

武汉邮电科学研究院硕士学位论文

**基于行车大数据的商圈建设研究与分析**  
**Research and Analysis of Business Circle**  
**Based on Traffic Big Data**

专    业： 通信与信息系统

研究方向： 互联网技术

导    师： 巴继东

研 究 生： 彭晨伟 学号： 20140036

二〇一七年三月

## 独创性声明

本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果，除了文中特别加以标注的地方外，没有任何剽窃、抄袭、造假等违反学术道德、学术规范的行为，也没有侵犯任何其他人或组织的科研成果及专利。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。如有任何侵权行为，本人愿意为此独立承担全部责任。

作者签名： 彭晨伟 签字日期： 2017.3.8

## 关于论文使用授权的说明

本人完全了解武汉邮电科学研究院（烽火科技集团）有关保留、使用学位论文的规定，本文知识产权归武汉邮电科学研究院所有，武汉邮电科学研究院有权保留送交论文的复印件和电子版本，允许论文被查阅和借阅。同意将本人的学位论文提交中国学术期刊（光盘版）电子杂志社全文出版并收入《中国学位论文全文数据库》。

☒ 公开      ☐ 保密一年      ☐ 保密两年

(注：保密的学位论文在解密后遵守此协议)

作者签名： 彭晨伟 签字日期： 2017.3.8

导师签名： 巴继东 签字日期： 2017.3.8

## 摘要

商圈建设是城市现代化建设的重要内容，是城市商业文明和经济发展最具代表性的城市名片。重庆市响应国务院提出的智慧城市发展战略，将智慧城市建设正式引入规划，重点发展智能交通、智慧商圈等方面的建设。智能交通和智慧商圈的建设相辅相成，通过智能交通项目所采集的大量行车数据来为商圈的建设提供数据支撑，帮助商圈向着持续高效的方向发展，同时克服了经典商圈理论指导实际商圈建设中存在的数据迟滞的问题。

一方面大规模的行车数据为商圈理论分析提供了强有力的数据支撑，另一方面处理海量的行车大数据对系统平台也是一个巨大的挑战，引入云计算、Hadoop 大数据处理平台等先进技术，使处理大规模行车数据的问题迎刃而解，为商圈建设提供了技术支撑。

本文在发展智慧城市建设的大背景下，首先介绍了经典商圈理论内容，并分析了其在实际应用中存在的缺陷；然后根据城市交通物联网系统的建设内容，将行车大数据运用到经典的商圈理论中，丰富了商圈理论的研究内容；接下来针对经典商圈理论中存在的缺陷，提出了基于行车大数据的商圈辐射范围分析方法，该方法充分运用了大数据处理技术，在估算商圈辐射范围方面取得了良好的效果；最后对比分析了两种技术路线计算得到的商圈辐射范围的结果，同时对新技术的应用范围边界作出说明，总结了基于行车大数据计算商圈辐射范围方法的不足之处和后续的研究工作。

**关键词：**商圈；吸引力模型；数据挖掘；大数据平台

## Abstract

Business circle construction is an important content of urban modernization, is the most representative city card of urban commercial civilization and economic development. To response of the State Council put forward the smart city development strategy, Chongqing City introduce smart city construction planning, focusing on the development of intelligent transportation, intelligent business circle and other aspects of the building. The construction of intelligent transportation and intelligent business circle are complementary to each other. Through the big traffic data gathered by the intelligent transportation project, it can provide the data support for the construction of the business circle, and help the business circle to develop in a sustainable and efficient way. Meanwhile, it overcomes the problem of the existence of data delay in the Circle construction.

On the one hand, big traffic data for the theoretical analysis of the business circle provides a strong data support, on the other hand, dealing with massive traffic data on the system platform is also a huge challenge, the introduction of cloud computing, Hadoop large data processing platform and other advanced technology, So that the problem of handling large-scale traffic data be solved, technical support be provided for the construction of the business circle.

In this thesis, in the background of the development of smart city, first introduced the classic business circle theoretical content, and analysis its shortcomings in practical application; and then according to the construction of the urban traffic IOT system content, traffic data will be applied to the classic. It enrichs the study of business circle theory content. Then, according to the defects in the classical business circle theory, the radiation range analysis method of the business district based on traffic data is put forward. This method makes full use of the large data processing technology, In the evaluation of the radiation range of business circle has achieved good results; at the last, the result of two kinds of technical route is compared and analyzed, to explain the new technology

application range of the border at the same time, the shortcomings of the calculation method based on traffic data and the follow-up research work is summed up in the end.

**Key Words:** Business circle; Attractive models; Data mining; Big data platform

## 目 录

<b>第 1 章 绪论</b>	<b>1</b>
1.1 交通数据挖掘背景	1
1.2 国内外研究现状	3
1.2.1 交通数据挖掘的研究现状	3
1.2.2 商圈建设研究现状	4
1.3 本课题研究目的与意义	6
1.4 本课题主要内容及创新点	7
<b>第 2 章 商圈理论及数据基础</b>	<b>8</b>
2.1 商圈理论	8
2.1.1 商圈的定义	8
2.1.2 雷利法则	9
2.1.3 康维斯断点理论	11
2.1.4 哈夫概率法则	11
2.2 数据基础	12
2.2.1 城市交通物联网系统数据源分析	13
2.2.2 静态数据分析	15
2.2.3 动态数据分析	16
2.3 大数据平台	18
2.3.1 大数据平台应用的必要性	18
2.3.2 大数据平台介绍	19
2.4 本章小结	21
<b>第 3 章 商圈建设的分析与建模</b>	<b>22</b>
3.1 相关定义	22
3.1.1 商圈区域	22
3.1.2 采集点	23
3.1.3 采集方向	23
3.1.4 消费事件	24

3.2 数据集描述 .....	24
3.3 雷利法则估算商圈辐射范围 .....	27
3.3.1 时间距离候选集提取方法 .....	29
3.3.2 关键算法实现 .....	31
3.3.3 雷利法则确定商圈辐射范围流程总结 .....	36
3.4 基于行车数据估算商圈辐射范围 .....	37
3.4.1 商圈消费者数据集提取 .....	38
3.4.2 分析消费者居住点 .....	40
3.4.3 划定商圈辐射区域 .....	43
3.5 本章小结 .....	45
<b>第 4 章 商圈辐射范围计算过程与结果分析 .....</b>	<b>46</b>
4.1 解放碑商圈辐射范围计算 .....	46
4.1.1 通过雷利法则计算商圈辐射范围 .....	46
4.1.2 通过行车数据计算商圈辐射范围 .....	49
4.2 计算结果的分析比较 .....	50
4.2.1 解放碑商圈结果 .....	51
4.2.2 从时间角度分析商圈辐射范围变化 .....	52
4.2.3 从总体上分析商圈辐射范围 .....	53
4.2.4 总结两种计算方法优缺点 .....	54
4.3 基于行车大数据商圈建设应用举例 .....	55
4.4 本章小结 .....	55
<b>第 5 章 总结与展望 .....</b>	<b>57</b>
5.1 全文总结 .....	57
5.2 进一步研究 .....	57
<b>参考文献 .....</b>	<b>59</b>
<b>致谢 .....</b>	<b>61</b>
<b>附录 1 攻读硕士学位期间参与的项目和发表的论文 .....</b>	<b>62</b>

## 第1章 绪论

### 1.1 交通数据挖掘背景

近几年来,随着物联网技术的迅猛发展,物联网成为当前研究热点,甚至被誉为继计算机、互联网技术之后的第三次信息革命<sup>[1]</sup>。目前,我国已经把物联网明确列入《国家中长期科学和技术发展规划纲要(2006~2020年)》和国家“十二五”期间重点发展的战略性新兴产业。随着我国经济水平的不断提升,人们的生活水平也随之提升,私家车保有量呈现强有力的上升趋势。当前,我国机动车保有量已达2.64亿,其中汽车1.54亿,驾驶人达到3.02亿,汽车驾驶人2.47亿。目前,机动车仍以年均1500万辆,驾驶人2000万人的增长趋势在发展<sup>[2]</sup>。

智能交通(ITS, Intelligent Transportation Systems)是对车联网技术的更进一步的发展,它结合了信息科技产业现阶段更前沿的技术手段,如数据挖掘、机器学习、移动互联网和大数据分析等高新技术,在车联网的基础上探索交通路网的高效运行方式,帮助人们合理有效的规划和管理城市交通路网建设和城市的总体规划。一个成熟的智能交通系统可以采集所有在网车辆的各项基本数据,包括车辆的基础数据和海量的行车轨迹数据,特别是对行车数据的采集体现了一个智能交通系统的实用性,通过对行车数据的分析可以获得当前路网状态的各项指标,比如交通拥挤、车流量预测等<sup>[3][4]</sup>,这类基础性的交通问题已经研究得相当成熟。

智能交通与车联网产业作为传统交通行业与现代信息通信技术的结合,既符合人民群众对交通运输日益增长的迫切需求,也是信息通信技术和产业、经济社会深度融合发展的必然结果,对带动汽车、信息通信、交通等行业的产业转型升级具有重要意义。交通数据的增长为智能交通行业带来新的生命力的同时,也给交通领域带来了新的挑战。重庆市于2006年在地方立法的支撑下,全面开通大数据服务平台建设,重庆交通信息卡(电子牌)实现了重庆籍机动车和驾驶人证照电子化全覆盖<sup>[5]</sup>。经过十年深耕,智能交通物联网大数据服务平台将重庆带进智能交通新时代,该平台的成功应用使得重庆在公安部交通管理科学研究院主持编撰的汽车电子标识(RFID)国家标准以及大数据国家标准的制定工作中占有一席之地。显然,如此庞大的行车数据背后



隐藏着巨大的商业价值,利用数据挖掘的知识,研究数据与数据之间隐藏的内在联系,最大程度的发掘其中的有用信息可以为人类社会的发展起到促进性作用。对这些数据的合理挖掘将为各行各业的发展提供良好的驱动力,保险行业通过车辆的基础情况向车主提供针对性的保险服务,金融行业通过车主的生活消费圈提供合理的理财方案,汽车制造业通过收集车辆的基本状况调整生产研发计划等等<sup>[6][7]</sup>都是数据挖掘应用的重要体现。

随着智能交通系统的普及,使得用户的出行数据的获取变得轻而易举,其实这些出行数据不仅仅真实的反映了用户在日常生活中的位置移动,也从侧面表现了用户的就餐、购物、出游等日常生活状态,能够准确和完整的反映用户的兴趣习惯<sup>[8]</sup>。例如车辆的行车数据可以记录用户出行的位置信息,将之与用户的消费行为联系起来就可以推断出该用户在出行过程中的经济活动。再结合大数据技术,整合驾驶人、机动车、城市道路等信息资源,对海量的交通数据进行分析挖掘。其中通过采集车辆在购物中心及周边的停车情况,对停车量、商圈周边行车数据、机动车数据等进行分析,总结商圈经营与周边行车数据之间的关联关系,得出该商圈经济发展情况,继而分析在未来几年的发展潜力,可为商圈商家策划营销计划提供参考意见,为政府职能部门对各区域建设的协同发展及对重点经济发展区域的规划提供参考依据。出行数据的基本研究方向如下图所示。

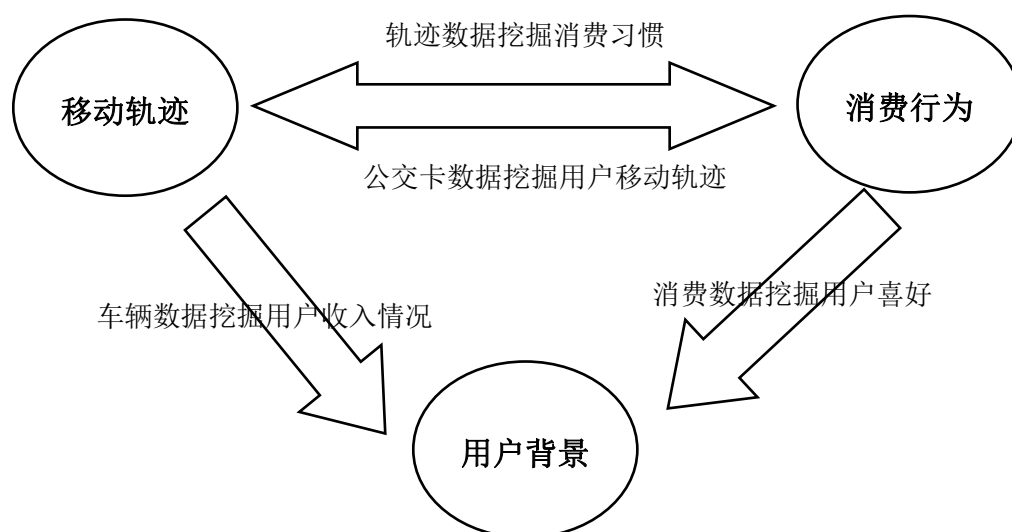


图1.1 大规模出行数据的基本研究方向

城市发展是人类社会文明发展的集中体现，同时也是社会发展的引擎，城市的发展水平往往通过其经济发展水平来衡量。城市商圈的规划建设是人类在商业文明与社会经济发展中不断摸索出来的普遍规律，遵循规律就可以促进城市向着有利的方向发展，反之则将错过发展的有利先机，甚至造成巨大的经济损失。重庆市自从加入我国直辖市以来，凭借其有利的政策条件和创新型发展的思想，不仅在交通领域大力推动智能交通系统在实际中的应用，还结合具有重庆特色的“商圈文化”来集中力量发展重庆的经济建设。2015年，重庆市政府发布了《重庆市智慧商圈建设实施方案》<sup>[9]</sup>，其中提出，重庆商圈是城区商业商务资源富集、集约化程度较高的商业商务核心区，统计数据指出，2014年底纳入统计的20个商圈实现零售额2837亿元，全市商圈占社会消费品零售额比重达到56%，全市过半的消费在商圈实现。随着商圈的快速发展，信息网络化时代的到来和互联网经济的迅猛崛起，商圈传统的商业模式已经越来越不适应形势发展的需要，智慧商圈的建设是利用物联网、云计算和大数据等先进技术建设商圈，是重庆智慧城市建设的重要内容，更是重庆创新型发展策略的具体体现。

## 1.2 国内外研究现状

### 1.2.1 交通数据挖掘的研究现状

21世纪互联网的迅速发展将人们带进了“信息时代”，数据作为信息的载体，在短短十几年间数据量呈现爆发性增长，“大数据”进入人们的视野，如何利用这些大数据为人类的发展作出贡献成了国内外科学家们的研究热点，数据挖掘就是一门研究数据与数据之间隐藏的内在联系，将数据转换成有用的信息和知识的学科。

交通大数据来源广泛，卡口抓拍、视频监控的非结构化数据占据存储空间大，同时提取有价值的信息比较困难，所以数据挖掘一般从结构化的数据着手，非结构化数据仅作为辅助数据集。以结构化数据为对象的研究，早期主要是对公交IC卡数据的研究，慢慢由于科学技术的发展，GPS、RFID等各类传感器数据逐渐占据主导地位。

国外方面，对于IC卡数据的研究开始比较早，主要集中在公交出行OD的推算和公交出行行为分析<sup>[10]</sup>，这类分析具有基础价值，对公交系统的调度、运营和管理具有应用价值。Martic<sup>[11]</sup>通过轨迹数据来构建人类动力学特征模型，对136个志愿者的轨

迹数据进行分析处理后,建立模型来对个体移动的下一个地点进行预测,最后从理论上证明了预测的可行性。Bazzani<sup>[12]</sup>等从保险公司获得私家车的轨迹数据,通过分析后得出车辆每天的出行距离分布是服从指数分布的,而出行时间则是服从复合幂律分布。Wang<sup>[13]</sup>等利用安装在出租车上的GPS定位工具获取轨迹数据,再与地图匹配得到道路拥挤的时间序列,通过模型得到城市道路每日的拥挤熵,发现每天的拥挤程度可预测性与时间和车辆的起止点无关。Luong<sup>[14]</sup>等运用基于EIDBSACN的算法找出轨迹数据中的显著位置,通过论证指出显著位置的确定对分析人类活动模式具有相当重要的作用。Chen<sup>[15]</sup>等基于车辆的轨迹数据建立了3种不同的马尔科夫模型,对不同模型预测个体下一个位置的效果不同进行分析。

国内方面,王明生<sup>[16]</sup>等在对石家庄公交IC卡数据进行实证研究中发现,公交乘客与地铁乘客的乘车距离分布形式非常接近。RFID技术在交通领域应用也很广泛,李学龙<sup>[17]</sup>根据RFID采集数据构建交通参数模型,计算交通流量、流速密度等交通参数,提出了交叉口转移流量的计算方法,可以依此预测交通状态。张晨<sup>[18]</sup>对RFID智能交通系统中的交通流参数检测进行建模分析,同时应用到高速路OD矩阵估计模型中,用拉格朗日乘数法对OD矩阵元素的概率极大似然估计进行推导,给出了基于EM算法的OD矩阵估计详细流程,通过仿真模拟验证了模型有效性。曹洁<sup>[19]</sup>等使用联合卡尔曼滤波器提高RFID技术采集交通量信息的精度,通过对各个节点间的同一车牌数在周期内进行统计,获取动态OD数据,实验证明了该方法可行性。徐秋平<sup>[20]</sup>利用RFID技术对交通信息采集与交通事故自动识别进行研究,包括有防碰撞问题的算法分析与仿真,车联网可测参数的检测模型仿真模拟,以及基于排队论思想的交通事故自动识别算法的研究,实验结果表明事故检测算法效果良好且能够适应不同道路交通状况。

