

图 2-6 数据展示图

2.2.2 数据处理

本次调查对象为成都市绕城城区内的摩拜车辆,被监控车辆超过 30 万。获取的数据样式为某时间点的坐标数据,尚不能反应出单车的 OD 出行关系。但由于正在被使用的单车将无法被扫描记录,因此同一辆单车在相邻几个表中的位置将互为 OD 点。基于以上分析,本文将通过 MATLAB 对单车数据进行重新配对,获取每辆单车的 OD 数据。另外,由于本次调研所获取的数据样本量极大,且没有实现全时段的实时跟踪,因此为保证后续研究的严谨性,需对数据样本进行如



下处理，从而排除一些人为因素所导致单车移动及其他一些特殊情况。

2.2.2.1 时间窗阈值

根据摩拜公司公布的调查数据显示，在人均单次骑行时间最长的城市里，用户每人单次骑行平均时长为 25 分钟，而成都地区的用户每人单次骑行平均时长为 18 分钟，因此本文将共享单车完成 OD 的时间阈值设置为 30 分钟。也就是说，将骑行时间超过 30 分钟的单车数据视为无效数据。为了清洗数据的便捷，现将 30 分钟设置为固定时间段，通过 MATLAB 软件编程，寻找每个时间段内的单车出行 OD 坐标及其时间（由于数据量极大，因此不同时间段之间不做交叉），并把落在同一个时间段内的单车 OD 出行信息整理成一个新的表格。

2.2.2.2 速度阈值

为满足不同区域的共享单车需求量和不同时间段共享单车的需求量，共享单车公司负责人会根据市场调查制定合理的人工调度方案。为确保本文所研究的共享单车 OD 分布仅为用户行为而不是管理者安排的车辆调度行为，现计算同一辆单车在相邻两次扫描时间内的平均移动速度。根据调查一般普通人骑单车的平均速度为 15-20 公里/小时，由于数据限制只能计算出两点的直线距离，所计算的平均速度将小于实际值。另外，多数单车运营商都使用电动三轮车或小型货车完成调度任务，据调查电动三轮车的平均车速为 25-30 公里/小时。综上，现将相邻两次扫描时间内的平均速度阈值设置为 20 公里/小时。

2.2.2.3 天气设置

根据 2.1.1.2 的分析可知天气条件是影响共享单车使用的重要影响因素，极端天气对共享单车使用情况影响较大。所以本次只统计晴天的骑行数据，而在中雨以上天气的数据将予以剔除。根据历史天气预报^[62]查询结果，现剔除 8 月 15 日、8 月 20 日、8 月 25 日、8 月 29 日、8 月 30 日的骑行数据。

2.2.3 出行数据大数据分析

共享单车现如今已成为城市居民完成日常出行的重要交通工具，根据单车的空间位置、行驶时长及轨迹都可以在一定程度上反应人的出行行为特征。例如：在工作日中的上班时段，人们多将使用共享单车完成从居住区到工作区的转移，或从居住区骑行到附近的交通枢纽站进行换乘，这时共享单车就主要集中在以办公职能或换乘职能的区域；在结束工作后的下班时段，根据人们的生活习惯，居民区或娱乐区的共享单车数量将有所上升。而在非工作日期间，共享单车的用途多为休闲娱乐。因此，此时工作区域内的共享单车数量相对较少。根据分析可知，共享单车的移动数据对分析城市空间布局、功能区划分、公共交通布局等问题有



积极作用。

基于对摩拜 APP 用户的使用数据的挖掘，现将 2017 年 8 月 14 日~9 月 4 日的成都摩拜单车数据分为工作日和非工作日两部分分别进行处理。由于数据为单车的 OD 点位置坐标，只能计算出两点间的直线距离，并不能很好的反映单车具体的骑行距离。因此，考虑到直线距离和折线距离之间的差值，故将基于经纬度计算出的直线距离放大 1.5 倍。具体分析结果如下：

2.2.3.1 骑行距离分布

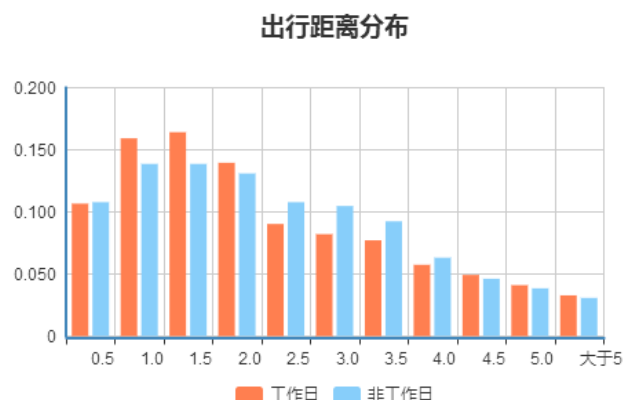


图 2-7 出行距离分布

根据大数据统计所绘制的分布图可知，共享单车的骑行距离分布无论是在工作日还是非工作日都有规律可循。工作日期间，共享单车的骑行距离约 50%集中在 0.5-2km 的区间范围内，其中骑行距离在 1-1.5km 之间的频率最高，而超过 2km 骑行距离的频率出现陡降。由此可见，共享单车在工作日的被使用趋势更多的为短距离出行。造成此现象的原因可能是居民在工作日期间，更倾向于短距离的使用共享单车来完成公共交通的换乘。这种使用共享单车作为衔接工具的通勤模式能有效降低出行时间成本，帮助出行者更快更早更方便的到达交通枢纽。

非工作日期间，共享单车的骑行距离约 40%集中在 0.5-2km 的区间范围内，骑行距离分布相较于工作日更均匀，且 2.0-3.5km 的骑行距离占比有明显增加。由此可见，出行者的平均骑行距离有所增长。长距离使用共享单车是不过分计较时间成本的表现，即非工作日的出行者出行时间价值低于工作日，由此推测出行者在非工作日多使用共享单车多出于休闲娱乐等目的。

同时，从数据的详细分析中还可以发现，拥有 75%的出行都是超过了 900m 范围内的，因此，对于共享单车的分析和建模而言，900m 将成为一个具有统计意义的分割点。



2.2.3.2 出行距离的时间分布

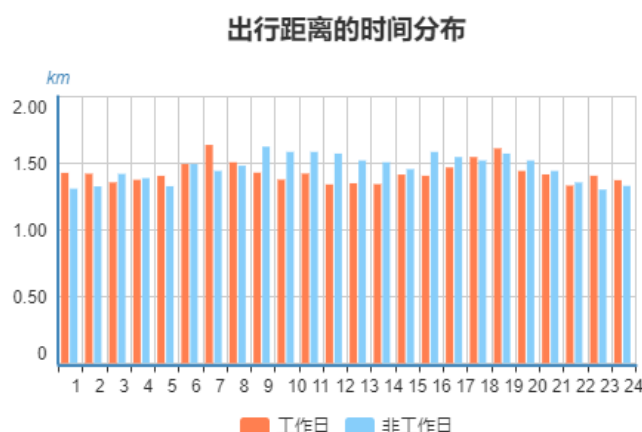


图 2-8 出行距离时间分布图

根据出行距离的时间分布图显示,共享单车的被使用情况在工作日出现了明显的潮汐现象。共享单车的上班早高峰出现在 06:00-07:00,而下班晚高峰时段出现在 16:00-18:00,都略早于公共交通的早晚高峰时间。同时,共享单车在这段时间的骑行距离有明显的上升趋势。根据经验,这段时间使用共享单车的出行者多出于公共交通换乘或者去往工作地点。这些连接距离大多集中在 1.2~1.7km 的区间上。而非高峰期时的出行距离明显下降,可见出行者一般都是在工作地附近小范围活动。

非工作日期间,共享单车的骑行距离分布相对均匀,没有明显的潮汐现象。骑行距离较长的时间主要分布在 9:00-13:00、15:00-18:00 之间。可见,出行者在非工作日使用共享单车受时间影响较小,目的多为休闲娱乐。

2.2.3.3 出行频率分布

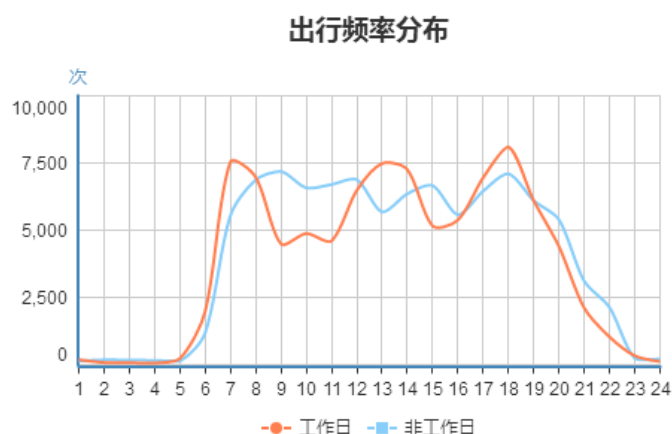


图 2-9 出行频率分布图

根据出行频率分布图可知,工作日共享单车在 01:00-06:00 期间基本上没用



出行者使用共享单车，这段时间共享单车基本进入一个休眠期。随着早高峰的开始，共享单车的使用频率在 06:30-08:30 左右达到第一个峰值，该时段共享单车多被使用为通勤工具。12:00-14:00 共享单车达到非通勤时段的高峰期，该时间段的出现目的多为午餐或休息的短距离出行。最后一个波峰出现在晚高峰 16:30-19:00，相较于早高峰，晚高峰的开始时间有所提前且持续时间也更长。同时，在 21:00 以后，共享单车的使用频率也处于一个不低的水平，这也侧面的说明当代居民生活越来越休闲化，或更多的出行者选择错峰回家。

而非工作日，共享单车的出行频率没有明显的潮汐现象，都处于一个较高且稳定的水平。和工作日相比，非工作日在日间被开始大量使用的时间有所推迟，而夜间出行（20:00-23:00）的频率有明显增加。

2.2.3.4 骑行时间分布

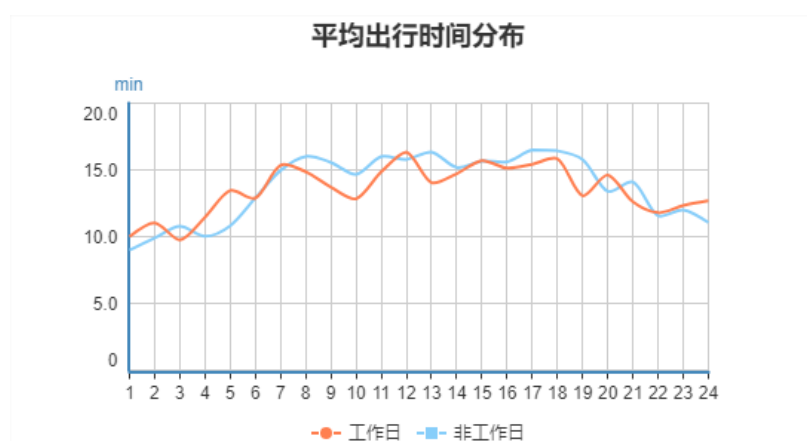


图 2-10 平均出行时间分布图

从平均出行时间分布上可以看出：在工作日期间，平均出行时间在 10-15min 之间。从早上 6 点开始，骑行时间呈现出了较为明显的增长趋势，在早上 7-8 点之间到达峰值，这些出行大量集中在住宅与公共交通方式之间。而到中午 12 点左右，骑行时间经历了第二个峰值，对比图 2-8 出行距离时间分布图可知，中午人们的平均出行距离并没有明显增加，而时间的增加很大可能是由于午间出行者由于时间紧迫感较低，一定程度上放慢了骑行速度。晚高峰出行在 18 点左右，大多也以中短距离出行或者接驳换乘为主；

而在非工作日时，骑行时间在 15 分钟以上的比例较高，结合图 2-8 出行距离时间分布图可知，除了距离的增加导致骑行时间增加以外，周末出行者的出行速度也有所减慢，这也侧面说明出行者周末骑共享单的目的多为休闲娱乐，对时间成本的在意程度较低。另外，非工作日峰值出现的时间较晚，大约在 08:00 以后会出现持续的峰值，在 19:00 左右骑行时间才会逐渐下降。

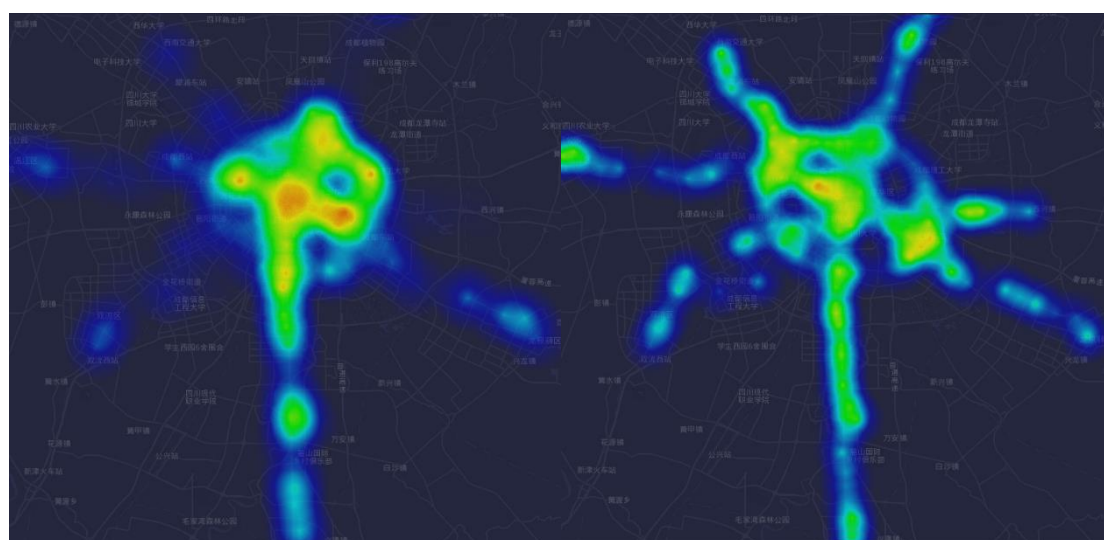


2.2.3.5 地铁周围 OD 分布

为更深入的了解共享单车在换乘过程中起到的重要作用及以共享单车为依托挖掘城市功能区分布情况，现对成都市 1 号、2 号、3 号、4 号和 7 号地铁线中每个站点方圆 1500m 范围内的共享单车 OD 数量进行统计。

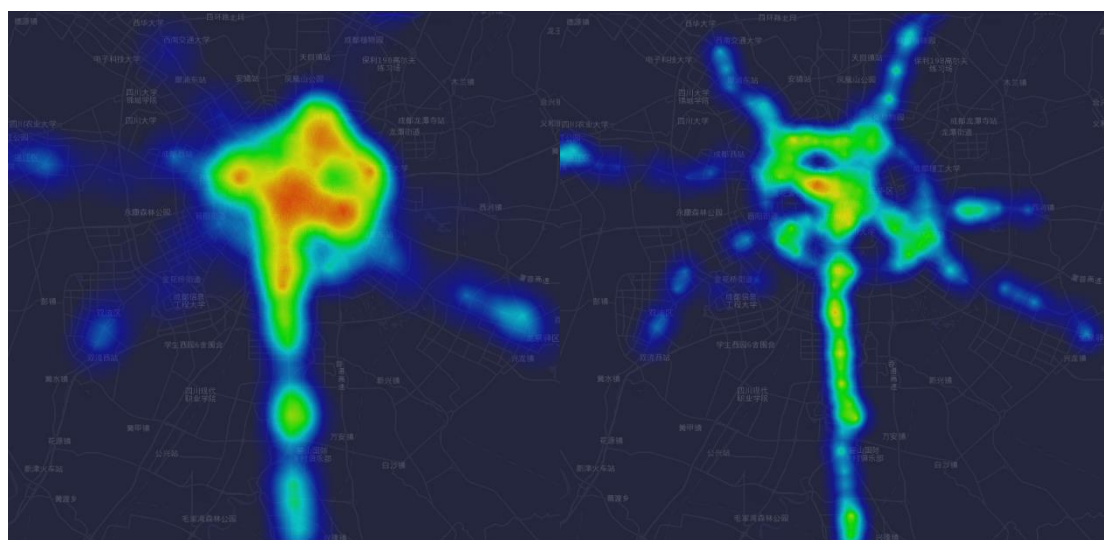
(1) 工作日

根据经验，共享单车在工作日多被使用为上下班换乘公共交通的接驳工具，因此选取 07:00-08:00、08:00-09:00 为统计时间。



(a) 7:00 单车出发分布图

(b) 7:00 单车到达分布图



(c) 8:00 单车出发分布图

(d) 8:00 单车到达分布图

图 2-11 工作日地铁周围 OD 分布图

工作日期间，成都市地铁站周围（1500m）的共享单车出行与到达数量分布我们也可以发现很多关于共享单车潮汐行为和地铁站所属功能区的相关信息：



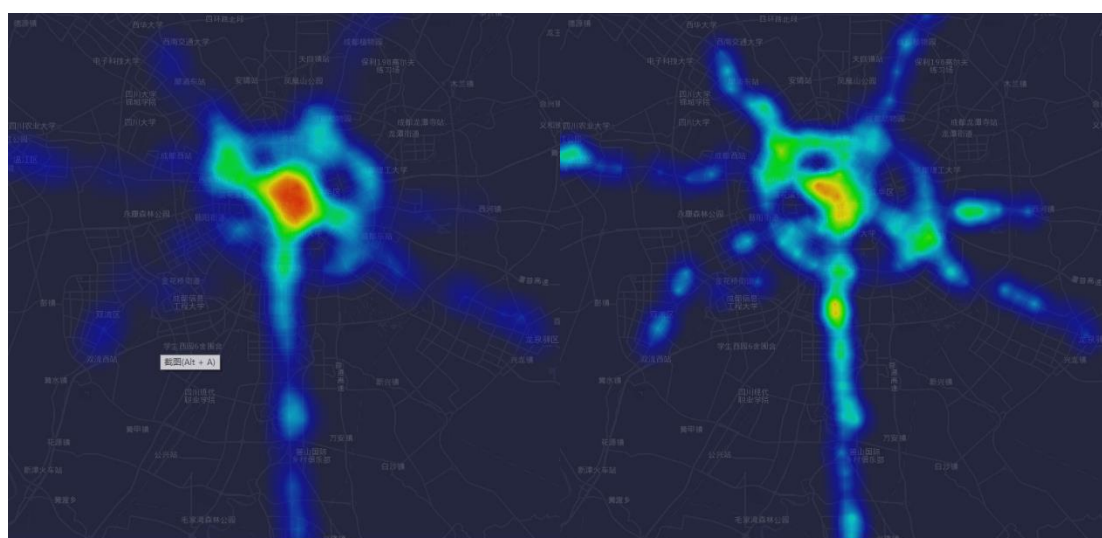
1) 从早高峰 07:00-08:00 和 08:00-09:00 的出行分布可以发现, 在统计的两个小时的早高峰时段里, 以住宅功能为主或者郊区地铁站的区域明显缺乏早高峰单车出行用户; 而出行者大多集中在以办公或者混合功能为主的地铁站区域, 其峰值出现在 08:00 左右, 这与通勤规律基本相符, 人们大多会在出行的最后阶段选择共享单车进行接驳。

