
SOUTHWESTERN UNIVERSITY OF FINANCE AND ECONOMICS



金融智能

餐饮用户满意度因素研究：什么在影响消费者的选择？
——基于结构方程模型和自然语言处理技术的餐饮商家分析平台

作者

结课快乐！

谭贺娟

宋佳

周炫怡

金锐丽

左丹

肖云彤

JUNE ,29, 2021

目录

一、项目介绍	3
1.1 项目背景.....	3
1.2 国内外研究现状.....	3
1.3 项目概述	7
1.4 研究意义	8
二、功能介绍	9
2.1 结构方程模型	9
2.2 观点挖掘	9
2.3 情感分析	10
2.4 语义网络	10
2.5 同类竞品分析	11
2.6 餐饮年度总结	11
三、结构方程模型	11
3.1 模型假设	11
3.2 变量假设	13
3.3 研究结果	17
3.4 研究结果	22
四、算法	25
4.1 数据收集	25
4.2 数据预处理	28
4.3 模型与算法介绍	28
4.4 结果分析——语义网络结果展示	31
4.5 结果分析——情感分析及关键词提取算法（以江南道为例） ..	35
五、可视化设计	42

5.1 网页设计	42
5.2 年度报告	51
六、总结	59
6.1 主要研究工作及创新点	60
6.2 工作成果	61
6.3 改进	62
参考文献	64

一、项目介绍

1.1 项目背景

互联网技术的快速发展，网络用户规模的迅速扩大，推动电子商务快速的过渡到成熟期，使得线上市场也成为企业新的快速增长点。据中国互联网络中心 CNNIC（China Internet Network Information Center）第 43 次发布的《中国互联网络发展状况统计》显示，截止至 2018 年 12 月，我国网络用户达到 8.29 亿，用户互联网普及率达 59.6%，互联网设备的普及对餐饮行业也产生了巨大的影响，相较之前的店内堂食，口口相传进行店铺宣传，越来越多的人选择了网络这种方式来点餐、评论和分享美食。

美团网作为中国成立最早、最具影响力的团购网站，已经成为人们生活中的重要组成部分。在美团平台上，每个餐厅都存在大量的消费者评论，消费者评论是消费者在该餐厅用餐后所形成的，其评论内容主要是针对菜品或者服务质量方面，是消费者表达自身在餐厅用餐整个过程中的感受。餐厅通过消费者评论可以快速的了解消费者的关注点、需求点，以及商品本身的不足点和店铺服务相关的问题点，并通过闭环管理，对菜品和店铺进行调整改善，提升消费者的满意度，但基于海量的消费者评论数据，企业需要一套能够快速实现海量消费者评论中的价值挖掘的方法。

本文通过问卷调查的方式获取用户调研数据，并通过结构方程模型探究影响成都市温江区消费者进行餐厅选择的因素，同时通过情感分析、观点挖掘等技术对成都市温江区 60 家美团商家消费者评论进行文本分析挖掘，从评论中分析出消费者的关注点和需求点，根据消费者需求点去评价餐厅的好坏，并根据消费者抱怨点去优化产品，改善服务，从而达到餐厅产品优化、服务改善的目的。

1.2 国内外研究现状

文本情感倾向分析是对一个带有情感色彩和主观性的句子进行分析、总结和

处理的过程。基于互联网的社交工具和应用工具的飞速发展，产生了海量的文字记录，这也使得文本情感倾向分析领域快速发展起来。这些海量的文字信息成为 21 世纪珍贵的资源，可是如何使用这笔资源，如何将信息资源转化为经济效益、社会效益却是一个亟待解决的问题。

1.2.1 美团平台评价的研究现状

互联网的快速发展，人们的生活习惯也随之发生着改变，2010 年美团网的诞生促使中国深厚的餐饮文化和美食文化产生了巨大的变革。美团作为国内成立最早、综合实力最强的团购网站，每天都会接受到大量订单以及顾客给予的反馈评论。如何将这海量信息利用起来，为顾客和商家提供更好的服务，是美团平台亟待解决的问题。通过丰富的网络评论数据挖掘人们对美食偏好的显著差异性，不仅可以揭示人们饮食习惯的差异，还可以揭示美食在各菜系的传播和发展，岳子静等（2017）以美团网作为研究平台，以北京地区餐饮在线点评数据为研究对象，对菜品及其相关属性进行频度统计分析。从文本情感分析的角度出发，在线评论数据可以在一定程度上揭示不同地区用户对不同菜系的偏好，以及不同菜系在不同地区的传播和发展。分析了评论情感倾向与菜品关注程度和满意程度之间的关系，在一定程度上可以用来反映用户的偏好和烹饪风格的传播发展情况。

1.2.2 情感倾向分析研究现状

自然语言的情感分析发展到现在大概可以分为两个研究方向，一个是基于统计和情感词典对文本进行情感倾向分析，另一个是基于机器学习使用已知的标记数据训练模型，然后使用模型对文本数据的情感倾向进行识别。

基于情感词典的情感分析是事先准备好情感词典，然后设置相应的匹配规则，将文本按照既定的规则分析，即可得到分析结果。封丽(2018)利用情感词典结合情感分析的方法对微博文本的舆论内容进行情感分析，不仅建立了文本的情感词典，而且增加了表情的情感词典。定义的情感词典和文本数据匹配的规则，可以比较迅速、实时地表现出随着时间变化的舆情状况。Xu, Guixian 等(2019)构造了一个扩展的情感词典。扩展的情感字典里有基本的情感词、领域情感词和多义情感词，提高了情感分析的准确性。朴素贝叶斯分类器用于确定多义情感词所在的文本区域，

从而得出多义情感词在该领域的情感价值。利用扩展的情感词典和设计的情感评分规则，实现了文本的情感倾向。实验结果证明了基于扩展情感词典的情感分析方法具有一定的可行性和准确性。这个研究成果对于评论文本的情感识别具有重大意义。

机器学习主要是利用大量的算法来让机器模拟人类学习的行为。基于机器学习的情感分析方法就是利用一定数量的数据来训练算法模型，能够使模型识别短语句中人类的情感。Dos Santos C 等(2014)提出了一种新的深度卷积神经网络，利用从字符到句子级别的信息来对短文本进行情感分析。他们将这种方法应用于两个不同领域的语料库，其一是包含电影评论中句子的斯坦福情感树语料库(SSTb)，另一个是包含 Twitter 信息的斯坦福 Twitter 情感语料库(STS)。对于 SSTb 语料库，他们新的深度卷积神经网络的方法在二元正、负分类和细粒度分类中都获得了最先进的单句情绪预测结果，准确率分别为 85.7%和 48.3%。对于 STS 语料库，他们新的深度卷积神经网络的方法实现了 86.4%的情绪预测精度。McDonald R 等(2007)研究了一个结构化的模型，用于在不同粒度级别上对文本情感进行联合分类。模型中的推理是基于标准的序列分类技术，使用 Viterbi 约束来确保结果的一致性。这种模型的主要优点是它允许文本中一个级别的分类决策影响另一个级别的决策。实验表明，与孤立训练的模型相比，该方法能显著降低分类误差，使情感的分类更加精确。

考虑到情感字典和机器学习的两种方法都已经达到了瓶颈，将两种方法相结合的研究发展成为一种趋势。Lin Q 等(2018)在基于知识的方法中，使用词共现法和词嵌入相似度法自动构建和扩展了通用情感词典，结合多种机器学习模型和多词汇特征，最终获得了比使用一般情感词典的准确度高 12.11%的机器学习模型，提高了教师评价领域的情感分析的准确性。

1.2.3 观点挖掘研究现状

观点挖掘，是在大量的数据中识别和提取主观信息，这些信息可以是积极的，中立的或消极的。利用适当的机制和技术，可以将大量数据处理成信息，以支持相应的运营、管理和战略决策。通常，用户评论中有两种类型的文本信息：代表事实