随着含有地理位置信息的数据规模不断的扩大,不同的设备和业务产生的数据包含的上下文信息内容更加丰富,形式也是纷繁多样,如何对这些海量的、上下文丰富的数据进行有效挖掘值得深入研究。通过对商圈周边的交通数据深入挖掘,可以分析商圈的发展情况,对商圈建设有良好的促进作用。

### 1.2.2 商圈建设研究现状

消费者的购买行为和商业企业的经营行为共同决定了市场的商业运营,而商圈就

是商业企业吸引顾客前来消费的一个空间范围，所以对商圈的研究显得尤为重要。商圈理论起源于西方的区域差异理论，其核心问题是如何确定商圈的最佳位置，以保证在该区域能以最小的成本充分发掘市场的利润。

美国的学者 Reilly<sup>[21]</sup>用三年时间对超过 150 个城市的商圈零售情况进行调查，并结合牛顿力学中的万有引力定理，提出了“零售引力法则”，总结了城市人口与零售引力之间的关系，也被称为雷利法则。他认为一个城市能吸引周围用户来消费的几率是与城市的人口数量成正比，而与它们之间距离的平方成反比，它从数理分析的角度对消费者在购物时选择商店的行为给出一般性的解释。

随后在 1949 年，美国学者 Convers<sup>[22]</sup>同样是在调查分析了 100 多个美国中小城市商品的购买量与附近大城市的购买量之间的距离关系后，提出了修正的雷利法则，后被称为康帕斯断点理论。他提出了商圈界限的概念，在两个或多个存在竞争关系的商圈之间，必然存在共同的重叠区域，那么将这一区域中对所有周边商圈具有相同吸引力的平衡点找出来，并将它们连接起来就可以确定各个商圈的辐射范围。

1963 年 Huff<sup>[22]</sup>提出消费者选择商场的概率吸引力模型，他将研究重点从商圈转移到对消费者的购买行为分析上，并指出消费者光顾某一商圈的概率由商圈的商业区规模和消费者到商圈消费的距离来决定的。日本学者针对概率吸引力模型计算量大的局限性对其进行修正，提出消费者到商圈消费是与商圈内卖场的面积成正比，与消费者到商圈的距离的二次方成反比。

美国的研究人员 Brubaker 通过对小区购物中心的选址问题进行研究，提出了除考虑交通、商业规模、距离等传统因素之外，还要将消费者的年收入与职业、同小区的竞争者等因素作为标准列入考虑范围。

近几年我国城市化进程加快，国内购物中心的建设逐步加快，相关领域也吸引越来越多的学者进行研究，研究成果也逐步丰富起来。

孙良静<sup>[24]</sup>分析商圈特征、商圈供应以及商圈消费者需求，提出影响商业地产项目定位的关键因素，同时以铜梁印象项目的商业定位为例进行了应用模拟。

王先庆<sup>[25]</sup>等提出 9 个影响因素和 11 个影响购物中心聚客能力的检验假设，并建立了理论研究模型，为实证研究奠定了理论基础。

刘瑞霞<sup>[26]</sup>通过分析经济欠发达地区的中小型城市影响商圈布局的因素，将影响因

素分为外部因素、时间因素和内部因素三大类。通过分析，其中外部因素与内部因素是影响商圈的两大主要因素，而时间因素是两大主要因素作用下的自然延续。

吴忠才<sup>[27]</sup>对商圈理论的各种分析方法和模型进行了总结归纳，指出城市商圈和企业商圈之间存在差异，并且认为消费者在某城市商圈购物的机率与城市商圈规模大小成正比，而与达到该处距离和该处与城市中心商圈的距离成反比。

通过对商圈建设的发展现状和应用情况的分析，对商圈建设的深入分析是市场经济大环境下商业企业提高自身竞争力的切实需求，也是城市现代化建设、各区域协同发展的需求。从海量行车数据分析商圈，可以针对商圈的自驾车消费者做精准分析，该研究帮助商圈经营者在精细化运营方面起到创新性的探索作用。

### 1.3 本课题研究目的与意义

本课题的研究目的是在现有的城市交通物联网系统采集的海量历史行车数据基础之上，通过数据挖掘技术来分析城市商圈建设与行车大数据之间的内在关系。具体包括：(1)通过行车数据来确定商圈的辐射范围，为商圈各商户制定销售计划提供决策支持；(2)以私家车行车数据为依据，分析现有的商圈理论模型存在的缺陷并提出修正方案；(3)从行车大数据分析影响商圈体量的因素，为今后的商圈建设奠定理论基础。

20 世纪 40 年代至今，有关商圈的理论研究由最初的雷利法则发展到哈夫概率法则，以及众多研究学者提出的各种修正理论，这些理论研究成果离实际应用都有不同程度的偏差，一方面是理论成果的先进性没有跟上社会发展的节奏，另一方面是研究的数据来源不够丰富，数据量达不到要求，从而制约了商圈理论的实际应用。经过几十年的发展，到 21 世纪信息技术的广泛应用，各种数据源和数据量增长迅速，在智慧城市的发展大潮下，城市交通物联网系统可以采集到与机动车相关的各类数据，将与商圈关联的行车数据收集起来对商圈研究意义重大，不仅数据种类丰富并且数据量庞大，能够有力支撑本文的商圈建设研究，为商圈理论的实际应用起到促进作用。

## 1.4 本课题主要内容及创新点

本文主要介绍了商圈理论的研究情况,尤其是如何通过行车大数据来确定商圈的辐射范围,还分析了消费者的消费行为习惯,课题主要研究内容包括以下五点:

(1) 详细介绍了商圈理论中的雷利法则、康维斯断点理论和哈夫概率法则经典法则,并分析了各自优缺点和在实际应用中的局限性;

(2) 介绍了城市交通物联网系统的数据源组成情况,分析了其中与商圈有关联关系的数据特征,为行车大数据在商圈建设研究中得到应用提供数据支撑;

(3) 以行车大数据为基础,结合经典的修正雷利法则来确定商圈的辐射范围,详细介绍了如何从海量的行车数据中提取雷利法则中所需的商圈时间距离参数的方法;

(4) 提出基于行车大数据划分商圈辐射范围的实现方法,与经典商圈理论不同的是,该方法直接以停车场数据源为基础,关联行车数据、车辆基本数据、车主信息等数据源,充分利用 Hadoop、Hbase、MapReduce 等大数据存储、处理技术,通过分布式集群强大的云计算特性可以计算得出商圈的消费用户分布情况,从而划分商圈的辐射范围边界;

(5) 分析了通过两种不同方法得到的商圈辐射范围存在差异的可能因素,为商圈理论的进一步研究奠定基础。

本研究的创新点在于:(1)以行车大数据代替传统的问卷调查,为商圈理论研究的引入新的数据来源,通过行车大数据所得到的商圈分析结果比问卷调查更具应用价值,技术上更有优势;(2)提出一种基于行车大数据的商圈辐射范围计算方法,该方法充分利用大数据处理技术,得到的商圈分析结果更具代表性;(3)通过对比文中两种商圈辐射范围计算结果,分析现时商圈理论中计算法则存在的问题,并提出改进意见。



## 第2章 商圈理论及数据基础

### 2.1 商圈理论

学术界学者们逐渐关注对商圈理论的研究源于19世纪30年代美国学者雷利提出了著名的商圈零售吸引力法则,随后不断有学者研究在不同应用场景下雷利法则的局限性,并提出修正意见,诸如康维斯断点理论、哈夫概率法则等都是对雷利法则的扩展研究成果。

#### 2.1.1 商圈的定义

商圈,早期的研究将其定义为单个零售商业企业所能吸引的顾客在空间上的分布范围,这里强调了是某一个零售商业企业的辐射范围或者商业地域,其本质是一个与零售企业距离不等的消费者需求圈。这类以单个零售商业企业视角定义的商圈也被称作企业商圈,这是从企业微观层面上出发的定义。这也是国外早期研究者普遍采用的微观层面的企业商圈概念,但是随着研究人员对商圈研究的不断深入,从不同的研究角度延伸出各种商圈概念<sup>[27]</sup>。

现在人们谈论与运用更多的是诸如东京的银座商圈、北京的王府井商圈、上海的徐家汇商圈、重庆的解放碑商圈、香港的铜锣湾商圈、台北的西门汀商圈等等。这里的商圈内涵明显与上面的企业商圈不同。这类商圈一般是城市内部区域性的商圈,由若干商业企业集聚形成的商业中心所能吸引来消费的顾客所分布的地理范围,这类商圈称之为城市商圈,也是人们实际运用最多的一类商圈概念,为城市规划设计、城市商业布局等方面所广泛应用。相对于微观层面的企业商圈来说,城市商圈具有中观层面属性。而宏观层面的商圈则体现的是城市间协同发展的理念,具体表现为国家区域发展战略,以促进现代城市化建设的进程,对这类宏观都市商圈的研究也具有现实意义。现在研究的热点和难点问题主要集中在中观层面的城市商圈,本文也将以城市商圈为研究对象,以下为方便表述,将城市商圈统一称作商圈。

通过以上对商圈的定义,可以从三个方面来认识商圈:(1)商圈是一个具体的区域空间,它是可以在地图上大致标定的地理区域;(2)商圈,对于商业企业来说是一

个销售空间，对于购买者来说是一个购买空间，合理分析两者之间的关系可以做到双方利益最大化；(3)影响商圈建设的因素有很多，一般从商业企业角度去分析商圈的建设规范，将购买者的购买行为作为研究对象。

商圈是一个地理概念，其范围的确定对商业企业的发展有着至关重要的作用，可以通过分析商业覆盖的空间范围来了解市场销售机会，从而制定相应的市场营销策略，取得商业上的成功。

### 2.1.2 雷利法则

商圈理论早期的研究中，影响力最大、最著名的是雷利法则<sup>[21]</sup>，1929 年雷利通过研究 150 多个城市的零售业的销售情况，创新性的在商业经济领域引入牛顿力学中的万有引力定律，提出零售吸引力法则。其具体表述是：两个销售区域 A、B 同时对它们之间的任意一点 C 处的居民的购买行为具有吸引作用，那么区域 A、B 对 C 点顾客的吸引力大小与该区域的人口数量成正比，而与该区域距离 C 点的距离成反比。该公式在一定的假设前提下具有理论指导意义，假设 A、B、C 三个区域的人口数分别为  $P_a$ 、 $P_b$  和  $P_c$ ，C 点是 A、B 之间的区域的人口聚集区， $P_c$  是聚集区的人口数。同时 A、B 两地所提供的商品对 C 点的居民顾客来说是具有同等价值的，即在不考虑其它影响因素的情况下，C 点的居民到任何一个地区消费获得的商品价值和服务质量等都是是一致的，那么它们满足： $V_a=V_b=V$ ， $V_a$ 、 $V_b$  分别为 A、B 两区域平均顾客价值。再假设 C 点居民到 A、B 区域消费时，主要道路的交通易达性是一样的，即从 C 点到消费地点的时间只与两点的物理距离相关。在以上约定的条件下，将万有引力定律运用进来，把承担商品销售角色的 A、B 区域看作万有引力定律中的物体，万有引力定律中的质量用区域的规模大小（人口数的多少）来衡量，用  $F$  来表示两个地区之间吸引力的大小。根据万有引力定律的公式，A、B 区域吸引 C 点居民到该区域购物的吸引力可以用公式(2.1)和公式(2.2)表述如下：

$$F_a = V_a \times k \left( \frac{P_a \times P_c}{D_a^2} \right) \quad (2.1)$$



$$F_b = V_b \times k \left( \frac{P_b \times P_c}{D_b^2} \right) \quad (2.2)$$

其中，

$F_a$ 、 $F_b$ ：表示两个销售区域对点 C 处居民的购买吸引力；

$V_a$ 、 $V_b$ ：表示 A、B 两地区平均顾客价值，并且满足  $V_a=V_b=V$ ；

$k$ ：常系数，相当于牛顿万有引力公式中的万有引力常数  $g$ ；

$D_a$ 、 $D_b$ ：表示 C 点到 A、B 两个区域的距离。

将公式(2.1)，(2.2)相除就可以得到一般意义上通常见到的雷利法则的表达形式，如公式(2.3)所示：

$$\frac{F_a}{F_b} = \frac{P_a}{P_b} \times \left( \frac{D_b}{D_a} \right)^2 \quad (2.3)$$

单一的表达方式很难将该理论运用与实际商业环境，为了在更多场合合理的表述这一理论，雷利法则的引入了经验常数的表达方式，如下公式(2.4)所示：

$$\frac{F_a}{F_b} = \left( \frac{P_a}{P_b} \right)^N \times \left( \frac{D_b}{D_a} \right)^n \quad (2.4)$$

其中，

$N$ ：表示人口修正系数，通常取 1；

$n$ ：表示距离的修正系数，取值范围 1.5~2.5，典型值取 2。

通过分析雷利法则的假设前提，结合最终得到的理论公式可以看出，雷利法则在商圈建设初期无法获得周边相关资料的情况下，利用官方数据（地区人口数）来作为该地区商圈吸引力的判断依据，这样做数据取得很容易，运算也很简单，但是在应用方面还存在一些局限性，具体分析如下<sup>[26]</sup>：

- (1) 商业区域若想吸引更多的消费者到本地区消费就必须具有较高的人口密度，单一以该地区的人口密度来衡量购物中心对周围居民的吸引力显得太单薄；
- (2) 居民居住点到购物地点的地理上的距离对吸引力大小作用显著，但是未考虑其他交通状况（如交通费用、交通工具不同等）；
- (3) 各个区域的商家为顾客提供的价值差异被忽略，实际上商圈自身的建设（基

基础设施完备程度、服务态度好坏、商品的多样性等)对用户的最终选择具有很大的影响。

### 2.1.3 康维斯断点理论

康维斯断点理论<sup>[22]</sup>是在雷利法则的基础之上提出来的,是对雷利法则在应用方面更进一步的扩展。通过计算可以得到两个或多个区域之间的商圈吸引力相同的点,这些点被称为康维斯平衡点,把这些平衡点连接起来恰好可以将各个不同商圈区隔开来,从而确定商圈的范围。

康维斯平衡点可以通过雷利法则直接推导得出,康维斯断点是 A、B 区域商圈吸引力相等的 C 点的集合,即令  $F_a/F_b=1$ , 设 A、B 两地之间的距离为  $D$ , 则  $D=D_a+D_b$ , 代入雷利法则中可以得到康维斯平衡点的表达式, 如下(2.5)式所示:

$$D_a = \frac{D}{\left(1 + \sqrt{P_b/P_a}\right)} \quad (2.5)$$

式中  $D_a$  表示与地区 A 商圈相距  $D_a$  的点,该点居民到 A、B 两地消费吸引力相等,通过康维斯平衡点公式,可以得到一个平衡点的合集,根据这些点就可以大致确定商圈的辐射半径,以平衡点合集的平均值为半径作一个圆形区域,该区域即为商圈辐射范围。

### 2.1.4 哈夫概率法则

雷利法则及其修正法则在理论研究中得到广泛关注,但是在实际应用中往往受到人口密度相差巨大、道路交通条件等因素影响,使其得到的结果与现实并不相符,另外随着经济不断发展,汽车的普及程度越来越高,雷利法则中的距离因素不再是主导因素,商圈理论也在这种大环境下得到更进一步的发展。

美国的经济学博士哈夫<sup>[23]</sup>提出了哈夫概率法则,使得雷利法则衍生出来的商圈理论模型更加接近于实际应用。在哈夫之前的学者对零售商圈的理论研究均以城市为对象,没有具体到城市中的商业街、购物中心、百货中心等商业聚集区,哈夫以此为出发点,将商圈理论模型的研究对象聚焦到城市中的某一具有销售性质的商业聚集区。

在综合考虑了消费者到商圈消费的空间距离、区域人口数等基础因素之外，针对雷利法则的局限性提出以商店营业面积作为吸引消费者到该商圈消费的关键因素。他认为在数个商业聚集区集中于一地时，居民到哪一个商业聚集区发生消费行为的概率是由商业聚集区的规模和居民到商业聚集区的距离所决定，其中商业聚集规模用商场的营业面积来表示，距离用时间距离表示。其公式表达公式(2.6)所示：

$$P_{ab} = \frac{\frac{S_b^\mu}{D_{ab}^\lambda}}{\sum_{b=1}^n \frac{S_b^\mu}{D_{ab}^\lambda}} \quad (2.6)$$

其中，

$P_{ab}$ ：表示 A 地区的居民选择到 B 商圈购物的概率；

$S_b$ ：表示 B 商圈的营业面积；

$D_{ab}$ ：表示居住在 A 地区的消费者到 B 商圈购物所需要的时间距离；

$\mu$ ：表示商店规模对消费者的吸引力的影响指数；

$\lambda$ ：表示消费者从 A 地区到 B 商圈路程中障碍指数；

$n$ ：表示相互竞争的零售商圈的数量。

哈夫概率模型也是从雷利法则演化而来，在雷利法则的基础上进一步分析了影响消费者做出最终决策的各类因素，使其更贴近实际，但是也存在一些缺陷<sup>[28]</sup>：