2) 从早高峰 07:00-08:00 和 08:00-09:00 的到达分布比较, 同样可以发现一些有价值信息: 以住宅功能为主或者郊区地铁站的达到高峰更早地出现在 07:00 左右, 人们通常需要预留较长的时间进行通勤与接驳, 而在市区的办公区域和混合区域的地铁站点周围, 人们使用共享单车到达地铁站的时间分布相对比较均匀, 这与较短的通勤过程和复杂的出行目的均有一定关系。

3) 在郊区以居住功能为主的地铁站点, 站点周围的共享单车以到达为主, 人们从居住区周围骑行单车到地铁站点进行换乘。而在市中区以办公和混合功能为主的地铁站点, 共享单车的出行与达到则比较平均, 人们乘地铁之后使用共享单车到达办公或者服务区域, 同时也从这些区域去往其他地铁站点。这既反应了明显的地铁站功能区之间的相互关系, 又反应了明显的基于共享单车与地铁的潮汐交通现象。

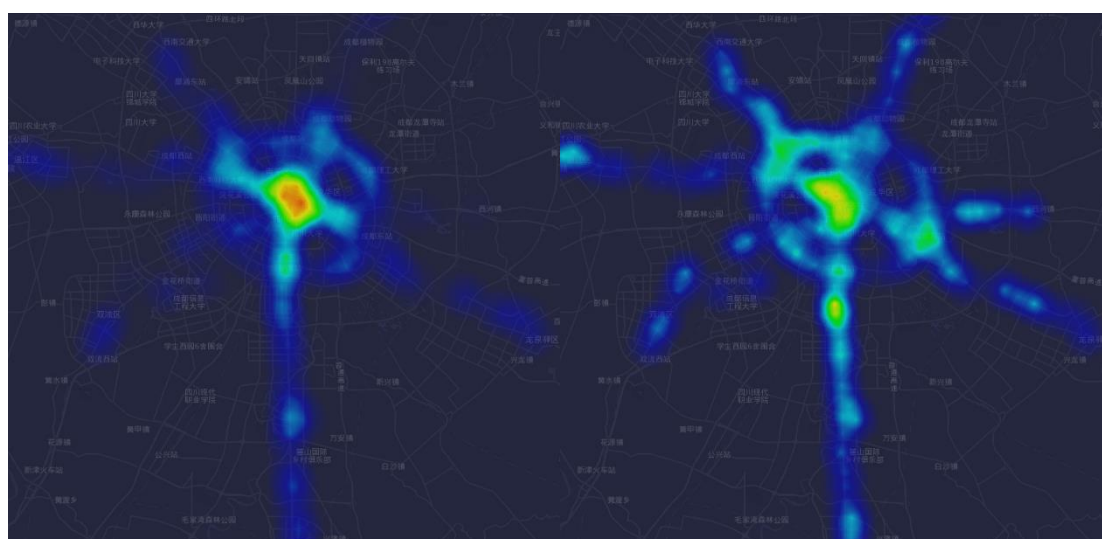
(2) 非工作日

根据前几小节的分析可知, 共享单车在工作日及非工作日期间被使用的目的和时间都呈现出不同的特点。因此, 本文除了分析工作日期间, 还将分析非工作日下午不同功能区域地铁站的共享单车出行行为, 这对于分析功能区的分布和居住功能、商业功能及休闲功能之间的出行链有十分重要的意义。在非工作日中, 人们的休闲和娱乐出行将大大延后, 根据大量文献中统计的休闲娱乐出行时间, 本文将基于休闲娱乐的共享单车+地铁出行(居住功能为主的地铁站区域-休闲娱乐及商业为主的地铁站区域)的统计时间区间定为 11:00-12:00 和 12:00-13:00。在周末的这个时间段, 人们的大多出行需求均为休闲娱乐出行。



(a) 11:00 单车出发分布图

(b) 11:00 单车到达分布图



(c) 12:00 单车出发分布图

(d) 12:00 单车到达分布图

图 2-12 非工作日地铁周围 OD 分布图

根据上图可分析得出：

1) 从非高峰时段 11:00-12:00 和 12:00-13:00 的出行分布我们可以发现：在 11:00-12:00 和 12:00-13:00 的时间区间内，以休闲娱乐功能为主或市中心的地铁站附近的共享单车出行量明显高于以住宅功能为主或郊区地铁站的单车出行量。这说明人们在这个时间点大多都会选择从居住区域出行，在乘坐地铁后选择共享单车去往休闲娱乐功能区域，这就造成了休闲娱乐功能为主的地铁站点出行量大量增加。通过对比 11:00-12:00 和 12:00-13:00 这两个不同时间段的出行量我们还可以发现，11:00-12:00 之间的出行量略微大于 12:00-13:00 之间的出行量，这可能是由于大多数人会选择在午餐前到达休闲娱乐区域，这也符合实际的出行需求。



2) 从 11:00-12:00 和 12:00-13:00 的到达分布上比较可以发现: 11:00-12:00 这个时间段, 出行者使用共享单车到达地铁站周围的数量要略高于 12:00-13:00 这个时间段内的达到数量, 这种现象在以住宅功能为主或郊区地铁站区域尤其明显, 这个时间段共享单车的使用目的通常是接驳到地铁站去往休闲娱乐区域, 而休闲娱乐需求多发生的午餐或者晚餐前。

3) 非工作日的 11:00-13:00 为休闲娱乐出行时段, 在郊区以居住功能为主的地铁站点, 站点周围的共享单车以到达为主, 人们从居住区周围骑行单车到地铁站点进行换乘。而在市中区以休闲娱乐和商业功能等混合功能为主的地铁站点, 共享单车的使用需求则以出行为主, 人们使用到达休闲娱乐与商业区域。这反应了非工作日时地铁与共享单车的潮汐交通现象和功能区之间的相互关系。

此外, 将工作日和非工作日的共享单车与地铁的接驳出行范围进行统计分析后还能发现, 工作日时共享单车对地铁的平均接驳距离要明显大于非工作日的平均出行距离, 这必然和不同的出行目的有关, 通勤出行对单车的依赖性要明显高于休闲娱乐出行, 通勤出行对距离的敏感程度也要低于其他出行目的。

2.2.3.6 功能属性的 OD 分布

随着共享单车在城市日常出行与日常生活中的作用日益凸显, 其与城市各项功能和各个区域之间融合也在不断加深。城市是一个统一的有机体, 在不同区域会出现不同功能的区分, 这种基于功能区域的划分是城市发展的潮流。而共享单车作为城市中短途出行和换乘与接驳出行的重要交通工具, 这种基于不同功能区时空特征将在共享单车的出行特征中得到较为明显的体现。共享单车在不同时间点和日期的出行将在很大程度上反映了该区域所属的功能特征。现选取成都市人民北路地铁站所属的区域进行单车出行与功能区划分的研究, 所选区域充分保证了其人口密度、区域特征、经济社会发展程度的一致性和功能的多样性。首先将所选区域按照高德地图兴趣点分布分为了住宅区(蓝色)办公区(绿色)和混合功能区(红色)。数据挖掘的时间范围分为工作日 7:00-8:00、12:00-13:00、18:00-19:00 和非工作日 11:00-12:00、12:00-13:00、18:00-19:00。在此基础上通过出行(绿色柱)和到达(蓝色柱)的分析来研究共享单车使用行为和功能区划分之间的联系。

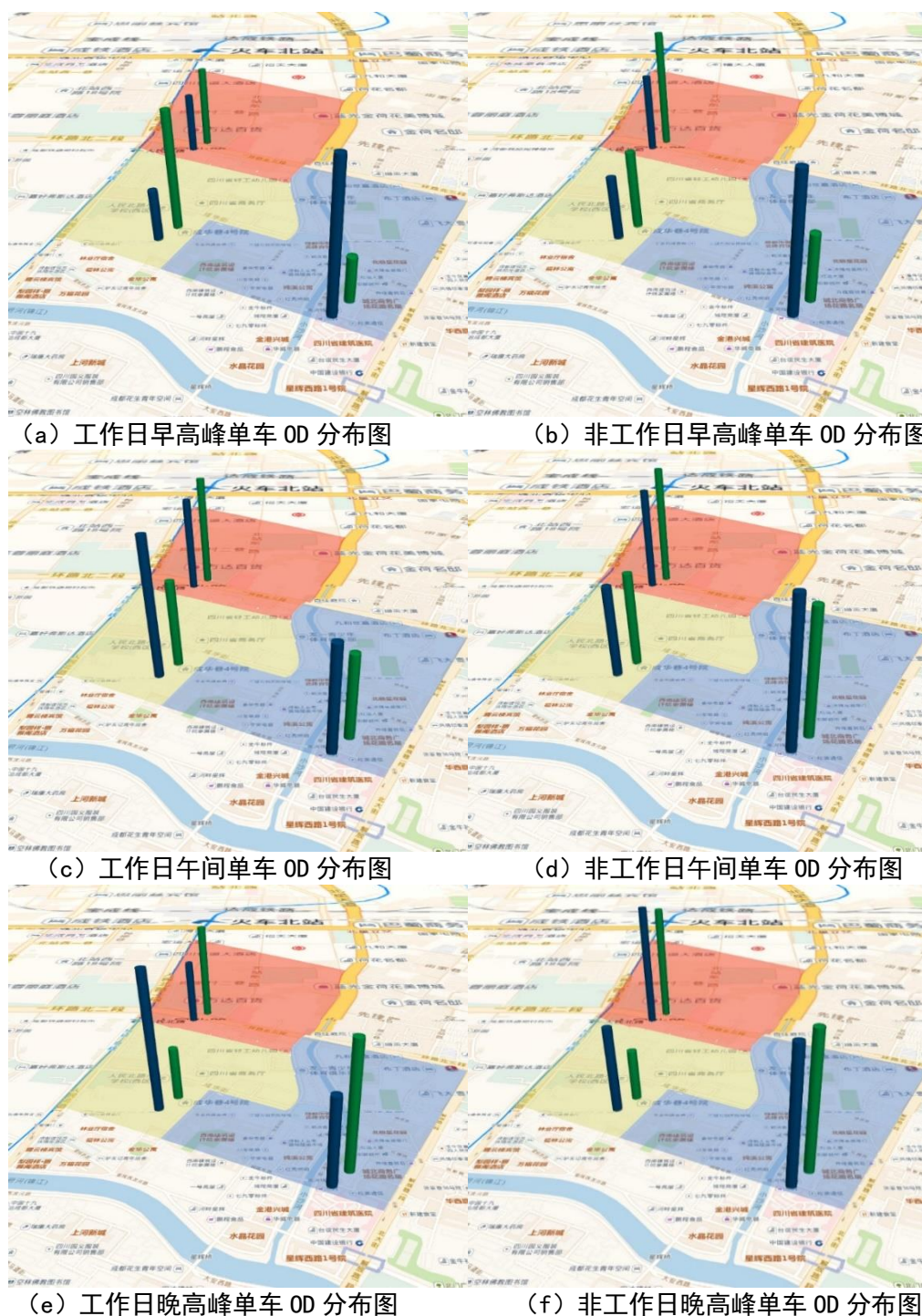


图 2-13 功能区属性的 OD 分布图

1) 从工作日的早中晚三个时段的纵向对比来看 (7:00-8:00、12:00-13:00、18:00-19:00): 早高峰时段, 以住宅功能为主的区域共享单车的使用行为以出行为主, 而办公区域和混合功能区域则刚好相反; 在中午的平峰期时, 办公区域的出行量大量增加且到达量下降, 而混合功能区的出行和到达都有了明显的增加; 在晚高峰时期, 住宅区域和混合功能区域的到达量明显增加且高于出行量, 而办公区域则呈现相反的趋势, 主要以出行需求为主。这体现了不同功能区的共享单



车使用的潮汐现象。

2) 从非工作日的早中晚三个时段的纵向对比来看(11:00-12:00、12:00-13:00、18:00-19:00): 早高峰时段, 以住宅功能为主的区域共享单车的使用行为以出行为主, 而混合功能区域则以达到为主, 办公区域的出行和达到量都相对较少; 在中午的平峰期时, 办公区域和混合功能区域的出行需求都有了一定程度的增加, 而住宅区域的到达量也呈现出明显上升趋势, 所有区域的出行和达到量都呈现出较为均衡的趋势; 在晚高峰时期, 混合功能区域的出行量明显上升, 而住宅区域的到达量也呈现出上升趋势。非工作日的共享单车使用行为在住宅和混合功能区呈现出明显的潮汐交通流行为。

3) 从工作日和非工作日的横向对比来看: 在各自的早高峰时段(7:00-8:00、11:00-12:00), 工作日的到达主要出现在以办公为主的功能区域, 而非工作日则出现在以休闲娱乐为主的混合功能区域; 在中午平峰期时段(12:00-13:00), 工作日的办公区域出行量明显高于非工作日, 非工作日时住宅区域的出行和到达也出现一定程度的增长; 在晚高峰时段, 工作日的共享单车出行主要出现在以办公功能为主的区域, 而非工作日出行则较多发生下住宅区域和混合功能区域。从横向对比可以看出, 工作日和非工作日的共享单车出行分布在不同功能区的特征区别尤为明显。

2.3 本章小结

本章以共享单车为研究对象, 分析了影响出行者选择共享单车的诸多因素, 并提出了研究共享单车出行行为的特征要素。通过对大量数据进行挖掘, 并使用数据可视化的方法, 更直观的展示了出行者使用共享单车的出行行为。

由于获取的大数据并不包含用户特征, 为更完整的展现共享单车的被使用情况, 本文通过了 iiMedia Research 对用户层面进行补充分析。其中, 性别和收入对共享单车的使用影响较小, 而用户群体年轻化趋势的较为明显。从基于骑行距离分布、出行距离时间分布、出行时长分布、出行时段分布、地铁周围 OD 分布、功能区属性 OD 分布的出行数据分析来看, 无论是工作日还是非工作日共享单车的出行距离基本保持在 2km 以内的中短距离为主, 但(非)工作日出现的早晚高峰现象及潮汐交通现象的时间略有不同。

另外, 在住宅功能区或办公功能区的地铁站附近的共享单车出行行为有明显差异, 这也侧面反映出不同功能区的地铁站之间有着一定的交互作用。然而, 这种差异不仅体现在不同功能区的地铁站附近的共享单车, 整个功能区中的单车出行行为都有不同的规律。因此后文的所有研究都将基于不同功能区的划分。



第 3 章 基于城市功能区的共享单车出行时空预测

随着大数据挖掘技术的不断成熟,共享单车数据作为城市实时动态数据,将能为城市的功能区识别及规划布局提供海量数据基础。共享单车出行数据主要包括出行者在时间及空间两个维度上的移动轨迹。因此,共享单车数据将蕴含丰富的时空信息,这将为研究者在研究城市交通流方面提供大量的有效信息。另外,共享单车主要是以中短距离的短时流动为主,便能较好的揭示跨越功能区的用户出行行为。

海量性、动态性及多维性是不同功能区下共享单车数据的显著特征,因此用传统的时空分布预测模型将很难准确抓起其时空特点,更难以揭示数据背后隐藏的运动规律和内在联系。作为人工智能的深度学习模型将使用逐层训练的方法构造更深层的模型,使整个模型对时空特征的获取能力明显提高,有效的弥补了传统模型的不足。近年来,深度学习在图像识别、语音识别等领域有较广泛的应用,这也加快了模型对时空特征的学习速度及能力。

对城市交通流分布进行预测是交通问题中的核心之一,准确性和及时性则是对该工作提出的基本要求^[63]。目前,共享单车投放量大且用户的出行路径灵活多变,由用户行为引发的区域性供需不平衡将对共享单车循环系统产生极大影响。若管理者不能及时进行调配调整,将导致用户的出行效率及单车的整体使用频率大打折扣。因此,准确的共享单车短期时空预测分布能为管理者提供决策辅助,使其在满足短距离出行需求的同时,与其他公共交通方式也形成高效互补。

3.1 城市功能区的划分

为城市更高效健康的发展,为满足城市居民的各种生活需求,城市中各类功能区应运而生。功能区的基本定义是指土地的功能、使用强度、利用方向、基准价格都基本一致的区域。既有研究中,城市功能区主要分为以下几种:居住区、工业区、商业区、行政区、文教区、修养疗养区等。合理划分城市功能区能使城市空间的功能及结构更明确,各区域的定位及发展目标更清晰。而城市功能区的有效识别能加深管理者对城市土地利用、空间组织的理解和认识,为今后的城市规划奠定基础。

目前,利用遥感影像技术并结合空间分析可对城市土地利用分类变化进行监测^[64]。遥感技术主要是提取城市的物理特征,例如获取地表反射率及城市空间结构。但该技术有明显的局限性,即无法很好的识别社会经济环境^[65]。POI (Point of Interest) 数据是一种城市静态数据,能反应城市空间实体的不同类型。与遥感技术相比,POI 能更准确的探测城市空间,更精细的划分城市功能区^[66-67]。