的客观陈述和象征意见或看法的主观陈述。观点挖掘可以在三个不同的层面进行研究，即文档、句子和特征水平（Pang 和 Lee, 2008）。文档级情感分析基于整个文档的整体情感，是对固定的文档（例如，产品评论）进行情感的分类。它假定整个文档表达了单一情感。同样，句子级别的情感分析对句子级别的情感进行分类。然而，并非每一句话都是主观句（Wilson, 2005）。而且，即使在文档和句子级别的意见挖掘是有价值的，也不会发现用户喜欢和不喜欢什么。关于特定实体的固定意见的文本并不意味着该文本的作者喜欢或不喜欢该实体的各个方面。基于特征的观点挖掘用于识别和提取意见及其指向的产品特征。基于特征的观点挖掘有三个主要任务：首先，识别和提取产品特征；然后，确定观点词及其极性；最后，对应特征和观点之间的关系。选择合适的技术手段，从产品的在线评论中获取关于产品特征的用户意见，将对产品缺陷的识别与产品的完善起到重要作用。

1.2.4 结构方程模型研究现状

1) 餐厅感知价值

感知价值是个体基于获得和付出的知觉对一个产品效用做出的总体评价，是理解顾客行为的重要因素。就餐饮业而言，感知价值对食客行为有同样重要的影响。目前只有少数学者系统地探索过餐厅感知价值的具体内容。事实上，即便是麦当劳这样的标准化快餐，其顾客感知价值也是多方面的，包括食物品质、服务质量及环境。对其他类型餐厅而言，除了食物品质和服务质量，其他因素如就餐环境和餐厅卫生，也在感知价值中占一席之地。

2) 顾客感知价值

现有研究多关注普通顾客对餐饮功能价值的感知，然而顾客群体的感知价值远不止如此。HUANG 等(2019)认为，澳门美食节的顾客感知价值包括快乐感受、食物多样性、干净卫生、价格合理、食物味道等。杨春华等(2019)发现顾客感知成都美食形象是种类繁多、食材丰富且品牌知名度高。MUSKAT 等(2019)对山屋休闲民族餐厅的游客进行调查，发现其感知价值包括食物品质、服务质量、环境质量、价格公平和原真性。KIM 等(2019)通过调查到访韩国餐厅的国际游客，构建了测

量目的地餐厅服务价值的五维度量表，包括服务质量、原真体验、情感体验、社会及效用价值。由此可见，餐厅顾客感知价值内涵丰富。

1.3 项目概述

本文通过问卷调查的方式获取用户调研数据，并通过结构方程模型探究影响成都市温江区消费者进行餐厅选择的因素，同时本文通过情感分析、观点挖掘、语义网络等自然语言处理技术对成都市温江区 68 家美团商家消费者评论进行文本分析挖掘，从评论中分析出消费者的关注点和需求点以及情感得分。并对比结构方程模型和文本挖掘所得出的结果，探究通过两种不同的方法所得出的影响消费者选择的因素的异同。通过以上的分析，提出餐厅专属改进意见，生成餐厅年度报告，同时，为了给商家提供更好的改进服务，开发餐饮商家评论分析网页平台，展现商家的基本信息、同类餐馆分析、算法分析结果、竞品分析结果，从而更好地助力餐厅优化产品，改善服务。

本文的主要内容概况如下：

1) 数据收集：本文数据分为两类，第一类是问卷数据，对成都市温江区进行用户调研，探究影响消费者餐厅选择行为的因素，共计收集用户问卷数据 208 份；第二类为文本数据：利用八爪鱼软件爬取美团商家评论数据，共计 69959 条。

2) 结构方程模型：通过结构方程模型对问卷进行分析，探究影响消费者做出餐厅选择的具体因素，在该分析中本文将不可量化的消费者选择行为用可以量化的价格，口味等变量进行衡量，并通过消费者感知价值理论构建四个维度的结构方程模，最后通过回归分析对模型进行验证，发现口味，服务质量，价格和评分对消费者行为产生了显著正向影响。

3) 文本数据处理：本文基于非结构化数据自身的特点，对用户评论数据进行一系列的预处理工作，包括去重、去停用词、分句分词等。

4) 语义网络分析：本文基于武汉大学开发的反剽窃系统 ROSTCM6，通过对在线评论进行高频词提取、无意义词过滤、行特征提取、共现矩阵构建，最终得到十四个类别餐饮评论的语义网络图。

5) 情感分析：针对美团商家消费者评论的无监督文本分析，使用知网的消极词典、积极词典与多个程度词典对文本进行情感分析，在对原始文本进行基础的预处理操作后，再进行否定词处理、感叹符号处理、程度词处理后进行加权计算，最后得出情感值。

6) 观点挖掘：采用基于 PageRank 的 TextRank 对文章（即评论数据）进行关键词提取，将文本化成带权无向图，将词语看作是图上的节点加入图中，并识别不同词之间的关系绘制边后，基于图的排序算法进行迭代，后根据节点得分进行排序。

7) 结果可视化网页平台：平台的主要页面为排行榜和商家画像展示。

在排行榜中，本文展现了温江区商家的情感得分排行以及不同类型的餐厅在各自领域的情感排行。

在商家画像展示板块，本文对商家的基本信息、同类餐馆分析、算法分析结果、竞品分析结果等进行了可视化，展现了商家的情感得分，以及在温江区同类餐馆中的排名情况，并展示评论观点聚类结果，帮助商家进行投资决策，并为他们提供了获取有效信息的新渠道。

8) 商家年度报告：依托微信公众号平台，为商家生成具有特色的专属年度报告，包括商家年度热评、年度热议菜品、年度死忠粉、年度提升指南等内容，通过这些内容和形式输出，商家可以对自家美团评论进行年度简短回顾，并在一定程度上根据内容进行改进提升。

1.4 研究意义

1) 理论意义

通过数据收集，通过构建结构方程模型，对消费者评论进行语义网络分析和情感分析，本文探究出餐饮消费者选择和满意度的主要影响因素，即消费者对餐饮的关注和需求，从而在一定程度上对餐饮商家提供服务有一定的理论指导作用。

2) 现实意义

通过可视化平台构建，将研究数据对商家直观展示，商家画像、同类竞品分析等功能的实现，为餐饮商家管理提供现实依据，对商家服务定位及提升改进有直接

的现实意义。

二、功能介绍

2.1 结构方程模型

在社会科学研究领域，有很多概念是无法直接测量和观测的，例如人们的焦虑、态度、动机、工作压力、满意度等等，它们都是抽象的概念，这些无法直接测量的变量在结构方程中成为潜在变量或构成变量；而像身高、工作和学历等可以直接测量和量化的变量，称为测量变量或观察变量。通过构建结构方程模型可以将不能直接观察的概念，通过潜变量的形式转换成由多个观测变量构成的可以量化的概念。结构方程分析不仅可以估计测量过程中的误差，还能够评估测量的信度与效度。

我们小组严格按照“提出假设—构建模型—模型检验—模型评估”的步骤，针对消费者餐饮消费行为，利用顾客价值模型的五个维度，选取与研究主题有关的相关变量，构建研究模型，探究哪些因素会切实影响到消费者选择餐厅的行为，并利用 SPSS 相关软件对所收集的量表数据进行计算，得出可靠的研究结果，进而为餐厅进一步改善自身经营状况提出可行性建议。通过结构方程分析，我们将不可量化的消费者选择行为通过餐厅环境、菜品口味等观察变量转换成可以由多个观察变量构成的可以量化的概念。从结构方程的结果和数据来看，消费者选择餐厅时更加注重服务质量、菜品口味、平台评分和菜品价格。这四个因素对消费者选择行为有显著影响，餐厅可以结合自身经营情况，着重在上述四个方面进行改进和提高。