(1) 哈夫概率模型中，由于只考虑了商业聚集区营业面积和时间距离这些数据，相对于雷利法则来说更接近实际情形但是仍然存在一定的误差；

(2) 哈夫概率模型公式中引用了  $\mu$ 、 $\lambda$  指数来修正模型在不同情形下的误差，但是  $\mu$ 、 $\lambda$  的具体取值难以数量化，往往需要大量的市场调查数据来支撑，这无形的增加商圈建设的时间成本和人力成本，很大程度上限制了哈夫概率模型的应用场景。

哈夫概率模型对商圈理论在实际中的应用起到了巨大的促进作用，各国学者也在做致力于改善其在适用性方面的努力。其中日本学者用吸引力与购物距离的平方成反比的关系，将  $\lambda$  的典型值取为 2，将  $\mu$  的典型值取为 1。

## 2.2 数据基础

智能交通系统中大数据来源可以分为交通流数据、位置数据、非结构化视频数据、

多源互联网及政务系统数据,从不同的数据来源获取车辆各个维度的数据帮助研究人员分析机动车驾驶人的各种行为特征,深度挖掘就可以在各个行业各个方面中应用。

通过基于微波雷达、超声波、感应线圈、视频监控等检测器,北京市公安局交通管理局建立了交通信息采集、处理、发布系统,北京市道路交通流预测预报系统等。通过移动通信技术,使位置数据的获取变得容易,通过分析公交智能卡的数据为公交基础设施建设和运营线路规划提供支持,通过分析出租车车载终端的数据,研究用户加油消费的规律。物联网的出现提高了城市交通信息综合获取的水平,极大的丰富了数据的来源于发布途经。这些维度丰富的数据不再仅仅应用于交通相关领域,海量位置数据通过关联分析可以引申到消费领域,对挖掘用户的消费行为也大有裨益。

由以上的理论分析得知,对商圈建设的分析需要建立在对影响商圈建设因素的大量历史数据分析的基础上,对商圈整体现状的评估和将来发展情况的准确预测离不开大量历史数据的支持,因此,研究的首要条件是获取商圈运营过程中采集到的相关的大量历史数据,本文研究的商圈建设命题是依托重庆市面向城市交通管控的 RFID 交通物联网系统项目,研究的数据来源于城市交通物联网系统的数据库,这些数据通过城市交通物联网系统的感知层设备采集后,经过清洗过滤等操作后存储到关系型数据库、分布式数据库以及空间数据库,为业务层的各种应用提供数据支撑。基于 RFID 的城市交通物联网系统建立了车辆出行在城市道路网中所产生的数据的全面采集体系,并通过地方法的支持和各部门的通力合作获得了极其有价值的多源数据,从用户提车上牌照到最终报销销毁的整个生命周期,包括中间的车辆买卖转让、行车违法记录等数据都将记录在本系统中,这些数据为本文的研究提供了全面支持。

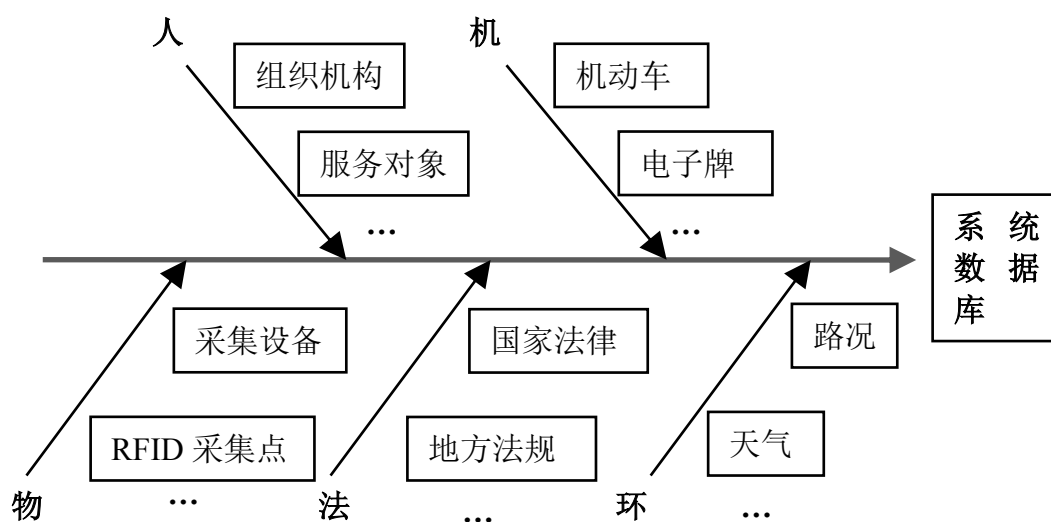
下面将对城市交通物联网系统的数据做一些简单分析,鉴于项目涉及业务面相对庞大,数据项覆盖范围广,这里主要分析与商圈建设相关的系统数据。

### 2.2.1 城市交通物联网系统数据源分析

城市交通物联网系统的数据可以分为两个大类,其一是城市道路交通管网中采集到车辆日常行车数据,这类数据通常特征是数据量大,冗余数据多,需要经过一系列再处理才能为分析人员使用;其二是与交通运输管理局、车辆管理所等相关部门交换所得到的各类基本数据,这类数据通常具有规范的格式,处理起来容易,但是由于各

个部门的业务类型不同导致数据库设计所遵循的标准规范也不尽相同，所以在交换得到多源数据之后也需要对数据进行再处理才能整合到系统中，从而高效利用数据，充分挖掘有用信息。

城市交通物联网系统的数据可以以人、机、物、法、环五个方面来进行分类，人是指与人相关的数据，如，服务对象以及组织机构等的信息；机在本系统中是指与车辆相关的数据，如，机动车基本信息、电子车牌等；物是指系统正常运行所涉及到的基础设备信息，如，RFID 采集点信息、采集设备信息等；法是指与车辆行驶的标准规范相关的数据，如，国家交通法律法规、地方及企业规章制度等；环是指环境相关数据，如，天气、路况等。下图 2.1 以鱼骨图的形式展示了城市交通物联网系统数据库的数据组成。



2.1 城市交通物联网系统数据组成分类图

商圈建设分析所涉及到的车辆行车数据、驾驶员数据、停车场数据等就是来源于人、机、物、法、环，统称为系统数据，系统数据可以分为静态数据和动态数据两大类<sup>[29]</sup>，静态数据（或称固定信息）是指在生产活动开始之前要准备的数据，如车辆基本数据、采集点基本数据、驾驶人基本数据等；动态数据（或称流动信息）是指在生产活动中发生的数据，不断发生且经常变动，如采集设备获取的车辆位置数据、卡口抓拍数据等。

根据城市交通物联网系统设计要求，系统数据库的基础数据包括以下 3 个部分：

(1) 重庆交通信息卡（机动车卡）由公安交管局“机动车卡”发行系统产生，规模：机动车卡现有 191 万辆，年增 21%，设计要求 340 万辆，静态数据 100 字节/卡。

(2) 重庆交通信息卡（驾驶人卡）由公安交管局“驾驶人卡”发行系统产生，规模：驾驶人卡现有 292 万人，年增 14%，设计要求 517 万人，静态数据 140 字节/卡。

(3) 重庆交通信息卡（电子牌）由公安交管局“电子牌”发行系统产生，规模：电子牌现有 156 万辆，年增 21%，设计要求 280 万辆，静态数据 90 字节/卡。

通过“机动车卡”，系统可以采集机动车在收费站、停车场等产生的通行记录数据；通过“电子牌”，系统可以采集该车牌号对应机动车在 RFID 采集点产生的通行记录数据。实际采集量 1000 万次/日，设计采集量 2000 万/日。按照以上标准估算，不算视频、图片等占用存储空间极大的非结构化数据，单算电子牌采集点产生的行车记录等结构化数据，一年的数据量是 4.1TB，按年增 21% 计算五年的数据量约为 30.5TB，具体数据下面做详细介绍。

### 2.2.2 静态数据分析

上面对商圈建设的理论分析可以看出，车辆相关的数据扮演者十分重要的角色，为了提取出相关度高的数据，首先对静态信息进行了归纳和分类，相关静态信息如下表 2.1 所示。

表2.1 相关静态数据项统计表

序号	数据表名称	详细数据项
1	驾驶人信息	姓名、身份证明、性别、出生年月、国籍、行政区划、联系地址、联系电话、驾驶人档案编号等
2	车辆信息	车辆档案编号、号牌种类、车牌号、制造厂商名称、车辆品牌、车辆类型、使用性质、车辆所有人、排量、车身颜色、车主联系电话、车主住所地址等其它车辆基本信息
3	车牌信息	车辆识别代号、车辆档案编号、车牌号

附注：数据来源项目数据库设计文档

通过相关静态数据项的观察可知，静态数据主要来自机动车卡、驾驶人卡和车牌信息三个方面，下面对各自的数据量计算结果统计如下表 2.2 所示。

**表2.2 相关静态数据量统计表**

序号	数据名称	记录数	单条数据量	合计	备注
1	机动车卡	340 万	0.1KB	332MB	要求：可容纳全市登记注册机动车数 340 万，数据量： $3400000 \times 0.1 / 1024 (\text{MB})$
2	驾驶人卡	517 万	0.14KB	706MB	要求：可容纳全市登记注册驾驶人数量 517 万，数据量： $5170000 \times 0.14 / 1024 (\text{MB})$
3	电子牌	280 万	0.09KB	246MB	要求：可容纳全市登记注册驾驶人数量 280 万，数据量： $2800000 \times 0.09 / 1024 (\text{MB})$
	总数据量			1.25GB	数据量： $(332+706+246) / 1024 (\text{GB})$

附注：数据来源项目数据库设计文档

### 2.2.3 动态数据分析

各不相同的数据源所描述的对象也各不相同，静态数据是分析的不同对象之间关联关系的基础，也是连接商圈理论与交通数据的关键钥匙，而动态数据是通过长期、持续采集的行车数据得到的庞大数据集，所以分析动态数据可以观察研究对象沿时间线方向的变化情况，对未来的预测起到重要作用。

下面表 2.3 是有关动态数据项的统计情况，这些数据项只是众多数据项中很小一部分，但和本文分析商圈建设的主题结合非常紧密。

表2.3 相关动态数据项统计表

序号	数据表名称	详细数据项
1	采集点信息	采集点名称、采集点类型、采集点经度、采集点纬度、采集设备情况
2	卡口采集结构化数据	车辆信息、卡口信息

附注：数据来源项目数据库设计文档

城市交通物联网系统在设计初期需要考虑未来几年数据量可能成倍增长的情况，所以对动态数据量的分析显得尤为重要，下面表 2.4 列出了相关动态数据量的估算情况。

表2.4 相关动态数据量统计表

序号	数据名称	记录数	单条数据量	合计	备注
1	电子牌采集通行记录	8 千万条/天	0.15KB/条	11.44GB/天	全市 2000 个 RFID 采集设备，每个设备采集 2 个车道，采集频度 20000 次/天，数据量： $80000000 \times 0.15 / 1024 / 1024 (\text{GB})$
2	卡口采集结构化数据	124.5 万条/天	0.05KB/条	0.06GB/天	数据量： $1245000 \times 0.05 / 1024 / 1024 (\text{GB})$
	总数据量			11.5GB/天	

附注：数据来源项目数据库设计文档

上面是按静态数据与动态数据的分类对数据库的设计作了基本介绍，下表 2.5 是对数据周期存储量分析，可以看到存储周期为一年的数据量就达到了 4.12TB，而随



随着社会的发展，每年的数据量将会以 21% 的增长量继续增多，其数据量的大小可想而知。

**表2.5 相关数据周期存储量统计表**

序号	数据名称	记录数	单条数据量	合计	备注
基础数据	机动车卡	340 万	0.1KB	332MB	静态数据，全量存储
	驾驶人卡	517 万	0.14KB	706MB	
	电子牌	280 万	0.09KB	246MB	
动态采集数据	电子牌采集通行记录	8 千万条/天*365 天	0.15KB/条	4.1TB	存储周期为一年的数据量
	卡口采集结构化数据	124.5 万条/天*365 天	0.05KB/条	21.7GB	存储周期为一年的数据量
	每年总数据量			4.12TB	存储周期为一年的总数据量

附注：数据来源项目数据库设计文档

## 2.3 大数据平台

### 2.3.1 大数据平台应用的必要性

根据前文的分析可知，本文研究的商圈建设数据来源于系统的静态数据和动态采集数据，通过对数据量的估算可知，涉及到的数据量相当庞大，并且上述列出的数据还不包括其它与商圈分析无关的数据，如果算上所有的数据，包括卡口抓拍图片数据，标清视频采集点数据等，那么数据量会成倍的增加，因为数据类型不一致，包括结构

化数据、半结构化数据和非结构化数据，整个数据库的数据结构也变得更加复杂，如何从中提取价值数据成为关键，大数据处理技术提供了解决之道。商圈建设需要的是与车辆出行轨迹、出行时间、停车时长等可以与商圈建设建立联系的数据，那么从数据构成如此复杂、数据量如此庞大的系统数据中获得有价值的信息，并对这些数据进行分析处理，除了合理的数据库设计和良好的硬件支持之外还需要依靠大数据分析处理技术来支撑才能完成。具体来说，城市交通物联网系统对数据的存储和处理需要依靠大数据技术主要体现在以下几个方面<sup>[29]</sup>：

(1) 城市交通物联网系统中行车数据量极大，再有车辆基础数据、驾驶人基础数据以及视频图片等大容量数据对数据库的要求非常高，需要 Hbase 分布式数据库进行存储；

(2) 在对海量行车数据进行提取、筛选、清洗等操作的时候，往往需要占用大量的硬件资源，为了不影响即时性操作，保证良好的用户体验，提高工作的效率，选择 Hadoop 大数据平台存储管理数据是必要的选择；

(3) 对行车大数据的分析需要从城市交通物联网系统海量的历史数据中抽取特定的数据集进行分析处理，这个过程需要用到 Hadoop 平台的 Hive 数据仓库和 MapReduce 数据处理模式才能良好的完成。

### 2.3.2 大数据平台介绍<sup>[31][32]</sup>

Hadoop 是 Apache 基金会下的一个开源分布式计算框架，主要负责分布式存储（HDFS）以及分布式运算（MapReduce）。

Hadoop 数据处理平台的优点是：

(1) 灵活可扩展，包括存储空间的扩展和计算节点的扩展，适应大数据发展的需求；

(2) 经济廉价，Hadoop 平台对硬件设施的要求低，可以降低运营成本；

(3) 数据存储可靠，采用 MapReduce 任务监控机制和数据容灾备份机制保证的服务正常运行；

(4) 高效处理数据，HDFS 实现了分布式处理的高效交互，克服了传统数据处理的缺陷。

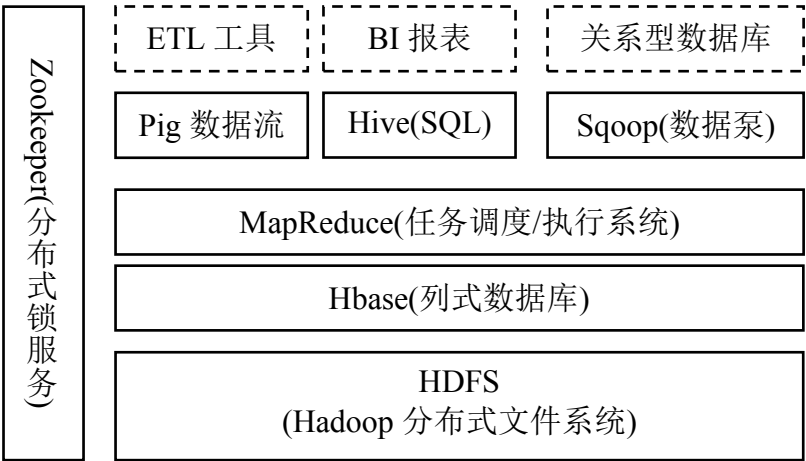
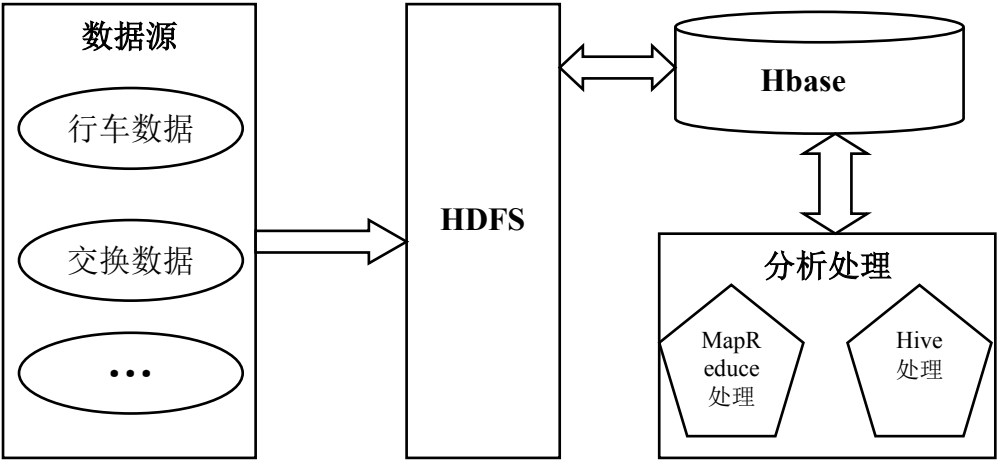


