

智能交通领域前沿理论与应用论著

公共交通大数据 挖掘与分析



马晓磊 丁川 于海洋 刘剑锋◎著

GONGGONG
JIAOTONG
DASHUJU WAJUE YU FENXI



人民交通出版社股份有限公司
China Communications Press Co., Ltd.

智能交通领域前沿研究论著

Gonggong Jiaotong Dashuju Wajue yu Fenxi
公共交通大数据挖掘与分析

马晓磊 丁川 于海洋 刘剑锋 著



人民交通出版社股份有限公司
China Communications Press Co., Ltd.

内 容 提 要

本书为智能交通领域前沿研究论著丛书之一,主要针对国内多模式公共交通系统信息化和智能化发展现状,基于广泛使用的公共交通大数据(IC卡和GPS数据),探讨了公共交通大数据的分析和挖掘方法,并综合利用人工智能、运筹学、交通规划、交通行为学等理论和方法,对公共交通大数据典型应用场景进行探讨,用于支撑城市多模式公共交通规划和运营的目的。主要研究内容包括:交通大数据基本类型与处理技术、公共交通客流OD矩阵推导方法、公共交通到站信息服务优化、公共交通短时客流预测、公共交通出行行为分析、公共交通网络性能评价。本书可为全方位、多层次、大范围、大数据化、高效运行的综合交通运输体系建设提供理论与技术支持。

本书适用于交通工程、智能交通、交通管理等相关专业的技术人员、管理与决策人员、科研工作者、研究生、教师与高年级本科生使用。

图书在版编目(CIP)数据

公共交通大数据挖掘与分析/马晓磊等著. —北京:
人民交通出版社股份有限公司,2017.12
ISBN 978-7-114-14110-2

I. ①公… II. ①马… III. ①公共交通系统—数据处
理—理论 IV. ①U491.1

中国版本图书馆CIP数据核字(2017)第204153号

书 名:公共交通大数据挖掘与分析
著 者:马晓磊 丁 川 于海洋 刘剑锋
责任编辑:李 良
出版发行:人民交通出版社股份有限公司
地 址:(100011)北京市朝阳区安定门外外馆斜街3号
网 址:<http://www.ccpres.com.cn>
销售电话:(010)59757973
总 经 销:人民交通出版社股份有限公司发行部
经 销:各地新华书店
印 刷:北京市密东印刷有限公司
开 本:787×1092 1/16
印 张:10.75
字 数:248千
版 次:2017年12月 第1版
印 次:2017年12月 第1次印刷
书 号:ISBN 978-7-114-14110-2
定 价:30.00元

(有印刷、装订质量问题的图书由本公司负责调换)

序

随着我国城市规模的不断扩大、机动车保有量与居民出行需求的急剧增加,城市交通拥挤和尾气排放等问题愈发突出,已成为我国诸多大中型城市可持续发展过程中亟待解决的关键问题。国内外对城市交通的实践经验表明,优先发展城市公共交通,是解决城市交通问题的有效途径。特别是近年来,城市公共交通在城市交通战略地位的不断提高,构建服务水平更高、运营效益更好的城市公共交通系统是城市交通发展的迫切需求。

随着大数据时代的来临,智能公共交通 IC 卡、GPS 定位和移动互联等信息技术在公共交通领域深度应用。在公共交通运营过程中所产生的数据量非常庞大,这些数据具有连续性好、覆盖面广、信息全面且动态更新快的特点,无论是在公共交通线网规划,还是在公共交通运营管理上都有着非常大的应用价值。可以说,大数据在公共交通领域的应用是一次深刻的变革,它将对公共交通行业的服务理念、管理机制、运营模式等产生革命性的转变,其具有深远的理论与现实意义。

马晓磊副教授等人在国家自然科学基金课题“数据不完备条件下基于个体行为的大规模公交网络出行矩阵估计研究”(课题编号:51408019)和“考虑空间异质性与中介效应的建成环境对通勤出行方式选择的影响研究”(课题编号:71503018)等的支持下,对公共交通大数据挖掘与分析进行了深入而细致的研究。这部学术著作是马晓磊副教授等人系列研究成果的系统总结,从交通大数据的处理技术、公交客流推断与服务优化、公交网络性能评价等几个方面,结合实际城市案例数据,系统阐述了公共交通大数据挖掘与分析的方法。希望本书能够为致力于交通大数据研究的科研人员提供有益的参考。

这部学术著作是公共交通大数据研究成就的重要见证。著者们是交通领域充满激情和富有远见的青年学者,其在公共交通大数据挖掘与分析等方面的理论研究及实践必将有利于我国交通大数据研究水平的不断提高,成为该领域重要的学术参考资料。

中国工程院院士

施仲衡

前 言

近年来,随着智能交通、智慧城市等概念的出现与不断推进,城市交通领域已经具有较为成熟的数据采集和技术应用基础。随着大数据时代的来临,以手机数据、浮动车数据及社交网络等为代表的海量、多源时空数据不断应用于城市规划实践中。2015年国务院印发的《促进大数据发展行动纲要》系统地部署了大数据发展工作,提出率先在交通等领域,实现公共数据资源合理适度向社会开放,带动社会公众开展大数据增值性、公益性开发和创新应用,充分释放数据红利。