2.2 观点挖掘

对于一家产品或服务的提供者而言，成功的关键之一是了解客户对自己产品的感觉。客户经常在互联网上表达自己的观点和意见，提供关于产品的想法和感受，这些观点中充满了丰富的信息。但随着信息技术的发展，数据快速增长，人工分析如此多的信息是不可能的，因此需要观点挖掘的技术来自动化理解客户的感受和

情绪。

观点挖掘通常是以文本分析技术为主体，利用计算机从客户提供的文本中自动识别其情感与观点（积极的、消极的或中性的）。这样可以让产品提供者快速了解客户的想法，找出他们喜欢什么、不喜欢什么，以及为什么，这样就能对自己的产品做出改进，使其更符合客户的需求。观点挖掘的技术节省了产品提供者的时间和金钱，也减少了人为手动处理所带来的主观性与不确定性。对客户进行准确和客观的分析，对产品提供者的成功而言十分重要。

2.3 情感分析

情感分析是观点挖掘中最重要的方面之一。它的本质上是利用自然语言处理工具对词语背后的情感(即情绪、态度、意见、思想等)进行分析。情感分析可以帮助产品提供者快速了解客户对产品的感受。客户是对产品非常满意，下次还会再来？还是非常不满？如果满意，是什么方面使得大多数客户满意？如果不满意，是什么方面影响了客户的感受？产品提供者做出了改进措施后，客户的反应是怎样的？随着时间变化，客户的态度还是与以前一样吗？情感分析让产品提供商了解以上问题，使其能更好地改进自己，为客户提供更贴切需求的产品。



图 2-1 情感分析作用

2.4 语义网络

语义网络分析是一种基于词频共现的文本分析技术，可以直观的反应复杂概

念之间的关联。在对评论进行情感分析时，需要进行切词处理，然后通过情感分析模型对文本进行分析，但是切词后的语句之间结构相对松散，而语义网络分析可以将词与词间关系进行整合，清楚的反映评论信息。通过构建语义网络，可以将评论数据以网络的形式连接为一个整体，在一定程度上实现数据可视化，直观地反映出评估对象与评论之间的联系，挖掘用户对商家服务的关注点。

2.5 同类竞品分析

基于本文算法得出的结果，本文对成都市温江区 68 家餐厅进行了分类，并分别对每一种类别下的餐厅评论进行文本挖掘，得出每家餐厅的情感得分。在该同类竞品分析板块，本文选取餐厅所属类别的情感得分最高的餐厅，以及情感得分最低的餐厅，与当前餐厅进行比对，分别展示三家餐厅的观点标签结果和历史情感得分波动情况，通过横向对比分析帮助餐厅了解市场情况。

2.6 餐饮年度总结

餐饮年度报告是以一年为研究区间，为每个商家提供年度评论信息总结的特色服务。餐饮年度报告是通过一定的数据整合处理和相关算法，向商家直观展示年度热评、年度热评菜品、年度死忠粉、年度提升指南等内容，有利于商家对过去一年评论内容进行总结回顾。

三、结构方程模型

3.1 模型假设

针对感知价值具体内容，SHETH 等(1991)构建了五维度顾客价值框架，包括功能价值、社会价值、情感价值、认知价值和条件价值。SWEENEY 等(2001)基于该模型，构建了感知价值量表，识别出情感价值、社会价值以及价格和质量两类功能价值等维度。基于前面对餐厅游客感知价值的讨论，四维度分类法更适用于本研

究，具体包括功能价值、认知价值、情感价值和社会价值。

第一，功能价值，如上文所述，包括服务质量、价格、菜品种类、环境等因素，是顾客对餐厅的基础价值要求，应是餐厅游客感知价值维度之一。

第二，认知价值是消费者对产品满足其好奇心、新奇感和求知欲的感知。其中，猎奇是认知价值的重要因素，有学者甚至认为认知价值就是猎奇。猎奇行为是个体在某种情境下尝试新鲜事物的倾向性。在美食消费中，满足食客对菜肴及用餐过程的猎奇心很重要。已有研究发现，陌生的原料、独特的食物名称以及背后的故事会提升顾客对美食的原真性感知。因此，认知价值应是餐厅游客感知价值维度之一。

第三，社会价值是顾客通过与一个或多个特定社会群体联系而感知到的效用。旅游是人类在长期交往中，建立人际关系和显示自身优势的心理适应器的反映。就餐饮而言，社交性饮食是创建和维持社会关系的重要方式。餐饮服务是一种特殊情境，食客对于餐厅质量的评价不仅基于产品本身，而同时考虑社会交往的因素。事实上，顾客大多结伴而行，并愿意将餐厅作为社交场所。除社会交往外，形象塑造也是社会价值的重要部分，丰富多样的餐厅经历常被人们用以塑造形象。HA 等 (2012)发现人们认为在高档餐厅就餐能够突显社会地位。因此高档餐厅具有较高的社会价值。因此，社会价值应是餐厅游客感知价值维度之一。

第四，情感价值是顾客从由产品带来的情感状态中获得的效用。在学界对情感价值的分类中，积极情感和消极情感双维度分类法认可度较高。就餐饮而言，DESMET 等(2008)认为人们品尝美食而产生的积极情感包括开心、惊喜、满意、放松等，消极情感包括失望、恶心、厌倦、生气等。因此，情感价值应是餐厅游客感知价值维度之一。

由此，本研究建立模型如下：

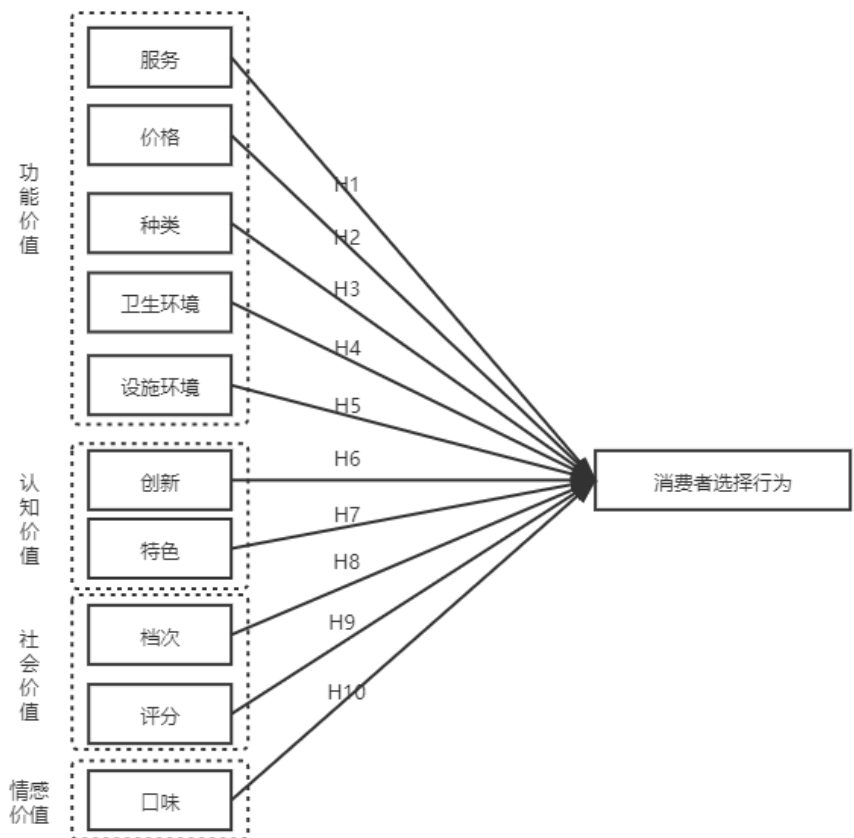


图 3-1 结构方程模型图

3.2 变量假设

3.2.1 服务质量

服务质量从本质上是顾客在服务传递的过程中形成的一种感知,它由顾客的服务期望与实际服务经历的比较决定。在实际运营中，一家餐厅可制订详细地服务创新策略，或从服务人员入手，加强服务人员的服务意识，使其在工作过程中不断总结更好的服务方式；或从整体服务方式入手，提出全新的服务概念，以此来招揽更多顾客。因此，在进行服务的过程中，餐厅对消费者的了解已经较为深入，而这将有利于更精准地把握消费心理，从而让消费者餐厅产生好感，增加到店次数。