3.1.1 POI 数据的获取及类型判定

POI (Point Of Information) 是城市大数据群体中非常重要的一环, 涉及城市生活中的方方面面。从获取途径来看, POI 数据可以从各地图网站的相应接口获取, 这也使得它被广泛地采用在城市规划、城市建设、功能划分等相关领域的研究工作中, 并成为了不可获取的部分^[68]。在这之中, 搜索服务 API 是一类常用的 HTTP 接口, 开发者可以通过相关协议, 从 API 接口获取对应的 POI 数据, 通常返回 JSON 和 XML 格式的数据结果, 同时, API 接口还可以提供搜索和查询的功能, 搜索机制包括关键字、定位周边、集合搜索、ID 查询等等。本文通过高德地图的 API 接口来搜索研究区域的相应 POI 数据。

一般来说, 通过 API 获取的兴趣点数据包括了 5 个大类, 23 个中类和上百个小类^[69]。

(1) 住宅相关

在城市的所有用地类型划分中。住宅区域是覆盖面积最广, 人员流动最复杂的一种用地类型, 对于共享单车而言, 住宅用地的 POI 点将对其时空分布产生最深远的影响, 本文将选取普通住宅、公寓、商务住宅作为住宅 POI 数据的挖掘对象。

(2) 商业相关

对于城市居民而言, 人们在日常生活中会产生各式各样的消费需求, 而承担这些消费需求的商业场所则分布在城市中的各个角落, 这些基于商业服务的出行将时刻影响着共享单车出行行为特征。本文将选取商场、市场、小型超市、餐厅、酒店、金融服务场所、生活服务场所等作为商业 POI 获取范围。

(3) 公共服务

在城市生活中, 公共服务也是日常生活中必不可少的一环, 对共享单车的时空行为也会产生一定的影响。本文将选取各种政府、企事业单位, 运动健身场馆、医疗服务保障以及其他公共服务场所作为公共服务的 POI 搜索范围。

(4) 交通设施

交通出行需求是人们生活中最基本的需求, 地铁站、公交站、机场、火车站、公共汽车站等交通服务设施将产生巨大的交通流, 这些交通流将有很大机会与共享单车产生出行关联。本文将选取以上交通设施作为 POI 点的挖掘对象

(5) 绿地与广场绿地

城市绿地、公园、广场、和相应的旅游景点都是居民休闲娱乐出行常出现的区域。

计算每个城市功能区的指标频数密度和类型比例以判定功能区类型^[70], 计算公式为:



$$F_i = \frac{n_i}{N_i} (i=1, 2, \dots, 6) \quad (3-1)$$

$$C_i = \frac{F_i}{\sum_{i=1}^6 F_i} \times 100\% (i=1, 2, \dots, 6) \quad (3-2)$$

其中， i 代表某种 POI 数据的类型； n_i 代表某单元内第 i 种 POI 的数量； N_i 代表整体内第 i 种 POI 的数量； F_i 代表第 i 种 POI 的数量占整体内 POI 总数的频数密度； C_i 代表第 i 种 POI 的频数密度占某单元内所有 POI 频数密度的比例。

按公式依次计算每个区域每个 POI 类型的频数密度和比例，并将类型比值 50% 定为判断功能区性质的标准值。当区域内的某种 POI 的类型比值超过 50% 时，设该区域为单一功能区；若所有 POI 类型的比例均不达 50%，则设为混合功能区^[71]。

3.1.2 区域划分

由于通过调用高德地图接口一次所获得的数据有 900 条的限制，通过关键字搜索等方式获得的兴趣点数据将不完整，且结合后文 OD 预测的需要，现将本文使用区域划分的方法来获取研究对象的兴趣点数据。

本文拟将待研究城市划分为若干个边长不超过 1km 的正方形方块(图 3-1)，其原因如下：(1) 如果对整个城市直接进行多次关键词搜索，得出的 POI 数据由于数量过多将显示不完整，同时面积太大也难以统计，因此必须采用划分区块的方法。(2) 功能区的划分最终是为后文的共享单车 OD 预测做铺垫，后文所用的深度学习模型必须为空间矩阵，因此需将待研究城市分割为若干同等大小的方块。

(3) 根据第三章的分析可知，共享单车的平均骑行距离为 1-1.5km。在保证能较好的研究相邻功能区间骑行行为的同时，尽可能避免共享单车跨越两个以上功能区，从而提高预测精度。由此，初步规定每个正方形的边长不超过 1km。

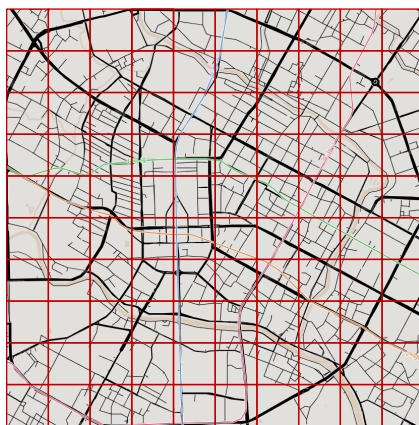


图 3-1 矩形栅格示意图



3.2 深度学习模型

深度学习其实是代表了一类算法，同时也属于机器学习的一个部分。最早是由机器学习界的鼻祖 G.E.Hinton 教授提出了深度学习这一基本概念，在之后的一段时间内快速的被运用于各个行业。深度学习的原理就是学习样本数据的内在规律和表示层次，而最终目标则是让模型模拟人的思维过程，获取一定的分析学习能力，比如识别文字、图像和声音等数据^[72]。

3.2.1 卷积神经网络（Convolution Neural Networks）

作为一种前馈神经网络的卷积神经网络^[73-74]，其神经元的连接方式受到动物视觉皮层的启发，设置最少的多层感知机识别变体。LeNet 是一种早期的卷积神经网络，是 Yann LeCun 在进行多次迭代验算后的成功之作。卷积神经网络不仅能完成读取邮编、数字等工作，还能准确的处理图像和识别视频资料。其中，卷积神经网络的结构主要为以下四部分：卷积层（Convolutional Layer）、激活函数（Activation Function）、池化层（Pooling Layer）、全连接层（Fully Connected Layer）。

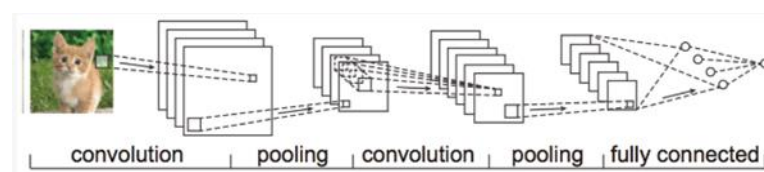


图 3-2 卷积神经网络示意图

（1）卷积层

卷积层（Convolutional Layer）则是对输入的变量进行一种数学操作，例如对变量 $f * g$ 进行卷积，则会产生第三个新的变量。一般会将高度及宽度为 a ，通道数量为 q 的图形输入进卷积层，经过多个滤波器（filters）的共同处理，完成对整个图像的特征提取。一般滤波器的大小为 $b \times b \times r$ ，其中 b 代表比图像维度小的任意值， r 可比图像通道数量小也可以相等。而特征提取过程中之所以需要通过多个滤波器，是因为不同滤波器与数据窗连接的权重是不同且固定的，因此每个滤波器关注图像的内容则各不相同。



图 3-3 卷积过程示意图

卷积神经网络中第 l 的第 j 特征图的卷积计算过程为:

$$x_j^l = \sum_{i \in M_j} x_i^{l-1} * k_{ij}^l + b_j^l \quad (3-3)$$

其中 x_j^l 表示在 $l-1$ 层中通过第 j 个通道提取的 i 个特征图 x_i^{l-1} 的特征。 k_{ij}^l 代表卷积核矩阵， b_j^l 表示偏置项， M_j 表示特征图集合。

(2) 激活函数

激活函数 (Activation Function) 则是为了体现模型中较复杂的映射关系，即将已获取的图像特征映射到新的特征图中去。激活函数的使用有非常重要的意义，因为由激活函数产生的非线性结构可以使函数获得更好的拟合结果，同时也可以避免大的数值在高层次中不断累加。目前，常见的激活函数有 sigmoid, tanh 和 ReLU。但卷积层通常采用 Relu 作为其激活函数，因为它可以减轻过拟合问题，使计算更容易。

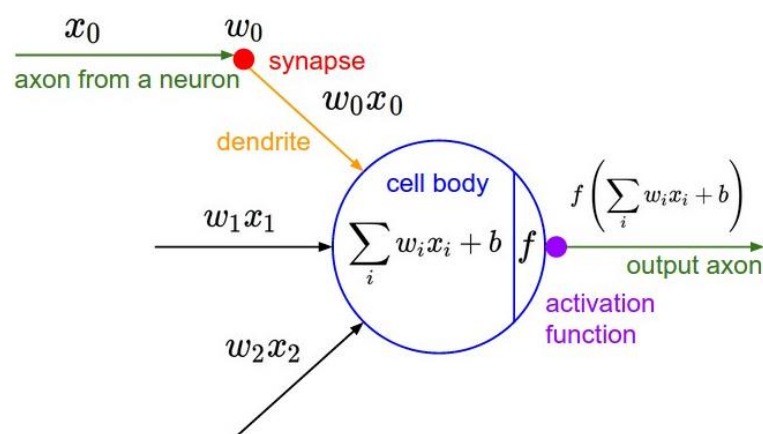


图 3-4 激活函数示意图

(3) 池化层

池化过程 (Pooling Layer) 就是一个样本离散化的过程，即对输入的样本数据进行再整合，在保证不减少样本特征的基础下降低其维度。常见的池化过程多



数是用一种抽象的形式来减少参数的数量并降低计算难度。例如一个 4×4 特征图与一个取最大值的滤波器（ 2×2 ）进行池化的过程，如下图所示：

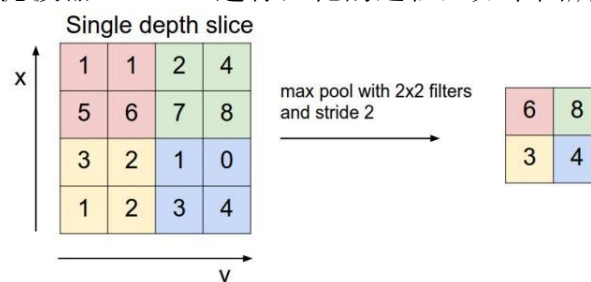


图 3-5 池化过程

（4）随机失活

作为人工智能的深度学习模型有着传统模型无法比拟的优点，但由于参数设置过多，很容易导致模型过拟合。随机失活（Dropout Layer）解决过拟合问题时，并不需要增加新的约束条件或是训练更多的模型，只需要将隐藏层中的特征值按照一定比例改为 0 即可。这种直接改变模型中神经元连接方式的方法能有效减少特征对模型结构的依赖性，从而进一步提高模型的泛化能力。

（5）全连接层

作为一种多层感知器的全连接层是指前一层里的每个神经元与后面一层中的每个神经元一一连接。同时，全连接层（Fully Connected Layer）将模型的隐藏层通过线性函数进行计算，得到最后的输出结果。

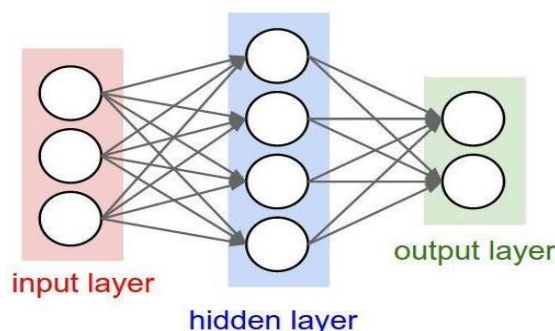


图 3-6 全连接层

3.2.2 长短时记忆网络（Long Short Term Memory Network）

一般的，研究者将传统神经网络中的各项输入设置为彼此独立，但这无形中增加了模型的复杂性，也为后续的工作增加了难度。若是能在已知上一个单词的基础上预测下一个单词，那么预测的难度将会的大大降低。循环神经网络（Recurrent Neural Network, RNN）的所有计算都是基于先前的结果，并且该模型在演算的过程中能保存之前计算的所有结果，但模型对记忆长时间信息的效果欠佳。



长短期记忆网络是循环神经网络中的一种，但由于其结构较为特殊，更能对时间维度的特征进行准确的提取。长短期记忆网络^[75-76]（LSTM）是由 Hochreiter and Schmidhuber 最早提出的，最初的设计目的是使模型能记住长时段的信息以解决长期依赖问题。但在之后不断运用及改进过程中，发现该模型在解决其他问题上亦有惊人效果。

长短期记忆网络作为深度学习模型的一种，整体结构也呈链式。单元之间是由时间戳来完成彼此的连接，这样能有效的使记忆细胞的记忆能力有所增强。而每个记忆细胞都是由以下三部分联合控制的：输入门（input gate），遗忘门（forgetting gate）和输出门（output gate）。输入门主要是负责将输入的向量保存到记忆细胞中。在此过程中，有一些信息会被遗忘门有选择的删除。经过一番处理后的新信息将成为下一个记忆细胞的输入信息，并在多次迭代后通过输出门输出最终结果。LSTM 模型的全过程示意图如下图所示。

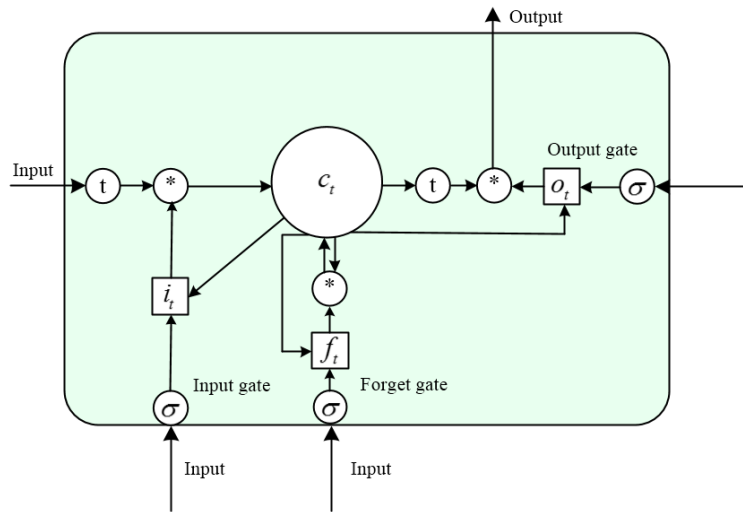


图 3-7 LSTM 结构图

从上一状态的输入中开始的输入过程如下所示：

$$i_t = \sigma(\mathbf{w}_{xi} \mathbf{x}_t + \mathbf{w}_{ir} \mathbf{h}_{t-1} + \mathbf{w}_{ci} * \mathbf{c}_{t-1} + \mathbf{b}_i) \quad (3-4)$$

$$f_t = \sigma(\mathbf{w}_{xf} \mathbf{x}_t + \mathbf{w}_{lf} \mathbf{h}_{t-1} + \mathbf{w}_{cf} * \mathbf{c}_{t-1} + \mathbf{b}_f) \quad (3-5)$$

$$\mathbf{c}_t = f_t * \mathbf{c}_{t-1} + i_t * \tanh(\mathbf{w}_{xc} \mathbf{x}_t + \mathbf{w}_{hc} \mathbf{h}_{t-1} + \mathbf{b}_c) \quad (3-6)$$

$$o_t = \sigma(\mathbf{w}_{xo} \mathbf{x}_t + \mathbf{w}_{oh} \mathbf{h}_{t-1} + \mathbf{w}_{co} * \mathbf{c}_t + \mathbf{b}_o) \quad (3-7)$$

$$\mathbf{h}_t = o_t * \tanh \mathbf{c}_t \quad (3-8)$$

$\mathbf{x} = (\mathbf{x}_1 \cdots \mathbf{x}_{t-1})$ 代表的是输入向量序列， $\mathbf{h} = (\mathbf{h}_1 \cdots \mathbf{h}_{t-1})$ 代表的是输出向量序列； \mathbf{x}_t 代表的是一维向量，而 \mathbf{h}_t 代表的是一个不同维度的向量； $\mathbf{w}_{xi} \sim \mathbf{w}_{co}$ 代表的是向量间做线性变换时的加权参数矩阵； $\mathbf{b}_i \sim \mathbf{b}_o$ 代表的是截距参数。“*”是一种运算符，即 Hadamard 乘积； σ 和 \tanh 则是两个非线性激活函数：

$$\sigma = \frac{1}{1 + e^{-x}} \quad (3-9)$$



$$\tanh = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (3-10)$$

LSTM 模型对输入与输出之间的关系把握十分准确, 原因是因为模型总是将上一层的输出结果作为下一层的输入变量。且位于模型不同层的记忆细胞共同学习, 相互协同能有效提高整个模型的识别能力。但 LSTM 对于识别空间特征表现不是很理想, 因为其不具备抓取空间相关性特征的基本结构。

3.3 城市功能区下的单车时空建模框架

3.3.1 关联性分析

(1) 空间依赖性

区域内的单车数量及分布情况都能影响共享单车在某区域内的时空出行分布, 同时分布情况还受到该区域及相邻区域功能区属性的影响。根据统计, 共享单车的自发迁移主要是发生在相邻功能区内, 而跨功能区的流通概率明显较小。综上所述, 共享单车的出行分布与网络内的功能区有明显的关联性。

(2) 时间依赖性

单车的出行分布还具有明显的时间依赖性。即一天之中, 共享单车的分布规律有一定的潮汐现象。不仅如此, 单车的时空分布情况与前一时间点的单车状态也息息相关。

3.3.2 变量分析

3.3.2.1 时空变量

(1) 区域出行数量分布

该功能区域在某一个时间段内的出行分布反映了该功能区在该时间段内的活跃程度与出行规律, 且出行与到达之间也存在必要的联系。通过时间维度和空间维度的引入, 区域单车出行分布将转化为时空变量的形式。本文分别定义 $S_{(i,j)}^{t,\Delta t}$ 为第 t 个时间间隔 $(t, t + \Delta t)$ 下区域栅格 (i, j) 的出行数量分布变量。

