商圈(Retail Trade Area)引力模型分析

# 简介

# 相关工作

# 商圈理论与数据

## 零售商圈引力模型

零售商圈是零售交易区域的辐射范围，但是商圈的概念并没有很明确的定义，本文中，我们认为零售商圈的商业企业聚集所形成的空间范围。同样，在商圈级别划分标准下，本文中所研究的商圈是指由核心商业圈和次级商业圈组成的空间范围。

商圈理论中应用最广的是Reilly Rule 和Huff Model 以及其的演化模型。

由于大多数情况下，企业很难获取详细的商业信息，那么如何选择投资地成为了一个难题，而Reilly Rule 最早为企业提供了容易实现的决策指导。Reilly Rule认为商业也具有相互吸引的特性，它以万有引力定律为核心，来确定商圈吸引力临界范围。但是Reilly Rule是以商圈为中心的研究，并且需要有较严格的前提才能使结果有效。在我们对上海十九个大型商圈进行研究之后，我们发现使用Reilly Rule 是很难确定商圈范围的，由于商圈差异和人群行为等因素。

与Reilly Rule 有所不同，Huff Model 是从顾客的角度进行研究，通过模型计算出一个概率值，这个概率值能够代表当前用户去往某个商圈的概率，但是模型计算概率的时候用到的阻力和魅力因素仅仅包含距离和商店规模。在大都市商圈引力研究中，这两个因素依旧很重要，但是其他因素，例如地理位置，商品档次等，对概率的计算同样占据很重要的位置。

除此之外，还有很多演化模型，我们在5.1中进行了详细的对比研究。

## Reilly Rule and Huff Model

Reilly Rule 是W. J. Reilly 最早在1931年提出，它的核心观点是：具有零售中心地机能的两个都市，对位于其中间的一个都市或城镇的零售交易的吸引力与两都市的人口成正比，与两都市与中间地都市或城镇的距离成反比（查原句）。模型如下：

其中Ba是商圈A对商圈A、B中间某地C处顾客的吸引力，Bb为商圈B对C处顾客的吸引力。Pa为商圈A附近人口，Pb为商圈B附近人口，Da和Db为商圈A和B与C处顾客的距离。

Reilly Rule 在一定程度上能够确立商圈的范围以及商圈吸引力范围，但是由于它考虑的因素过少，导致误差十分大，它的误差如表所示。同样，应用Reilly Rule 还需要具备几个前提：（1）两个城市（商圈）交通情况类似；（2）两个城市（商圈）属性类似；（3）两个城市（商圈）人口（人群类型）类似。而在实际的研究中，这些因素很难测定，并且具有很大的差异性，我们对满足这些前提的商圈进行研究，与实际的误差如表所示。

Huff Model 是由David. L. Huff 于1963年提出，它认为：从事购物行为的消费者对商店的心理认同是影响商店[商圈](https://baike.baidu.com/item/%E5%95%86%E5%9C%88)大小的根本原因，商店商圈的大小规模与消费者是否选择该商店进行购物有关（查原句）。我们对商圈吸引力的研究以及对大都市零售商圈引力模型的建立主要是基于Huff Model，模型如下：

其中 为 地区顾客到商圈 消费的概率， 是商圈 的魅力， 地区顾客到商圈 的阻力， 和 是以经验为基础所估计的修正值， 是互相竞争的商圈数。

Huff Model 从某种程度上得到和实际值很相近的商圈吸引力概率值，但是如果相互影响的商圈数目过多，那么它的计算精度会有一定程度的下降（如图所示），同时在当前经济背景下，更多的因素会对商圈魅力产生影响，而阻力的测定也不仅仅是基于空间距离。在本文中，我们根据文献并且与企业市场经理多次交流，总结了十几个可能对商圈魅力和阻力造成影响的因素，并进行了相关性分析和效应量计算（4.2和4.3），之后对商圈吸引力模型进行了设计（4.5）。

## Data

本文所进行的研究使用了大量数据，极少包含主观因素的影响，因此在一定程度上说具有很高的客观性，所用数据集如下：

交通卡刷卡数据：我们使用了2015年4月上海市轨道交通刷卡数据进行分析研究，数据集包含30天共计4亿多行数据，每一行数据包含刷卡时间，刷卡地点，消费金额，进出站标记等属性。