故本文提出下述假设：

H1：服务质量对消费者选择行为有正向影响。

3.2.2 菜品价格

餐饮定价是销售和成本控制的一个重要环节，价格将会直接影响到企业的经济效益，体现了一个餐厅的产品定位，价格还是企业营销的重要手段，在餐饮企业营销组合的诸多因素中，价格是作用最直接、见效最快的一个，在扩大市场占有率和推广新产品时，价格是一个常用的营销战略。不同的消费者拥有不同的消费能力，商家在定价时需充分考虑餐厅定位以及其受众群体。

故本文做出如下假设：

H2：价格对消费者选择具有正向影响

3.2.3 菜品种类

产品特性指的是该店铺产品的一系列属性，如价格、质量、种类等等。菜品种类作为一家餐厅的产品特性，对消费者的选择起到一定的决策作用。国外研究学者Lindquist(1974)提出了影响消费者商店选择的功能因素，即消费者可以感受到的、会对自身选择行为产生影响的、商店特征方面的属性，其中就包括商品种类。消费者在选择餐厅的时候，会综合考虑店中的菜品种类是否满足自己的选购要求。在这个过程中，消费者逐渐形成自己的心理预期，对其感知利得与感知利失之间进行比较和权衡，影响着消费者购买动机的形成和最终的购买决定。

故本文作出如下假设：

H3：菜品种类对消费者选择行为有正向影响。

3.2.4 店铺环境

店铺环境包括内、外两部分，内环境主要指餐厅桌椅布局、照明、店内装修等。它是消费者接触菜品及获取消费体验的主要场所。特别地，餐厅的店铺内环境中，卫生情况是十分重要的衡量因素。一家餐厅如果布局合适，卫生状况良好，会提高消费者的用餐体验。外环境主要指店铺门头和店铺橱窗，它是吸引顾客进店的重要

因素。合理有序的空间布局和明亮舒适的照明可以营造良好的购物环境，延长顾客的停留时间，增加商品售出几率。基于此，本文提出如下假设：

H4：店铺卫生环境对消费者选择行为有正向影响。

H5：店铺设施环境对消费者选择行为有正向影响

3.2.5 菜品创新

创新是指餐饮企业能否及时推出新的菜品，以满足消费者的需求。餐饮企业所提供的产品，就其功能而言，能永远满足人们在饮食上的需求，因而它是一种长线产品。但是，人们在不同的时期、不同的季节以及不同的经济条件下，对饮食的内容、方式及环境有着不同的追求，即人们饮食习惯的稳定性与求同存异性并存。此外，创新菜品还能使消费者对产品满足其好奇心、新奇感和求知欲的感知。

故本文做出如下假设：

H6：创新对消费者选择具有正向影响。

3.2.6 特色菜

特色菜是指店家有自己独特的秘方，味道特别好并且其它餐厅都没有的菜品，或者是店家主打的，能够鲜明体现餐厅特色的菜品。每一个餐厅或多或少都有着自己的一些特色菜品，特色菜不仅能够鲜明定位自己的餐厅，让顾客一来便知，还能和其它餐厅差异化竞争，以其特色菜赢得顾客的光临。在我们的前期调查中了解到某些顾客会因为某家店特色菜品相当吸引人而且选择前往其就餐，特色菜对消费者的选择行为有一定的影响。

故本文做出如下假设：

H7：特色菜对消费者选择具有正向影响。

3.2.7 档次

消费者对不同档次餐厅的选择反映了消费者的收入水平、生活质量和消费能力等状况，可用于判定个体所属的社会阶层。在不同档次的餐厅中，消费者对于实

体环境、员工态度及其行为都会有不同的期望与定位，继而产生不同的反应与评价。根据马斯洛需求层次理论，人们一直在寻求分属不同层次需要的渐次满足，在低层次需要得到满足的基础上，人们渴望沟通交流、获得尊重、实现自我价值等高层次需求。档次更高的餐厅更能满足消费者对于社会地位、情感交流等需求。基于以上逻辑。

故本文做出如下假设：

H8：店铺档次对消费者选择行为有正向影响。

3.2.8 店铺评分

数值型在线评论的效价作为一种最为简单常见、在电商平台上长期客观存在且容易获取的在线评论特征属性，是影响消费者行为意愿的重要因素之一。数值型在线评论的效价所表示的是众多其他消费者所给出的评分的平均数，以此来表示消费者对于该产品或服务所持有的态度是正面的、负面的还是中性的。电商平台会在统计所有消费者给出的评级后进行计算并给出一个总体评价的平均值供消费者参考。积极的在线评论效价会对消费者行为产生积极的影响，消极的在线评论效价会对消费者行为产生消极的影响。

故本文提出如下假设：

H9：店铺评分对消费者选择行为有正向影响。

3.2.9 口味（味觉感知）

味觉是人与生俱来的感觉，人类通过分布在舌头、鄂部以及喉咙处的味蕾来感知味道。相比视觉与听觉，味觉感知被认为是更“具体”的，因此也更容易接近。味觉所获得的体验更容易对人们一些社会判断产生具身效应，例如，喝了苦味饮料的被试比喝了甜味或无味饮料的被试表现出更高的厌恶情绪，并且对他人的道德评判更加严苛，而甜味糖果则让人更愿意帮助他人。因此，消费者在一家餐厅获得更好的味觉感知（即菜品口味更好），更容易获得正向的心理体验，对餐厅更容易有正向评价。

故本文提出如下假设：

H10：菜品口味对消费者选择行为有正向影响。

3.3 研究结果

3.3.1 描述性统计

表 3-1 描述性统计

	描述统计							
	最小值 统计	最大值 统计	均值 统计	标准 偏差 统计	偏度		峰度	
	统计				统计	标准 错误	统计	标准 错误
A1	1	5	3.85	.843	-.386	.169	.087	.336
A2	1	5	3.86	.839	-.515	.169	.328	.336
A3	1	5	3.90	.828	-.539	.169	.418	.336
A4	0	4	2.54	1.111	-.503	.169	-.494	.336
B1	2	5	4.31	.623	-.569	.169	.630	.336
B2	2	5	3.99	.729	-.136	.169	-.692	.336
B3	2	5	4.19	.682	-.631	.169	.676	.336
B4	0	4	2.43	1.136	-.471	.169	-.667	.336
C1	2	5	3.84	.724	-.513	.169	.406	.336
C2	2	5	3.78	.761	-.269	.169	-.181	.336
C3	1	5	3.78	.748	-.733	.169	1.338	.336
C4	0	4	2.42	1.028	-.383	.169	-.627	.336
D1	2	5	3.95	.741	-.356	.169	-.089	.336
D2	2	5	3.94	.733	-.355	.169	-.031	.336
D3	1	5	3.78	.834	-.575	.169	.447	.336
D4	0	4	2.55	1.120	-.497	.169	-.675	.336
E1	2	5	3.86	.728	-.075	.169	-.479	.336
E2	2	5	3.82	.771	-.056	.169	-.606	.336
E3	2	5	3.83	.673	.028	.169	-.422	.336
E4	0	4	2.21	1.074	-.195	.169	-.675	.336
F1	1	5	3.18	.920	.005	.169	-.192	.336
F2	1	5	3.20	.926	.029	.169	-.161	.336
F3	1	5	3.11	.942	.207	.169	-.223	.336
F4	0	4	1.77	1.088	.288	.169	-.631	.336
G1	2	5	3.90	.645	-.345	.169	.520	.336
G2	2	5	3.91	.654	-.330	.169	.429	.336
G3	2	5	3.86	.687	-.528	.169	.652	.336
G4	0	4	2.64	1.002	-.869	.169	.327	.336
H1	2	5	3.75	.745	-.548	.169	.268	.336
H2	2	5	3.63	.751	-.227	.169	-.201	.336
H3	2	5	3.75	.758	-.621	.169	.313	.336
H4	0	4	2.37	1.077	-.349	.169	-.612	.336
I1	1	5	3.24	.938	-.143	.169	-.358	.336
I2	1	5	3.08	.908	.081	.169	-.281	.336
I3	1	5	3.03	.927	.163	.169	-.416	.336
I4	0	4	1.86	1.006	.351	.169	-.417	.336
J1	1	5	3.63	.880	-.331	.169	-.152	.336
J2	1	5	3.62	.826	-.218	.169	.068	.336
J3	1	5	3.49	.857	-.016	.169	-.176	.336
J4	0	4	2.32	1.114	-.273	.169	-.713	.336
K1	2	5	4.10	.742	-.728	.169	.708	.336
K2	2	5	4.06	.693	-.429	.169	.257	.336
K3	2	5	4.04	.636	-.259	.169	.283	.336
K4	0	4	2.87	1.121	-1.019	.169	.295	.336