(2) 区域单车密度

区域单车的平均密度主要反映的是在区域栅格 (i, j) 中, 第 t 个时间间隔 $(t, t + \Delta t)$ 的平均单车数量, 即 $m_{(i,j)}^{t,\Delta t} \in M^t$ 。本研究中, 将用栅格内的单车总数除以栅格面积以获取单车的平均密度。

基于以上分析, 共享单车的出行数量分布和单车密度具有明显的时空依赖性, 因此使用时空变量来对模型进行描述。



3.3.2.2 空间变量

(1) 功能区属性

在城市中,功能区的属性和特征会对出行行为产生深远的影响。对于共享单车而言,位于不同的功能区域,其出行和到达所呈现出的时空分布也截然不同,因此将区域栅格 (i, j) 所属的功能区属性纳入到模型进行分析。

本章节研究的是基于城市功能区来预测共享单车 OD 出行,因此在功能区属性的划分上尽量体现出共享单车区域性的潮汐规律。所以,将 3.1.1 提到的功能区类型中的商业相关和公共服务整合为办公属性。根据 2.2.3.5 地铁周围 OD 分布图可知,不同区位不同属性的地铁站对共享单车的分布有着极为重要的影响。而本文待研究城市的市内常规公交网络非常发达,相邻两站台间的距离通常为 200-500 米左右,几乎能覆盖到每个功能区块,因此交通设施区主要考虑地铁站。另外,为突出体现地铁站对共享单车的直接影响,如果单元内有地铁站则不考虑其他功能属性,一律设定为“地铁功能区”。基于以上分析,本文定义区域栅格 (i, j) 的功能区属性为:

$$\delta(i, j) = \begin{cases} 1, & (i, j) \text{ 属于办公区域} \\ 2, & (i, j) \text{ 属于住宅区域} \\ 3, & (i, j) \text{ 属于地铁区域} \\ 4, & (i, j) \text{ 属于混合功能区域} \end{cases} \quad (3-11)$$

功能区的属性是由功能区所处的空间区位所决定的,同时还受到相邻功能区的属性影响,具有很强的空间依赖性,因此用空间变量来对模型进行描述。

3.3.2.3 时间序列变量

(1) 时间戳属性

根据第二章中对单车出行规律的初步探索,不难发现单车分布的时间序列特征明显。因此,本文首先将时间分为工作日和非工作日两大板块,再把每天分为早高峰、平峰及晚高峰三个时段。具体定义如下:

$$\beta^t = \begin{cases} 1, & t \text{ 属于工作日早高峰} \\ 2, & t \text{ 属于工作日平峰期} \\ 3, & t \text{ 属于工作日晚高峰} \\ 4, & t \text{ 属于非工作日早高峰} \\ 5, & t \text{ 属于非工作日平峰期} \\ 6, & t \text{ 属于非工作日晚高峰} \end{cases} \quad (3-12)$$

由此,模型将基于共享单车的在前段时间戳内的历史信息 $\{S^{t-1}, M^{t-1}, \delta, \beta^t\}$, 对下一时段的每个功能区的单车达到分布 A^t 进行预测。



3.4 基于 CNN-LSTM 的 OD 分布预测

共享单车数据同时具有较强的空间依赖性和时间依赖性。因此对共享单车进行 OD 分布预测时,需使用能兼顾时间和空间两个维度的特性,且对外延特征抓取效果良好的模型。单纯使用长短期记忆网络(LSTM)或者卷积神经网络(CNN)是无法达到上述要求。因此,本文将构建新的组合式深度模型,使该模型能兼顾全局特征的同时对细节特征也有较好的识别能力。

在本章,我们将采用一种新的深度学习架构进行单车分布的临近预测^[77-78]。作为深度学习的常用的两种模型,卷积神经网络在针对空间特征有较好的学习能力,可以针对局部特征进行组合抽象生成更高层次的特征。而长短时记忆网络则更适合进行时间序列方面的特征提取,同时兼具长期记忆功能。为了更好地获得模型的时空特征,我们将传统的单一模型思想拓展为 CNN 与 LSTM 的并行架构进行学习,通过利用 CNN 的空间特征抓取能力弥补 LSTM 在这方面的不足。

在这一部分中,我们采用一种融合空间变量和时间序列变量的 CNN-LSTM 的深度学习架构去预测用户的到达分布。分别利用 CNN 和 LSTM 来捕获空间和时间的临时性特征,最后融合这两类变量实现短期预测的输出。

(1) 时空变量

共享单车在不同区域中的单车密度和出行分布会对下一个时段的单车到达分布产生非常明显的影响。因此,将针对这两个时空变量使用不同的架构,每个变量对应一组独立的 CNN 架构。堆叠的 CNN 层数分别为 l_s, l_d 。独立体系结构的公式如下:

$$\mathbf{X}_t^S = g_{l_s}^C \cdots g_1^C(\mathbf{S}_{t-1}) \quad (3-13)$$

$$\mathbf{X}_t^M = g_{l_d}^C \cdots g_1^C(\mathbf{M}_{t-1}) \quad (3-14)$$

其中 \mathbf{S}, \mathbf{M} 是分别对应共享单车区域出行数量分布和单车密度。 g_l^C 对应不同输入变量的卷积神经网络架构。两种时空变量的分别对应的输出 $\mathbf{X}_t^S, \mathbf{X}_t^M$ 可以通过上述结构实现。

(2) 空间变量

在本文中,共享单车所属的空间区位的功能区属性被定义为空间变量,位于不同功能区的共享单车将呈现不同的时空特征。则对应的 CNN 层数为 l_δ , 其结构公式如下:

$$\mathbf{X}_t^\delta = g_{l_\delta}^C \cdots g_1^C(\delta) \quad (3-15)$$

其中 δ 是区域的功能区属性。 g_l^C 对应不同输入变量的卷积神经网络架构。



(3) 时间序列变量

时间戳属性是一种时间序列的变量集合。本文定义时间戳变量为 β ，则 LSTM 结构可以表示为

$$\mathbf{X}_t^\beta = g_{l_\beta}^L \cdots g_1^L(\beta_{t-K_p}, \beta_{t-K_p+1}, \cdots, \beta_{t-1}) \quad (3-16)$$

最终的输出变量 \mathbf{X}_t^β 可以通过上述结构实现。

(4) 时空变量、空间变量与时间序列变量的融合

通过多种变量的时空特征提取更深层特征，在全连接层中增加权重矩阵来实现 LSTM 与 CNN 的特征融合，这种通过 tanh 函数的融合方式可以避免数据信息损失。预测 t 时隙的共享单车的到达分布 \mathbf{X}_t 为：

$$\mathbf{X}_t = \tanh(\mathbf{w}_s * \mathbf{X}_t^S + \mathbf{w}_p * \mathbf{X}_t^M + \mathbf{w}_d * \mathbf{X}_t^\delta + \mathbf{w}_r * \mathbf{X}_t^\beta) \quad (3-17)$$

式中： \tanh 是双曲正切函数，确保输出值在 $[-1, 1]$ ； $\mathbf{X}_t^S \sim \mathbf{X}_t^R$ 分别对应每个独立深度架构的输出； $\mathbf{w}_s \sim \mathbf{w}_r$ 是对应不同输出变量的权重，从而使模型具有时间、空间的特征表达能力。

通过最小化估计值和实际值之间的均方误差，可以很好地学习加权和截距参数，该体系结构的目标函数如下所示：

$$\min_{w,b} \|\bar{\mathbf{X}} - \mathbf{X}\|_2^2 + \beta \|\mathbf{W}\|_2^2 \quad (3-18)$$

其中， $\beta \|\mathbf{W}\|_2^2$ 是 L2 范数正则项则是为了防止过拟合。 \mathbf{W} 表示加权参数集。

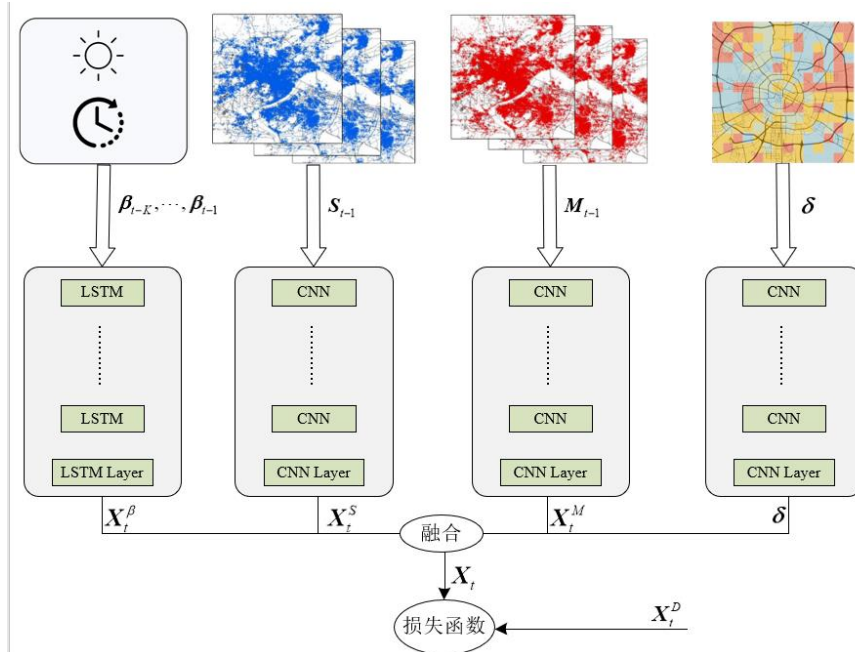


图 3-8 CNN- LSTM 结构图

CNN-LSTM 的训练步骤如下：



Input 区域的出行分布变量 $\{S_1, \dots, S_n\}$

区域的单车密度分布变量 $\{M_1, \dots, M_n\}$

区域的功能区属性变量 $\{\delta\}$

时间戳变量 $\{\beta_1, \dots, \beta_n\}$

Output CNN-LSTM 的架构参数

初始化一个空集: L

For $K \leq t \leq n$ 循环

$\{\beta_{t-K}, \dots, \beta_t\} \rightarrow E_t^\beta$, 时间窗为 K

将 $\{S_{t-1}, M_{t-1}, \delta, E_t^\beta, D_t\}$ 输入到 L

end for

初始化所有权重和参数

repeat

随机从 L 抽取一批样本到 L'

估计参数通过在 L' 中最小化目标函数

until 收敛停止

end procedure

3.5 实例分析

成都是西南地区唯一的一个副省级市，是西部地区重要的中心城市。成都的整体面积约 1.46 万平方公里，其中常住人口多达 1600 余万人，城镇化率约 70.6%，GDP 超 1.3 万亿元，排名全国第八。

根据相关报道，成都市区内共享单车总数约 123 万辆。其中，单车 72 小时内的活跃率大约为 52%，单车日周转率大约为 1.8 次，与城市公共交通系统接驳的比例约为 40%。另外，据统计成都市每天共享单车的骑行次数超过了 200 万次，若每辆车每天出行距离为 10 公里，而按普通汽车的百公里油耗 7 升计算，那么市内共享单车就等同于代替了 36 万辆机动车出行。由此可见，由共享单车所带来的“绿色、低碳、环保”出行方式在缓解“大城市病”中起到重要作用。

由于本文所涉及的样本量极大，因此完成一次单车 GPS 数据的收集时间大约为 5-8 分钟，同时本文待研究的区域为以二环路为界限的正方形区域，具体示意图如下。

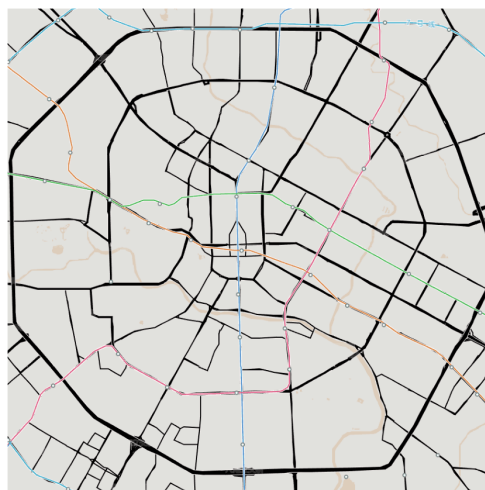


图 3-9 采样区域示意图

3.5.1 划分功能区

根据 4.1 中对城市功能区划分的介绍及具体步骤, 现将成都市二环路以内区域划为 10×10 栅格。经过计算, 可知每个栅格的面积约为 $0.9 \times 0.9 \text{ km}$, 如下图 3-10 所示。

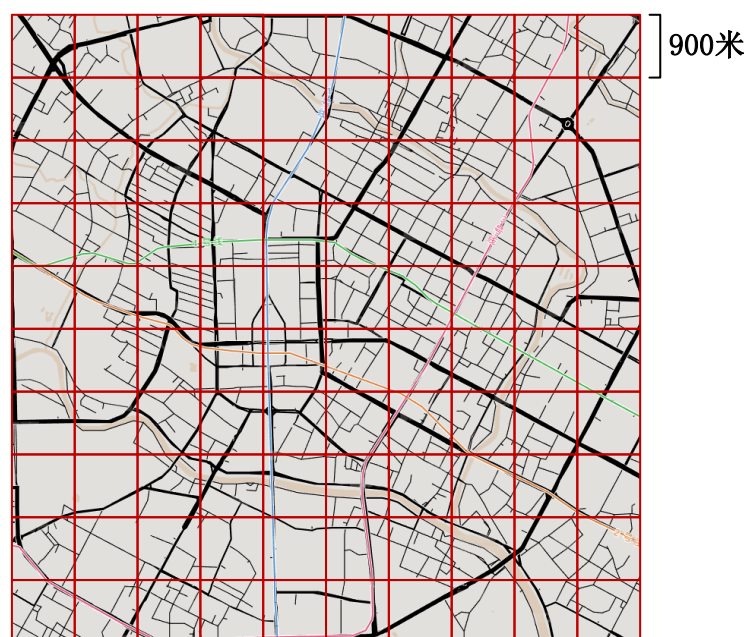


图 3-10 成都二环路栅格示意图

本文通过调用高德地图中的 API 上获取的成都二环路内的兴趣点数据, 在数据获取过程中难免会出现异常, 故在整合时需进行去重和删除异常点的处理, 经过以上处理, 共获取 POI 数据 31 万个, 如图 3-11 所示。其中, 高德地图的 POI 数据有较详细的分类信息, 获取后的数据类型一般分为 3.1.1 中所提及的六大类。但本文之所以将城市进行功能区划分, 是为了保证后文共享单车时空 OD



预测的准确性。因此将只统计分析住宅、办公和交通，根据公式 3-1、3-2 计算后，得出成都市二环路内共有 25 个住宅区（绿色）、11 个办公区（黄色）、38 个地铁区（蓝色）和 26 个混合区（红色），如图 3-12 所示。

name	type	cityname	adname	citycode	adcode	typecode	address	bdx	bdy	gpsx	gpsy
美的顾客生活服务		成都市	锦江区	28	510104	71200	锦东路555	104.113	30.64586	104.1039	30.64255
HONMA(三)购物服务		成都市	锦江区	28	510104	61101	总府路15号	104.0864	30.66438	104.0774	30.66094
成都锦江区政府机构及社会团体		成都市	锦江区	28	510104	130408	宏济巷21号	104.1041	30.64712	104.0949	30.64382
双林路26号地名地址信息		成都市	成华区	28	510108	190403	成华区	104.1192	30.65736	104.1101	30.65401
金牛绿野才交通设施服务		成都市	金牛区	28	510106	150905	二环路西三	104.0435	30.69308	104.0346	30.68924
成都铭亨利公司企业		成都市	锦江区	28	510104	170206	一环路东四	104.1021	30.6538	104.0929	30.65051
嘉源亚洲餐饮服务		成都市	成华区	28	510108	50202	双庆路8号	104.1218	30.65554	104.1127	30.65216
太阳驿站市住宿服务		成都市	金牛区	28	510106	100200	二环路北二	104.0842	30.70156	104.0552	30.6978
东较场街3地名地址信息		成都市	锦江区	28	510104	190202	锦江区	104.0968	30.67017	104.0877	30.66684
美团外卖 生活服务		成都市	锦江区	28	510104	70000	穿巷子11号	104.089	30.66263	104.0799	30.65922
金牛区永有购物服务		成都市	金牛区	28	510106	60603	府河中心楼	104.0682	30.69898	104.0593	30.69527
茶楼 餐饮服务		成都市	成华区	28	510108	50600	泰兴路17号	104.1006	30.69092	104.0915	30.68761
中国联通 生活服务		成都市	锦江区	28	510104	70604	静康路180	104.1215	30.62521	104.1123	30.62184
如果酒吧 体育休闲服务		成都市	武侯区	28	510107	80304	紫瑞大道2	104.0464	30.61958	104.0375	30.61575
三月 购物服务		成都市	青羊区	28	510105	60000	草堂东路8	104.0421	30.66379	104.0333	30.65997
海发内农村交通设施服务		成都市	金牛区	28	510106	150905	人民北路二	104.0794	30.69977	104.0704	30.69622
ZW扶危 生活服务		成都市	成华区	28	510108	70000	双成二路4	104.1207	30.64911	104.1115	30.64574
佳欣房屋住生活服务		成都市	武侯区	28	510107	71600	肖家河正街	104.0488	30.63799	104.0399	30.63415
苏苏日化店购物服务		成都市	金牛区	28	510106	61400	香樟巷与香	104.0466	30.69346	104.0377	30.68963
必胜客(科)餐饮服务		成都市	武侯区	28	510107	50303	科华中路2	104.0831	30.62472	104.074	30.62124
四川义陆村公司企业		成都市	金牛区	28	510106	170200	一环路北三	104.0886	30.68737	104.0795	30.68395
速安酒店(住)宿服务		成都市	金牛区	28	510106	100100	蜀汉路259	104.0291	30.69362	104.0203	30.68994
巴黎婚纱摄影购物服务		成都市	锦江区	28	510104	61100	正科甲巷4	104.0858	30.66233	104.0768	30.65888
KD定制家具购物服务		成都市	武侯区	28	510107	60600	二环路西一	104.0356	30.64503	104.0268	30.64126
金鱼街宠物购物服务		成都市	金牛区	28	510106	60000	金牛座金	104.0469	30.68997	104.0381	30.68612
泸天化大厦通行设施		成都市	青羊区	28	510105	991401	西康路街2	104.0577	30.67772	104.0489	30.67391
煎饼侠(东)餐饮服务		成都市	成华区	28	510108	50300	下东大街3	104.1051	30.67443	104.096	30.67114
中国石油分公司企业		成都市	金牛区	28	510106	170100	马鞍山北路5	104.0959	30.68486	104.0868	30.68152
博亿铝业 购物服务		成都市	成华区	28	510108	60603	成华大道南	104.1183	30.66743	104.1092	30.66409
金源商贸(仁)购物服务		成都市	武侯区	28	510107	61210	玉林东路3	104.07	30.63559	104.061	30.63191
旭瑞茶艺 购物服务		成都市	金牛区	28	510106	61200	金沙路西巷	104.0479	30.69337	104.039	30.68951
洛姆叔叔住生活服务		成都市	成华区	28	510108	70000	府青路三	104.1109	30.6927	104.1018	30.68939
四川宏义分公司企业		成都市	武侯区	28	510107	170200	高升桥路9	104.0443	30.64577	104.0354	30.64194
西安北路3地名地址信息		成都市	金牛区	28	510106	190302	金牛区	104.0522	30.67785	104.0434	30.67401
泊广传媒 科教文化服务		成都市	锦江区	28	510104	140000	SOHO商务	104.0986	30.64672	104.0894	30.64341
成都市博2公司企业		成都市	武侯区	28	510107	170200	口	104.0832	30.64124	104.0741	30.63776
冒烤鸭酒餐饮服务		成都市	金牛区	28	510106	50000	肖家村二	104.0815	30.69452	104.0724	30.69101
Attaste Kat餐饮服务		成都市	锦江区	28	510104	50500	盐道街9号	104.0738	30.65486	104.0648	30.65124
公共厕所 公共设施		成都市	武侯区	28	510107	200300	永盛路与	104.0306	30.63774	104.0217	30.63405