图2.2 Hadoop大数据平台架构

Hadoop 平台主要组件包含：MapReduce（并行数据处理框架）、HDFS（Hadoop 分布式文件系统）、HBase（Key-value 数据库）、ZooKeeper（分布式协调系统）、Hive（构建在 MapRudece 之上的数据仓库软件包）、Pig（高级数据处理层）、Sqoop 可改进数据的互操作性和其余部分。它们之间的关系如上图 2.2 所示。



2.3 Hadoop 关键组件间关系结构图

上图 2.3 是 Hadoop 平台关键组件之间的关系结构图。由 HBase 数据库对数据源进行管理，底层是 HDFS 支持数据分布式存储，根据处理任务的特征，执行具体任务，首先从 Hbase 中提取数据建立 Hive 数据仓库，然后进行分布式计算，其中运用的最为广泛的有 MapReduce、Hive 等任务。

Hive 是基于 Hadoop 的一个数据仓库工具, 可以将结构化的数据文件映射为一张数据库表, 并提供类似 sql 查询功能, 可以将 sql 语句转换为 MapReduce 任务进行运行。Sqoop 是一个用来将 Hadoop 和关系型数据库中的数据相互转移的工具, 可以将一个关系型数据库 (例如: MySQL, Oracle, Postgres 等) 中的数据导入到 Hadoop 的 HDFS 中, 也可以将 HDFS 的数据导入到关系型数据库中。HBase 利用 Hadoop HDFS 作为其文件存储系统, 利用 Hadoop MapReduce 来处理 HBase 中的海量数据, 利用 Zookeeper 作为协同服务。HBase 以表的形式存储数据, 表有行和列组成, 列划分为若干个列族/列簇(column family)<sup>[33]</sup>。

## 2.4 本章小结

本章介绍了商圈理论的相关内容, 包括有商圈的基本定义, 商圈经典理论, 如雷利法则, 康维斯断点理论和哈夫概率公式, 分析了商圈辐射范围计算所需要的数据进行了分析, 数据来源于城市交通物联网系统, 整个系统的数据量极大, 只取与商圈研究结合紧密的部分数据。大的数据量离不开好的大数据处理平台, 计算过程对数据的读写速度和存储性能都有较高的要求, 所以引入了 Hadoop 的大数据处理技术来作为后台, 为庞大的计算量作支撑。

## 第3章 商圈建设的分析与建模

本章将对商圈辐射范围的界定做出进一步的分析,由第二章理论部分的讨论可以知道,商圈的辐射范围边界可以通过雷利法则公式来计算得到,还有一种方法是本文提出的基于行车大数据的商圈辐射范围计算。下面分别对这两种方法进行详细介绍。

### 3.1 相关定义

在本小节中,涉及到几个常用的术语,为方便描述和理解,作如下定义。

#### 3.1.1 商圈区域

商圈区域:如图 3.1 所示,商圈区域是一个具有营业性质的商业企业聚集区域,从地理上来说,这个区域通常可以在地图上以城市路网中的道路为界限,形成一个封闭的不规则多边形,多边形所覆盖的区域即可明晰的界定出商业企业聚集区域,该区域包含了该商圈的几乎所有商业企业。这里的所指的商圈区域是商业企业的聚集区,可以在前期的规划阶段确定其范围,随着商圈的发展和后期规划,这个区域也会发生变化,所以在应用时要取对应时间段的数值。



图3.1 重庆解放碑商圈示意图

### 3.1.2 采集点

采集点：如图 3.2 所示，采集点是一个逻辑概念，由在地理位置上相近的多个采集方向聚合而成，比如采集点“科园六路（科园二街路口）”就包含 2 个方向：分别是奥体路方向和龙腾大道方向。有些采集点包含的采集方向更多，比如一些转盘路口，如采集点“火炬大道（云湖路路口）”有 4 个采集方向，当然也有一些采集点只有一个采集方向。

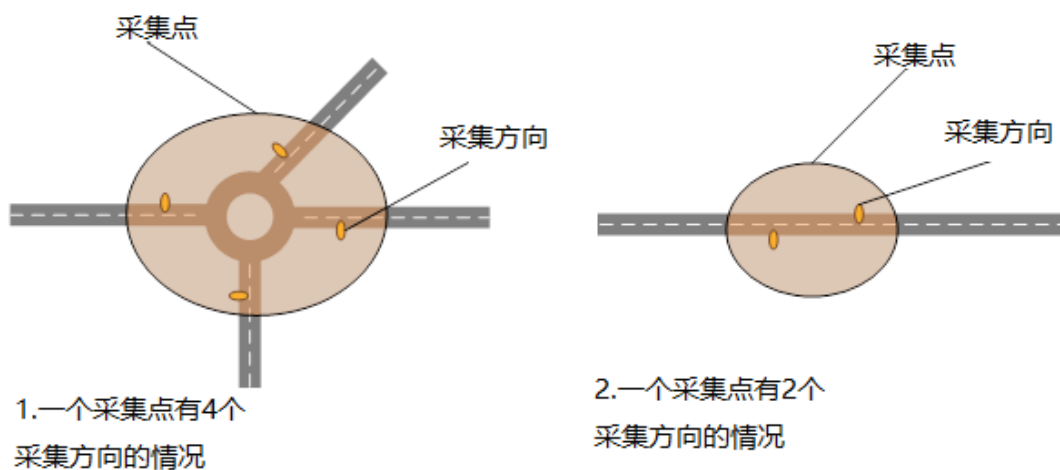


图3.2 采集点与采集方向示意图

### 3.1.3 采集方向

采集方向：如图 3.3 所示，采集方向也是一个逻辑概念，由一个或者多个 RFID 阅读器及其所带天线组成，采集方向负责采集某条道路一个过车方向的车流量。

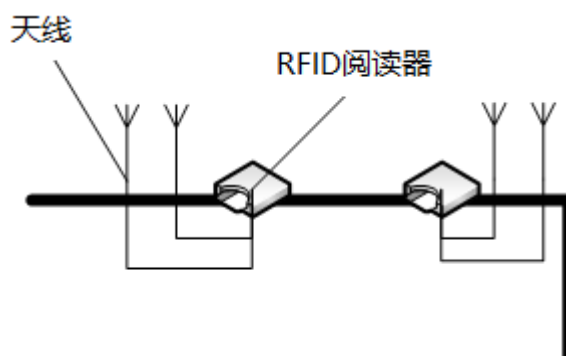


图3.3 采集方向设备简图

### 3.1.4 消费事件

消费事件：一个消费事件指的是用户在商圈内发生了消费行为。在本系统中，以停车时长代表该车辆拥有者在商圈消费时长，若车辆在商圈内停留时长在一定范围内，则判定该车辆在商圈消费（消费事件的判定方法在下文“商圈消费者数据集提取”章节详述），通过计算可以将商务办公、居民居住地等停留时间过长的数据集过滤，同时可以将路段拥挤、出租载客等停留时间过短的数据集过滤，从而得到相对准确的判定结果。消费事件描述包含用户到达商圈的时间，离开商圈的时间，还包含用户所选择的消费商圈。

## 3.2 数据集描述

城市交通物联网系统数据库中存储了大量的历史行车数据，还包括商圈附近停车场、过江大桥车流量等数据，系统后台通过大数据运算可以将历史数据中，满足一定筛选条件的数据找出来供统计分析使用。此外，由于重庆直辖市在商圈建设规划上有地方法的支持，相关组织机构也建设相当完善，所以有关商圈建设方面的数据，比如商圈年度社会消费品零售额、商圈商业商务面积大小、主营商业种类等数据可以通过查询官方网站和相关研究资料查询到。

(1) 行车数据：这个数据集包含了重庆市约 190 万辆机动车的日常行车记录，该数据通过安装在城市路网中关键结点处的采集点设备所采集，车辆经过采集点时触发采集设备生成一条与电子车牌和采集点信息相关联的行车数据，记录车辆行驶信息，也称过车数据，每天采集车辆通行记录约 8 千万条。根据对历史数据的统计，五大商圈范围内的数据约 2.2 千万条，约占全市机动车通行记录的 27.5%，节假日高峰时期可占到全市机动车通行记录的一半以上。

(2) 停车数据：这个数据集包含了从商圈范围内部分停车场停车系统的数据，从中取得具体停车记录信息，还包括了其他停车信息，如未实施信息化停车系统的停车场（路边停车、老旧停车场等），以及部分未参与本项目合作建设的停车场数据。上述中的其他停车数据将通过一定的过滤条件，从行车数据中分析得到，从而更加严谨的分析在商圈中与消费事件相关的机动车数据集的组成，下文中将作详细说明。

(3) 路网数据：这个数据集包含了重庆市的路网信息，本系统中通过采集点分布的属性来描述路网信息，其中在早期建设的有 234 个采集点以及 788 个采集方向，且根据计划，在未来两年采集点将增加到 1000 个，采集方向达到 2700~3000 个。在这些路网信息中，目前包含描述商圈的采集点数量有 22 个，采集方向的数量有 51 个，因此本文将重点关注这 22 个采集点和 51 个采集方向。现将部分数据呈现如下表 3.1 到 3.5 所示。

**表3.1 观音桥采集点信息**

序号	行政区	采集点名称	采集点方向
1	江北区	建新东路（324 医院）	鲤鱼池社区方向
			观音桥方向
2	江北区	建新东路茶园	鲤鱼池社区
			五里店
3	江北区	观鸿大道（燃气大厦）	观鸿大道方向
			小苑路方向
			观鸿大道往小苑路方向

附注：数据来自于项目数据库设计文档

**表3.2 解放碑商圈采集点信息**

序号	行政区	采集点名称	采集点方向
1	渝中区	打铜街（民族路路口）	东水门大桥方向
			新华路方向
			小什子轻轨站方向
2	渝中区	八一路（中华路路口）	民权路方向
3	渝中区	民生路（乐海家私）	解放碑方向
			两路口方向
4	渝中区	和平路（火药局路口）	金汤街方向
			中兴路方向
5	渝中区	临江门（重医附二院）	邹容广场方向
			重庆美术馆方向
6	渝中区	中兴路（二手市场）	南区路方向
			较场口方向

附注：数据来自于项目数据库设计文档



表3.3 南坪商圈采集点信息

序号	行政区	采集点名称	采集点方向
1	南岸区	南城大道	单行道方向
2	南岸区	南坪东路（现代女子医院）	女子医院方向
			福红路方向

附注：数据来自于项目数据库设计文档

表3.4 杨家坪商圈采集点信息

序号	行政区	采集点名称	采集点方向
1	九龙坡区	直港大道（天宝路口）	杨家坪横街方向
			九滨路方向
2	九龙坡区	石坪桥正街（港城静园）	石坪桥立交方向
			杨家坪环道方向
3	九龙坡区	西郊路（区红十字医院）	动物园东门方向

附注：数据来自于项目数据库设计文档

表3.5 沙坪坝商圈采集点信息

序号	行政区	采集点名称	采集点方向
1	沙坪坝区	汉渝路市一中	三峡广场
			肿瘤医院
2	沙坪坝区	渝碚路	杨公桥方向
			沙坪坝方向

附注：数据来自于项目数据库设计文档

(4) 商圈数据：这个数据集包含了从重庆市商务委信息发布平台、商圈管理办公室、重庆统计信息网等官方网站获得的商圈基础建设信息，包括了商圈商业面积、主营商业种类、年度社会消费品零售总额等信息。统计结果如下表 3.6 所示。

表3.6 重庆市五大商圈基础信息表

商圈名称	行政区（面积/常住人口）	商业面积	人流量（日均/节假日）	主营商业种类	社会消费品零售总额（2014年）
观音桥	江北区(221平方公里/84.98万)	核心区域 1.5 平方公里，商业 125 万平方米	35/80 万	“五大”主力百货和“三大”购物中心；国际知名品牌	330 亿元
解放碑	渝中区(23.24平方公里/64.95万)	解放碑 CBD 硬核区域 0.92 平方公里，商业面积 212 万平方米	30/100 万	高、中档及适合大众消费的多层次零售企业；世界顶级品牌高档品牌的至高点	414 亿元
南坪	南岸区(263平方公里/85.81万)	核心区域 1 平方公里，商业面积 130 万平方米	30/50 万	各类商业业态纷呈共荣，购物、会展、金融、休闲、娱乐、商务、旅游、餐饮和专业市场为一体的商贸中心；会展经济效益明显	247.1 亿元
杨家坪	九龙坡区(432平方公里/118.69)	核心区域 1 平方公里，商业面积 112 万平方米	30/50 万	一线品牌、百货等各大商场。规模小、档次低	240 亿元
沙坪坝	沙坪坝区(396.2平方公里/112.83)	核心区域 1.42 平方公里，商业面积 100 多万平方米	30/50 万	现代商贸服务业和高端商务服务业作为重点产业，零售、餐饮、住宿等相关行业将走向中高端层次	187 亿元

附注：数据来自于重庆市统计年鉴（2014年）、重庆市商务委公众信息网、重庆市商管理办公室网站

### 3.3 雷利法则估算商圈辐射范围

根据上文对雷利法则的介绍可以知道，雷利法则提出的背景是小型化城市占主要地位的上个世纪 40 年代，所以它对零售商圈的划分是按照城市边界来划分，换言之其研究主体是城市，核心思想是零售商圈吸引力与该城市的人口数量成正比，与居民

去该地区的距离是成反比的。从雷利法则的提出到如今，由于经过了几十年的现代化城市建设，雷利法则中的小型化城市商圈已经不再适用，且经过研究人员的实证研究，商圈的吸引力因素也在随着人类社会的发展而发生变化<sup>[34]</sup>，一些学者针对雷利法则中的吸引力分子只引入了人口密度因素，而没有考虑商圈自身建设的相关因素，提出以商圈销售面积、商店数目等新因素来完善雷利法则的实际应用，所以本文将现代化都市商圈的商业面积来作为雷利法则中城市人口的代替<sup>[35]</sup>。其一般性的公式表达如下(3.1)式所示：

$$T_a = \frac{\Delta T_{ab}}{\left(1 + \sqrt{S_b/S_a}\right)} \quad (3.1)$$

其中，

$T_a$ ：表示居民点到 A 商圈的时间距离；

$\Delta T_{ab}$ ：表示从 A 商圈到 B 商圈的时间距离；

$S_a$ 、 $S_b$ ：表示 A 商圈和 B 商圈的商业面积。

其推导过程与雷利法则到康维斯断点公式相似，由于本文篇幅有限，不再赘述。

为了在地图上得到居民点对应的地理位置，需要将时间距离  $T_a$  转换成实际的地理距离  $D_a$ ，其推导公式如下(3.2)式所示：

$$D_a = T_a \times \frac{L_{ab}}{\Delta T_{ab}} \quad (3.2)$$

其中，

$D_a$ ：从 A 商圈居民点的实际地理距离；

$L_{ab}$ ：表示商圈 A、B 之间的某一典型路径总长度；

$\Delta T_{ab}$ ：表示 A、B 商圈的时间距离。

商圈辐射范围的确定需要大量的市场数据的支撑，通常的做法是采用填写调查问卷的方式来采集用户的购买意图信息，然后做汇总分析得到大致的商圈范围。调查的内容包括住址、性别、年龄、年收入、购买力、职业、喜欢光顾的商圈、品牌种类等等。通过这类调查可以分析具有竞争力的两个商圈在地理辐射范围上面的优劣，界定它们之间的商圈界限。但是这类调查通常有以下 5 个方面的缺陷：

- (1) 需要耗费的人力物力成本巨大，时间周期长；
- (2) 需要专业市场分析人员设计详尽的调查问卷，如果因为问卷不具针对性而导致最终分析结果缺乏准确性，则往往需要重新进行整套流程；
- (3) 目标人群难以区分，不能保证采集到的数据对本商圈的分析是否有贡献，又或者是有害数据；
- (4) 用户参与度不高，产生大量的冗余数据，筛选困难；
- (5) 用户填写问卷主观因素影响较大，有害数据难以区分，对分析结果产生很大干扰。

综上所述，人工调查的方法成本巨大，而且调查结果很难达到预期，所得到的数据不能动态反映长期变化的趋势，对商圈长远建设只具有阶段性的指导意义，不具备长期性。若以城市交通物联网系统采集的行车记录作为数据源，提取与商圈有关的数据来研究商圈建设问题，则能很好的弥补上述人工调查方法的缺陷。

### 3.3.1 时间距离候选集提取方法

通过观察公式 3.1 可知，雷利法则推断商圈辐射范围与两个商圈 A、B 之间的时间距离  $\Delta T_{ab}$  呈正相关，与两个商圈的商业面积比值  $S_b/S_a$  呈负相关。根据上一小节的商圈数据集就可以很容易知道商圈的商业面积，所以利用本公式的重点在于如何获得 A、B 商圈的时间距离  $\Delta T_{ab}$ ，本文拟采用的处理流程如下图 3.4 所示。

以城市交通物联网系统中数据库存储的海量历史行车数据为根据，通过一系列的筛选控制，可以得到经过指定路径由起始商圈到达目标商圈所花费的时间，即雷利法则中代表两商圈之间的时间距离  $\Delta T_{ab}$ 。