3.3.2 测量模型分析

表 3-2 变量对应表

服务环境	A
口味	B
菜品种类	C
卫生环境	D
设施环境	E
档次	F
评分	G
价格	H
创新菜品	I
特色菜	J
消费者选择	K

3.3.3 信度分析

信度是指所测量数据与结论的可靠性程度，即重复测量，结果是否基本一致。本研究采用目前最常用的 Cronbach's Alpha 系数、组合信度（Composite Reliability,C.R.）和 AVE（Average Variance Extracted，平均方差提取值）三个值来分析各建构变量的信度。

表 3-3 模型结构的内部一致性

	Cronbach's alpha	Composite Reliability	AVE
A	0.775	0.8956	0.741
B	0.698	0.8369	0.6318
C	0.773	0.8893	0.7282
D	0.732	0.7947	0.6594
E	0.791	0.8418	0.6407
F	0.850	0.8356	0.6289
G	0.721	0.7893	0.556
H	0.818	0.881	0.712
I	0.834	0.6216	0.5511
J	0.778	0.8151	0.5963
K	0.773	0.7154	0.5567

从表 3 可以看出，所有的 Cronbach's Alpha 均大于 0.7，AVE 都大于 0.5，组合信度都大于 0.6，表明该模型测量的变量具有较好的信度。

3.3.3 效度分析

本研究采用聚合效度和区别效度对问卷的结构进行了检验。结构效度是指测量结果体现出来的某种结构与测值之间的对应程度。结构效度分析所采用的方法是因子分析和相关性分析，其包括聚合效度 (Convergent Validity) 和区分效度 (Discriminant validity)。聚合效度是指测量同一概念多重指标之间的关联程度。Gefen 等认为，测量变量在对应建构变量上的因子载荷大于 0.6 时，测量量表具有较好的聚合效度；Fornell 等认为，测量量表要有很好的区分效度，各建构变量的平均抽取方差 (Average Variance Extracted, AVE) 应该大于 0.5。测量结果表明，本研究数据具有较好的聚合效度和区分效度。

	服务质量	口味	菜品种类	卫生环境	设施环境	档次	评分	价格	创新菜品	特色菜	消费者选择
A1	0.881										
A2	0.854										
A3	0.847										
B1		0.837									
B2		0.735									
B3		0.809									
C1			0.848								
C2			0.86								
C3			0.852								
D1				0.818							
D2				0.806							
E1					0.845						
E2					0.834						
E3					0.716						
F1						0.798					
F2						0.779					
F3						0.802					
G1							0.704				
G2							0.735				
G3							0.795				
H1								0.823			
H2								0.894			
H3								0.812			
I2									0.684		
I3									0.659		
J1										0.801	
J2										0.815	
J3										0.695	
K1											0.638
K2											0.728
K3											0.658

图 3-2 测量模型的因子分析

相关性												
		A	B	C	D	E	F	G	H	I	J	K
A	皮尔逊相关性	1	.345**	.410**	.645**	.496**	.266**	.582**	.244**	.145*	.408**	.454**
	Sig. (双尾)		0	0	0	0	0	0	0	0.037	0	0
	个案数	208	208	208	208	208	208	208	208	208	208	208
B	皮尔逊相关性	.345**	1	.406**	.460**	.347**	0.097	.520**	.434**	0.025	.421**	.566**
	Sig. (双尾)	0		0	0	0	0.162	0	0	0.725	0	0
	个案数	208	208	208	208	208	208	208	208	208	208	208
C	皮尔逊相关性	.410**	.406**	1	.410**	.325**	.345**	.409**	.254**	.393**	.414**	.236**
	Sig. (双尾)	0	0		0	0	0	0	0	0	0	0.001
	个案数	208	208	208	208	208	208	208	208	208	208	208
D	皮尔逊相关性	.645**	.460**	.410**	1	.723**	.391**	.626**	.289**	.217**	.565**	.453**
	Sig. (双尾)	0	0	0		0	0	0	0	0.002	0	0
	个案数	208	208	208	208	208	208	208	208	208	208	208
E	皮尔逊相关性	.496**	.347**	.325**	.723**	1	.569**	.513**	.149*	.386**	.482**	.306**
	Sig. (双尾)	0	0	0	0		0	0	0.032	0	0	0
	个案数	208	208	208	208	208	208	208	208	208	208	208
F	皮尔逊相关性	.266**	0.097	.345**	.391**	.569**	1	.264**	0.087	.610**	.491**	-0.014
	Sig. (双尾)	0	0.162	0	0	0		0	0.213	0	0	0.845
	个案数	208	208	208	208	208	208	208	208	208	208	208
G	皮尔逊相关性	.582**	.520**	.409**	.626**	.513**	.264**	1	.437**	.179**	.455**	.572**
	Sig. (双尾)	0	0	0	0	0	0		0	0.01	0	0
	个案数	208	208	208	208	208	208	208	208	208	208	208
H	皮尔逊相关性	.244**	.434**	.254**	.289**	.149*	0.087	.437**	1	.250**	.395**	.605**
	Sig. (双尾)	0	0	0	0	0.032	0.213	0		0	0	0
	个案数	208	208	208	208	208	208	208	208	208	208	208
I	皮尔逊相关性	.145*	0.025	.393**	.217**	.386**	.610**	.179**	.250**	1	.444**	0.006
	Sig. (双尾)	0.037	0.725	0	0.002	0	0	0.01	0		0	0.932
	个案数	208	208	208	208	208	208	208	208	208	208	208
J	皮尔逊相关性	.408**	.421**	.414**	.565**	.482**	.491**	.455**	.395**	.444**	1	.309**
	Sig. (双尾)	0	0	0	0	0	0	0	0	0		0
	个案数	208	208	208	208	208	208	208	208	208	208	208
K	皮尔逊相关性	.454**	.566**	.236**	.453**	.306**	-0.014	.572**	.605**	0.006	.309**	1
	Sig. (双尾)	0	0	0.001	0	0	0.845	0	0	0.932	0	
	个案数	208	208	208	208	208	208	208	208	208	208	208
**. 在 0.01 级别（双尾），相关性显著。												
*. 在 0.05 级别（双尾），相关性显著。												