图 3-11 成都二环路内 POI 点部分截图

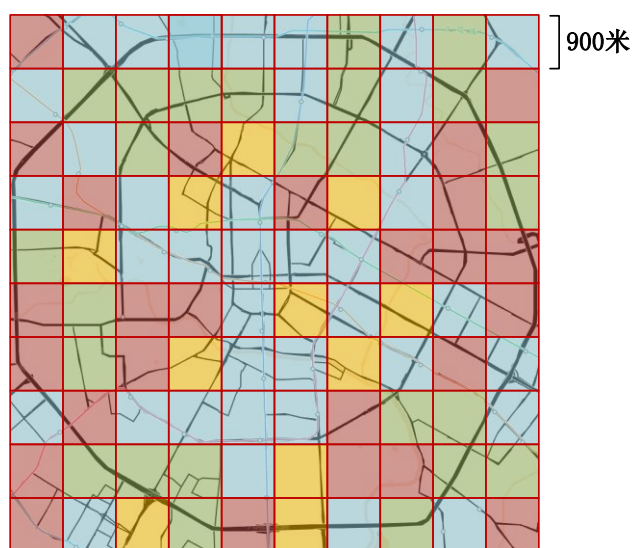


图 3-12 成都二环路功能区示意图

3.5.2 共享单车 OD 分布预测

设置模型的输入和输出的变量单位统计时间段为 20min，位置区域将会被划分为 10×10 栅格，每个栅格的面积为 $0.9 \times 0.9 \text{ km}$ 。

LSTM 和卷积神经网络 (CNN) 等深度学习方法将会被同时进行训练用作对比。数据集分为训练数据集和验证数据集，并以同样的方式测试。上述两种模型的定义如下所示^[79-80]：

(1) LSTM：我们采用双层 LSTM 框架，模型反向传播通过时间最小交叉



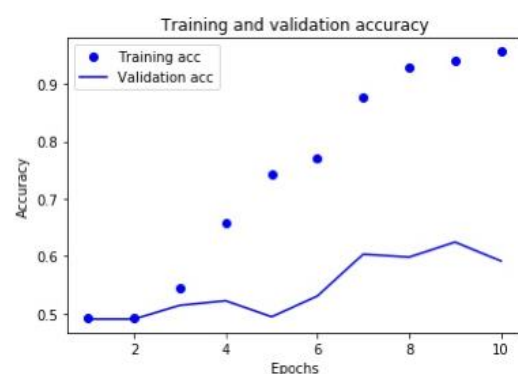
熵损耗 (BPTT) 和 RMSProp, 学习率为 10^{-3} , 衰减率为 0.9。

(2) 卷积神经网络: 采用一种 3 层的全连接卷积结构, 第一层拥有 24 个滤波器, 后两层拥有 64 个滤波器。前两层滤波器长度设置为 3, 最后一层滤波器长度为 2。激活函数类型为 SReLU。

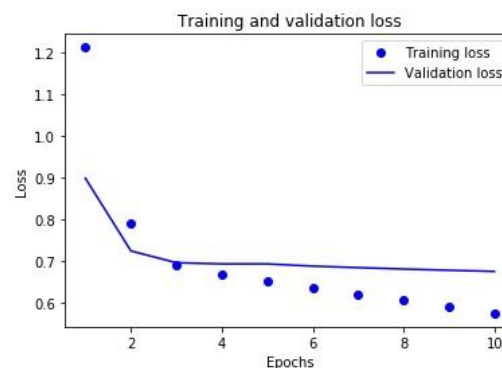
将 2017 年 8 月 14 日~9 月 4 日三周数据进行挖掘与处理, 形成单位时间段的区域出行分布、区域单车分布这两种时空变量的样本空间; 对应时空变量的样本空间, 生成对应时间戳变量的序列数据和功能区属性的矩阵数据。将数据前 80% 作为训练集, 剩下 20% 作为验证集, 并输入到端到端的深度学习框架中进行训练, 训练效果如图 3-13 所示。模型 CNN-LSTM 具有良好的空间和时间特征的识别性, 在训练过程中, 可以快速地实现收敛过程。

```
Train on 3999 samples, validate on 1000 samples
Epoch 1/10
3999/3999 [=====] - 1s 197us/step - loss: 1.2123 - acc: 0.4911 - val_loss: 0.8990 - val_acc: 0.4900
Epoch 2/10
3999/3999 [=====] - 0s 53us/step - loss: 0.7907 - acc: 0.4911 - val_loss: 0.7253 - val_acc: 0.4900
Epoch 3/10
3999/3999 [=====] - 0s 51us/step - loss: 0.6910 - acc: 0.5451 - val_loss: 0.6970 - val_acc: 0.5140
Epoch 4/10
3999/3999 [=====] - 0s 51us/step - loss: 0.6693 - acc: 0.6569 - val_loss: 0.6943 - val_acc: 0.5220
Epoch 5/10
3999/3999 [=====] - 0s 50us/step - loss: 0.6526 - acc: 0.7417 - val_loss: 0.6943 - val_acc: 0.4940
Epoch 6/10
3999/3999 [=====] - 0s 50us/step - loss: 0.6375 - acc: 0.7699 - val_loss: 0.6889 - val_acc: 0.5300
Epoch 7/10
3999/3999 [=====] - 0s 50us/step - loss: 0.6224 - acc: 0.8762 - val_loss: 0.6854 - val_acc: 0.6030
Epoch 8/10
3999/3999 [=====] - 0s 49us/step - loss: 0.6070 - acc: 0.9285 - val_loss: 0.6820 - val_acc: 0.5980
Epoch 9/10
3999/3999 [=====] - 0s 49us/step - loss: 0.5914 - acc: 0.9412 - val_loss: 0.6791 - val_acc: 0.6240
Epoch 10/10
3999/3999 [=====] - 0s 49us/step - loss: 0.5760 - acc: 0.9565 - val_loss: 0.6763 - val_acc: 0.5910
```

(a) 计算过程图



(b) 精度曲线



(c) 损失曲线

图 3-13 模型误差分析图

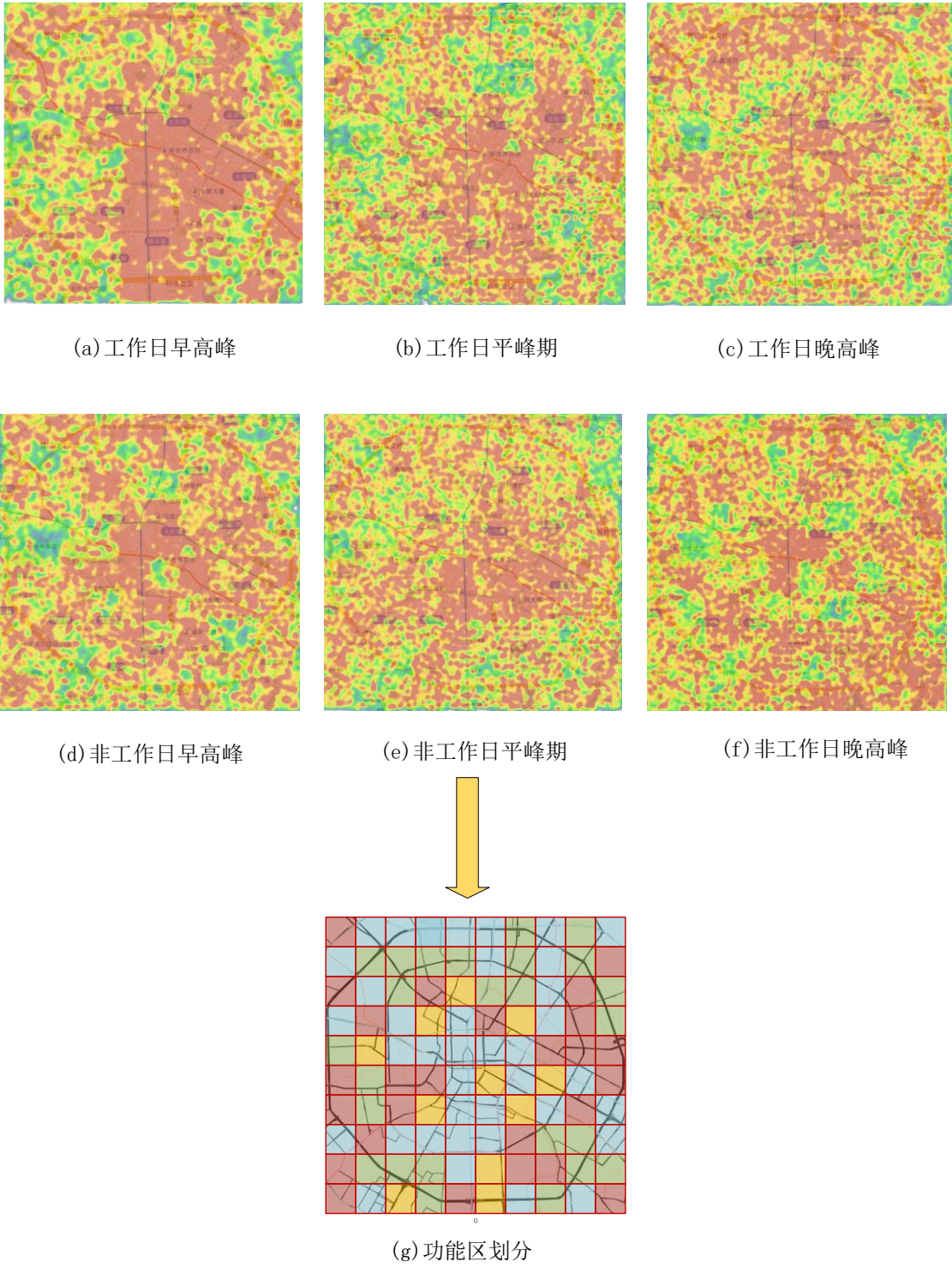


图 3-14 共享单车到达的实际热力图

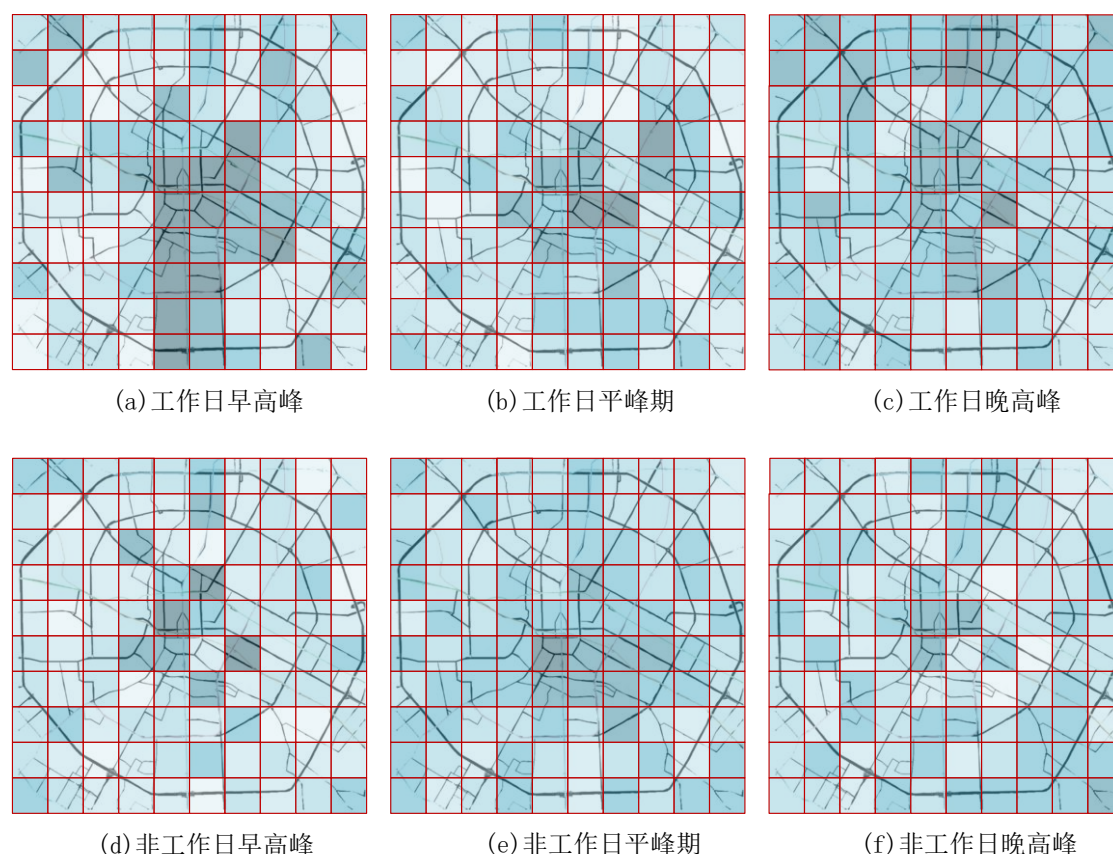


图 3-15 共享单车到达的预测栅格图

用于预测数据的时间范围分为工作日 7:00-7:20、12:00-12:20、18:00-18:20 和非工作日 11:00-11:20、12:00-12:20、18:00-18:20。图 3-14 和图 3-15 分别显示了共享单车在工作日时早高峰（7:20-7:40）、平峰期（12:20-12:40）及晚高峰时期（18:20-18:40）及非工作日时早高峰（11:20-11:40）、平峰期（12:20-12:40）及晚高峰时期（18:20-18:40）的实际到达的热力图和预测到达的栅格图，其中颜色越深意味着强度越大。通过图 3-14 和图 3-15 的对应比较，可以看出 CNN-LSTM 模型能较好地捕捉到共享单车分布的时空特征，模型的预测效果良好。

根据预测的栅格图可得以下结论：

1) 工作日的早高峰期间，地铁功能区的到达量明显高于其他功能区，且与区域内住宅数量与商业办公性质有明显的关系。而其他区域的到达量相对比较均衡。预测结果也较符合实际情况，说明模型在对地铁功能区有较强的识别能力。

2) 工作日的平峰期间，从整体上看，各个功能区的到达分布相对均匀。而混合功能区的到达量较早高峰来说有上升趋势，这些区域通常是中午休闲出行的首选，模型对于时间戳序列下的特征变化也展现了极强的学习能力。



3) 工作日的晚高峰期间,从整体上看,市中心和相对外围区域的到达量相差不多。而在住宅区及住宅区周围的其他功能区到达量都保持一个较高水平,也就是说在这个时段住宅区周围的功能区界限相对不明显,这个时间段时间戳变量和功能区变量对模型的作用尤为明显。

4) 非工作日的早高峰期间,从外围区域来看,混合区的到达量相对于其他功能区的数量偏多。而在市中心区域的地铁区和混合区的到达数量明显增加。办公区整体都处于较低水平,时间戳变量关于工作日和非工作日的定义在模型中发挥了非常重要的作用。

5) 非工作日的平峰期间,从整体上看,各个功能区的单车到达数量都相对均衡且维持在一个相对较高的水平,而市中心的各个功能区到达量都相对较高。由此可见,功能区的界限在此时相对模糊,单车的到达数量主要受地理位置影响。

6) 非工作日的晚高峰期间,从整理上看,各个功能区的达到数量也相对均衡。主要体现在市中心各功能区到达量整体有所下降,而住宅区及其周边的到达量整体有所增加。由此可见,地理位置对共享单车到达量的影响有所减弱,而主要是受住宅区的影响。而非工作日和工作日的划分对于模型的敏感性十分显著。

3.6 本章小结

在本章中,首先介绍了城市功能区划分的基本步骤,即通过高德地图所提供的 API 获取城市的所有 POI 点,并根据相关计算结果判断区域的基本属性。同时,本章构建了符合本文研究对象的深度学习模型,并基于城市不同功能区完成了对共享单车 OD 分布的预测。CNN-LSTM 模型将出行数量分布、单车密度、城市区域功能属性及时间作为输入变量,输出变量则为出行到达分布,并在具体实例中展现出较好的栅格预测性能和预测结果。这也说明, CNN-LSTM 在短期分布预测中能较好的捕捉单车的时空特征,为运营商提供再分配计划提供一定理论依据。