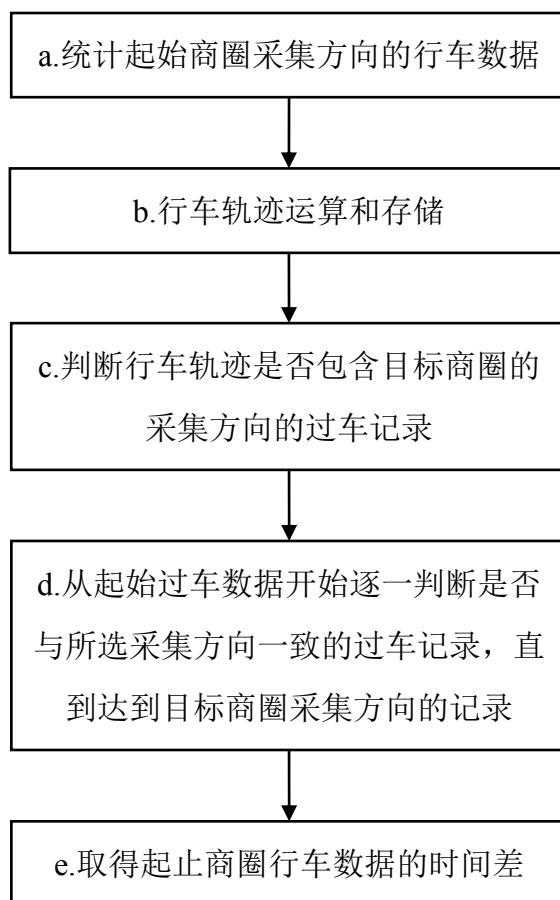


图3.4 两个商圈之间时间距离分析方法流程框图

下面对“基于行车数据的商圈间时间距离分析方法”的具体流程作详细分析如下：

a) 统计起始商圈采集方向的行车数据

起始商圈 A 边界路口的采集方向可分为进入商圈和出商圈两种，这里以 A 商圈为起始商圈所以取得的数据是出商圈的采集方向所采集的行车数据，得到一个经过 A 商圈的机动车辆数据集。

b) 行车轨迹运算和存储

以车牌号码为依据对全部行车数据进行分组，并按过车时间排序，通过对每辆车的行车数据进行聚合操作，可以计算得到该车辆的完整行车轨迹，将行车轨迹以“时间-采集方向-车牌号”为行键，以“时间-采集方向-行车数据”为列，建立行车轨迹数据序列并用 Hbase 数据库进行储存。

c) 判断行车轨迹是否包含目标商圈的采集方向的过车记录

通过判断每一辆车的轨迹数据中是否包含目标商圈 B 的过车记录，同步骤 a 中所

述，B 商圈的过车记录必须是进商圈的采集方向所采集的行车记录。

d) 从起始过车数据开始逐一判断是否与所选采集方向一致的过车记录，直达到目标商圈采集方向的记录

判断该车辆是到达目标商圈 B 的行车轨迹后，再按时间顺序从头开始逐一判断每一个采集方向的行车记录是否与选定的路径中分布的采集方向相匹配，该匹配规则为相对顺序匹配，即在组成该条行车轨迹的众多过车记录中，只要顺序包含指定路径的采集方向的信息即为匹配成功，当行车记录最后与 B 商圈的采集方向相匹配时，则判定该行车轨迹是按照指定路径由起始商圈 A 到达目标商圈 B。

e) 取得起止商圈行车数据的时间差

判定车辆的行车轨迹与指定路径相吻合后，分别取 A、B 商圈采集方向产生的行车记录，两者的时间之差即为 A、B 商圈之间的时间距离样本。

### 3.3.2 关键算法实现

商圈间时间距离的提取，从技术实现角度来看主要分为三个关键步骤：(1)行车轨迹的提取；(2)行车轨迹数据清洗；(3)商圈时间距离的计算。下面分三个部分介绍：

#### (1) 行车轨迹的提取

上述基于行车数据的商圈间时间距离分析方法，步骤 a 中主要是对海量的行车数据进行数据清洗，将 A 商圈的出商圈方向的行车数据过滤出来，得到一个可以以机动车为对象的数据集，每一个对象可以通过车牌号码唯一识别，从而得到 b 步骤中的每一个机动车的行车轨迹数据，该数据集中的机动车包含了关联 A、B 商圈的机动车。数据清洗的策略较为简单，可以通过 MapReduce 计算模型进行处理，MapReduce 编程模式可以充分利用 Hadoop 平台集群分布式计算的优势，提高数据处理的时效性，按照指定的分类方式和筛选条件，高效的实现对数据的分类与整合。步骤 a 的数据处理通过一次 MapReduce 操作即可完成。Map 阶段，选择行车数据的车牌号码作为 key，每一条行车记录的其余属性作为 value，将原始行车数据拆分成若干记录。再检查行车记录的 value 是否满足条件，即采集方向编号是否为指定的出 A 商圈方向的行车记录，对于不满足条件的数据项进行遗弃，将正常的数据发送至管道，进入 Reduce 过程。Reduce 阶段将清洗过的数据按照车牌号码进行归并，车牌号相同的记录被归并

成一条记录，最终得到步骤 a 的输出结果。

步骤 a、b 中所述的行车数据与行车轨迹是一一对应，步骤 a 中的机动车辆数据集所描述的车辆都可以通过商圈 A 中采集到的行车数据计算得到一条行车轨迹数据，该行车轨迹数据包括当前过车时间之后的所有行车数据。

步骤 b 中通过 MapReduce 算法模型对行车数据进行归并，以车牌号码为维度计算出步骤 a 得到的数据集中每一辆车的行车轨迹，利用分布式数据库 Hbase 面向列的特点，充分利用 Hbase 数据库列可以无限扩展的优势，将行车轨迹数据按车牌号为主键存储到 Hbase 中。步骤 b 所得到的以“时间-采集方向-车牌号码”为行键，以“时间-采集方向-行车数据”为列的行车轨迹，将行车轨迹按照时间从小到大的顺序排列得到的数据格式如下表 3.7 所示。

**表3.7 行车轨迹数据存储格式表**

行键	列族			
	列 1	列 2	列 3	列 ...
时间 T11_采集 方向 K1_车辆 C1	时间 T11_采集方向 K1_行车数据主键 KEY11			...
时间 T12_采集 方向 K2_车辆 C1	时间 T11_采集方向 K1_行车数据主键 KEY11	时间 T12_采集方向 K2_行车数据主键 KEY12		
时间 T13_采集 方向 K3_车辆 C1	时间 T11_采集方向 K1_行车数据主键 KEY11	时间 T12_采集方向 K2_行车数据主键 KEY12	时间 T13_采集方向 K3_行车数据主键 KEY13	
时间 T21_采集 方向 K1_车辆 C2	时间 T21_采集方向 K1_行车数据主键 KEY21			
时间 T22_采集 方向 K4_车辆 C2	时间 T21_采集方向 K1_行车数据主键 KEY21	时间 T22_采集方向 K4_行车数据主键 KEY22		

表3.7 行车轨迹数据存储格式表（续表）

时间 T31_采集 方向 K1_车辆 C3	时间 T31_采集方向 K1_行车数据主键 KEY31			
时间 T32_采集 方向 K3_车辆 C3	时间 T31_采集方向 K1_行车数据主键 KEY31	时间 T32_采集方向 K3_行车数据主键 KEY32		
...				

由于每辆车在行驶的过程中所经过的采集方向是不确定的，所以车辆每经过一个采集方向时，该车当前时间在当前采集方向所形成的行车轨迹也是与之前经过其他采集方向时的行车轨迹是有差别的。所以车辆在经过每一个采集方向时，都会计算得到一条长度各异的行车轨迹，在上表中的表现就是每一行数据的列数是大小不一的。并且如前所述，每一条行车轨迹与过车数据是一一对应的，即采集方向每采集一次过车数据都计算出一条这样的行车轨迹：包含经过时间小于等于当前时间的所有过车数据。以车辆 C1 为例进行说明，车辆 C1 在 T13 时间经过了采集方向 K3，那么将生成一条行车轨迹数据（表格中行键为“时间 T12\_采集方向 K2\_车辆 C1”的记录），该行车轨迹包含了它之前经过的采集方向的过车数据，可以看出车辆 C1 在经过采集方向 K2 之前，在 T11 时刻经过了采集方向 K1，将这些采集方向所在的点以时间顺序连起来即可完整的描述车辆 C1 在 T11 到 T12 时间内的行车路线图。

## (2) 行车轨迹数据清洗

由行车轨迹数据存储格式表不难看出，步骤 b 得到的车辆轨迹数据集包含了大量的冗余数据，为了得到目标商圈之间的行车轨迹，需要对行车轨迹数据再次清洗，即步骤 c 的操作，这个处理过程与步骤 a 中类似。对步骤 b 产生的行车轨迹数据集采用 MapReduce 方式进行处理。Map 阶段，以行车轨迹数据的行键“时间-采集方向-车牌号”为 key，轨迹数据的其他属性作为 value，将行车轨迹数据拆分成若干记录，这里同样需要设置过滤条件，只取以目标商圈为起点-终点的行车轨迹数据，并以时间顺序从小到大进行排序。在过滤掉异常数据之后将满足条件的数据发送至管道。接下来



Reduce 阶段，将轨迹数据以行键“时间-采集方向-车牌号”中的车牌号分组。经过步骤 c 后产生新的数据集描述了以目标商圈为起点-终点的所有行车轨迹，其中包含了指定路径的行车轨迹。

下面举例说明，假设车辆 C1 是步骤 c 中得到的行车轨迹数据，满足起点-终点为目标商圈 A、B，车辆 C2 为任意其他轨迹数据，由流程框图描述可知，C2 只满足起点为 A 商圈，不满足终点是 B 商圈，对比分析车辆 C1、C2 的行车轨迹数据，如下表 3.8 所示。

表3.8 行车轨迹数据表一

	车辆 C1	车辆 C2
行键	时间 T1b_采集方向 Kb_车辆 C1	时间 T1n_采集方向 Kn_车辆 C2
列族	时间 T11_采集方向 K1_行车数据主键 KEY11	时间 T21_采集方向 K1_行车数据主键 KEY21
	时间 T12_采集方向 K2_行车数据主键 KEY12	时间 T22_采集方向 K3_行车数据主键 KEY22
	时间 T13_采集方向 K3_行车数据主键 KEY13	时间 T23_采集方向 K5_行车数据主键 KEY23
	时间 T14_采集方向 K4_行车数据主键 KEY14	时间 T24_采集方向 K6_行车数据主键 KEY24
	...	...
	时间 T1b_采集方向 Kb_行车数据主键 KEY1b	时间 T2n_采集方向 Kn_行车数据主键 KEY2b

表中车辆 C1 的行车轨迹的起始点是采集方向 K1 所在的位置，根据假设 K1 是商圈 A 的出商圈采集方向，终点是采集方向 Kb 所在的位置，那么 Kb 就是进商圈 B 的采集方向。行车轨迹数据是按照时间顺序排列，所以可以看出车辆 C1 先后经过了采集点 K1、K2、K3、K4...最终到达 Kb，满足步骤 c 数据清洗的条件；对于车辆 C2，可以由行车轨迹数据知道，它从商圈 A 出发，依次经过采集方向 K1、K3、K5、K6...最终到达 Kn，Kn 不在商圈 B 内，所以该条记录将被遗弃。

### (3) 商圈时间距离的计算

步骤 d 中将获得的目标商圈行车轨迹数据与指定路径上的采集方向进行对比判断, 得到所有通过指定路径到达目标商圈的轨迹数据。首先依次将轨迹数据提取、加载到计算节点的缓存中, 然后将行车轨迹数据中的每一条过车记录包含的采集方向与指定的一组采集方向按顺序进行对比判断, 如果两者吻合则断定该条行车轨迹满足指定路径由起始商圈到达目标商圈, 经过步骤 d 又可以获得一个新的数据集, 该数据集描述了以目标商圈为起点-终点的经指定路径行驶的行车轨迹。

下面举例说明, 假设车辆 C1 是经由指定路径从 A 商圈到达 B 商圈, 车辆 C2 是经由其他路径从 A 商圈到达 B 商圈, 与一组指定的采集方向对比, 分析车辆 C1、C2 的行车轨迹数据, 如下表 3.9 所示。

表3.9 行车轨迹数据表二

	车辆 C1	车辆 C2	指定路径的采集方向 序列
行键	时间 T1b_采集方向 Kb_车辆 C1	时间 T2b_采集方向 Kb_车辆 C2	
列族	时间 T11_采集方向 K1_行车数据主键 KEY11	时间 T21_采集方向 K1_行车数据主键 KEY21	采集方向 K1
	时间 T12_采集方向 K2_行车数据主键 KEY12	时间 T22_采集方向 K3_行车数据主键 KEY22	采集方向 K2
	时间 T13_采集方向 K3_行车数据主键 KEY13	时间 T23_采集方向 K5_行车数据主键 KEY23	采集方向 K4
	时间 T14_采集方向 K4_行车数据主键 KEY14	时间 T24_采集方向 K7_行车数据主键 KEY24	采集方向 K6
	...	...	
	时间 T1b_采集方向 Kb_行车数据主键 KEY1b	时间 T2b_采集方向 Kb_行车数据主键 KEY2b	采集方向 Kb

由上表中可以看出车辆 C1 的采集方向的时间序列是 K1-K2-K3-K4...Kb, 对比指定采集方向序列 K1-K2-K4-K6...Kb 可以发现两者并不相符, 但是 C1 车辆的这条行

车轨迹符合筛选条件，即满足采集方向的相对顺序匹配，这样做的原因是采集方向在某些情况下会产生误差数据，例如同一个采集点由于地理条件复杂，设置了多个采集方向，车辆经过 K4 采集方向时会触发 K3 采集方向也产生一条行车记录。再看车辆 C2 的采集方向的时间序列是 K1-K3-K5-K7...Kb，与指定采集方向序列相差很大，作为垃圾数据被遗弃。

步骤 e 的处理相对较简单，直接将步骤 d 获得的数据集中的数据通过加权取平均值计算得到目标商圈之间的时间距离，计算公式如下 3.3 式所示：

$$\Delta T_{ab} = \frac{\sum_{i=1}^n (T_{ib} - T_{i1})}{n} \quad (3.3)$$

### 3.3.3 雷利法则确定商圈辐射范围流程总结

本小节介绍了行车大数据结合修正的雷利法则划分商圈的辐射范围的方法，主要介绍了如何利用行车大数据计算雷利法则中的两商圈时间距离的实现步骤。下面将商圈辐射范围的确定过程用流程图表示如下图 3.5 所示。

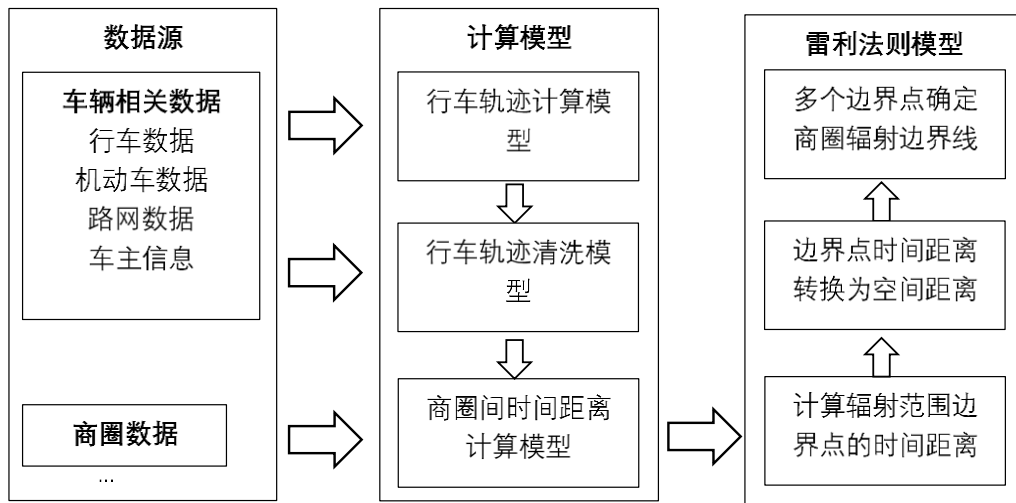


图3.5 雷利法则确定商圈辐射范围流程图

为了更直观的显示商圈辐射范围，需要在得到边界点集之后在地图上将边界点依次连接起来，形成阻断两个商圈的边界线，但是这只能确定商圈间的边界线，对于外延方向的边界点，该公式是无法确定的，通常的做法是取已经获得的时间距离的平均

值为半径，将外延方向的商圈辐射边界确定下来。示例图如下图 3.6 所示。



图3.6 解放碑商圈辐射范围边界示例图

### 3.4 基于行车数据估算商圈辐射范围

上面介绍了如何利用城市交通物联网系统的行车数据来估算商圈的辐射范围，理论依据还是商圈理论中经典的雷利法则。下面提出一种新的估算商圈辐射范围的方法，该方法通过对数据源关联关系的分析，以及对行车数据的运算汇总等一系列处理，可以直接计算得到经常光顾该商圈的消费者居住点分布情况，从而分析得到商圈的辐射范围。

根据上面提出的思路，可以很容易的发现该过程分为 3 个部分：(1)确定商圈的消费者数据集；(2)确定消费者居住点数据集；(3)确定商圈辐射范围边界。其中重点在于如何确定商圈消费者数据集和商圈消费者居住点分布两个问题，下面分别对以上三个问题的解决思路作详细描述。