图 3-3 相关性

3.3.4 结构方程分析

我们用 SPSS 来检测我们的模型，通过线性回归来处理服务质量、口味、菜品种类、卫生环境、设施环境、档次、评分、价格、创新菜品、特色菜 10 个自变

量对消费者行为选择因变量的影响。

表 3-4 模型摘要

模型摘要					
模型	R	R 方	调整后 R 方	标准估算的错误	更改统计
					R 方变化量
1	.773 ^a	.597	.577	.41176	.597
a. 预测变量: (常量), I, H, A, J, C, B, E, G, F, D					

表 3-5 系数

系数 ^a								
模型		未标准化系数		标准化系数	t	显著性	B 的 95.0% 置信区间	
		B	标准错误	Beta			下限	上限
	(常量)	.564	.235		2.399	.017	.100	1.028
	A	.162	.057	.181	2.863	.005	.051	.274
	B	.249	.065	.233	3.808	.000	.120	.379
	C	-.065	.058	-.065	-1.124	.262	-.178	.049
	D	.051	.080	.052	.635	.526	-.107	.209
	E	.104	.075	.106	1.387	.167	-.044	.252
	F	-.123	.052	-.157	-2.355	.020	-.227	-.020
	G	.205	.076	.182	2.700	.008	.055	.354
	H	.391	.053	.420	7.417	.000	.287	.495
	I	-.057	.054	-.069	-1.058	.292	-.162	.049
	J	-.051	.057	-.058	-.890	.375	-.163	.062
a. 因变量: K								

通过线性回归来分析验证自变量对因变量的影响：

表 3-6 结构方程分析结果

X→Y	P	是否有显著影响	是否支持	路径系数
A→K	P<0.01	是	是	0.162
B→K	P<0.01	是	是	0.249
C→K	P>0.05	否	否	
D→K	P>0.05	否	否	
E→K	P>0.05	否	否	
F→K	P<0.05	是	否	-0.123

G→K	P<0.01	是	是	0.205
H→K	P<0.01	是	是	0.391
I→K	P>0.05	否	否	
J→K	P>0.05	否	否	-

从表中我们可以看到，A→K、B→K、G→K、H→K 的 P 值小于 0.05，且具有显著正向影响，因此假设 H1、H2、H7、H8 成立，假设 H3、H4、H5、H6、H9、H10 不成立。

3.4 研究结果

3.4.1 得到支持的假设

（1）服务质量

根据收集到的样本数据，进行回归分析后得到结果：

产品价格的 $p<0.01$ ，这表明服务质量对消费者选择行为有显著影响； $\beta=0.162$ ，说明服务质量对消费者选择行为有正向影响，假设 H1 成立。餐厅的服务质量越高，人们越愿意在该餐厅就餐。

产生影响的原因如下：

服务质量从本质上是顾客对餐厅的实际服务产生了特定的感知，并且和预先的服务期望进行比较的过程，一家餐厅可以通过不断对自身服务细节和服务人员进行改进来加深消费者的好感程度，最终增加消费者的消费次数。如果餐厅的服务质量无法达到消费者的心理预期，消费者会对餐厅产生诸如失望、生气等负面情绪，最终影响到自身的选择。

（2）口味

根据收集到的样本数据，进行回归分析后得到结果：

产品价格的 $p<0.01$ ，这表明服务质量对消费者选择行为有显著影响； $\beta=0.249$ ，说明菜品口味对消费者选择行为有正向影响，假设 H10 成立。餐厅菜品的口味越好，人们越愿意在该餐厅就餐。

产生影响的原因如下：

味觉是消费者获得美食体验的重要来源，人与生俱来的感觉。相比视觉、听觉

等，味觉感知在餐厅消费中更容易获得，也更容易成为消费者衡量一家餐厅好坏的影响因素。在实际中，不同的菜品口味会带来不同的消费体验，例如一家餐厅口味不好，会让消费者产生更多的负面心理，很难获得回头客，只能在短期获得一定的发展。

（3）评分

根据收集到的样本数据，进行回归分析后得到结果：

评分的 $p < 0.01$ ，这表明相关美食平台的评分对消费者选择行为有显著影响； $\beta = 0.205$ ，说明评分对消费者选择行为有正向影响，假设 H9 成立。美食平台的评分越高，人们越愿意在该餐厅就餐。

产生影响的原因如下：

平台的流量和用户都是餐厅有力的宣传者。美团等平台早已深入人们的生活，这些平台自身的知名度、信誉等特性在一定程度上会转化成平台内商家的信誉基础。当一家餐厅在平台上拥有更高的星级或是评分时，消费者对该餐厅会产生一定的初始信任，消费者对该餐厅的选择意愿也会增强。因此，评分对消费者选择行为有正向影响。

（4）价格

根据收集到的样本数据，进行回归分析后得到结果：

产品价格的 $p < 0.01$ ，这表明价格对消费者选择行为有显著影响； $\beta = 0.319$ ，说明价格对消费者选择行为有正向影响，假设 H2 成立。餐厅价格越合适，人们越愿意在该餐厅就餐。

产生影响的原因如下：

价格是消费者衡量该餐厅是否值得消费的主要因素。当消费者在选择餐厅时，会考虑该餐厅的价格对自己来说是否合理。对消费者来说，不同类的餐厅有不同的心理价位，如果不同种类餐厅放在一起进行选择，哪家餐厅更符合自身产品的合理价位，哪家餐厅就更容易被选择；如果是在同类餐厅中进行选择，价格合理的餐厅更能打动消费者。如果一家餐厅的定价超出了自己的定位，那么就算菜品做得再好，价格也会“劝退”一批消费者。

3.4.2 被否定的假设

(1) 菜品种类

根据收集到的样本数据，进行回归分析后得到结果。菜品种类的 $p>0.05$ ，这表明菜品种类对消费者选择行为没有显著影响，不符合假设。可能的原因如下：

菜品种类是一家餐厅可供消费者选择的多样性体现，菜品种类的丰富程度在一定程度上能够影响消费者是否进行消费。但在实际当中，餐厅的菜品种类一般是根据自身特性，比如火锅餐厅的菜品种类是依据菜品荤素划分。虽然每个餐厅划分的种类多少有差异，但不同种类餐厅的菜品划分规则不同，很难在数量少进行比较，因此消费者对菜品种类的重视度不高。

(2) 餐厅环境（卫生环境、设施环境）

根据收集到的样本数据，进行回归分析后得到结果。餐厅环境的两项度量指标， p 值都大于 0.05，这表明餐厅环境对消费者选择行为没有显著影响，不符合假设。可能的原因如下：

因为问卷中对于地域存在一定限制，大部分参与问卷调查的消费者生活在成都市温江区，也就是说，可能绝大多数餐厅对于消费者而言是曾经见过甚至熟悉的。因此，消费者对于餐厅所处的环境大致了解，已经无法对消费者选择行为产生显著影响。

(3) 创新菜品

根据收集到的样本数据，进行回归分析后得到结果。创新菜品的 $p>0.05$ ，这表明餐厅的创新菜品对消费者选择行为没有显著影响，不符合假设。可能的原因如下：

餐厅实质上是一种长线产品，它需要不断地满足消费者在饮食上的需求，与时俱进、推陈出新是餐厅必备的技能之一。相应地，餐厅的创新菜品也是一种周期较长的产品，在消费者流动性较强的地方（如大学周边），很难有大规模的消费者能明确感知到一家餐厅是否有创新菜品，因此无法对消费者选择行为产生显著影响。

(4) 档次

餐厅档次的 $p<0.05$ ，这表明档次对消费者选择行为有显著影响；但是 $\beta=-0.123$ ，说明菜餐厅档次对消费者选择行为有负向影响，假设 H8 不成立。餐厅档次越高，