本章通过一种新的深度学习模型完成了对成都市二环路内共享单车到达分布的预测,为后文单车的再调度问题做好铺垫。在实例的预测结果中可以看出,在不同功能区的共享单车存在规律不同的潮汐现象,且工作日和非工作日分布也有明显差异。另外,功能区属性及地理位置对共享单车到达分布所产生的影响具有较强的时间关联性。



第 4 章 基于功能区划分的共享单车时空调度模型

根据统计,共享单车使用者总是在不同功能区之间来回的出行,这就导致了共享单车在空间上也发生了转移。由于功能区的定位和出行规律不同,不同功能区的共享单车在不同时间段的出行数量也会存在一定的潮汐规律,这也是导致不同时间不同功能区发生供需不平衡的原因。供需不均将极大程度的影响单车的使用率及用户的使用体验,因此解决单车的调度需求迫在眉睫。

本章主要探讨的是基于不同功能区的出行规律,确定共享单车的调度方案。以功能区为对象进行区域间的车辆调整,以解决单车自发转移下的分配不均,同时促进整个系统的单车运行形成完整闭环,达到良好的运作状态。

4.1 基于功能区特征的共享单车调度问题分析

共享单车的使用目的较为广泛,但被用于通勤或接驳依然是单车的主要功能。在工作日的早高峰时期,大量的出行者使用共享单车前往地铁站、公交站等交通枢纽;工作期间,大量共享单车被滞留于 CBD 等商业区及工作区;而晚高峰期间,共享单车又大量涌入居民住宅区及混合功能区。基于以上分析可知,不同功能区会存在非常明显的潮汐现象。

导致不同功能区共享单车分布不均的根本原因是共享单车的归还车地点不一致,但间接原因大概有以下几种^[81]:(1)区域人口密度的差异,这将导致共享单车在不同区域的需求量存在较大悬殊,但运营管理者却不能用科学的办法准确的预测需求量的真实情况,均匀或随机投放将一定程度上导致潮汐现象;(2)功能区的属性差异,不同功能区在不同时段的需求不同,若不能准确把握单车在时间及空间上的规律也将导致供需不平衡。(3)用户具有群体方向性,虽然每个用户的出发点存在随机性。但对整个群体里来说,在某些特定的时间点,某个群体往往会向同一个方向聚集,导致该区域的单车数量陡增。

共享单车调度问题往往涉及到一个较广的范围,有时甚至需要整个市统一协调安排,这种大范围的调度问题往往只能由运营公司聘请专业调度员进行统一调度。但对于只涉及相邻几个功能区的片区式调度,雇佣专职人员就不是最经济的选择了。

不同时间段的共享单车调度特性将有所区别,因此将其调度优化问题分为两种类型:不含时间限制的网络优化问题和含时间限制的网络优化问题

(1) 不含时间限制的网络优化问题

不含时间限制的优化问题是指不考虑规定时间内完成调度,也就是不考虑调度的及时性。该类型的优化问题主要针对的是夜间车辆,其服务时间为每天的 0:00-6:00。在夜间时段,使用单车人数少、调度需求小、道路通畅、调度时间也



宽裕。因此这段时间内不需要考虑时间的限制,可将需求看为定值,且目标函数一般设置为总成本最低。

(2) 含时间限制的网络优化问题

含时间限制的优化主要是针对共享单车使用较频繁的白天时段,在这个较长的时段里,不同时间节点的调度任务是不同的,可见此时用户对单车的需求存在一定的时效性。

含时间窗限制的优化模型是指在规定的时间内完成既有的调度任务,即对模型增加时间上的约束条件。为不影响客户满意度的基础上,保证调度任务按时完成,模型将在运营成本最低的基础上,增加关于时间的惩罚函数,即若不能按时完成调度将产生高额的罚款。本文也将采取此种模型来完成基于功能区划分的共享单车时空调度的建模。

4.2 基于功能区的共享单车时空调度模型构建

4.2.1 模型的假设

共享单车的调度模型涉及参数较多且过程较为复杂,但不同时段模型求解的问题基本相同。基于此,对基于功能区的共享单车时空调度模型做出以下假设:

- (1) 由于功能区的直线范围只有 900 米,所以本文不考虑功能区内部的车辆调度问题,只考虑功能区与功能区之间的车辆调度问题;
- (2) 为方便计算,功能区与功能区之间的距离默认是功能区中心之间的直线距离。同时忽略天气,特殊交通状况等因素对调度过程的影响;
- (3) 整个调度区域会有若干个调度中心,每个调度中心只有一辆调度车辆,且调度车具有相同的类型、容量、最大行驶里程;
- (4) 每辆单车的装车成本和卸车成本相同,且行驶成本与调度车辆所装载单车数量无关;
- (5) 每个功能区只能由单一的调度中心服务,每个调度中心可以覆盖多个功能区;
- (6) 调度中心可以得到未来 60 分钟的单车出发量分布;
- (7) 单车数量恒定,所有区域均不在短时间内再重新投放单车;
- (8) 调度车辆在每个区域的调度服务时间与取放车辆数量成正比,不考虑其他时间消耗。

4.2.2 模型的符号说明

$DC^{t,t+\Delta t}$ 表示在 $(t, t+\Delta t)$ 时间段所有车辆的调度成本的总和;

c_1 , c_2 分别表示调度车的搬运成本和行驶成本;



c_3 , c_4 分别表示区域“供不应求”数量的单位惩罚成本和“供过于求”数量的单位惩罚成本;

$EN_{i,j}^{t,t+\Delta t}$ 表示在 $t \sim t + \Delta t$ 时间段从功能区 i 调入功能区 j 的共享单车数量, 其中, $EN_{i,j}^{t,t+\Delta t} = EN_{j,i}^{t,t+\Delta t}$; (反转过来生成一个对称矩阵)

$DL_{i,j}$ 表示功能区 i 和 j 之间的最短路径长度;

$x_{ijk}^{t,t+\Delta t}$ 表示调度车辆 k 是否参与路径 (i, j) 之间的运输;

$PC^{t,t+\Delta t}$ 表示在 $t \sim t + \Delta t$ 时间段所有功能区的惩罚成本之和;

M_i 表示功能区 i 的共享单车保有量;

D_i, Q_i 分别表示功能区 i 的到达需求量和骑行需求量;

N 和 NN 分别表示功能区数量和调度车辆数量。

tt_1 和 tt_2 分别表示调度车辆在每个功能区的调度服务时间和单位行驶时间;

$\alpha_d(t, t + \Delta t)$ 功能区 i 在 $t \sim t + \Delta t$ 时间段的调度优先级系数

4.2.3 模型的构建

共享单车在整个运营过程中同时具备了商业性和服务性, 因此管理者在提高自身服务水平的同时更注重追求集团收益。基于以上分析, 建模过程中首先考虑的是调度成本最低, 其次尽可能的提高使用者的出行体验, 满足用户的出行需求。

(1) 目标函数——综合成本最低

本文构建的共享单车车辆调度模型的综合成本函数由两个目标函数组成, 一个是调度成本函数, 另一个是惩罚成本函数。

其中调度成本函数主要包括参与调度的车辆数、调度过程中产生的服务成本及行驶过程中的行驶成本共同构成。用公式 4-1 表示:

$$DC^{t,t+\Delta t} = \sum_k \sum_i \sum_j (c_1 EN_{i,j}^{t,t+\Delta t} + c_2 DL_{i,j}) x_{ijk}^{t,t+\Delta t} \quad (4-1)$$

对于惩罚成本函数, 关键是判断功能区在前一时间段的单车保有量能否满足下一个时间段的单车出行需求, 其次是否有大量单车在某个区域被闲置。单车的供需关系判定是调度服务时间窗和确定调度需求量的基础和前提, 只有满足区域单车供需关系才能保证区域的共享单车数量处于科学合理的水平。一旦功能区 i 在调度后的 $t + \Delta t$ 时段的单车保有量低于下一个时间段的单车出行数量, 那么就会判定运营商承担相应的惩罚成本, 同时, 如果有大量共享单车被闲置在了该区域, 该区域也要承担一部分惩罚成本。但“供大于求”的造成的损失仅仅是导致该区域的单车数量过剩, 产生一定的管理费用。而“供不应求”极大可能导致用



户的流失, 对公司造成深远影响, 因此“供大于求”惩罚成本要低于“供不应求”的惩罚成本。本文定义供不应求和供大于求的惩罚函数可以被定义为:

$$PC_i^{t,t+\Delta t} = \begin{cases} c_3(Q_i^{t,t+\Delta t} - M_i^t), M_i^t < Q_i^{t,t+\Delta t} \\ 0, & Q_i^{t,t+\Delta t} = M_i^t \\ c_4(M_i^t - Q_i^{t,t+\Delta t}), Q_i^{t,t+\Delta t} < M_i^t \end{cases} \quad (4-2)$$

同时, 根据第三章的结论, 不同类型功能区在不同时间段的出行需求和出行规律是不同的。根据其在不同时间段的出行类型, 设置不同功能区的出行优先级系数 $\alpha_d(t, t+\Delta t)$, $d = \{1, \dots, 4\}$ 代表 4 种功能区划分, 而每个时间段都会有其时间戳属性:

$$(t, t+\Delta t) \in \left\{ \begin{array}{l} \text{工作日早高峰, 工作日平峰期, 工作日晚高峰,} \\ \text{非工作日早高峰, 非工作日平峰期, 非工作日晚高峰,} \end{array} \right\}$$

对于不同功能区而言, 在不同时间戳下所对应的优先级是不同的。

综上, 在 t 至 $t+k\Delta t$ 时间段内综合成本最低 C 的目标函数为:

$$\min C = \sum_i \alpha_d(t, t+\Delta t) PC_i^{t,t+\Delta t} + \sum_i \sum_j DC_{ij}^{t,t+\Delta t} \quad (4-3)$$

(2) 约束条件

1) 调度阈值

设功能区 i 在 $(t, t+\Delta t)$ 时段的共享单车需求量为 $Q_i^{t,t+\Delta t}$, 在未来 k 个时间段最大出行需求为 $\max(Q_i^{t,t+\Delta t}, \dots, Q_i^{t+(k-1)\Delta t, t+k\Delta t})$ 。对于某个功能区而言, 必须要保证该区域的单车保有量在现阶段或者未来短时间内能满足一定的出行需求, 当车辆保有量过低或严重无法满足未来需求时, 会导致大量乘客集中等待, 最终使乘客对共享单车服务的极度不满。因此将需求匹配下限为 0.3, 强制调度可避免某区域内车辆严重不足。考虑到调度的调节作用, 本文设定预满足的时间段 $k=3$, 则 t 时段共享单车的调度阈值为: $0.3\max(Q_i^{t,t+\Delta t}, \dots, Q_i^{t+2\Delta t, t+3\Delta t})$, 功能区 i 在 t 时段的共享单车保有量与调度阈值的关系为:

$$M_i^t > 0.3\max(Q_i^{t,t+\Delta t}, \dots, Q_i^{t+2\Delta t, t+3\Delta t}) \quad (4-4)$$

2) 调度车辆容量约束

依照现实, 每辆调度车辆的调度容量是有限的, 且每个区域的调度车辆数量是固定的, 调度范围也是固定。即:

$$EN_{i,j}^{t,t+\Delta t} \leq U \quad (4-5)$$



3) 调度车辆的服务时间约束

对于调度车辆而言, 其服务时间由两部分组成, 第一部分是车辆的装卸服务时间, 第二部分是车辆的行驶服务时间, 在一定的时间段内, 这两部分时间之和必然小于时间窗:

$$\sum_k x_{ijk}^{t,t+\Delta t} \leq NN \quad (4-6)$$

$$\sum_i \sum_j x_{ijk}^{t,t+\Delta t} (tt_1 + DL_{i,j} tt_2) \leq \Delta t, \quad (k=1, \dots, NN) \quad (4-7)$$

4) 区域单车数量约束

由于考虑不再新投放车辆, 所以对于整个区域而言, 单车总量是守恒的, 而且每个功能区的单车保有量由上一个时段出行到达数量和调度数量共同决定:

$$M_i^t = M_i^{t-1} + D_i^{t-\Delta t, t} - Q_i^{t-\Delta t, t} - \sum_{j \neq i} EN_{i,j}^{t-\Delta t, t} + \sum_{j \neq i} EN_{j,i}^{t-\Delta t, t}, \quad (i=1, \dots, N) \quad (4-8)$$

综上所述, 其模型为:

$$\begin{aligned} \min C = & \sum_i \alpha_d(t, t+\Delta t) PC_i^{t,t+\Delta t} + \sum_i \sum_j DC_{ij}^{t,t+\Delta t} \\ & \left\{ \begin{aligned} DC_{ij}^{t,t+\Delta t} = & \sum_k \sum_i \sum_j (c_1 EN_{i,j}^{t,t+\Delta t} + c_2 DL_{i,j}) x_{ijk}^{t,t+\Delta t} \\ PC_i^{t,t+\Delta t} = & \begin{cases} c_3 (Q_i^{t,t+\Delta t} - M_i^t), & M_i^t < Q_i^{t,t+\Delta t} \\ 0, & Q_i^{t,t+\Delta t} = M_i^t \\ c_4 (M_i^t - Q_i^{t,t+\Delta t}), & Q_i^{t,t+\Delta t} < M_i^t \end{cases} \\ M_i^t > & 0.3 \max(Q_i^{t,t+\Delta t}, \dots, Q_i^{t+2\Delta t, t+3\Delta t}) \end{aligned} \right. \quad (4- \\ s.t. & \left\{ \begin{aligned} EN_{i,j}^{t,t+\Delta t} & \leq U \\ \sum_i \sum_j x_{ijk}^{t,t+\Delta t} (tt_1 + L_{i,j} tt_2) & \leq \Delta t, \quad (k=1, \dots, NN) \\ M_i^t = M_i^{t-1} + D_i^{t-\Delta t, t} - Q_i^{t-\Delta t, t} - \sum_{j \neq i} EN_{i,j}^{t-\Delta t, t} + \sum_{j \neq i} EN_{j,i}^{t-\Delta t, t}, & (i=1, \dots, N) \\ \sum_k x_{ijk}^{t,t+\Delta t} & \leq NN \end{aligned} \right. \end{aligned}$$

9)



4.3 本章小结

由于功能区的定位和空间属性的差异,用户在不同功能区之间展现了不同的时空规律。在这些不同属性的功能区之间,共享单车有着非常明显的潮汐现象,同时较多区域存在共享单车供需不匹配的现象,这就需要在相应的时间点通过调度行为加以干预。

本章在充分分析功能区特征和调度问题的关联性的基础上,通过创新性地引入功能区和时间戳等控制变量,建立一种多车场多车辆的动态共享单车调度路径优化模型。在目标函数的构建中,为了更好地平衡目标搜索方向,本文构建了调度成本函数和惩罚成本最小化的双目标优化函数,并考虑了不同时间戳的功能区优先级。在约束条件的构建中,本文基于调度阈值、调度车辆容量、调度车辆的服务时间、区域单车数量等四个方面分别建立单位时间段的调度行为与车辆行为的多重约束,以此增加了模型的适应性。



第 5 章 模型的求解及算例分析

5.1 共享单车调度问题的算法分析

车辆调度路径优化基本上都属于 NP-Hard 问题，而目前的研究表明没有一个算法能精确的求解该类问题。因为随着输入进模型的数据不断增加，将导致计算量也不断增长，使得求解过程不断复杂。近年来，有效解决车辆路径优化问题的算法主要有精确算法和启发式算法^[82]。而启发式算法又由传统启发式算法和现代化启发式算法组成。

(1) 传统启发式算法

该算法的基本步骤是给出一个初始解，再将该解进行迭代改进以获得最优解。传统启发式算法计算效率低，因此常于其他算法组合使用。

(2) 现代启发式算法

多种常见现代启发式算法的描述及特点分析如表 5-1 所示。

表 5-1 现代启发式算法

算法名称	算法描述	算法特点
遗传算法	该算法的搜索方式受启发于自然进化规律的“优胜劣汰，适者生存”。得到的初始解将经过选择、交叉、变异等处理产生新解，并将其中较优的特性遗传给下一代。在多次迭代中得到最优解。	全局搜索能力强 收敛速度慢 容易产生局部最优解
禁忌搜索算法	禁忌表不仅可避免搜索的重复性，还能打破规则获取一些被禁的良解，从而实现解的全局性优化。	搜索速度快 对初始解的依赖性强
模拟退火算法	通过干扰初始解来产生新解，并设置条件以一定比例接受劣解以避免造成最优解的局部性。	计算简单，算法鲁棒性强 收敛速度慢，运算时间长 易受参数值设定的影响

5.2 算法原理及求解思路

5.2.1 蚁群算法原理

蚁群算法的基本原理为模拟蚂蚁的觅食行为。蚂蚁在觅食时，将在经过的路径上留下信息素，同时蚂蚁还能感受到这些信息素的强弱并通过这些物质进行信息交互及方向指引。当某条路径较优时，蚂蚁就会大规模的聚集在该路径



上, 因此留在该路径上的信息素就越多, 然后蚂蚁很大概率的选择这些信息素浓度较高的路径, 如此循环最终得到最优路径^[83]。

其算法的基本过程如下:

首先, 对符号做以下的定义:

d_{rs} 表示路径 (r, s) 长度; η_{ij} 表示路径 (r, s) 的能见度; τ_{rs} 表示路径 (r, s) 的信息素浓度; $\Delta\tau_{rs}^a$ 表示蚂蚁 a 在路径 (r, s) 上留下的单位长度轨迹信息素量; p_{rs}^a 表示蚂蚁 a 由节点 r 转移到 s 的概率。