商圈数据：本文中使用了上海市所有核心商圈和部分大型次核心商圈进行研究，共计19个商圈与142个大型商场的信息。

实际吸引力概率：我们根据交通卡刷卡数据区分出工作人群和购物人群，其中我们认为在五天工作日出现四次以上并且路径完全一致，时间大致相同的cardid为工作人群。而周末两天这些工作人群出现并和工作日出行路径一致，时间大致相同的人认为这些人在周末也在工作。这样，我们就得到了购物人群数，并对这些人群进行分析，得到了不同人群聚集地以及公司聚集地信息。之后我们通过对一个月人流的统计分析，得到了不同地区顾客到不同商圈购物的概率值，这个值我们认为能够代表商圈对其的吸引力程度。

时间成本：由于上海市的交通发达，传统意义上的空间距离对计算结果会产生一定误差，因此我们引入时间成本这个概念来代替空间距离作为阻力因素的一部分。最初我们对上海地铁官方网站数据进行爬取。在初步计算之后我们发现有很大误差，在实际测算之后我们认为这是由于在交通系统中，总是会以最坏情况（即最大时间）进行测算，因此我们对刷卡数据进行统计分析，把任意站和商圈间的乘客以时间排序，然后取前80%进行平均值计算，因为会有部分乘客由于某种原因长时间停留在轨道交通系统中。

商业面积：在商圈吸引力研究中，商业面积也具有很高的地位。但是对商业面积的测定十分困难，我们对每个商圈商场进行统计，并对每个商场的官方主页中得到其公布的商场总面积和营业面积两个数据，之后进行计算，得到了每个商圈的商业面积。但是由于只能得到大型商场的营业面积，无法得到步行街等小型商场的营业面积，在对南京东路、豫园等小型商场集中的商圈的计算会出现误差，因此在0.0中，我们进行了对比研究。

人口总数、人口密度、GDP、人均GDP：这些可能的影响因素我们从政府的年度报表中得到，因此具有很高的可信度。

房屋均价：在与企业经理讨论后，我们认为房屋均价可以一定程度上代表这个地区的经济程度，而经济程度可能会对此地区顾客选择商圈有所影响。这部分数据来自互联网房地产企业公布的信息。

商圈等级，商品档次，枢纽等级：其中商圈等级和枢纽等级来自于上海市政府公布数据，商品档次通过我们对商圈内商场店铺类型和数量进行统计并分类得到。

商圈的活力指数：这部分数据来自于复旦大学城市发展研究院发布的《上海市商圈活力指数月报》(查原句)，是由人流数据、手机信令数据、Pos机刷卡数据等计算得到。

# 人群与聚集地分析

每天生成的上千万条刷卡记录中包含着不同的群体, 如上班族, 老人和游客等. 这部分主要分析上班族群体的移动行为特征与聚集地划分。如果一个人的刷卡记录满足：

其中, *W*表示所有满足条件的上班族集合；*Wi*表示对第*i*个乘客若满足至少在工作日连续四天存在刷卡记录并出现在相同的始发站和目的站，则认为该乘客属于上班族群体。

此外，为了进一步分析上班族的地铁移动行为特征，这部分定义了常规的上班族和非常规的上班族。常规的上班族是指那些一周最多连续工作五天的人；非常规的上班族是指连续工作大于五天的人，通常会在周六或者周日加班。

之后，通过可视化的形式，我们得到了上班族的居住地点和工作地点（如图）和购物人群的居住地点（如图），有助于进一步分析该群体的出行行为特征。

商业行为和居民成分有着密切的关系，通过研究我们发现，商圈附近必然会有工作区存在，但是工作区却不一定会依托商圈存在。本文中对具有相似居民组成的区域进行对比研究，验证了模型（如图）。

# 引力模型研究

## 计算误差对比

## 相关性与相关系数

## 效应量(Effect Size)

## 大都市零售商圈现状

## 商圈引力模型设计

# 案例研究

## 模型对比研究

## 模型有效性验证

# 讨论

# 总结与未来工作