促进大数据发展,为“互联网+”时代下城市交通提供了发展新思路:一方面,大数据技术的应用为提升公共服务水平、促进产业创新提供了新的机遇;另一方面,大数据技术也为政府提升交通治理能力、缓解城市交通资源压力提供了新途径。随着大数据时代的来临,城市公共交通 IC 卡的推广与卫星定位技术的使用,在为广大乘客提供便利的同时,也提供了信息化的客流调查统计方法。基于海量多源公共交通数据提取并分析公共交通客流信息,可为公共交通企业的决策、运营管理提供可靠的数据依据,从而提升公共交通企业的管理水平,提升公共交通吸引率,推进公共交通优先发展战略实施,缓解城市交通拥堵。

本书系统构建了公共交通大数据在运营与规划研究中的分析框架,以实际城市作为案例全面介绍了公共交通大数据挖掘与分析方法,为公共交通规划与运营提供数据支撑及技术方法,主要包括交通大数据基本类型与处理技术、公共交通客流 OD 矩阵推导方法、公共交通到站信息服务优化、公共交通短时客流预测、公共交通出行行为分析、公共交通网络性能评价等几方面的应用。希望本书能够给更多致力于公共交通大数据研究的科研人员提供有益的参考,促进公共交通理论与实践的发展。

本书由马晓磊、丁川、于海洋和刘剑锋撰写,其中马晓磊主要撰写了第 1~3 章,丁川主要撰写了第 5、6 章,于海洋主要撰写了第 4 章,刘剑锋主要撰写了第 7 章。博士研究生陈汐、代壮,硕士研究生刘从从、陈栋伟、栾森、张继宇、杨洁、段金肖、李屹等参与了撰写工作,在此一并表示感谢。

本书的出版得到了国家自然科学基金(课题编号:61773036, 71503018, 51408018, U1564212)、北京市自然科学基金(课题编号:9172011)、北京市科技新星(课题编号:z151100000315048)以及中国科协“青年托举”人才工程(课题编号:2016QNRC001)等项目的支持。

著者

2017年9月于北京

目 录

第1章 绪论	1
1.1 背景与意义	1
1.2 国内外研究现状	3
1.3 公交大数据分类	12
1.4 本书主要内容	16
1.5 章节结构	18
本章参考文献	19
第2章 交通大数据基本类型与处理技术	25
2.1 交通大数据基本类型	25
2.2 数据处理工具与数据清洗	32
本章参考文献	38
第3章 公交客流 OD 矩阵推导方法	39
3.1 上车站点识别	39
3.2 下车站点识别	49
本章参考文献	61
第4章 公交到站信息服务优化	62
4.1 公交到站信息概况	62
4.2 公交到站信息影响因素分析	62
4.3 基于相关向量机的乘客候车时间范围预测	63
4.4 基于 LS-SVM 的公交串车预测	69
本章参考文献	82
第5章 公共交通短时客流预测	83
5.1 考虑动态波动性的轨道交通站点短时客流预测	83
5.2 基于多尺度径向基函数神经网络模型的地铁短时客流预测	90
5.3 预测地铁短时客流并基于梯度提升算法的决策树模型分析客流影响因素	108
本章参考文献	117
第6章 公交出行行为分析	120
6.1 乘客时空规律性挖掘	120
6.2 通勤人群识别与职驻地估计	125
6.3 通勤指标分析	132

6.4 城市通勤出行方式与出发时刻联合选择的交叉巢式 Logit 模型	134
本章参考文献	145
第7章 公交网络性能评价	147
7.1 交通 E 科学理论框架	147
7.2 公交大数据平台搭建	152
7.3 本章小结	162
本章参考文献	162

第1章 绪 论

1.1 背景与意义

随着城市规模不断扩大、机动车保有量与居民出行需求的急剧增加,导致城市道路交通流量日趋饱和,城市交通拥堵频繁发生,城市交通问题已成为我国许多大中型城市发展过程中亟待解决的问题^[1]。国内外对城市交通的实践经验表明,优先发展城市公共交通,是解决城市交通问题的有效途径。特别是近年来,城市公共交通在城市交通战略地位的不断提高,构建服务水平更高、运营效益更好的城市公交系统是城市交通发展的迫切需求。

近年来,随着智能交通、智慧城市等概念的出现与不断推进,城市交通领域已经具有较为成熟的数据采集和技术应用基础。随着大数据时代的来临,以手机数据、浮动车数据及社交网络等为代表的海量、多源时空数据不断应用于城市规划实践中。2015年9月5日,国务院印发《促进大数据发展行动纲要》,系统地部署了大数据发展工作,并提出:率先在交通等领域,实现公共数据资源合理适度向社会开放,带动社会公众开展大数据增值性、公益性开发和创新应用,充分释放数据红利。因此,促进大数据发展,为“互联网+”时代下城市交通提供了发展新思路:一方面,大数据技术的应用为提升公共服务水平、促进产业创新提供了新的机遇;另一方面,大数据技术也为政府提升交通治理能力、缓解城市交通资源压力提供了新途径^[2]。

一直以来,公交系统的供需关系分析及运营规划方面的工作基本以人工调查数据作为主要的数据来源。我国大中城市建设和发展日新月异,出行需求在时间和空间上呈现出差异大、变化快的特征^[3]。因此,通过传统的交通调查来获取全面的公交客流时空分布数据是非常困难的,传统的调查方法往往通过问卷或问询的方式来获得居民的出行信息,而这种方式获取的居民出行数据一般在样本选取的数据和质量上满足不了要求,存在调查样本比例小、精度低、成本高、时效性差、调查周期长、后期数据处理困难等问题,而且无法反映乘客出行需求的动态变化特征。此外,大型交通调查通常是五年才会进行一次,为此想要通过传统的交通调查方法来获取实时的公交供需及整体运行情况并不现实^[1]。因此,获取实时、全面的公交供需数据是城市公交系统科学决策和发展的迫切要求,也是提升公共交通规划与管理运营决策水平与效率的需要。