在初始阶段时, 各条路径上信息素相等, 设 $\tau_{rs}(0) = c$ (c 为常数)。在 t 时刻, 蚂蚁 a 选择路径 (r, s) 的转移概率 $p_{rs}^a(t)$ 为:

$$p_{rs}^a(t) = \begin{cases} \frac{\tau_{rs}^\alpha(t) \eta_{rs}^\beta(t)}{\sum_{m \in allowed_a} \tau_{rm}^\alpha(t) \eta_{rm}^\beta(t)}, & s \in allowed_a \\ 0, & s \notin allowed_a \end{cases} \quad (5-1)$$

$$\eta_{rs}(t) = 1/d_{rs} \quad (5-2)$$

其中, $allowed_a = \{0, 1, \dots, n-1\}$ 表示下一次能被选择的节点集合。参数 α 表示的是蚂蚁在移动过程中积累的信息浓度在路径中的相对重要性, 参数 β 反映的是移动过程中能见度在路径中的相对重要性。禁忌表 $tabu_a$ 记录了蚂蚁 a 已经走的节点, 且禁止蚂蚁 a 在该次循环中再选择这些节点。

当一次循环完成后, 各路径上信息素根据下面的公式进行更新:

$$\tau_{rs}(t) = (1 - \rho) \cdot \tau_{rs}(t-1) + \Delta\tau_{rs} \quad (5-3)$$

$$\Delta\tau_{rs} = \sum_{a=1}^n \Delta\tau_{rs}^a \quad (5-4)$$

其中: ρ 表示信息素的衰减系数 ($0 \leq \rho \leq 1$); $\Delta\tau_{rs}$ 表示本次循环中路径 (r, s) 的信息素的增量; $\Delta\tau_{rs}^a$ 表示本次循环中第 k 只蚂蚁在路径 (r, s) 中遗留下来的信息素数量。

$$\Delta\tau_{rs}^a = \begin{cases} \frac{Q}{L_a} & \text{如果第 } k \text{ 只蚂蚁经过路段 } (r, s) \\ 0, & \text{否则} \end{cases} \quad (5-5)$$

式中: Q 代表路径中信息素浓度, L_a 代表蚂蚁 a 在该次循环中所经过路径的长度之和。

当所有蚂蚁完成一次上述循环后即结束一次迭代过程, 记录当前所得到的路径信息。重复以上过程得出最终的最优配流方案为止。



5.2.2 算法求解思路

共享单车调度问题其实是一种 1-TSPPD (ONE-commodity pickup-and-delivery traveling salesman problem) 问题。1-TSPPD 问题指的是送货取货的对象为单一性质的商品的旅行商问题。共享单车调度问题的商品为共享单车, 通过调度车去往相应的功能区, 且调度车辆可以在不同功能区之间运输车辆。但是共享单车调度问题也不完全是旅行商问题, 主要表现在:

- (1) 在本文的调度问题中, 小区设置为可多次访问或不访问;
- (2) 达到相应的时间点即可触发调度任务, 结束时刻终止调度;
- (3) 调度中心可以存在多个。假设某个时间点开始触发调度阈值, 一辆容量有限的运输车由调度中心出发, 在功能区之间运输车辆, 不断保持综合成本最低的系统状态。

结合本文研究的情况复杂, 现对算法进行如下两个方面改进:

第一, 针对取车和送车两个环节针对转移规则进行更新;

第二, 将蚁群的投放位置设置为调度中心所在位置, 通过设置 $allowed_k$ 来规定相应蚁群的活动区域, 对应每辆调度车辆的调度活动范围;

第三, 考虑到调度车辆的取送问题, 因此对局部的更新规则也进行了改进。

5.2.2.1 转移规则

调度车辆将大概率去到调度成本越低的区域, 因此对路径 rs 在第 t_z 个时间段的能见度 $\eta_{rs}(t)$ 进行如下更新:

$$\eta_{rs}(t_z) = PC_s^{t_z} - c_2 DL_{rs} \quad (5-6)$$

DL_{rs} 表示通过从功能区 r 去到功能区 s 进行调度行为的最短路径长度。 $PC_s^{t_z}$ 分别表示功能区 i 在 t_z 时间段的惩罚成本。

将 w 个调度中心的蚁群设为 A_1, A_2, \dots, A_w , 初始阶段时, 各条路径上信息素相等, 设 $\tau_{rs}(0) = c$ (c 为常数)。则在第 t_z 个时间段。调度车辆 A_k 在 t_z 时间段选择路径 (r, s) 的转移概率 $p_{rs}^k(t_z)$ 为:

$$p_{rs}^k(t_z) = \begin{cases} \frac{\tau_{rs}^\alpha(t_z) \eta_{rs}^\beta}{\sum_{m \in allowed_k} \tau_{rm}^\alpha(t_z) \eta_{rm}^\beta} \theta(t_z, s), s \in allowed_k \\ 0, \dots, s \notin allowed_k \end{cases}, k = 1, \dots, w \quad (5-7)$$

式中 $allowed_k$ 表示调度车辆 A_k 的可调度范围; 参数 α 反映了信息素强弱的在路径选择中的相对重要性 ($0 < \alpha < 1$); 参数 β 反映了能见度在路径选择中的相对重要性 ($0 < \beta < 1$), 同时参数满足 $\alpha + \beta = 1$ 。 $\theta(t_z, s)$ 表示取送平衡函数, 对于调度车辆 k 而言, 每经历三次调度行为, 其中必有一次是取车行为:



$$\theta(t_z, s) = \begin{cases} M_s^{t_{z+1}} - Q_s^{t_{z+2}}, & \theta(t_{z-1}, r) + \theta(t_{z-2}, r) < 2 \\ 1, & \theta(t_{z-1}, r) + \theta(t_{z-2}, r) \geq 2 \end{cases} \quad (5-8)$$

且 $\theta(t_0, s) = M_s^{t_1} - Q_s^{t_2}$ 。

5.2.2.2 局部规则

改进局部规则可以防止调度车收敛到同一个路径, 因此调度车辆完成每一步搜索后, 路径的信息素浓度将被更新, 具体规则如下:

$$\tau_{rs}(t_z) = \left[(1 - \rho) \cdot \tau_{rs}(t_{z-1}) + \frac{\lambda}{\tau_{rs}(t_{z-1})} \Delta \tau \right] \quad (5-9)$$

$$\Delta \tau_{rs} = \sum_{a=1}^n \Delta \tau_{rs}^a \quad (5-10)$$

$\frac{\lambda}{\tau_{rs}(t-1)}$ 代表路径的承载力, 当调度车经过某条路径时, 该路径的吸收能力将会被适当削弱, 由此一来该路径对其他调度车的吸引能力也会降低。若某路径的信息素浓度已经较高, 此时该路径其吸收信息素的能力就会被减弱, 那么其他路径被调度车的概率就会加大。通过该设置可避免搜索太过集中而算法被迫停滞。

$$\Delta \tau_{rs}^k = \begin{cases} \frac{Q}{LS_k} & \text{如果第 } k \text{ 辆调度车辆经过路段 } (r, s) \\ 0, & \text{否则} \end{cases} \quad (5-11)$$

式中: Q 代表路径中信息素浓度, LS_a 代表蚂蚁 a 在该次循环中所经过路径的长度之和。

基于以上改进方法, 算法步骤如下:

Step1: 初始化共享单车在每个功能区的保有量 ($M_1^{t_0} \dots M_w^{t_0}$), 同时设置功能区 i 在未来 T 个时间段 ($(t_0, t_1), \dots, (t_{T-1}, t_T)$) 的需求量为 ($Q_i^{t_0, t_1} \dots Q_i^{t_{T-1}, t_T} \mid i = 1, \dots, w$);

Step2: 将未来 T 个时间段的区域共享单车需求量 ($Q^0 \dots Q^T$) 带入上一章 4.4 节中构建的深度学习框架中进行预测, 输出未来 $T-1$ 个时间段 ($(t_1, t_2), \dots, (t_{T-1}, t_T)$) 的共享单车达到量分布 ($D_i^{t_0, t_1} \dots D_i^{t_{T-1}, t_T} \mid i = 1, \dots, w$);

Step3: 设置算法所需要的最短路径距离矩阵 DL_{rs} 和惩罚成本 $PC_s^{t_z}$;

Step4: 初始化 τ_{rs} 、 $\Delta \tau_{rs}$ 和禁忌表 $tabu_a$, 将 A_1, A_2, \dots, A_w 辆调度车分别放置在 $1 \dots w$ 个功能区中的不同初始调度点中;

Step5: 每辆调度车均以概率 $p_{rs}^k(t_z)$ 尝试选择下目标点 s , 并将目标 s 添入至 $tabu_a$ 中, 直至禁忌表满;

Step6: 根据禁忌表 $tabu_a$, 更新 τ_{rs} 与 $\Delta \tau_{rs}$;

Step7: 对各路径 (r, s) 置 $\Delta \tau_{rs} = 0$;

Step8: 记录最佳的路径, 若不满足终止条件, 清空禁忌表, 转步骤 step4;



Step9:输出最佳调度方案。

5.3 算例分析

为了验证上文所构建的基于功能区划分的共享单车时空调度模型和改进型蚁群算法的有效性,选取第三章中所划定的城市范围和时空数据进行的调度问题的分析建模。

按照第三章的时间划定方法,将 7:00-8:00, 12:00-13:00, 18:00-19:00 三个小时分割为 9 个时间段,每个时间段为 20 分钟。为了研究模型在复杂需求下的适应性,将这 9 个时间段按照时间先后顺序拼接组成一个连续的时间序列— $(t_0, t_1), \dots, (t_8, t_9)$ 。

选取 2017 年 8 月 15 日的摩拜单车数据为基础,对上述 9 个时间段的各个功能区的共享单车保有量和出行量进行挖掘,得出 100 个功能区在 7:00 时初始共享单车保有量 $(M_1^{t_0} \dots M_{100}^{t_0})$ 和功能区 i 在未来 1~9 个时间段 $((t_0, t_1), \dots, (t_8, t_9))$ 的实际需求量 $(Q_i^{t_0, t_1} \dots Q_i^{t_8, t_9} | i=1, \dots, 100)$;

将未来 9 个时间段的区域共享单车需求量 $(Q_i^{t_0, t_1} \dots Q_i^{t_8, t_9} | i=1, \dots, 100)$ 带入第四章构建的深度学习框架中进行预测,输出未来 8 个时间段的共享单车达到量分布 $(D_i^{t_1, t_2} \dots D_i^{t_8, t_9} | i=1, \dots, 100)$;

默认几何中心距离代表两个矩形区域之间的实际距离,通过算法生成最短路径距离矩阵 DL_{rs} 。

通过查阅相关文献,确定每辆车的调度相关的单位成本如下表所示:

表 5-2 参数设置表

成本类型	单价
调度行驶成本	$c_1+c_2=3$ 元/每公里
供小于求的惩罚成本	$c_3=1.2$ 元/辆
供大于求的惩罚成本	$c_4=0.3$ 元/辆

初始化 τ_{rs} 和 $\Delta\tau_{rs}$, 禁忌表 $tabu_a$ 置空, 将 A_1, A_2, \dots, A_5 辆调度车辆置于 5 个功能区中, 默认车辆初始位置均在调度中心, 如下图所示。

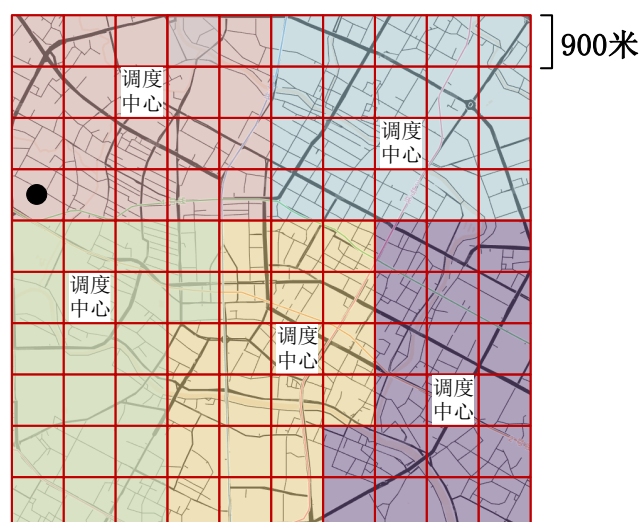


图 5-1 调度中心分布图

由于五个调度中心的计算过程完全相同,因此本文将选取左上角红色区域的调度区域作为分析对象。为了研究方便,现将该区域的每个功能区进行编号,区域的中心点坐标即为区域的坐标,坐标原点为图 5-1 的黑点处(区域 16 的中心点),绘制比例为 1: 1000m,具体坐标如下表所示。

表 5-3 区域坐标

区域	横坐标	纵坐标	区域	横坐标	纵坐标
1	0	2.7	11	0	0.9
2	0.9	2.7	12	0.9	0.9
3	1.8	2.7	13	1.8	0.9
4	2.7	2.7	14	2.7	0.9
5	3.6	2.7	15	3.6	0.9
6	0	1.8	16	0	0
7	0.9	1.8	17	0.9	0
8	1.8	1.8	18	1.8	0
9	2.7	1.8	19	2.7	0
10	3.6	1.8	20	3.6	0

基于单车的数量分布、出行分布及第三章的到达预测结果,可以根据公式 5-12 和 5-13 计算出每个区域的供需差值,其中负数代表可调出的车辆,具体如下图所示。

$$\Delta_i^{t,t+\Delta t} = M_i^t - Q_i^{t,t+\Delta t}, \quad (i=1, \dots, N) \quad (5-12)$$

$$M_i^t = M_i^{t-\Delta t} + D_i^{t-\Delta t,t} - Q_i^{t-\Delta t,t}, \quad (i=1, \dots, N) \quad (5-13)$$

18	27	-13	-21	-78
-82	95	-77	16	-20
-33	-64	-86	-80	-121
54	12	79	-25	-55

(a) 7:20-7:40

22	21	-16	-26	-82
-72	93	-61	21	-24
-28	-54	-72	-74	-151
41	12	74	-28	-42

(b) 7:40-8:00

23	17	-16	-33	-84
-92	103	-47	19	-23
-26	-47	-65	-82	-120
33	11	77	-27	-50

(c) 8:00-8:20

-61	-31	-24	15	-79
-47	18	-19	-22	20
-30	-96	-53	-88	-21
54	-35	87	-14	-9

(d) 12:20-12:40

-51	-27	-28	15	-88
-57	22	-23	-26	23
-33	-100	-38	-76	-17
68	-27	61	-15	-8

(e) 12:40-13:00

-58	-27	-24	13	-103
-63	21	-19	-26	25
-41	-113	-48	-65	-18
71	-20	76	-17	-9

(f) 13:00-13:20

54	-18	21	41	-32
-7	9	-30	16	-20
-44	-38	-65	-97	-34
-32	-16	8	-21	-45

(g) 18:20-18:40

52	-13	21	36	-37
-9	7	-38	16	-22
-33	-47	-83	-95	-40
-34	-15	11	-20	-47

(h) 18:40-19:00

图 5-2 供需关系图

本文基于功能区对蚁群算法进行了重新改进与设计，使蚁群算法在每个时间段都会对转移概率进行更新，保证了取送车的均衡性，在经历了 8 个时间段共 100 次整体迭代，最终获得稳定的调度方案。调度结果如下所示。

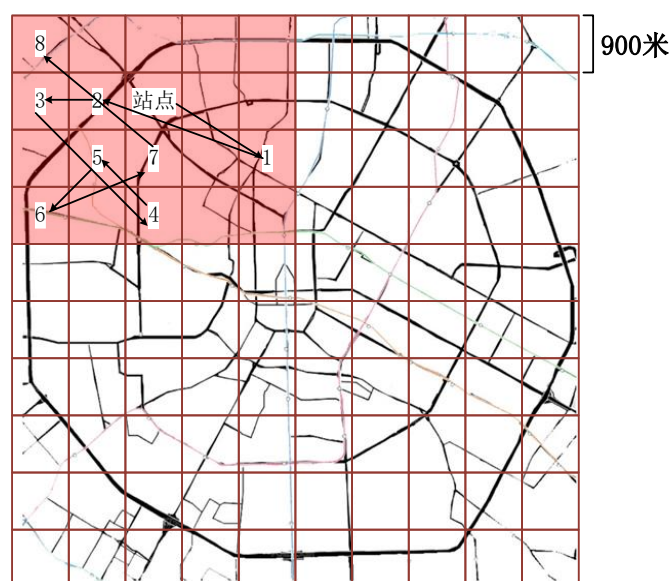


图 5-3 调度结果展示图

蚁群算法最终的结果为调度站 $\rightarrow 15 \rightarrow 7 \rightarrow 6 \rightarrow 18 \rightarrow 11 \rightarrow 16 \rightarrow 12 \rightarrow 1$ ，其中供大于求需要取车的站点为 15、6、11、12，而供不应求需要补给的站点为 7、18、16、1。而根据运输成本及惩罚成本的计算公式该结果的最小成本为 46.23 元。

5.4 本章小结

本章通过对基于功能区划分的共享单车时空调度模型进行剖析，分析了共享单车调度问题的特性，并根据各类算法的适用性特点设计了改进型蚁群算法进行求解，同时梳理了问题的求解流程；在已有车辆调度问题相关理论的基础上，对原始数据加以处理与分析，并对共享单车调度模型的相关参数进行设定及其调度需求量进行预测分析。根据设计的算法步骤，对成都市二环内的数据进行实例分析与求解，最终得出全局近似最优解，验证了模型的有效性。



结论与展望

主要结论

本文以城市共享单车为研究对象,探索了共享单车的大数据特征,构建了基于城市功能区的共享单车出行时空预测模型及其时空调度模型。并用实例验证了模型的可行性。具体研究结论如下:

(1) 共享单车的大数据特征

通过对成都摩拜单车大数据进行挖掘与分析,从数据来源和数据处理两个方面进行了一些新的探索,并在此基础上分析了共享单车的选择行为的影响因素。基于共享单车大数据的挖掘,本文除了发现一些比较基础性比如骑行距离、出行 OD、出行频率、骑行时间等方面的行为特征,还基于地铁站和功能区两个新的层面进行了一些特征挖掘,发现在不同区域的地铁站附近的共享单车在不同时间段存在截然不同的规律性行为。另外,共享单车出行行为的区域性差异不仅体现在地铁站附近,还体现在功能区属性不同的整个区域内。这为后续研究的开辟了一个新的思路。