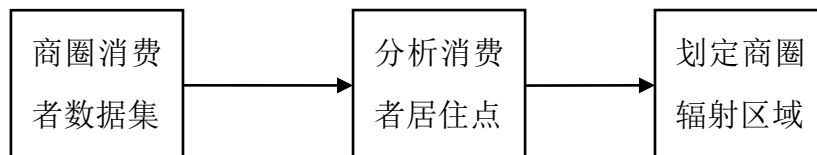


图3.7 从行车大数据确定商圈辐射范围

### 3.4.1 商圈消费者数据集提取

根据本章对数据集的描述可以知道,城市交通物联网系统数据库中包括了商圈范围内停车场的停车数据,该数据描述了某一号牌车辆在何时进入商圈停车场,以及在何时离开停车场,可以以该数据为基础数据来提取商圈消费者数据集。由于种种客观原因(如前 3.2 数据集描述小节所述),商圈停车场数据并不能代表整个商圈的所有参与消费事件的机动车辆数据集,所以需要扩大数据集,剩下这部分数据可以由行车数据经过一定的算法筛选出来补充到商圈消费者数据集中。

首先商圈停车场所停车辆必然可以看作是来商圈消费的用户,顺着这个思路,这一类用户的停车时长可以看作是该类用户(私家车用户)在商圈内的消费时长,那么可以推断只要是与该消费时长吻合的商圈内停车数据均可以看作是来商圈消费的私家车用户,换言之商圈消费者数据集得到扩充。

为了提取完整的商圈消费者数据集,本文拟采取的处理流程如图 3.8 所示下。

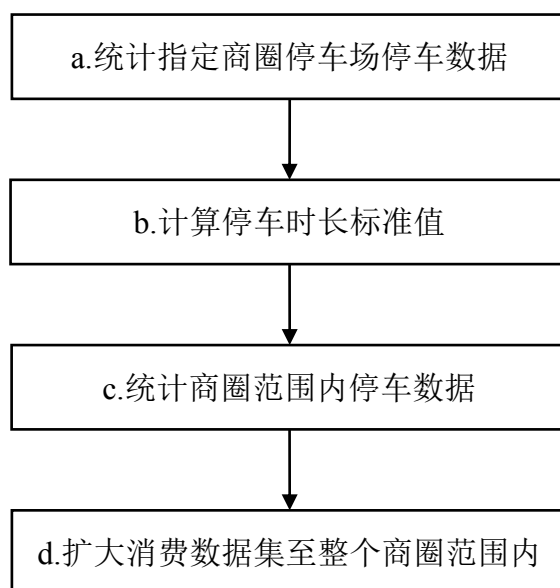


图3.8 商圈消费者数据集提取流程图

以城市交通物联网系统数据库中合作商圈停车场数据为据,并结合行车大数据,通过一系列筛选控制,可以得到整个商圈参与消费事件的所有消费者数据集,该数据

集描述了所有驾驶私家车到指定商圈消费的消费者。下面对“基于行车数据的商圈消费者数据集提取方法”的具体流程作如下描述：

a) 统计指定商圈停车场停车数据

本操作的目的是从所有的停车场数据中筛选出指定商圈停车场的停车数据，操作方法比较简单，首先要确定提取的停车场 ID 和停车数据的时间周期，一般时间周期为一个月，以停车记录的停车场 ID 作为分类条件，将原始数据按停车场归类，提取到指定的商圈停车场的停车数据，这个数据集描述了该时间周期内指定停车场的所有停车数据；

b) 计算停车时长标准值

首先要明确停车时长标准值是一个时间范围，这个时间范围代表了商圈消费者的消费时长取值范围。首先将停车数据集加载到计算节点，以停车记录的主键作为分类条件，计算每一条记录的停车时长，并存入缓存中，再将所有的停车时长累加除以停车记录的条数，得到一个停车时长标准值（根据不同消费季和商圈类型，范围处于 80-150 分钟），以这个标准值为中心分别向前后扩展 10 分钟得到一个标准停车时长区间，根据方法的描述，这个标准停车时长区间将作为商圈行车数据的筛选条件。

c) 统计商圈范围内停车数据

本操作需要先获得指定时间周期内的商圈范围内行车数据，然后以标准停车时长区间为筛选条件来区分该机动车是否参与了消费行为，抑或只是途径商圈、在商圈内的商务区上班等其他行为。本操作涉及数据量庞大，需要借助 MapReduce 编程模式来完成，首先需要得到车辆进出本商圈的时间间隔数据集，也即商圈内停车时长数据集，这个过程与上一小节的商圈之间时间距离本质上是一样的，所以这里不再赘述，需要说明的是本过程不需要经过指定路径进出商圈，且一个商圈的采集方向可以判别车辆是驶进商圈还是驶离商圈，所以省去了“商圈间时间距离分析方法中的步骤 d”。

d) 扩大消费数据集至整个商圈范围

在得到停车时长数据集后再以“车牌号”为 key，以“停车时间区间”为 value 进行 Map 操作，其中 value 需满足在 b 步骤中得到标准时间区间条件，Reduce 过程将清洗后的数据按照“车牌号”进行归并，车牌号相同的数据将被归并为一条数据，最后获得整个商圈内消费车辆数据集，该数据集描述了商圈参与消费事件的所有车辆

信息。

### 3.4.2 分析消费者居住点

在前一小节中得到了商圈消费者数据集，通过车牌号可以将行车数据串联起来形成一条车辆轨迹数据，如果能找到车辆轨迹数据的一般规律的话，就可以进一步分析用户的出行行为习惯。例如用户在一周内的工作日具有很强的规律性，每天下班的时间是行车数据最后一次出现在居住小区附近的采集点，每天的上班时间是行车数据第一次出现在居住小区附近的采集点，更进一步，假设用户每天的最后一条行车数据与第二天第一条行车数据相差大概八小时以上，并且两条记录的采集点相同或相邻很近，那么可以得出结论：该用户的日常居住点就是采集点所在的区域。

以此为思路，结合城市交通物联网系统数据库中的行车大数据，推断用户居住点的分析方法流程图如下图 3.9 所示。

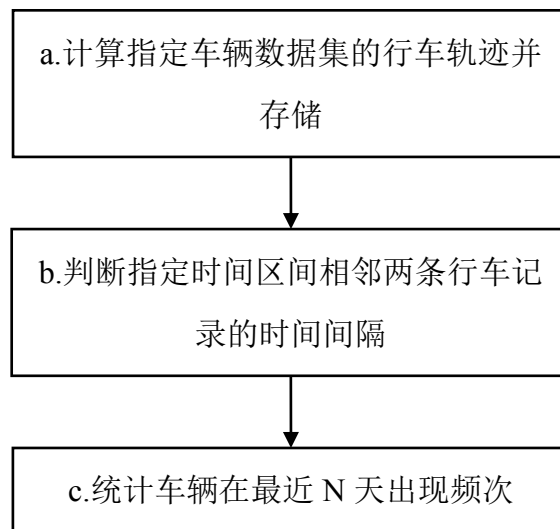


图3.9 判定消费者居住点流程图

以城市交通物联网系统数据库中的行车数据为根据，通过一系列筛选控制可以分析得到商圈消费者居住点分布情况。通过本章前面对数据集的描述可以知道，这里消费者居住点的分布是与消费者所在小区附近的采集点分布相对应，也就是说以小区附近采集点所在位置来代替消费者居住点。下面对“基于行车数据的商圈消费者居住点分析方法”的具体流程作如下描述：

a) 计算指定车辆数据集的行车轨迹并存储

这里的指定车辆数据集与上一小节得到的商圈消费者数据集是一致的,拿到所有参与消费事件的车辆信息后再计算它们的行车轨迹,这个行车轨迹是以天为单位的行车轨迹。这部分算法的实现可以参看“关键算法实现”小节的“行车轨迹提取方法”,需要说明的是,这部分的时间间隔  $T_{i1}-T_{i2}$  的取值必须满足:(1)区间长度能包含一天 24 小时;(2)区间能包括用户在当天第一条行车数据与前一天的最后一条行车数据。这里取值为当天中午 12 点( $T_{i1}$ )到第二天中午 12 点( $T_{i2}$ ),这样就得到了用户在 24 小时之内的行车轨迹,将冗余的数据清洗过后存储到 Hbase 数据库中,为下一步计算提供数据基础。

b) 判断指定时间区间相邻两条行车记录的时间间隔

在得到车辆一天之内的行车轨迹数据之后,就需要估算用户一天之内在家休息的时间段长度,这里假设一般用户在下班之后直到第二天上班的时间间隔在 8-14 小时,所以若当天第一条行车记录与前一天的最后一条行车记录的时间间隔大于等于 8 小时,且两条记录来自同一采集点或相邻采集点,那么就判定该用户可能居住在采集点附近,具体是否居住在此地还需要根据一个月内出现的频次来确定。示意图如下图 3.10 所示。

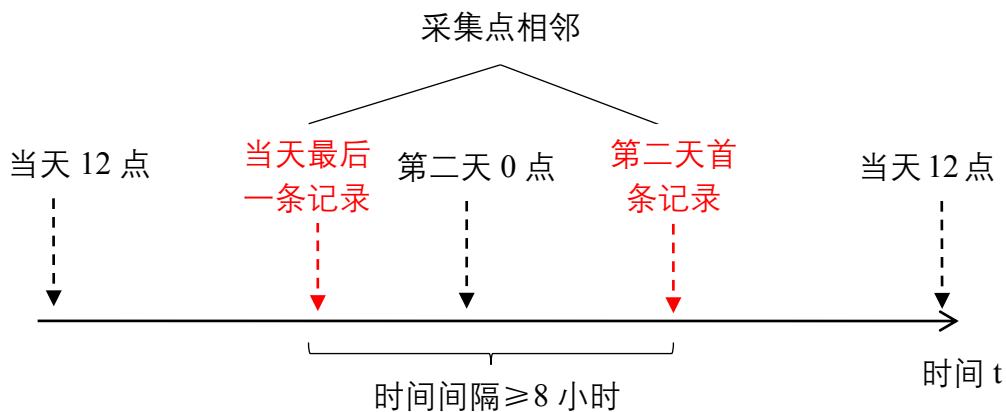


图3.10 行车轨迹判别用户居住点示意图

举例说明,假设车辆 C1 在不同日期的两条行车数据如下表 3.10 所示。



表3.7 行车轨迹数据存储格式表

	车辆 C1 (日期 D1)	车辆 C1 (日期 D2)
行键	时间 T_采集方向 K_车辆 C1	时间 T_采集方向 K_车辆 C1
列族	时间 T11_采集方向 K1_行车数据主键 KEY11	时间 T21_采集方向 K1_行车数据主键 KEY21
	时间 T12_采集方向 K2_行车数据主键 KEY12	时间 T22_采集方向 K3_行车数据主键 KEY22
	时间 T13_采集方向 K3_行车数据主键 KEY13	时间 T23_采集方向 K5_行车数据主键 KEY23
	时间 T14_采集方向 K4_行车数据主键 KEY14	时间 T24_采集方向 K7_行车数据主键 KEY24
	...	...
	时间 T1m_采集方向 Km_行车数据主键 KEY1m	时间 T2n_采集方向 Kn_行车数据主键 KEY2n

由上表可以知道车辆的行车数据已经按照“时间-采集方向-车牌”中时间大小顺序排列好，然后判断每条轨迹的第一条数据与最后一条数据的时间间隔  $T1m-T11$  时候超过 8 小时，且同时判断两条数据中采集方向  $K1$ 、 $Km$  所在采集点是否相邻或相同，若二者任意之一不满足，则该条行车轨迹被清除；若满足条件则该条记录被记录到出现次数统计队列中。例如车辆 C1 在 D1 日期不满足筛选条件，在 D2 日期满足，则统计队列中 D1 日期的轨迹数据被视为无效的 0，D2 日期的轨迹数据有效的 1。

这样就得到一个由 N 天行车数据经过分析筛选后的统计数据集，这个数据集表示车辆在 N 天中一共在采集点附近地区停留的频次。

#### c) 统计车辆在最近 N 天出现频次

在上一步中已经得到一个统计停留次数的统计表，为了判断用户是否真的定居在该区域，停留频次需要满足一个阈值条件，例如一周之内的工作日回家，休息日户外出行，那么此时的 N 取 7，阈值条件就是 5；现在以月为单位 N 值取 30，阈值条件可以取 20。也就是说统计一个月之内，判定用户在同一居住点的次数超过 20 次，那

么就可以断定该用户定居于此。

至此，已经完成了消费者居住点判定工作，消费者居住点数据的一般特点是居住点具有长期性，所以需要一张专门的表来维护，定期更新即可，这样可以减少计算节点的压力。

### 3.4.3 划定商圈辐射区域

本部分的思路比较简单，只需要将得到的商圈消费者居住点信息在地图上描出来，根据商圈分级理论在地图上大致圈出商圈消费者所在区域即可。流程图如下图3.11所示。

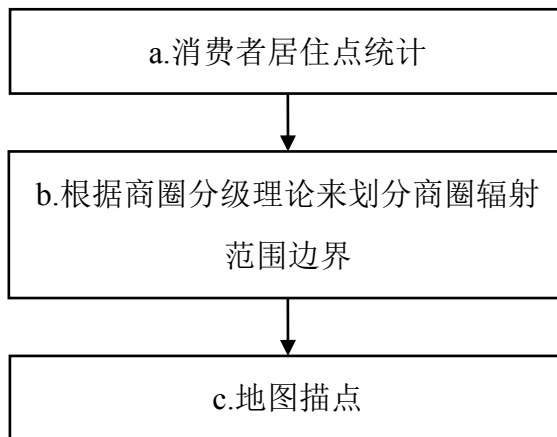


图3.11 界定商圈辐射范围流程图

以城市交通物联网系统数据库中行车数据确定出来的消费者居住点信息为根据，通过简单的筛选分类，结合商圈分级理论，可以得到商圈辐射范围的边界，从而确定商圈辐射范围。下面对“基于行车数据的商圈辐射范围确定方法”的具体流程作如下描述：

#### a) 消费者居住点统计

通过前面一系列的数据分析、数据处理过程，现在已经知道商圈消费者的居住信息，为了根据商圈分级理论确定商圈辐射范围，还需要对数据进行分类排序等统计操作。首先将居住信息表按照采集点分类，并统计同一采集点采集到的消费者数目，然后按照从大到小的顺序排序，得到每个采集点附近的消费者数量统计结果集。

## b) 根据商圈分级理论来划分商圈辐射范围边界

商圈分级理论<sup>[36]</sup>是按顾客总数占整个商圈的比例来划分商圈的影响范围的,可以分为3个层次,即中心商圈、次级商圈和边缘商圈。中心商圈占商圈消费者总人数的55%~70%,是最靠近商圈核心的区域,每个消费者的平均购买额也最大。次级商圈占到商圈消费者总人数的15%~25%,是位于中心商圈外围的环形区域,消费者分布较为分散。边缘商圈即是处于最外层的商圈,拥有的消费者最少,而且最为分散。

根据商圈分级理论,可以将步骤a中的消费者数量统计结果集按照一定比例取出排名靠前的记录,再将对应的采集点的位置信息记录下来就可以知道商圈的辐射边界了。下图3.12以A商圈消费者为例,横坐标表示采集点编号,共5个采集点,纵坐标表示每个采集点消费者数量,通过计算排名靠前的几个采集点消费者总数是否占整个商圈消费者总数的85%来判断该采集点是否被纳入商圈辐射范围。下表中的K1、K2、K3、K4选为辐射范围内采集点,K5在剩余的15%中,被当做边缘商圈采集点,故被遗弃。

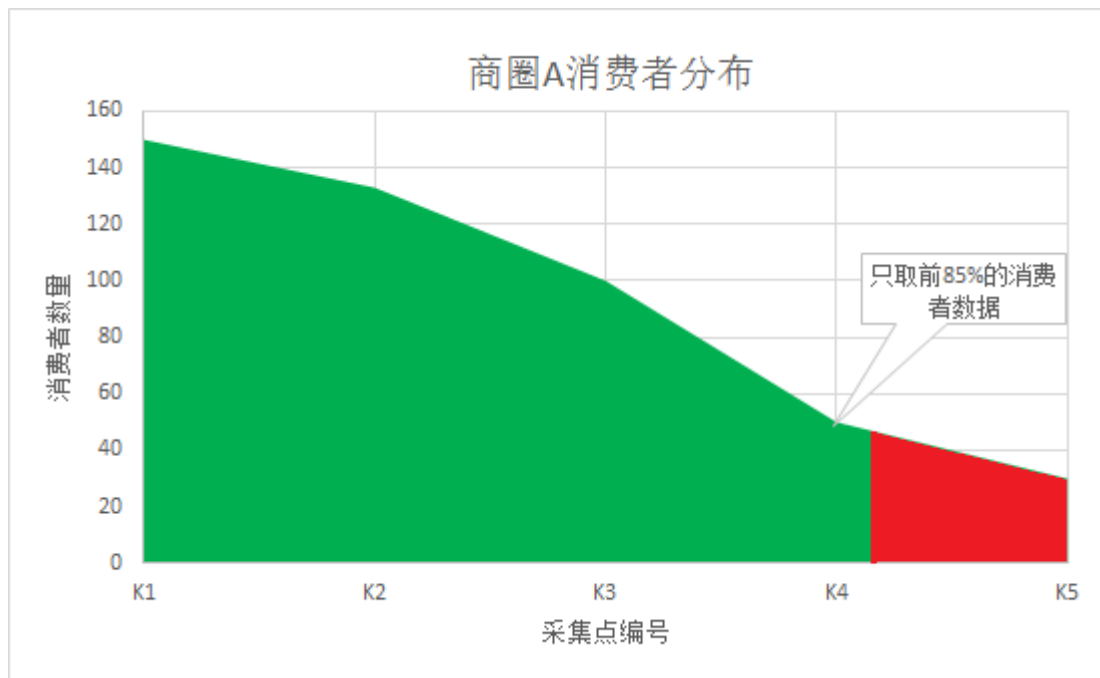


图3.12 商圈A采集点消费者分布统计图

## c) 地图描点

在得到商圈辐射范围内的采集点信息后,需要结合地理位置信息将采集点分布到

地图上，以图形界面形式将数据呈现出来可以更直观的反映商圈的辐射范围，示例图如下图 3.13 所示。



图3.13 基于行车大数据确定商圈辐射范围示意图

### 3.5 本章小结

本章介绍了两种确定商圈辐射范围的方法，一种是基于雷利法则的商圈辐射范围估算方法，另一种是基于行车大数据的商圈辐射范围估算方法。并对两种方法的整个处理流程作了详细介绍，介绍了如何将行车大数据与商圈建设结合起来，进一步分析如何通过行车数据求得商圈的辐射范围，是本文研究的重点。