人们在该餐厅就餐的意愿越小。产生影响的原因如下：

餐厅的不同档次可以反映消费者的收入、消费能力等情况，但是对本次参与问卷调研的群体进行简单的人口统计发现收入并没有太大差异，而且大部分都是学生身份。参与调研的群体对于饭店档次并没有必要的追求，还可能因为收入水平限制无法去更高档次的饭店就餐。因此，餐厅档次对消费者行为选择有负向影响。

（5）特色菜

根据收集到的样本数据，进行回归分析后得到结果。特色菜的 $p>0.05$ ，这表明餐厅的特色菜对消费者选择行为没有显著影响，不符合假设。可能的原因如下：

实际中，很多餐厅都会推出招牌菜、特色菜系列，这已经成为餐厅吸引顾客的常见营销方式，因此消费者对这类营销方式已经拥有了一定的熟知度。这在很大程度上导致餐厅的特色菜无法吸引新的消费者到店，而餐厅固有的消费者对于特色菜品已经有过尝试，因此，无法对消费者行为选择产生显著影响。

四、算法

4.1 数据收集

目前国内存在大量可用于评论观点挖掘的平台主要有两类：以京东、淘宝为代表的电商平台，以美团、大众点评、饿了么为代表的美食（外卖）点评平台。由于美食平台上的数据是大家更熟悉的，利用后续分析，因此选取此类平台作为数据来源。美团使用者广泛，且其数据能在八爪鱼数据采集器上方便爬取，因此选取八爪鱼采集器作为美团评论数据的采集工具。所采集数据部分预览如下：

表 4-1 商家评论采集数据预览

店铺名称	用户名	评论
韩呈一品韩式自助 (温江店)	iBg177 10682 7	和室友一起去的，环境很干净卫生，菜品多，有很多小零食（蛋糕，凉面，卤菜，蛋炒饭，泡芙，寿司等），各种饮料（雪碧，可乐，奶昔，果啤，光明酸奶等），水果（西瓜，哈密瓜，橘子，圣女果等），服务态度也不错，会随时来收盘子，换烤纸，性价比高，推荐

韩呈一品韩式自助 (温江店)	dPM46 68804 86	味道和别的烤肉差不多吧，不过餐厅的地面是真的滑，一不小心就可能摔跤，卫生间环境太差了，一开始去上了个厕所，然后胃口都没了，基本菜品还是有的，鸡翅，牛肉，脑花等等，那里的凉面应该算是我觉得最好吃的了，总体来说算是合格吧
韩呈一品韩式自助 (温江店)	VjT88 54284 31	干净卫生，味道不错，价格实惠，跟室友去过两次了，都觉得很棒，会多次去的，菜品不错，服务很好，烤肉滋滋发出声响，一滴热油顺着饱满的肉纹路慢慢滑下，令人心醉。细细的嗅，慢慢的闻，馋虫迅速被勾起。终于忍不住，顾不得烫，一咬就是一大口，哇！
韩呈一品韩式自助 (温江店)	匿名用户	素菜品种少，荤菜品种少，饮料难喝（说它是气泡水吧味道怪得很，说它是水又对不起水）油像注水的，烤盘设计不合理，油溅到处都是。衣服带着围裙还是被溅的脏兮兮的。手和脸都被烫到了，几个人吃了之后窜稀，室友投诉我们说上厕所过于频繁。「#套餐：单人自助烤肉」
韩呈一品韩式自助 (温江店)	吃货 妹纸 12	第一次去还不错，就是菜品太少了。服务态度也挺好的。和家人一起去的，环境也还可以，还是那句菜品太少，菜品太少，菜品太少。肉考出来要自己加点味道会更好吃一些。水果也样品也很少，如果朋友聚餐都可是比较不错的选择，图片总得美颜相机照的

最终采集所得数据有 14 个分类共 68 家餐厅(部分餐厅有多种分类)总共 69959 条，具体分类、评论数量如下：

表 4-2 数据分类及每类评论数

店铺名称	所属分类	评论数
蜀大侠火锅（温江店）	火锅	669
江南道自助烤肉海鲜涮锅（柳浪湾店）	烧烤烤肉，自助餐	2761
泰澜得泰国菜（温江店）	东南亚菜	1215
大芙蓉城自助火锅烤肉（温江店）	火锅，自助餐	2813
汉堡王（温江新尚天地店）	小吃快餐，西餐	380
雅府正红木桶鱼（杨柳河店）	火锅	2808
江南道自助烤肉海鲜涮锅（温江大学城店）	烧烤烤肉	866
乡村基（国色天香店）	小吃快餐	1570
渝味晓宇火锅（温江店）	火锅	1250
烤匠麻辣烤鱼	川湘菜，香锅烤鱼	790
华莱士（农大店）	小吃快餐，西餐	374
众里寻她·千百度自助烤肉（温江店）	烧烤烤肉	2527
官人桥蒙自过桥米线	小吃快餐	2787

屋头串串（临江路店）	小吃快餐	1920
鸡毛店·川菜（同兴西路店）	川湘菜	1234
猪圈小米辣（温江1店）	火锅	773
芭莉与彩虹西餐厅（成都合生汇店）	西餐	2260
尼克熊西餐（财大东门店）	西餐	1321
蛙三泡椒牛蛙（温江合生汇店）	川湘菜	466
韩婆婆小院烧烤	烧烤烤肉，自助餐	789
俭让佬妈市井串串	火锅，小吃快餐	2296
北木南烤肉和酒（西南财大店）	烧烤烤肉	440
芭夯兔（温江店）	火锅，聚餐宴请	2026
柳飘飘亚洲烤肉集合店（温江店）	烧烤烤肉	990
韩呈一品韩式自助（温江店）	自助餐	529
吉布鲁牛排海鲜自助（温江合生汇店）	自助餐	2810
诺曼汀艺术烘焙（财大店）	甜点饮品	777
麦香园（庆丰店）	甜点饮品	530
豪客来牛排（成都温江新尚天地店）	西餐	375
泥巴小院市井火锅（温江店）	火锅	765
五味鲜肉抄手（温江总店）	小吃快餐	552
川西坝子火锅（温江2.0店）	火锅	608
伍氏猪脚（柳浪湾店）	小吃快餐	1198
守柴炉烤鸭（温江新尚天地店）	京菜鲁菜	396
在首尔韩国料理·自助烤肉（温江店）	日韩料理	813
鱼契·烤鱼（西南财大总店）	川湘菜，香锅烤鱼	456
谦和清心素食（温江店）	自助餐	995
黄记煌三汁焖锅（温江合生汇店）	京菜鲁菜	2808
司乎.日式鲜吐司（新光天地店）	西餐	434
锦荷苑自助火锅	火锅，自助餐	1020
韩国小胖料理自助烤肉	自助餐，烧烤烤肉	600
三牦记.牛肉火锅（温江店）	火锅	827
龙腾庭院火锅（恒大新城店）	火锅	807
好利来（温江合生汇店）	蛋糕甜品	385
台新石头火锅（温江店）	火锅	1542
九孔莲藕汤锅（总店）	汤锅	681
码头故事火锅（温江店）	火锅	725
锦城大院火锅（大学城店）	火锅	1946
重庆辛麻道火锅	火锅	668
一只酸奶牛（财大店）	蛋糕甜品	832
书亦烧仙草（温江财大二店）	蛋糕甜品	210
牛蹄筋老面馆(财大店)	小吃快餐	883

嘎嘎鸭脑壳（温江南浦店）	小吃快餐	222
正新鸡排（繁华广场店）	小吃快餐	64
蜜语之恋（大学城店）	蛋糕甜品	866
青椒鱼（温江大学城店）	火锅	467
紫燕百味鸡（来凤店）	小吃快餐	450
宾 GO101 牛排披萨意面（温江合生汇店）	西餐	484
巴色鱼捞（温江直营店）	火锅	1093
奇味干锅馆焖锅乡	香锅烤鱼	293
独二家私房面（珠江广场店）	小吃快餐	483
爱达乐蛋糕（温江财大店）	蛋糕甜品	712
院坝记忆牦牛串串·龙虾·烧烤（鱼凫路总店）	烧烤、串串	680
王婆婆大牌档	川湘菜	771
烤鱼匠麻辣烤鱼	川湘菜，香锅烤鱼	521
香街坊汤锅馆（国色天乡店）	汤锅	583
余记全家福芋儿鸡（温江店）	火锅	828
裕昌火锅（德通桥路总店）	火锅	945