随着大数据时代的来临,以及城市公交 IC 卡的推广与卫星定位技术的使用,在为广大乘客提供便利的同时,也提供了信息化的客流调查统计方法。在公交运营过程中所产生的数据量非常庞大,这些数据具有连续性好、覆盖面广、信息全面且动态更新快的特点,无论是在公交线网规划,还是在公交运营规划上都有着非常大的应用价值。例如,北京公交集团运营的 1020 条线路将通过大数据进行线路优化,根据客流情况安排运力,减少乘客等待时间,

提高运营效率;北京公交集团目前已在所有公交车上安装了GPS,通过数据共享,可以通过手机APP进行公交查询,未来将更多基于“互联网+”提升公交服务水平。总体来说,依托大数据的迅速发展,一方面,IC卡、GPS数据、视频监控数据等公交大数据在数据仓储、挖掘与处理技术层面有着广泛的研究及应用前景,如分布式处理框架,包括分布式存储和计算;高效的实时数据处理,即根据实时分析数据来随时掌握公交车辆的运行状况,为实时决策提供依据;数据类型的多样化,即满足结构化数据、半结构化数据、非结构化数据的融合关联分析,实现文本分析、图形分析、空间分析等功能,为公交系统决策者提供不同角度不同形式的分析判断依据^[4]。另一方面,“大量的数据胜于好的算法”,大数据对于公交运营与规划中的优化问题,不仅可以改变传统的建模求解的优化思路,即将模型驱动研究转向数据驱动研究,还可以将成果应用于更人工智能化的运营调度系统及平台。传统公交的行车计划或调度安排由车队调度组长手工编制而成,全部依靠静态的客流数据和调度组长的个人经验,技术手段落后。同时这种运营调度管理模式时效性差,调度方式很不灵活,运营计划调整周期过长。此外,无法考虑实时客流量,乘客所获得的信息很少,因此,无法真正发挥城市公共交通汽车在缓解城市交通拥堵方面的作用^[5]。通过城市公交静态和动态数据,包括公交站点、公交线路数据、城市居民刷卡数据、卫星定位数据以及城市公交调度信息等数据,识别城市居民时空路径信息,并结合城市用地数据进行分析,包括城市公交流量分析、城市职住关系分析、公交运行时间分析及公交服务水平分析,可为公交线网调整优化及运营调度提供数据支撑及规划建议,以有效地缓解高峰时段客流拥挤情况,节约乘客出行等待时间,实现对社会资源的合理利用,并进一步提高公共客运交通的吸引力,为公交企业进行科学的运营决策提供理论依据和技术支持,保障城市交通的协调、有序和高效发展^[6]。

综上所述,大数据在公交领域的应用是一次深刻的变革,它将对公交行业的服务理念、管理机制、运营模式等产生革命性的转变,具有深远的理论与现实意义。

理论方面,首先,基于海量多源公交数据提取并分析公交客流信息,是对现有问卷调查这种传统数据获取方式的变革,是丰富居民公交出行数据来源的重要措施。第二,公交动态信息采集及特征分析技术是对现有的动态交通信息采集技术的补充和完善。第三,可以优化公交调度系统,使得公交运营者根据公交到站时间调整发车间隔、行车时刻表等,提高运营效率和服务质量。

现实方面,首先,提取并分析居民公交出行数据,可为公众出行提供公共交通信息服务,为公共交通管理部门提供数据支撑和决策支持,缩短居民公交出行时间,提高公共交通运营管理水平,实现公共交通社会效益与经济效益最大化。第二,实时、准确地识别公交运行状态是实现先进的公共交通信息服务系统的前提,这就需要获得实时、充分、准确的公众出行动态客流及公交运行信息。通过大数据的相关技术可改善现有的数据获取方式单一,耗时、耗资、耗力、调查信息不完整等问题,进而获取充分的公共交通动态客流信息。第三,通过对公交出行数据的深层次挖掘与分析可以得到可供发布的高效、准确、实时的公共交通出行服务信息,为公众出行服务提供多方位、多层次、多渠道的交通出行综合信息服务与支持。第四,可以为公交企业的决策、运营管理提供可靠的数据依据,从而提升公交企业的管理水平,提升公交吸引率,推进公交优先发展战略实施,缓解城市交通拥堵。

目前还少有专著对城市公交大数据在实践中的应用进行系统的梳理和总结。本书从公

交大大数据获取、处理到与其他数据结合,对公交大数据在乘客出行规律的分析以及公交运营中的应用进行梳理,结合作者的已有研究成果,介绍公交大数据在公交客流 OD 矩阵推导方法、公交到站信息服务优化、地铁短时客流预测、公交出行行为分析、公交网络性能评价等几方面的应用,并以不同城市作为实证研究案例,构建公交数据在运营与规划研究中的分析框架,总结公交大数据在实际问题中的应用,为公交规划与运营提供数据支撑及技术方法更新。