(2) 基于城市功能区的共享单车出行时空预测

在第三章中,考虑到共享单车出行 OD 预测的必要性和复杂性,我们使用了一种新的深度学习框架进行 OD 分布的预测—CNN-LSTM。这种模型在针对本文空间和时间属性同时存在的预测问题时展现了非常好的预测效果;同时,功能区的引入很好的解决了区域的时空划分问题,让模型可以更好地学习这种基于不同功能区的规律性特征。这种基于城市功能区划分的共享单车 OD 分布预测揭示了出行者在一天中不同的时段各种潮汐行为,为共享单车的调度和规划提供了特征基础。

(3) 基于功能区划分的共享单车时空调度模型

用户在不同功能区之间有规律地使用共享单车,使其在空间上发生了转移。由于功能区的定位和出行规律不同,不同功能区的共享单车在不同时间段的驶出数量和驶入数量也会产生非常大的潮汐波动,本文通过分析这种规律性行为,提出了基于功能区的出行特征来进行调度行为的优化,以功能区为单位进行区域车辆动态调度,实现资源的合理配置。这种基于时间序列和功能特征的时空调度模型可以很好地解决共享单车的调度问题,实现调度资源的良好利用。



未来工作展望

在共享单车的发展研究中,本文从时空特征和功能特征双重角度对共享单车的预测和调度问题提供了解决方案,揭示了共享单车的规律性特征和运行特征。但是,还存在诸多问题和方法需要进行进一步探索和研究:

(1) 在本文中,取送环节和单车数量都是动态变化的,本文所构建的算法是默认每次取送是相互交叉的,并没有考虑连续取或者连续送的情况发生,因此与实际情况有所差距;另外,模型默认取单车的数量由下次需要送的车辆数决定,也与实际情况有所偏差。

(2) 共享单车的 OD 行为和调度行为需要更为精确的框架和算法来研究。单纯地利用模型去研究数据和调度问题并无法得出更具价值的结论,在未来的研究中,可以从政府监管、企业自身发展方面提出一些较为系统的建议。

(3) 深度学习模型揭示了共享单车的规律性特征,但是没有深入挖掘其用户的个体特征,还可以结合天气、特殊事件、季节等多种因素进行机器学习框架的建立,提升对单车行为的匹配程度。



致 谢

研究生生活转瞬即逝，三年的学习生活以论文的即将完成而告一段落。三年的学习时间将为我以后的生活打下坚实的基础。交大给我的研究生生活带了很多不一样的惊喜，严谨的教学风格，一流的教学团队，还有学子们浓厚的学习氛围，这一切都让我们感受到科研是有意义的！在三年的学习时光中，这些都给我带了很多莫大的帮助，在这里要向他们表示真挚的谢意。

首先要感谢我的导师帅斌老师，在论文完成的整个过程，他总是积极和蔼地和我们交流讨论，在我不知道如何进行下一步工作时给我们的论文研究指点方向，利用他对专业几十年的透彻理解和经验给我提出了很多建设性的意见和建议。坚持做到每一周都跟我们进行面对面地交流，让我可以更清楚直观地表达自己的想法。在进行线下交流时，他也能做到第一时间回复我，消除我的疑问和不解。除此之外，他不仅在学术方面帮助我授业解惑，在做人做事的细节和对未来的展望中也经常鼓励开导我，让我能更自信地步入社会。帅老师专业严谨的教学态度、渊博的教学知识让我受益匪浅，再一次感谢老师在论文研究期间对我的关心和指导。

感谢同师门的各位兄弟姐妹，在我平时的学习和生活上给我带来很多帮助，还有和我一起共同生活的室友，和谐的宿舍生活环境让我能够更好地专心于学业。感谢我的父母，总是无条件地站在我身后默默地支持我，虽然有的时候会发生矛盾，但最后总是给我无限地包容和忍让。希望今后的日子里，我能努力做到让他们幸福，祝他们安康快乐。

最后感谢辛勤审阅论文的各位老师，还有出现在我生命中的所有人，希望未来我能成为更好的自己。愿论文答辩一切顺利！



参考文献

- [1] Chen, H., Jia, B., Lau, S. S. Y. Sustainable urban form for Chinese compact cities: Challenges of a rapid urbanized economy[J]. *Habitat International*, 2008, 32 (1) : 28-40.
- [2] Gakenheimer, R. Urban mobility in the developing world[J]. *Transportation Research Part A: Policy and Practice*, 1999, 33 (7) : 671-689.
- [3] Hayashi, Y., Doi, K., Yagishita, M., et al. Urban transport sustainability: Asian trends, problems and policy practices[J]. *European Journal of Transport and Infrastructure Research*, 2004, 4 (1) : 27-45.
- [4] Aguiléra, A., Grébert, J. Passenger transport mode share in cities: Exploration of actual and future trends with a worldwide survey[J]. *International Journal of Automotive Technology and Management*, 2014, 14 (3) : 203-216.
- [5] 北京交通发展研究中心. 2011 年北京市交通发展年度报告. 2011.
- [6] Midgley, P. Bicycle-sharing schemes: Enhancing sustainable mobility in urban areas. United Nations Department of Economic and Social Affairs, 2011.
- [7] Midgley, P. The role of smart bike-sharing systems in urban mobility[J]. *Journeys*, 2009, 2:23-31.
- [8] Dell'Olio, L., Ibeas, A., Moura, J. L. Implementing bike-sharing systems[J]. *Proceedings of the ICE-Municipal Engineer*, 2011, 164 (2) : 89-101.
- [9] Jensen, P., Rouquier, J. B., Ovtracht, N., et al. Characterizing the speed and paths of shared bicycle use in Lyon. *Transportation Research Part D: Transport and Environment*, 2010, 15 (8) : 522-524.
- [10] Buck, D., Buehler, R. Bike lanes and other determinants of capital bikeshare trips[C]. 91st Transportation Research Board Annual Meeting. 2012.
- [11] Harms, L., Bertolini, L., Brömmelstroet, M. T. Spatial and social variations in cycling patterns in a mature cycling country exploring differences and trends[J]. *Journal of Transport & Health*, 2014, 1 (4) : 232-242.
- [12] Fishman, E. Bikeshare: A Review of Recent Literature[J]. *Transport Reviews*. 2016, 36 (1) : 92-113.
- [13] Hu, J., Yang, Z., Shu, Y., Cheng, P., Chen, J. Data-Driven Utilization-Aware Trip Advisor for Bike-Sharing Systems[C]. *IEEE International Conference on Data Mining*, 2017: 167-176.
- [14] Tian, Z. P., Wang, J.Q., Wang, J., Zhang, H.Y. A multi-phase QFD-based hybrid fuzzy MCDM approach for performance evaluation: A case of smart bike-sharing programs in Changsha [J]. *Journal of Cleaner Production*. 2018, 171: 1068-1083.



- [15]Yahya, B. N. Overall Bike Effectiveness as a Sustainability Metric for Bike Sharing Systems[J]. sustainability. 2017, 9 (11) : 2070.
- [16]Shen, Y., Zhang, X., Zhao, J. Understanding the usage of dockless bike sharing in Singapore[J]. International Journal of Sustainable Transportation. 2018, 0 (0) : 1-15.
- [17]Wang, KL., Akar, G., Chen,YJ., Bike sharing differences among Millennials, Gen Xers, and Baby Boomers: Lessons learnt from New York City's bike share[J]. Transportation Research Part A-Policy and practice. 2018,116:1-11.
- [18]Liu, LX., Gong, DY., Guan, B., CF-Cluster: Clustering Bike Station Based on Common Flows[C]. 2017 IEEE 15th International Conference on Pervasive Intelligence and Computing. 2017:115-125.
- [19]薛强. 国人共享单车使用情况调查[J]. 金融博览: 财富, 2017 (2) : 24-26.
- [20]刘立巧. 共享单车现状调查分析与对策研究[J]. 交通科技与经济, 2017, 20 (1) : 42-44.
- [21]张雨宁, 严 捷. 共享单车市场现状及发展趋势分析[J].产业经济, 2018 (3) : 62-63.
- [22]刘硕秋, 叶国清, 邬博文, 周素素. 关于共享单车的时空分布研究与分析[J]. 中国高新区, 2018 (4) : 34.
- [23]赵曼. 共享单车网络分析及其优化调度研究[D]. 山东科技大学, 2017.
- [24]张俊英. 用户对 ofo 共享单车使用意愿和行为的影响因素研究[D]. 华中师范大学, 2017.
- [25]吴淑婷. 基于大数据下城市公共自行车管理模式的研究[D]. 徐州工程学院, 2018.
- [26]王宇洋. 基于大数据的共享单车移动趋势分析与预测[D]. 北京市育英学校, 2018
- [27]Álvaro, L., Paz, J. D., González, G. V., et al. Multi-Agent System for Demand Prediction and Trip Visualization in Bike Sharing Systems[J]. Applied sciences, 2018, 8 (1) : 67.
- [28]Chen, P. C., Hsieh, H. Y., Sigalingging, X. K., et al. Prediction of Station Level Demand in a Bike Sharing System using Recurrent Neural Networks[C]. IEEE Vehicular Technology Conference: vtc-spring, 2017: 1-5.
- [29]Zhou, Y., Wang, L., Zhong, R., et al. A Markov Chain Based Demand Prediction Model for Stations in Bike Sharing Systems [J]. Mathematical Problems in Engineering, 2018, 2018: 1-8.



- [30]Feng, C., Hillston, J., Reijsbergen, D., et al. Moment-based availability prediction for bike-sharing systems [J]. Performance Evaluation, 2018, (117) : 58-74.
- [31]Liu, Z., Shen, Y., Zhu, Y. Inferring Dockless Shared Bike Distribution in New Cities [C]. Eleventh Acm International Conference on Web Search & Data Mining. 2017.
- [32]Xu, CC., Ji, JY., Liu, P., The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets [J]. Transportation Research Part C: Emerging Technologies, 2018, (95):47-60.
- [33]Ai, Y., Li, ZP., Gan, M., A Deep Learning Approach on Short-Term Spatio-Temporal Forecasting of Dockless Bike Sharing System[J]. Neural Computing and Applications. 2018:1-13.
- [34]Li, YX., Zheng, Y., Yang, Q., Dynamic bike reposition: A spatio-temporal reinforcement learning approach [C]. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2018:1724-1730.
- [35]段文欣. 基于 BP 神经网络的共享单车时空分布模型[J]. 中国战略新兴产业, 2018 (4) : 23.
- [36]陶启萌. 基于聚类分析的共享单车时空分布模型[J]. 中国战略新兴产业, 2018, (4) : 61.
- [37]王忱. 基于 LM 神经网络的共享单车需求量分析模型[J]. 中国战略新兴产业, 2018, (4) : 145.
- [38]焦志伦. 大数据驱动下的共享单车短期需求预测——基于机器学习模型的分析[J]. 南开大学经济与社会发展研究院, 2018, (5) : 134.
- [39]Chemla D, Meulier F, Calvo R W. Bike sharing systems: Solving the static rebalancing problem [J]. Discrete Optimization, 2013, 10 (2): 120-146.
- [40]Tal Raviv, Ofer Kolka. Optimal inventory management of a bike sharing station [J]. HE Transactions, 2013, 45 (10): 1077-1093.
- [41]Forama I A, Raviv T, Tzur M. A 3-step math heuristic for the static repositioning problem in bike sharing systems [J]. Transportation Research Part B Mythological, 2015, (71): 230-247.
- [42]Contardo C, Morency C, Rousseau L M. Balancing a dynamic public bike sharing system [M]. Montreal: Cirrelet, 2012.
- [43]Caggiani L, Ottomanelli M. A dynamic simulation based model for optimal fleet repositioning in bike sharing systems [J]. Procedia-Social and Behavioral Sciences, 2013, 87:2013-210.



- [44] Alvarez-Valdes R, Belenguer J M, Benavent E, et al. Optimizing the level of service quality of a bike sharing system [J]. Omega, 2016, 62: 163-175.
- [45] Fricker C, Gast N. Incentives and redistribution in homogeneous bike sharing system with station of finite capacity [J]. Euro journal on transportation and logistics, 2016, 5 (3): 261-291.
- [46] Pal A, Zhang Y. Free-floating bike sharing: Solving real-life large-scale static rebalancing problems [J]. Transportation Research Part C: Emerging Technologies, 2017, 80: 92-116.
- [47] 秦茜. 公共自行车租赁系统调度问题研究[D]. 北京交通大学, 2013.
- [48] 张辉. 基于蚁群算法的城市公共自行车调度研究[J]. 科技与管理, 2015, 17(6):32-36.
- [49] 于文超. 城市公共自行车系统智能调度优化算法研究[D]. 上海交通大学, 2015.
- [50] 陈昕昀, 蒋永康, 李牧原, 等. 基于 B P 神经网络的公共自行车单站点调度需求量研究[J]. 交通标准化, 2016, 2(3): 30-35.
- [51] 刘路美. 城市公共自行车站点需求预测及调度优化方法研究[D]. 北京交通大学, 2017.
- [52] 刘兆仁, 徐冠宇, 尹航. 基于遗传算法的公共自行车调度优化[J]. 物流技术, 2017, 36 (2): 78-81.
- [53] 王璐, 李斌, 徐永龙, 马博超, 魏俊博. 基于共享单车数据的居民出行热点, 区域与时空特征分析[J]. 河南科学, 2018, 36(12): 2010-2015.
- [54] 范文婷. 城市交通方式间合作竞争机制研究[D]. 长沙理工大学, 2009.
- [55] 岑乐陶. 城市道路交通规划设计[M]. 机械工业出版社, 2006.
- [56] 罗赞. 城市公共自行车选择行为研究[D]. 长安大学, 2013.
- [57] 李玉. 城市共享单车选择行为及发展对策研究[D]. 西南交通大学, 2018.
- [58] 陈超, 高禹祺. 共享经济视角下的城市共享单车发展对策分析[J]. 产业与科技论坛, 2019, 18(15): 12-13.
- [59] 谷丛, 王菲. 城市公共自行车的“公共性”功能设计研究——以杭州为例[J]. 装饰, 2011(5): 110-111.
- [60] iiMedia Research. 共享单车人群分析报. <http://mi.talkingdata.com/report-detail.html?id=462>
- [61] 中商情报网. 共享单车用户分析 http://www.askci.com/news/chanye/20170601-/16421699416_2.shtml



- [62] 成都历史天气预报 http://tianqi.2345.com/wea_history/56294.htm
- [63] 苏影. 基于数据分析的共享单车动态调配优化研究[D]. 北京交通大学, 2019.
- [64] VOL. N. Inferring urban land use from satellite sensor images using kernel-based spatial reclassification[J]. Photogrammetric Engineering & Remote Sensing, 1996, 62(8): 949-958
- [65] HEROLD M, COUCLELIS H, CLARKE K C. The role of spatial metrics in the Analysis and modeling of urban land use change[J]. Computers Environment & Urban Systems, 2005, 29(4): 369-399
- [66] BANZHAF E, NETZBAND M. Monitoring Urban Land use Changes with Remote Sensing Techniques [M]. 2011: John Wiley & Sons, Ltd. 18-32
- [67] PEI T, et al. A New Insight into Land Use Classification Based on Aggregated Mobile Phone Data [M]. Vol. 28. 2013
- [68] 索超, 丁志刚. POI 在城市规划研究中的应用探索[C]. 新常态: 传承与变革 2015 中国城市规划年会, 贵阳, 2015, 634-643
- [69] Jing Yuan, Yu Zheng, and Xing Xie. Discovering Regions of Different Functions in a City Using Human Mobility and POIs[C]. KDD 12th Proceedings of 18th ACM SIGKDD international conference on Knowledge discovery and data mining, Beijing, China, 2012:186-194
- [70] Cibebe Q, Megan M, Kristina R, et al. Mapping bundles of ecosystem services reveals distinct types of multifunctionality within a Swedish landscape[J]. AMBIO, 2015, 44(1): 89-101.
- [71] 张慧杰, 王蓉, 陈斌, 侯亚芳, 曲德展. 基于轨迹和兴趣点数据的城市功能区动态识别与时变规律可视分析 [J]. 计算机辅助设计与图形学学报, 2018, 30(09): 1728-1740.
- [72] 蒙冰皓. 面向动态场景理解的时空深度学习算法[D]. 电子科技大学, 2017.
- [73] 曹敏杰. 浙江近岸海域海洋生态环境时空分析及预测关键技术研究[D]. 浙江大学, 2015
- [74] Ke, J. T., Zheng, H. Y., Yang, H., et al. Short-Term Forecasting of Passenger Demand under On-Demand Ride Services: A Spatio-Temporal Deep Learning Approach. Transportation Research Part C: Emerging Technologies, 2017, 85: 591-608.
- [75] Ai, Y., Li, Z. P., Gan, M., et al. A deep learning approach on short-term spatiotemporal distribution forecasting of dockless bike-sharing system [J]. Neural Computing & Applications, 2018, online.



- [76] Shi, X., Chen, Z., Wang, H., et al. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. *Neural Information Processing Systems* (NIPS2015), 2015.
- [77] WANG B, KIM I. Short-term prediction for bike-sharing service using machine learning[J]. *Transportation Research Procedia*, 2018, 34:171-178
- [78] CHEMLA D, MEUNIER F, WOLFLER CALVO R. Bike sharing systems: Solving the static rebalancing problem[J]. *Discrete Optimization*, 2013, 10(2):120-146.
- [79] Minsoo Cho, Jihwan Ha, Chihyun Park, Sanghyun Park. Combinatorial feature embedding based on CNN and LSTM for biomedical named entity recognition[J]. *Journal of Biomedical Informatics*, 2020, 103.
- [80] Chen Chen, Zhengchun Hua, Ruiqi Zhang, Guangyuan Liu, Wanhui Wen. Automated arrhythmia classification based on a combination network of CNN and LSTM[J]. *Biomedical Signal Processing and Control*, 2020, 57.
- [81] 史越. 共享单车需求预测及调度方法研究[D]. 北京交通大学, 2019
- [82] 金含笑. 基于摩拜出行数据的共享单车调度研究[D]. 北京交通大学, 2019.
- [83] 汪慎文, 徐亮, 杨锋, 李美羽. 基于蚁群算法的动态共享单车调度优化[J]. *南昌工程学院学报*, 2019, 38(03):71-76.