## 第 4 章 商圈辐射范围计算过程与结果分析

### 4.1 解放碑商圈辐射范围计算

前面介绍了计算商圈辐射范围的两种方法，现在通过城市交通物联网系统中的数据来分别用两种方法计算商圈的辐射范围，并将结果进行分析。

#### 4.1.1 通过雷利法则计算商圈辐射范围

数理模型是一种理想化的研究方法，需要在一定的假设条件下才能得到方便表达的理想情况下的研究结果，该结果对实际应用具有指导意义。由于理想化的研究方法与实际情况必然存在差异，所以在得到预期的研究结果时往往需要约定一些假设前提。假设前提：

- (1) 约定研究对象为重庆市拥有私家车的家庭，以该对象作为商圈的消费者，对象产生的数据主要包括行车数据、机动车数据和通过机动车关联的车主信息；
- (2) 约定商圈间的时间距离通过指定的典型路径来计算，典型路径需满足条件：  
a. 相邻商圈的若干路径呈均匀对称分布，尽量不要出现重叠情况；  
b. 典型路径需要选择车流量大的主干道，能代表从居民区到达商圈的普遍出行情况；
- (3) 外延方向的商圈辐射边界通过对已知数据集求平均值的方法来确定。

根据典型路径选取条件，考虑到与解放碑商圈相邻（存在竞争关系）的商圈主要是位于西北方向的观音桥商圈、位于南面的南坪商圈，那么分别以解放碑商圈-观音桥商圈组合和解放碑商圈-南坪商圈组合选取典型路径。

下图 4.1 是重庆五大商圈地理分布图，可以反映五大商圈在地理上的相对位置，便于选取合适的典型路径。

下图 4.2 是本小节的研究对象——解放碑商圈与观音桥商圈组合的 3 条典型路径，这 3 条路径能够反映驾驶者从解放碑到观音桥的普遍选择。



图4.1 重庆五大商圈地理分布图



图4.2 解放碑-观音桥典型路径选取示意图

下图 4.3 是解放碑商圈与南坪商圈组合的 2 条典型路径，这 2 条路径能够反映驾驶者从解放碑到南坪的普遍选择。





图4.3 解放碑-南坪典型路径选取示意图

选取好合适的典型路径后统计雷利法则计算公式所需要的数据，这里取本文第3.2 小节中表 3.6 “重庆市五大商圈基础信息表” 中的商圈数据。

商圈间时间距离的计算需要行车数据支持，选择一定时间内（比如 2014 年 12 月份）的行车数据计算得到  $\Delta T_{ai}$  的结果， $\Delta T_{ai}$  表示从解放碑到观音桥/南坪的平均时长。计算结果是根据 3.3 小节中介绍的方法，利用大数据平台强大的数据处理能力得到的。现将数据统计如下表 4.1 所示。

表4.1 解放碑商圈计算数据一览表一

商圈	典型 路径	商圈时间距 离 $\Delta T_{ai}$ (单 位: 分钟)	商业面积 $S_a(S_i)$ (单位: 万平方米)	距商圈的时 间距离 $T_a$ (单 位: 分钟)	商圈间空间 距离 $L_{ai}$ (单 位: 公里)	距商圈的空 间距离 $D_a$ (单位: 公 里)
解放碑(A)- 观音桥(B)	1	33	212(125)	18.7	9.3	5.3
	2	24		13.6	8.7	4.9
	3	27		15.3	9.2	5.2
解放碑(A)- 南坪(C)	4	23	212(130)	12.9	7.3	4.1
	5	20		11.2	6.8	3.8

根据表中数据可以计算出商圈辐射范围的边界点，在图中标记出来后可以连成一条曲线，所组成的覆盖区域如下图 4.4 所示。

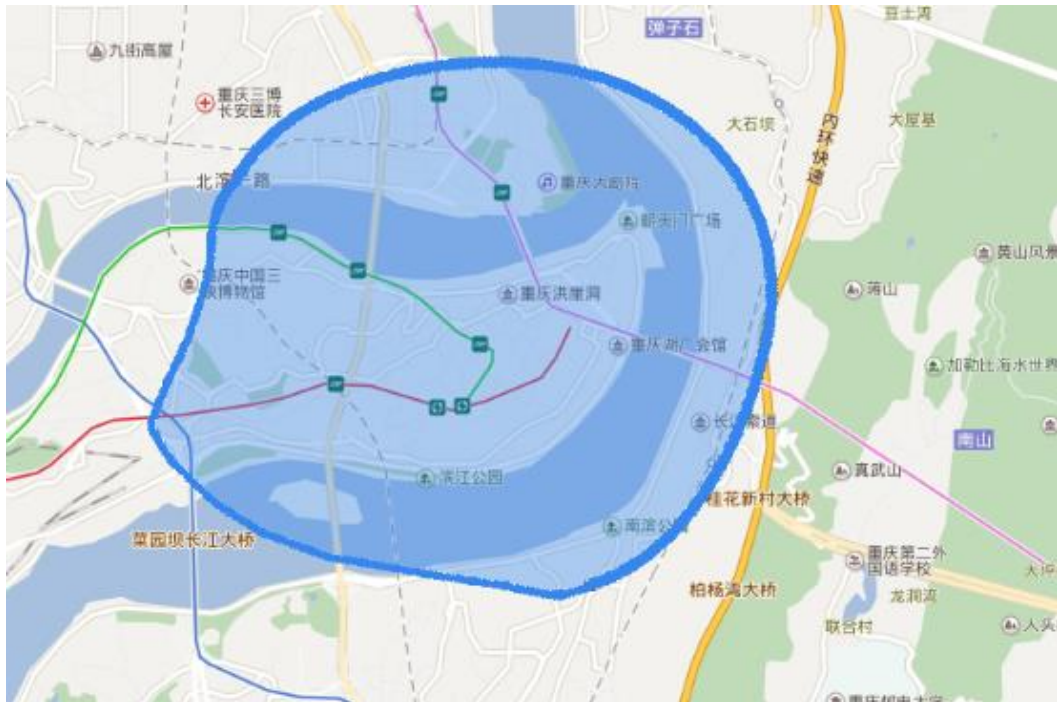


图 4.4 解放碑商圈辐射范围（雷利法则）

#### 4.1.2 通过行车数据计算商圈辐射范围

按照 3.4 小节中描述的商圈辐射范围计算方法，需要首先统计解放碑商圈消费者分布情况，这里同样选取 2014 年 12 月份的行车数据来做原始数据。

经过城市交通物联网系统的大数据后台的计算，一共选取了 865 辆轨迹记录来估算商圈的辐射范围，通过商圈分级理论，这里只选出了前 85% 范围内，大约 730 条记录的采集点分布情况，下面表列出了对应采集点的统计情况。

表4.2 解放碑商圈计算数据一览表二

采集点编号	居民数（以车辆计算）
A 小区	134
B 小区	121
C 小区	117
D 小区	101



表4.2 解放碑商圈计算数据一览表二（续表）

E 小区	92
F 小区	89
G 小区	51
H 小区	25
总计	730

接下来将这些采集点所在的地理位置对应的描到地图上，并通过热力图的形式表现出来，如下图 4.5 所示。

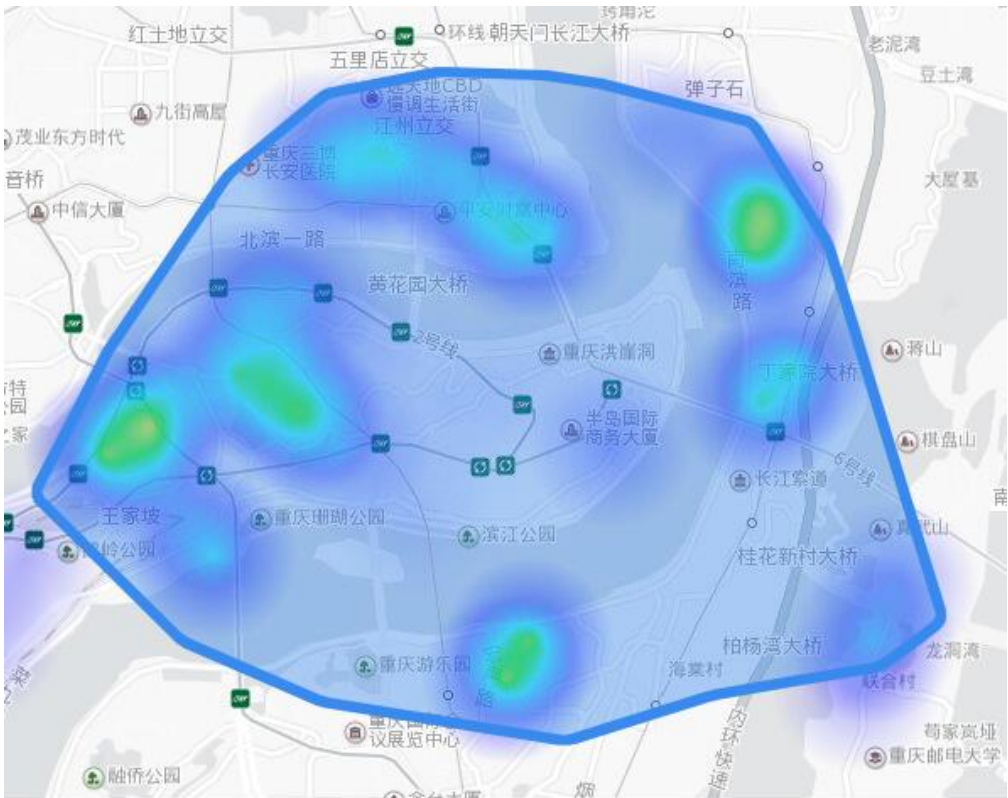


图 4.5 解放碑商圈辐射范围（行车大数据）

4.2 计算结果的分析比较

选不同时期的行车数据，用两种方法计算商圈的辐射范围可以得到众多的计算结果。下面只对几个典型结果作分析，能反映研究方法的差异对研究结果所产生的影响，也能说明两种方法各自的优缺点。

#### 4.2.1 解放碑商圈结果

上面详细介绍了通过两种方法计算出解放碑商圈的辐射范围，下面对它们得出的结果进行分析，下图 4.6 画出了两者结果的对比效果。

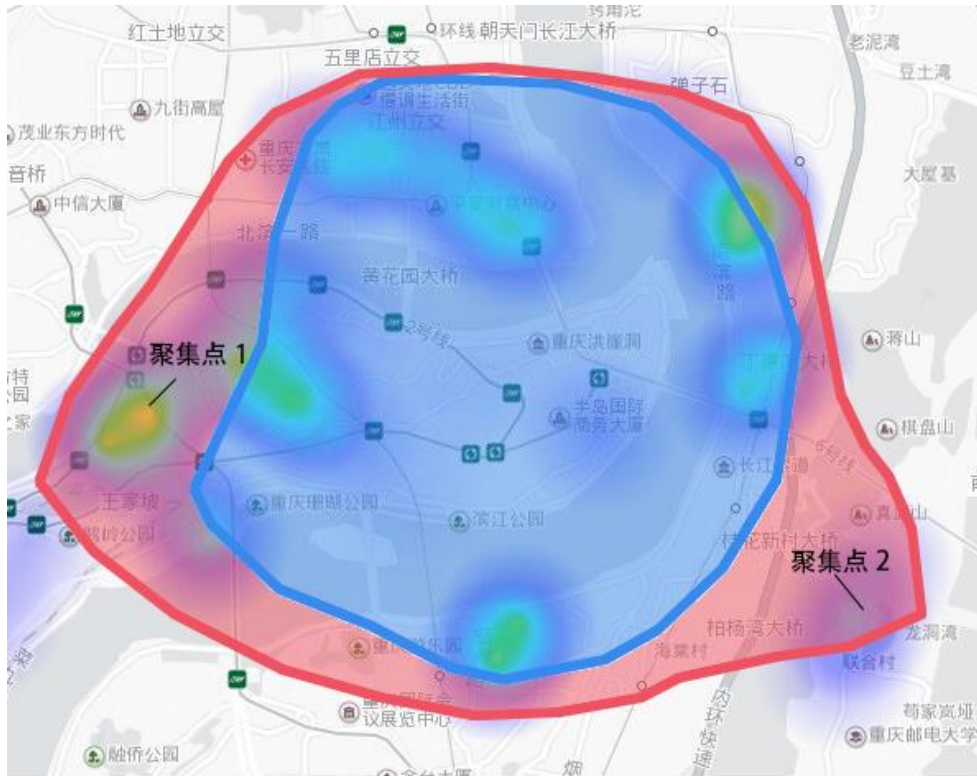


图4.6 解放碑商圈辐射范围对比分析图

上图 4.6 中的中间区域即是雷利法则计算得到的商圈辐射范围，外围区域是基于行车大数据计算得到的商圈辐射范围。

通过对上图的分析可以看出，外围区域包含了比中间区域更大的范围，而且图中左侧明显的聚集点 1 没能通过雷利法则表现出来，这与典型路径的选取有很大关系，由于观音桥与解放碑之间正好相望于长江，聚集点 1 没能体现在典型路径中，所以地理条件的原因，限制了雷利法则在实际场景的应用。本例中的聚集点 2 也属于特殊情况，这是由于商圈的相对地理位置分布（如图 4.1 所示），中间区域的右侧是商圈的外延方向，该方向没有竞争商圈的存在，无法直接计算分界点，该部分的划分是通过公式估算出来的，正好在该估算值的外面，所以其结果的参考意义很有限。

另一方面,虽然中间区域覆盖的范围明显要小于外围覆盖范围,但是能大概覆盖消费者分布点,这说明雷利法则在本例中能一定程度上反映商圈商业面积、时间距离等因素对商圈辐射范围的影响,但是这只能反映它们互相影响的大概趋势。

本例的结果正好说明了雷利法则在实际应用中存在的一些缺陷,而两者辐射范围的大致情况能表现一致,这与商圈间时间距离的计算方式有着很大关系,时间距离的计算直接影响辐射半径的大小,本例采用行车大数据来计算其值正好能弥补现时商圈理论计算方法的本质缺陷,更能反映交通状况这个重要因素在其中的作用。

#### 4.2.2 从时间角度分析商圈辐射范围变化

从时间维度,选取两个不同时间段的行车数据来计算商圈辐射范围可以发现商圈辐射范围并不是一尘不变的,这也从侧面说明商圈营销也需要考虑季节性因素,例如上图 4.7 是解放碑商圈在 2014 年 12 月份与 2015 年 2 月份的行车数据,通过第二种方法所计算出来的商圈辐射范围。

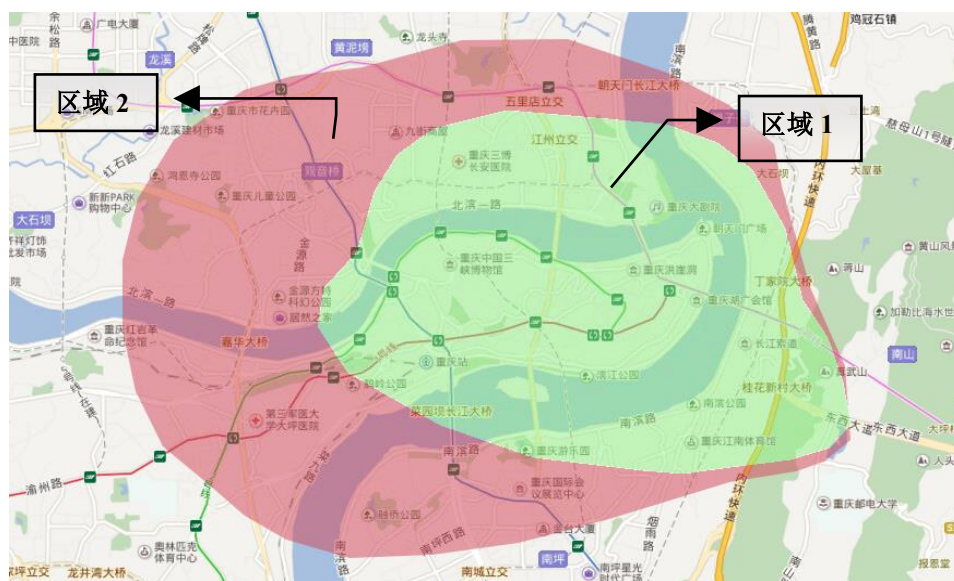


图4.7 解放碑商圈辐射范围随时间变化图

图中区域 1 是 2014 年 12 月份的结果,区域 2 是 2015 年 2 月份的结果,可以明显的看出 2 月份商圈的辐射范围突然骤增,分析其原因,其实不难发现是因为过年过节的原因才导致商圈辐射范围变大。解放碑商圈是全国知名商圈,主要业态是高中档多层次零售业,其中也包含众多顶级品牌,是重庆市高端市场的代表。根据重庆市渝