1.2 国内外研究现状

随着智能公交的发展,智能公交系统可以采集、传输、存储、处理大量的公交信息的相关数据。在大数据时代,如何利用智能公交系统的海量数据,通过高效的数据处理方式以获得正确、无噪声的有效数据,从中提取有效的基于公交客流、车辆信息,对其进行深层次的数据挖掘,使分析及预测的相关结果能够应用于公交的运营、规划与管理中一直是研究的热点之一。在此基础上,可以有效利用已有数据构建优化模型,提出相关的公交运营与管理措施,从而使得公交的出行满意度、公交的服务水平、公交运营的可靠性保持在较高的水平,以提高公交吸引率,进而对未来公交的运营调度进行决策指导。

公交数据有着特有的处理与分析方法,国内外学者在这方面已经取得了不俗的研究成果。下面着重对公交大数据在公共交通规划与运营两方面的研究现状进行梳理。

1.2.1 公共交通规划

1) 线网优化分析

城市公交线网规划与设计是一项较为复杂的工作,除了受到城市路网结构的影响,还与客流需求、车队规模、交通条件等各种因素相关。城市公交线网优化是解决目前城市交通系统存在的诸多问题的有效方法之一,科学合理的公交线网将有助于解决以下问题^[7]:

①科学合理的公交线网,对于提高公交系统运营效率、减少换乘次数、缩短居民出行时间等方面产生巨大影响。

②科学合理的公交线网,也是城市改善其公交服务水平,实现城市可持续发展的重要前提之一。

近年来,公交线网优化建立起了以公交出行 OD 与图论为理论基础的体系结构,体系十分完善,取得了十分丰富的成果。利用图论等经典理论使得优化具有直观性、准确性和科学性等特点,是线网优化相对成熟的体系,并被广泛地应用到实际生产中。然而,随着数据量的不断增加,这种算法的缺点也逐渐暴露出来。这类算法的计算量大,在数据量很大时效率极低,因此多数的研究仅局限于小区域的示例验算,没有大规模的场景应用^[8,9]。因此如何进行大规模的城市公交线网优化,并在整个城市路网中应用,不断有学者利用新方法和模型来进行求解。随着城市规模扩大,改善城市交通发展的不均衡,交通拥堵严重等现象,各大城市都提出相关政策积极进行公交线网的优化。因此如何进行大规模的实际线网优化是线网优化的最亟待解决的问题。

实际上,公交线网优化就是一个典型的 NP-Hard 问题^[10]。其本质就是在已知路网变

量、客流量和公交线网信息的前提下,按照一定的约束条件、优化目标来确定合理的公交线路,达到最终的交通需求。Ceder 对公交线网设计进行了系统的归纳,提出了线网优化的四个基本组成部分:需求特征,目标函数,约束条件和算法实现^[11]。公交线网设计是一种乘客出行的虚拟模型,利用数学规划的基础知识设计理想的线网走向^[12]。通过大量的理论和实践研究,可以将公交线网优化问题总结成三类:①以实际应用经验为基础的设计;②在理想假设情形下的最优化设计;③启发式算法的现实网络实际。表 1-1 总结了目前关于公交线网设计优化的研究。下面,本节将从优化目标、决策变量以及实现算法三个方面对目前的线网优化的研究进行简单的总结。

线网优化研究总结

表 1-1

作者	目标函数	决策变量	求解算法	应用场景
Daganzo ^[35]	总成本 (企业 + 乘客)	发车间隔、公交线间距、 公交线网覆盖率	解析法	算例
Estrada ^[36]	总成本 (企业 + 乘客)	站间距、水平与竖直方向 的线路间距、水平和竖 直方向中心区与城市总长 度比、发车间隔	解析法	巴塞罗那
Nourbakhsh ^[37]	总成本 (企业 + 乘客)	线网几何形状、灵活公 交的覆盖区域形状、发车 间隔	解析法	算例
Badia ^[38]	总成本 (企业 + 乘客)	中心区辐射线的角度、 中心区辐射线的站间距、 环行线的长度、环行线的 站间距、中心区域半径、发 车间隔	解析法	算例
王振报 ^[39]	企业、乘客、政府	公交主线线间距、快线 站距、线网覆盖区域大小	解析法	算例
Chen ^[40]	总成本 (企业 + 乘客)	线网几何形状、发车 间隔	解析法	算例

(1) 优化目标

公交线网优化设计的目标是以最小的成本投入实现公交效益的最大化。而这些目标根据设计目的的不同而有很大差异。总结目前的优化模型,最终的优化目标通常包括乘客利益最大化、运营利益最大化、总体利益最大化、节能减排最优化以及特定优化目标最大化中的一项或者几项。Ceder 等在线网优化中以乘客的时间和费用以及公交服务供给的最大化作为目标^[11]。Nes 等通过研究发现,乘客出行费用与运营商费用的总和最小是一种有效的优化目标^[13]。Fan 等指出线网优化的最终的目标应该是在有限的线网资源下使得公交发挥最大服务功能,实现利益最大化^[12]。随着时代的发展和绿色交通的提出,越来越多的学者

将环保的理念引入到公交线网的优化中,作为一个十分重要的指标^[14-16]。

(2) 决策变量

在线网优化中最常用的两类变量:决策变量和其他变量。其中决策变量中一些关键性的指标,直接决定了公交网络形态、线路的运营策略核心,通常包括公交线网的拓扑结构、公交车的时刻表以及公交的运载能力等核心信息。研究初期,决策变量的确定通常是在固定的公交线网结构中,求解最优化的线路走向以及发车的频率^[17,18]。随着研究的深入,优化线网的拓扑结构以及服务频率的优化成为研究的热点,同最小费用一起被引入了决策变量的确定过程中^[19,20]。在研究过程中,其他变量虽然不能直接决定线网的形态,但是却能够间接反映公交的运营环境、功能特征以及服务水平等,比如公交的运营模式、公交政策等。近年来,灵活公交、动态优化等理念的提出,要求公交的服务更加细化,同时也对研究提出更高的要求^[21]。Chien 等人提出在线网优化过程中的一些关键变量,比如出行时间、出行费用、发车间隔等能够灵活地引导公交乘客的出行^[22]。Lee 等人认为公交的线网优化应该以乘客的总出行需求为根本依托^[23]。Quadrioglio 等人以洛杉矶为示例研究动态的线网优化方法^[24]。而大站快车的灵活运营方式是当前研究的一个热点问题^[25]。

(3) 实现算法

目前,对于公交线网优化的算法大致可以分为两类。一种是传统依托分析模型的解析式算法,另一种是融合了人工智能以及计算机技术的启发式算法。对于传统的解析式算法,在研究线网间距、线路长度、车速、运载能力、发车间隔等因素的内在联系上应用广泛^[26-28]。这种算法适用于路网场景简单、较大范围的路网优化、偏重于公交的线网设计,并不注重于具体线路的走向。而启发式算法利用强大计算能力,根据经验构造算法,利用计算机进行实际的庞大空间网络的建立优化。对于解决现实的路网的公交线路走向、发车间隔等变量的最优化实际有很好的适用性。在公交线网优化方面,常用到的启发式算法有遗传算法、蚁群算法、禁忌算法以及多种算法的组合^[29-32]。

总结过去的研究成果,目前关于公交线网设计优化的研究大致可以分为两大类:第一类是单纯的理论研究,尝试利用单纯的定量模型解决公交线网的优化问题。这些模型大多以出行时间和距离、客流量等为指标,利用数学规划的理论方法建立模型进行线网优化。这类方法的缺点是目标函数比较单一,不能用于复杂的网络优化。随着研究的深入,很多复杂的理论模型被提出^[33,34],设计的线网也相对复杂,考虑问题也相对全面。因为理论性太强,因此这类方法往往设计很多假设条件,与实际的线网设计存在一定的差异,很难在实际中大规模地应用。另一类方法是以定性分析为主辅以定量分析。在线网优化设计过程中,一部分变量根据传统建模方法确定。在定线时,选取与客流方向一致的线路走向,通过线路长度、运载能力等约束条件的限制,提出乘客量最大化、客流均匀分布等目标,较为实用。而这类方法的缺陷就是没有考虑例如公交共线等复杂条件。

通过以上的综述可以看出,目前常用的两种线网优化思路都有自己固有的缺陷。第一类方法理论性过强,在计算过程中有大量的假设条件,虽然结论比较理想,但是应用于实际的难度过大。实际的公交线网情况复杂,很难符合模型的许多数据假设。而第二方法更多偏向于实际应用,因此,有很多参数的确定依靠实际经验,缺少理论支撑,因此,难以应用和解决复杂的问题。如何将理论与实际相互结合是目前公交线网优化的亟待解决的问题。而

随着信息技术的发展,越来越多的数据在运营过程中被收集起来。同时计算机技术迅速发展,人工智能和大规模计算理论方法越来越成熟。因此结合大数据与先进的理论算法,能够对公交线网进行更好的优化设计。利用数据进行网络优化紧紧依托实际,有十分重要的现实意义,同时先进的算法能够解决复杂的网络问题,充分地掌握内部规律。结合大数据挖掘的公交线网优化是未来研究的新方向。

2) 行为分析

(1) 国外出行行为分析研究现状

国外对于城市居民出行行为的研究开展较早,一般认为是起源于20世纪40年代的美国,早期的城市居民出行调查主要是为了给城市道路的建设提供参考信息,大多采用入户访谈调查的方式^[41]。调查内容主要有出行起始点、时间、出行终点、出行方式、人数以及其他信息。美国从1943年到1958年期间就已经对超过100个城市的居民出行情况进行了统计调查,1958年,Curran和Stegmaier对其中50个城市的居民出行调查资料进行了分析^[42]。最初的出行行为分析方法,一般是利用集计的数据,对一些简单的、只有几个参数的模型进行估计,从而达到对个体行为特征进行预测的目的。但是,随着人们出行模式、价值观念以及客观条件的改变,影响出行行为的各种因素之间的关系日趋复杂,传统的出行行为分析方法已不能清楚解释这些关系^[43]。因此,交通研究者们开始探讨新的方法,对个体出行行为特征进行分析。

传统的集计模型由于其集计的特点无法描述个体的出行行为,只能描述交通小区或者一个整体的出行特征,所以非集计模型得到了发展。非集计模型描述了决策者在不同的方案之间所做出的选择,因而所建立的模型更加贴近现实,更能确切地反映研究对象的行为特征和规律。1959年Luce等人首次对非集计模型中典型的logit模型进行推导^[44],随后Marshak和Suppes等人对Logit模型的一些理论基础进行了完善^[45],但最初的logit模型只能用于预测两种交通方式的选择,并且当时模型主要用于出行时间和出行费用之间的平衡分析。1974年McFadden对Logit模型及其特性进行了完整的论述^[46],逐步形成了非集计模型的理论体系,其中包括了MNL(Multinomial Logit)和NL(Nested Logit)等模型,以及对行为的解释和预测。Ben-Akiva等在20世纪70年代利用经济学的消费者行为理论,对非集计理论作了进一步完善,其中较为有代表性的是1985年Ben-Akiva出版的《Discrete Choice Analysis》一书^[47],随后,在Logit模型的基础上又出现了更加复杂的PCL(Paired Combinatorial Logit)模型、MNP(Multinomial Probit)模型、CNL(Cross-Nested Logit)模型等,使非集计模型的理论体系更加系统和完善。