中区政府数据显示，2015 年解放碑 CBD 上半年实现 209 亿的社会消费品零售总额，2015 年春节、情人节等节日集中在 2 月中旬，刺激了高端市场的消费。

另外，从上图 4.7 还可以看出，新的消费者增长点主要分布在江北区和渝中区以西，这也说明了商圈的主要消费人群来自重庆的中部地带，正好是重庆市的经济发达地区，与解放碑商圈的中高端定位相符。

#### 4.2.3 从总体上分析商圈辐射范围

通过上面对雷利法则的方法的介绍可以知道，一般来说通过雷利法则计算的商圈辐射范围是不会相交的，因为它们计算所得的边界点是近邻在一起的，而通过常理判断这明显是不符合实际的，如下图 4.8 所示。



4.8 雷利法则计算商圈辐射范围分布图

这里再通过第二种方法计算商圈辐射范围如下图 4.9 所示，图中五大商圈的辐射范围是相互交叠的，其中的原因不言而喻，同一个小区（采集点）的居民消费水平高低不一，居民零散的分布也从一定程度上能反映该小区居住着不同经济状况的居民。

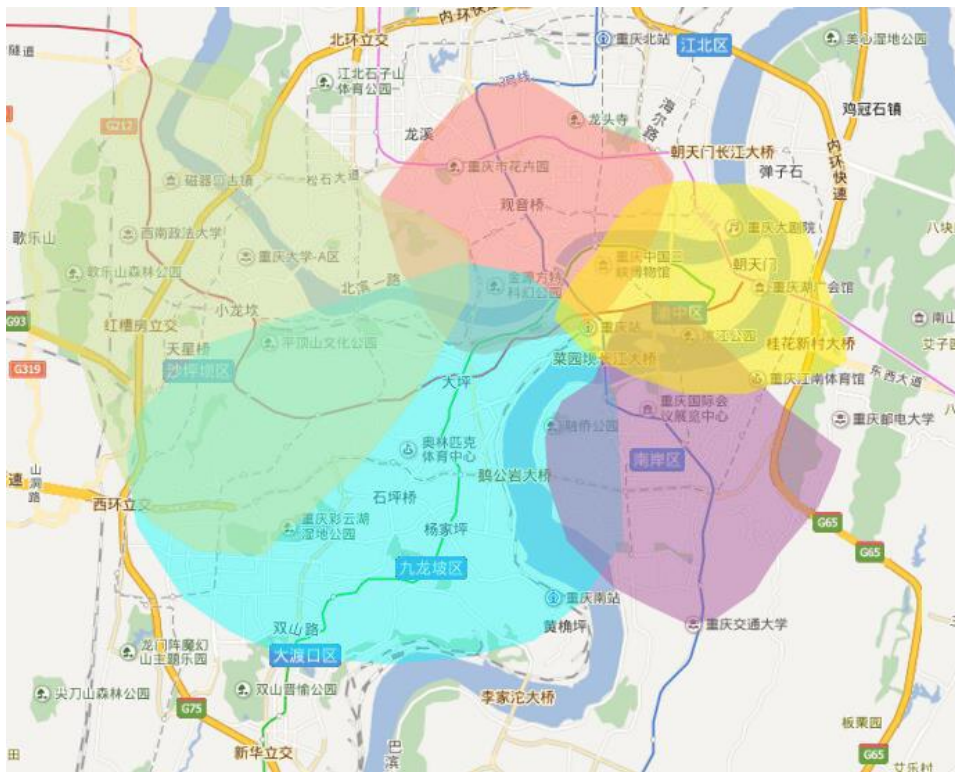


图4.9 行车数据计算商圈辐射范围分布

#### 4.2.4 总结两种计算方法优缺点

雷利法则计算商圈辐射范围方法的优缺点：

- (1) 通过公式计算简单明了，数据组成简单，处理起来方便，便于估算大致范围；
- (2) 商圈理论应用与实际结果偏差较大，灵活程度小，受地理环境限制其应用；
- (3) 不能反映商圈长期性变化趋势，数据来源更新缓慢；
- (4) 考虑的影响因素太过简单，不能良好的反映实际情况下商圈对消费者的吸引力和阻力的影响。

基于行车大数据的商圈辐射范围计算方法的优缺点：

- (1) 目的性明确，直接针对用户居住地分布来确定商圈辐射范围；
- (2) 行车数据实时更新，可以长期反映商圈辐射范围随时间的变化规律；
- (3) 便于后期统计分析商圈的经营状况，周期性的统计结果可以帮助商圈预测未来的经济发展走势；
- (4) 只针对私家车用户，对于乘坐公共交通或步行的用户无法追踪其动态；

(5) 采集点分布疏密程度直接影响估算结果，前期建设采集点不足很难反映真实的用户分布情况；

(6) 后期涉及到的数据源越来越多，各种关系关联起来越来越复杂，算法也会越来越复杂，需要不断更新。

### 4.3 基于行车大数据商圈建设应用举例

商圈建设分为多个方面，其中首要一点就是分析商圈的客源组成，只有知道了顾客的消费行为习惯才能更好的制定销售计划，帮助商圈朝着正确的方向发展，而商圈辐射范围的估算是分析消费者行为习惯重要的一步，相较于问卷调查的方式，通过采集用户日常出行数据可以减少人为因素的影响，得到的结果也更具参考意义。商圈辐射范围的估算对商圈建设的促进意义很大，下面举例说明。

(1) 商圈辐射范围的估算可以帮助商圈管理、经营者制定针对性的区域推广计划，降低推广成本效益比；

(2) 分析当前地区的商圈经营情况，推算地区市场容量，帮助商业地产开发商选择合适的新商业中心地址以及合适的经营业态；

(3) 结合商圈理论研究，指导商圈吸引力建设方向，增加商圈在同类型商圈的竞争力，如杨家坪商圈积极建设智能停车，提供预约等便民服务，提升用户感受可以吸引更多的驾车用户到商圈消费；

(4) 通过计算商圈辐射范围的大小和分析消费者组成，可以印证商圈体量与影响因素之间的关系，如体量大的商圈不一定辐射范围就大，消费者在商圈周围聚集度高可以通过热力图直接表现出来；

(5) 通过行车大数据可以分析影响商圈体量的因素，如基础设施的建设，商圈自身的定位，宣传所针对的特定人群等因素都会影响商圈的体量。

### 4.4 本章小结

本章主要介绍了两种方法计算商圈辐射范围的详细过程，从时间角度，空间角度方面分析了五大商圈辐射范围的各种情况，从实际角度出发解释了导致特殊情况的深

层次原因，还总结了以上两种计算方法的优缺点，最后举例说明了基于行车大数据估算商圈辐射范围在实际应用中可以有哪些方面的重要应用。

## 第5章 总结与展望

### 5.1 全文总结

本文介绍了数据挖掘技术在智能交通方面的应用背景和研究意义。阐述了行车大数据在各行业深入应用所取得的良好效果,分析了行车大数据在商圈建设方面的巨大潜力。调研了国内外关于商圈理论和交通大数据方面的研究现状,和目前商圈理论在实际应用中存在的一些问题。本文结合了本人工作中参与的重庆市城市交通物联网系统项目,提出了两种计算商圈辐射范围的分析方法,并总结了两种方法的优缺点,针对商圈理论在实际场景的应用存在的缺陷提出新的解决方案,可以更加真实的反映商圈消费者的分布情况,为重庆市智慧商圈的进一步建设,提供有力的理论支撑。

本文中针对商圈建设主题所做的工作如下:

(1) 通过对商圈理论的介绍,了解商圈辐射范围的定义,以及各种计算方法的优缺点,为下文提出新的解决方案奠定基础;

(2) 通过分析城市交通物联网系统的数据源组成,来说明基于行车大数据的计算模型合理性;

(3) 在商圈经典理论雷利法则模型的基础上,提出以行车大数据作为数据来源,建立新的估算模型,详细介绍了商圈间时间距离的提取方法;

(4) 针对商圈理论存在的问题,提出新的商圈辐射范围计算方法,详细描述了模型的建立过程,以及实现的技术与方法;

(5) 由经典理论和基于行车大数据这两种方法推算商圈辐射范围,对得到的结果进行分析,总结在实际应用中两者存在的优缺点,基于行车大数据的计算方法更具优势;

(6) 分析商圈辐射范围估算在实际商圈建设中的重要应用,说明商圈理论研究的重要性以及商圈理论在应用方面具有指导意义。

### 5.2 进一步研究

在智慧城市建设的浪潮下,城市智能交通发展得如火如荼,越来越多源的数据促



进数据融合时代的到来,关于数据挖掘的研究也向更深层次发展,跨领域的研究变得越来越多。基于行车大数据的商圈建设的研究就是在这个大环境下应运而生,以往关于交通的数据挖掘还只是停留在与交通出行相关的方面,而商圈理论研究方面的数据主要来自人工问卷调查,现在将两者互补,提出新的商圈建设研究内容,初步研究成果已经在上面说明,下面是对研究中还存在的一些问题总结和对今后的研究方向的说明:

(1) 以车辆为对象,代替真实的人类对象对研究结果有偏差,需要结合其他领域的人类行为研究做进一步的修正;

(2) 基于行车大数据的商圈辐射范围计算方法还只是初步模型,若将来数据源发生变化需要根据实际情况修改;

(3) 受限于项目中采集点的建设现状,研究成果不能更详尽的反映实际情况,为了达到更好的应用效果,需要一定数量的采集设备;

(4) 在商圈理论研究方面,还需要改进雷利法则的精细程度,参考第二个方法的分析结果可以帮助研究人员进一步研究商圈理论,将更多的影响因素考虑在内;

(5) 结合项目进展,后期将逐步纳入更多数据源,如商圈 WiFi 系统数据,商户营业数据,用户停车 APP 数据、机动车交易市场数据等,可以与用户消费行为进一步关联,探索消费者消费习惯,可以得到更加准确的计算结果,更进一步,还可深层次的挖掘商圈消费者社会身份情况。

## 参考文献

- [1] 戴刚, 刘沅杰, 郝俊华. 基于RFID的物联网研究与应用[J].现代计算机(专业版),2010(10): 62-64.
- [2] 朱建安, 巩建国, 戴帅. 公安交通管理大数据系统框架设计与应用展望: 2015年中国城市交通规划年会暨第28次学术研讨会, 中国浙江杭州, 2015[C].
- [3] 郑梦悦. 论大数据在智能交通建设中的应用[J].中外企业家,2016(02): 147.
- [4] 陆化普, 孙智源, 屈闻聪. 大数据及其在城市智能交通系统中的应用综述[J].交通运输系统工程与信息,2015(05): 45-52.
- [5] 重庆智能交通系统通过国家验收[J].中国公共安全(综合版),2012(20): 40.
- [6] 张梦轩, 冉斌. 车联网技术的发展沿革和发展趋势简析[J].科技传播,2015(23): 124-125.
- [7] 胡庆武, 王明, 李清泉. 利用位置签到数据探索城市热点与商圈[J].测绘学报,2014(03): 314-321.
- [8] 张富峥. 基于大规模位置和消费数据的用户行为理解[D]. 中国科学技术大学, 2015.
- [9] 重庆市人民政府办公厅关于转发市商委重庆市智慧商圈建设实施方案的通知[J].重庆市政府公报,2015(16): 17-19.
- [10] 中国公路学报编辑部. 中国交通工程学术研究综述·2016[J].中国公路学报,2016(06): 1-161.
- [11] Martic A. Predictability of Human Behavior using Mobility and Rich Social Data[J].Master's thesis [Academic thesis],2013.
- [12] Bazzani A, Giorgini B, Rambaldi S, et al. Statistical Laws in Urban Mobility from microscopic GPS data in the area of Florence[J].Journal of Statistical Mechanics Theory & Experiment,2009,2010(5): 823-831.
- [13] Wang J, Mao Y, Li J, et al. Predictability of road traffic and congestion in urban areas.[J].Plos One,2014,10(4).
- [14] Luong C, Do S, Hoang T, et al. A Method for Detecting Significant Places from GPS Trajectory Data[J].2015.
- [15] Chen M, Yu X, Liu Y. Mining moving patterns for predicting next location[J].Information Systems,2015,54(C): 156-168.
- [16] 王明生, 黄琳, 闫小勇. 探索城市公交客流移动模式[J].电子科技大学学报,2012(01): 2-7.
- [17] 李学龙. 基于RFID的城市道路交通状态检测与预测方法研究[D]. 重庆交通大学, 2015.
- [18] 张晨. 基于RFID的车联网在ITS中的应用研究[D]. 重庆大学, 2013.
- [19] 曹洁, 叶源潮. 运用RFID技术获取动态OD数据的方法[J].小型微型计算机系统,2015(04): 713-716.
- [20] 徐秋平. 基于车联网的交通信息采集与事故识别方法研究[D]. 北京交通大学, 2014.
- [21] Reilly W. Methods for the study of retail relationships. Studies in marketing, vol. 4[J].1929.
- [22] Converse P D. NEW LAWS OF RETAIL GRAVITATION[J].Journal of Marketing,1949,14(3): 379-384.
- [23] Huff D L. A Probabilistic Analysis of Shopping Center Trade Areas[J].Land Economics,1963,39(1): 81-90.
- [24] 孙良静. 基于商圈理论的商业地产项目定位研究[D]. 重庆大学, 2015.
- [25] 王先庆, 王晓春. 大型购物中心聚客力影响因素研究——一个理论模型[J].北京工商大学学报(社会科学版),2010(02): 34-38.

- [26] 刘瑞霞. 经济欠发达中小城市商圈布局研究[D]. 四川师范大学, 2011:
- [27] 吴忠才. 基于哈夫修正模型再修正的城市商圈区位模型[J]. 吉首大学学报(自然科学版), 2009(02): 108-111.
- [28] 王晓娜. 商业体量预测方法的应用研究[D]. 武汉理工大学, 2014.
- [29] 陆清华. ERP原理与实践[M]. 北京理工大学出版社, 2015.
- [30] 张红, 王晓明, 曹洁, 等. 基于大数据的智能交通体系架构[J]. 兰州理工大学学报, 2015(02): 112-115.
- [31] 程豪. 基于Hadoop的交通大数据计算应用研究[D]. 长安大学, 2014.
- [32] 王兴武. 基于Hadoop的海量交通数据研究与应用[D]. 浙江工业大学, 2015.
- [33] 吴小军. 基于Hadoop的石油运输车辆油耗预测模型研究[D]. 武汉理工大学, 2015.
- [34] 柳思维. 城市商圈论[M]. 中国人民大学出版社, 2012.
- [35] 曾锵. 零售商圈吸引力: 基于雷利法则和赫夫模型的实证研究[J]. 财贸经济, 2010(04): 107-113.
- [36] 祎石佳. 基于重庆主城区商圈发展问题的实证分析[J]. 西江月, 2013(1): 257-258.

## 致谢

两年半的研究生学习时光飞逝而过，在这期间的学习和生活中，我感悟很多，也收获了不少，不仅仅是收获了知识储备，同时也收获了友情，这和一直关心帮助我的人是分不开的，对我来说将会受益终生。

首先，要感谢我的导师，巴继东老师。在研究生学习实习期间，他给予了我大量的指导帮助，其渊博的知识，严谨的科学态度和刻苦钻研的精神使得我在这两年多的学习和生活中受益匪浅。论文在巴老师的悉心指导下完成，同时在论文的完成过程中我还受到了项目经理陆义平和同事李金振的巨大帮助，对此表示感谢，两年半的学习期间能和钟弘杰同学一起学习互相帮助，使我体会到真正友谊的力量，感谢一起工作的同事和一起学习的同学们。在论文即将完成之际，在此对所有帮助过我的人表示由衷的感谢！

然后，我还要感谢学院的老师们对我的帮助，谢谢你们在我学习生活中给予我的鼓励和帮助，指导我走上社会，走向工作岗位，老师们孜孜不倦的育人精神尤其的感动我，感谢陶智勇院长、王先兰老师、程雯老师、陈智老师、余耘老师和众多授课老师对我的教导。

最后，我要感谢我的父母。感谢他们在我研究生学习期间给予我的支持和鼓励，使我可以全身心投入到学习和工作中。正是他们在精神上和生活上对我的支持帮助，使得我能够顺利完成论文，谨以此文作为向他们致礼。

## 附录 1 攻读硕士学位期间参与的项目和发表的论文

### 发表的论文:

[1] 彭晨伟, 巴继东. 基于交通大数据的智能信息服务平台的研究与设计[J]. 计算机系统应用, 2017(6).