到了20世纪90年代后,随着GPS技术在民用方面的大力发展,一些学者开始将GPS技术引入居民出行行为的调查当中,进行了一些相关的实验和研究。如美国肯塔基州的列克星敦市,通过征集志愿者,在其私家车上安装GPS和PAD结合的装置进行家庭出行调查,出行者在PAD上人工输入出行终点和车上人数等信息,而GPS装置则自动收集时间、出行位置等信息^[48]。最后由科学家进行信息的核对和匹配,并进行统计分析。这里虽然引入了GPS技术,但是由于接受调查的家庭数目较少(100户不到),不具有代表性,只能作为一种探索实验。随着GPS技术的进一步发展,GPS数据凭借易于获取、数据量大、精确度高等特点,得到越来越广泛的应用,一些研究开始考虑全面地采用GPS调查来研究居民出行行为,

从而代替传统的调查方式。

在韩国实行的分段计价智能卡地铁费和公交费可以合并支付,其不仅能记录公交的上下车站点,还能记录乘客的换乘次数^[49],从而准确获取出行者的换乘信息。Jang^[50]利用韩国智能卡数据条件优势,关注了乘客公共交通出行时间和换乘特征的分析,绘制了行程时间空间分布图,分析了换乘时间、换乘需求等空间分布,为提升公共交通线网服务水平提供参考。Seaborn等^[51]利用伦敦市公共交通智能卡数据,重点研究了公交与地铁、公交与公交及地铁与公交三种换乘关系间出行阶段连接的时间阈值,完成完整的出行链连接,并在此基础上,对每日公共交通出行链总数、每人每天出行链数、每条出行链包含的出行阶段数、地铁与公交混合出行数等进行了统计分析。Morency等^[52]基于加拿大魁北克市一个公交公司的智能卡数据,提出了面向对象的数据处理方法模型,分析了刷卡数据的时空分布特征,包括观测时段内公交站点的使用数量和使用频率变化特征和通过聚类分析的不同类别持卡者公交出行时间维度变化特征。

(2) 国内出行行为分析研究现状

20世纪80年代是我国国内对于居民出行行为研究的起始阶段,主要通过纸质形式的调查报告对城市居民出行行为进行分析,在借鉴了国外经验的基础上,探讨我国城市居民出行信息在交通规划中的应用。由于这一时期的研究工作大多处于探索阶段,因此研究基本基于传统的四阶段法,是单个的、独立的,以城市居民个体研究和定性分析为主,表象性地描述居民出行行为特征的各构成元素^[53]。另外,这一时期也开始有学者对城市规模和居民出行行为之间的关系进行了研究^[54]。

到了20世纪90年代,随着改革开放的顺利进行,我国城市化建设开始加速,城市交通规划部门越来越意识到城市居民出行行为研究工作的重要性。1991年李旭宏根据城市居住环境的差异,将城市划为三类:中心区、中间区以及城市外围区,他认为不同的地理区域属于不同的交通区,其出行环境存在差异,具有不同的出行分布状况。另外,他还研究建立了不同出行目的条件下的出行分布模型,指出居民出行分布模型的建立应该考虑城市居民出行目的以及居住区域的差异^[55]。随着社会的不断向前发展,对城市居民出行行为的研究变得更加深入,一些学者开始探讨城市居民出行行为背后深层次的原因,比如出行目的选择分析、出行方式选择心理等。1994年东南大学的杨涛、王琳等人以马鞍山市的城市居民为研究对象,通过大量咨询调查,再结合已有的资料,研究了该城市居民出行行为背后的心理学因素。杨涛等人指出,在居民出行选择决策,如出行方式、出行时间、出行目的等背后实际反映了居民的心理考量^[56]。2005年上海交通大学隽志才对基于活动链的出行需求预测方法进行了综述^[57],介绍了基于活动链出行预测理论的发展过程,阐述了基于活动链出行需求模型系统的框架和预测过程,指出了目前模型系统存在的问题,探索了改进出行需求预测方法的途径。

步入21世纪以后,随着现代信息技术的变革与兴起,大数据环境下的交通出行行为及基于行为实验的出行行为分析正成为该领域发展的前沿热点问题。在国内的研究人员中,陈学武等人对公交IC卡数据的采集、分析、应用进行了研究,提出了基于IC卡数据的公交客流与出行信息分析方法,分析了乘客出行起终点和换乘点的判断规则,并探讨了公交IC卡数据在发车间隔获取和线路客流预测方面的应用^[58]。东南大学的戴霄在基于公交IC卡

信息的公交数据分析方法研究中,结合下车站点与上车站点的站距和下车站点附近的土地利用性质,提出了基于站点吸引的下车站点判断方法;并归纳总结出用于公交 IC 卡数据分析的原始数据采集方法和获取途径,提出公交 IC 卡数据分析系统的框架结构^[59]。东南大学郭婕利用公交 IC 卡数据,通过将公交线路的相邻站点合并成站点区段,提出了基于 IC 卡数据的通勤乘客上车和下车站点区段的分析方法,进而获取通勤乘客的 OD 信息^[60]。华南理工大学的章威提出了利用公交车 GPS 定位与乘客 IC 卡刷卡信息联合分析来获取公交 OD 量的思路。通过公交车的 GPS 定位数据和 GIS 电子地图中的车站位置信息能够准确地获取公交车的到站时间,减小了以往根据行程时间分析到站时间所造成的误差^[61];华中科技大学的戴维在分析现阶段公交出行调查方法优缺点、研究公交客流出行时间分布特性的基础上,运用相关数据融合理论对公交 IC 卡数据及公交车 GPS 信息进行互补融合,获取公交客流起讫点信息^[62]。

综上所述,出行行为相关研究发展至今,各国学者在该领域都取得了一定的科研成果。但是,通过对居民出行行为的研究综述,本书发现目前研究中还存在有以下几点不足:

①大部分的研究只是针对通勤出行影响因素中的一类或者几类进行分析,缺少对影响因素的全面的、系统的梳理。

②目前的研究中,出行行为分析模型主要集中在居民出行方式选择问题上,对于出行时间影响模型的研究不足,同时也缺少对交通出行的联合选择模型的研究。

③对居民出行链的应用研究主要集中在出行链模式选择及出行行为选择方面,利用出行链分析居民出行特征的相关研究成果比较少,且缺乏针对公共交通出行链的研究。

④在利用多源数据分析居民公共交通出行特征时,大部分研究只分析了使用单一模式出行(如仅乘坐公交或仅乘坐地铁)的乘客的出行行为,而缺少对采用多模式出行的乘客的出行行为研究,同时无法获取到较完整的出行 OD 分布。

⑤在利用公交 IC 卡数据获取客流信息的研究中,均要面对国内目前公交 IC 卡数据上下车站点信息缺陷的问题,虽然很多学者建立或提出了多种推断模型和方法,但当进行大规模数据处理时,由于其中一些参数的确定较为复杂,不利于公共交通出行特征系统化的提取与分析。

1.2.2 公共交通运营

1) 车辆调度优化分析

公交调度是公交运营管理的核心内容,实现公交调度智能化和调度方案的优化是提高运营调度水平、增强公交吸引力的关键。调度优化问题的研究,主要包括静态调度与动态调度两个方面。

对于静态调度,初期是以成本函数为目标的优化^[63],Hurdle 提出了一个总成本费用函数(包含了运营商和出行者的费用),得到了多条相似接运公交线路的优化位置及其发车间隔^[64]。Yan 等人综合考虑乘客需求和车辆数,以运营商效益最大化为目标,建立了用于优化市内公交发车时刻表的模型^[65]。Bookbinder 等人从统计学角度,分析了公交车辆延误分布的统计学规律,并以乘客候车时间最短为目标建立优化模型,分别对高密度和低密度公交线路以及延误对策进行了讨论,并推广到同一站点的多条线路上^[66]。Haghani 等人对于线路行程时间进行约束,对大规模公交车辆调度问题进行了研究^[67]。Zhao 等人基于乘客的费用

最小,优化公交车辆的发车间隔,设计了适用于大规模网络的求解算法^[68]。陈芳等人以乘客等待时间和公交公司车辆出行总公里数最小为目标建立公交调度模型,并进行求解^[69]。

随着出行者对公交出行质量的关注,公交调度优化问题的目标也逐步转为对公交服务质量的优化。宋瑞等人建立了基于随机期望值规划的优化模型,该模型综合考虑了公交企业的经济效益和公交乘客所得到的公交服务水平^[70]。陈茜、牛学勤等人通过加权的数学方法对公交线路发车频率进行优化^[71]。Szeto 和 Wu 同时优化香港地区的公交线路及发车频次,可使乘客换乘次数和旅行时间分别降低 20.9% 和 22.7%^[27]。吴丽荣考虑乘客的“有限理性”心理和行为特征,构建柔性路径公交实时调度模型,并基于元启发式算法求解^[72]。邱丰等针对可变线路式公交,设计了一种可同时处理预约需求和实时需求的两阶段车辆调度模型。公交系统的管理智能化使得基于大数据的公交调度研究也逐渐增加^[73]。穆礼斌基于公交客流数据,以企业运营成本和乘客候车成本最小为目标函数,求解公交线路的发车时间间隔^[74]。赵骞基于公交 IC 卡数据的规律和特点,以乘客满意度和企业满意度最大的目标函数,提出了以发车时间参数为变量的公交调度优化模型^[75]。郭淑霞提出了公交 IC 卡数据质量控制方法,建立了基于时变二源数据的公交调度协调模型^[76]。沈吟东等提出一种基于 GPS 数据的改进 K-means 聚类算法对公交运营时段进行合理划分,以提高公交优化调度效率^[77]。

对于动态调度,其与静态模型的最根本的区别是考虑了乘客需求的不确定性及公交运行的不确定性^[63]。动态调度主要研究公交车辆运行中突发事件、交通事故等不确定性的影响,通过实施行车控制、发车控制等控制策略,保证公交车辆能够按照时刻表计划运行。主要的控制策略包括:滞站控制、速度控制和越站控制^[78]。Chowdhury 等提出了基于换乘站的车辆动态调度问题,通过控制换乘站车辆的停留时间和发车时间,减少乘客换乘时间,以达到提高换乘服务质量的目的^[79]。Zhao 等人通过最小化平均候车时间,包括车外时间费用、在车时间费用,建立了公交实时协调控制模型^[80]。Dessouky 等人对车辆在换乘站滞站时间的实时优化控制方法进行了研究^[81]。Zolfaghari 基于实时信息,以线路各站点乘客平均候车时间最小为优化目标,建立了车辆滞站时间模型^[82]。

综上所述,目前对于公交调度优化的研究大多集中于确定性问题。实际上,城市交通网络中存在大量不确定因素。如车辆故障、交通管制、交通拥堵、交通事故等突发事件,这些突发事件还会与不确定的交通需求产生相互影响。因此,已有理论和方法在实际应用中具有局限性。近年来,不确定环境下车辆调度研究已经成为交通研究领域中的热点问题之一。因此,在大数据背景之下,充分利用智能公交系统的海量数据,考虑乘客动态需求等不确定因素,结合优化理论与方法,基于动态调度策略的公交运营优化研究具有理论和现实意义。

2) 到站时间预测分析

公交到站时间信息是智能公共交通信息服务系统的重要组成部分,提供实时准确的公交到站时间一方面可以为出行者提供实时的公交信息,方便人们进行出行计划安排;另一方面可以优化公交调度系统,使得公交运营者根据公交到站时间调整发车间隔、行车时刻表等,以提高运营效率和服务质量。

根据预测方法不同,公交到站时间预测方法主要分为三类:基于时空变化规律的预测方法、基于影响因素的预测方法及基于数据融合的预测方法。基于时空变化规律的预测方法就是根据公交行程时间在时间与空间上的变化规律,根据已有的历史数据对车辆在未来时

间里经过某一路段的行程时间进行预测。目前这种预测方法主要有时间序列方法、卡尔曼滤波法、神经网络、支持向量机等^[83]。基于影响因素的预测方法就是通过建立公交行程时间与这些影响因素之间的数学模型来进行公交到站时间预测。这种预测方法主要包括统计回归理论预测方法与数学模型预测方法。由于公交车行程时间影响因素众多^[84],需要采集大量的交通数据,而且每条公交线路的情况不一样,所以预测模型需要根据实际情况进行不断的修正,因此,这种方法实用性较差。基于数据融合的预测方法是指通过对多源检测器采集的数据进行综合处理,弥补了单一数据源可能出现的误差,可以得到更全面、更准确的数据;进而结合多种预测方法的预测结果,从而达到提高预测精度、保证预测结果稳定性的目的^[85]。目前数据融合方法主要有以下几种:加权平均法、卡尔曼滤波法、贝叶斯方法、统计决策理论、模糊集理论法、神经网络等。

国外方面,公交到站时间预报在国外许多城市已经得到应用。在美国旧金山,利用 MUNI 系统将车载 GPS 设备接收的信息传送到控制中心,控制中心再根据公交车的实际位置、即将到达的站点以及车辆所行路线的交通状况等信息,估计到达时间,并将其通过车站电子站牌提供给乘客^[86]。在澳大利亚悉尼市已经投入使用的机场快速公交,利用路标定位系统采集车辆信息,采用基于统计分析的方法进行到站时间预测,将车辆到达时间提供给乘客^[87]。日本研究人员对日本横滨一号国道的车辆运行时间进行研究,利用车辆自动识别系统记录车流量、车道占用率等交通信息,通过寻找与实时交通情况最匹配的历史数据的方式预测当前车辆运行时间,通过互联网向用户的电脑或者手机发送公交车的预测到达时间^[88]。Chen 等考虑公交车驾驶人对车辆行程时间的调整因素,应用卡尔曼滤波模型研究公交车行程时间预测。该研究提出了一种通过自动车辆定位(AVL)数据与自动乘客计数器(APC)数据动态预测当前公交车到达下游每一站点的时刻的方法,而且此模型中还考虑了驾驶行为^[89]。Ojili 将公交车线路分为单独的“一分钟区段”用以计算公交车的到站时间。通过计算公交车当前位置距车站还有多少个一分钟区段从而确定其行程时间。由于该研究是通过历史平均行程时间数据来划分“一分钟区段”,因此,这种方法只适用于交通状态比较稳定的情况^[90]。Chien 等利用公交 GPS 数据分别建立四个基于神经网络的到站时间预测模型,这四个预测模型分别为基于路段信息的预测模型、基于站点信息的预测模型以及分别对这两个预测模型进行修正后的预测模型^[22]。Mazloumi 建立了基于神经网络的公交到站时间预测模型,预测模型中将路段交通量、车辆满载率、时刻表执行情况作为输入变量。另外该研究对公交到站时间预测中考虑的变量进行了统计,其主要有公交站停站时间、时间因素、时刻表、交通量、公交车辆因素、天气、车速、历史数据等^[91]。Ramakrishna 等采用公交车 GPS 数据、客流数据及路测车辆瞬时速度构建多元回归模型;另外作者还建立了只采用 GPS 数据的回归模型,结果显示在预测模型中加入客流数据和路测车速可以提高预测精度^[92]。

近年来我国在公交到站时间预测方面的研究也正在迅速发展。同济大学研制的“交通信息网络系统”依托交通信息网络平台,将采集的公交车和出租车的 GPS 数据,用高性能计算机进行实时处理,并通过智能导航系统为车载终端用户、手机用户提供公交到站时间、实时路况、最佳出行方案等信息。郑长江等^[93]使用支持向量机预测公交到站时间,并利用卡尔曼滤波对预测结果进行实时调整。此种方法可预测车辆到达多个站点的时间,但是模型