4.2 数据预处理

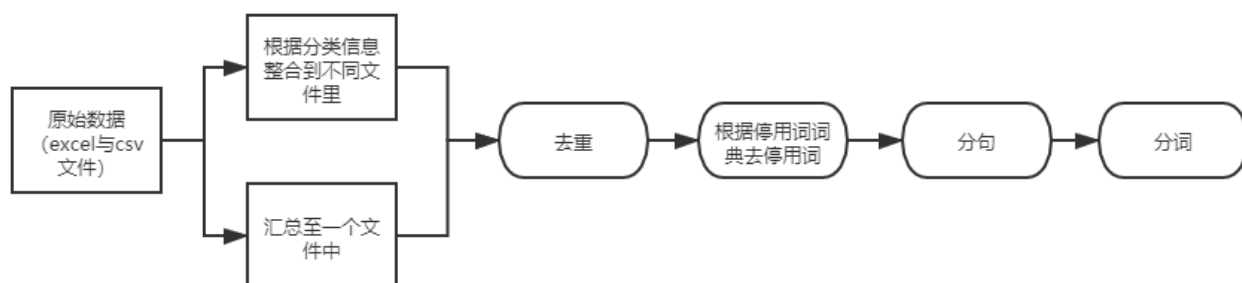


图 4-1 数据预处理流程图

4.3 模型与算法介绍

4.3.1 语义网络

为挖掘用户对商家服务的关注点，本文将对收集到的商家数据进行分类处理，依次对川湘菜、蛋糕甜品、东南亚菜、火锅、京菜鲁菜、聚餐宴请、日韩料理、烧烤烤肉、汤锅、甜点饮品、西餐、香锅烤鱼、小吃快餐、自助餐共十四个类别餐饮

评论进行语义网络分析。语义网络的构建主要通过武汉大学开发的反剽窃系统 ROSTCM6 来实现。在整个过程中，首先提取高频词，然后过滤无意义词、提取行特征、构建共现矩阵，最终得到语义网络图。

4.3.2 情感分析算法

使用知网的消极词典、积极词典与多个程度词典对文本进行情感分析。基于以下规则：

- 1) 一个积极词汇赋值为 1，一个消极词汇赋值为-1；
- 2) 若积极词或消极词前有程度副词，根据其程度的大小对 1 或-1 乘上一个权重；
- 3) 若积极词或消极词后有感叹符号等，同样对其进行加权；
- 4) 整篇文章的积极词得分与消极词得分汇总，大于等于 0 时将这篇文档判断为积极，小于 0 时将这篇文档判断为消极。

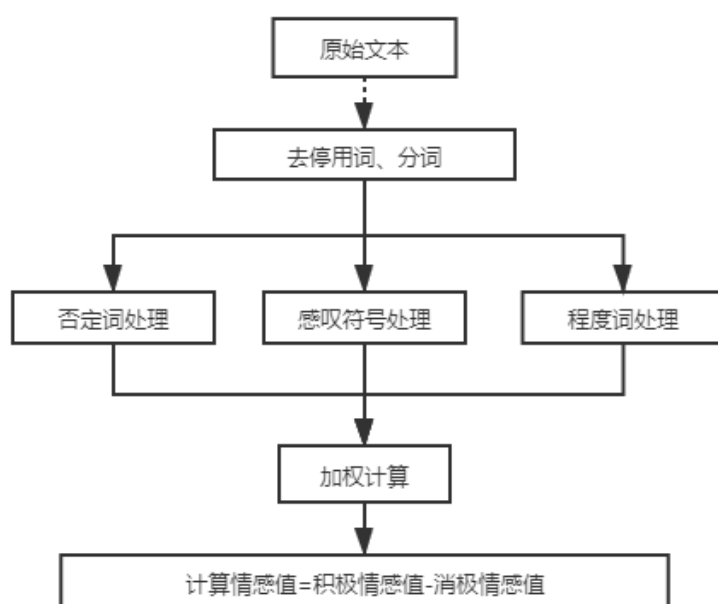


图 4-2 情感分析算法流程

4.3.3 关键词提取算法

采用 TextRank 对文章（即评论数据）进行关键词提取，此算法基于 PageRank。

PageRank 是由谷歌提出，用来计算网页重要性的算法。是将网页视为图上的阶段，用基于图的算法对网页得分进行计算。其基本思想有两条：

1. 一个网页被越多的其他网页链接，说明这个网页越重要。
2. 一个网页被一个越高权值的网页链接，也能表明这个网页越重要。

算法开始时会将所有网页的得分初始化为 1，然后通过多次迭代来对每个网页的分数进行收敛。收敛时的得分就是网页最终得分。若不能收敛，也可以通过设定最大迭代次数来对计算进行控制，计算停止时的分数就是网页的得分。

网页 V_i 的重要性得分计算方法为：

$$S(V_i) = (1 - d) + d * \sum_{j \in \text{In}(V_i)} \frac{1}{|\text{Out}(V_j)|} S(V_j)$$

其中 d 为阻尼系数，一般为 0.85。 $\text{In}(V_i)$ 表示能够跳转到网页 V_i 的页面，可以理解为图中入度对应的点。 $\text{Out}(V_j)$ 表示网页 V_j 能够跳转到的页面，可以理解为图中出度对应的点。

相比于 PageRank 将各个网页化成不加权的有向图，TextRank 是将文本化成带权无向图，将词语看成图上的节点，为了将权重考虑进行，使用一个新的计算公式来计算每个节点（词）的得分。

$$WS(V_i) = (1 - d) + d * \sum_{V_j \in \text{In}(V_i)} \frac{w_{ji}}{\sum_{V_k \in \text{Out}(V_j)} w_{jk}} WS(V_j)$$

将评论文本数据化成图包括以下步骤：

1. 对文本进行分词，将每个词作为节点加入图中；
2. 识别不同词之间的关系（本文中使用的词汇之间的共现关系，即先确定一个窗口大小，能够出现在同一个窗口内的词之间就存在关系），在词之间绘制边；
3. 进行基于图的排序算法直到收敛或达到最大迭代次数；
4. 根据节点的得分对其进行排序。

4.4. 结果分析——语义网络结果展示

烧烤烤肉高频词共现频数表和语义网络图分别如表 4-3 和图 4-3 所示，从表 4-3 和图 4-3 中可以看出，菜品、味道、服务是关键节点，以菜品节点为例，与之距离较近即共现频数较高的是服务、味道、新鲜等关键词，由此可以看出：消费者对于烧烤烤肉的菜品、味道、服务更为关注。

表 4-3 烧烤烤肉高频词共现频数表

语义关系		频数	语义关系		频数
味道	服务	53	好吃	新鲜	47
味道	菜品	51	味道	这家	47
味道	好吃	51	五花肉	菜品	47
菜品	服务	51	味道	态度	47
新鲜	菜品	51	五花肉	新鲜	46
味道	新鲜	50	好吃	环境	46
态度	服务	50	新鲜	环境	46
好吃	服务	49	好吃	五花肉	46
味道	环境	49	菜品	态度	46
味道	五花肉	49	菜品	这家	45
环境	服务	49	服务	这家	45
新鲜	服务	49	好吃	这家	45
好吃	菜品	48	环境	态度	45
环境	菜品	48	新鲜	态度	45
五花肉	服务	47	服务	朋友	45

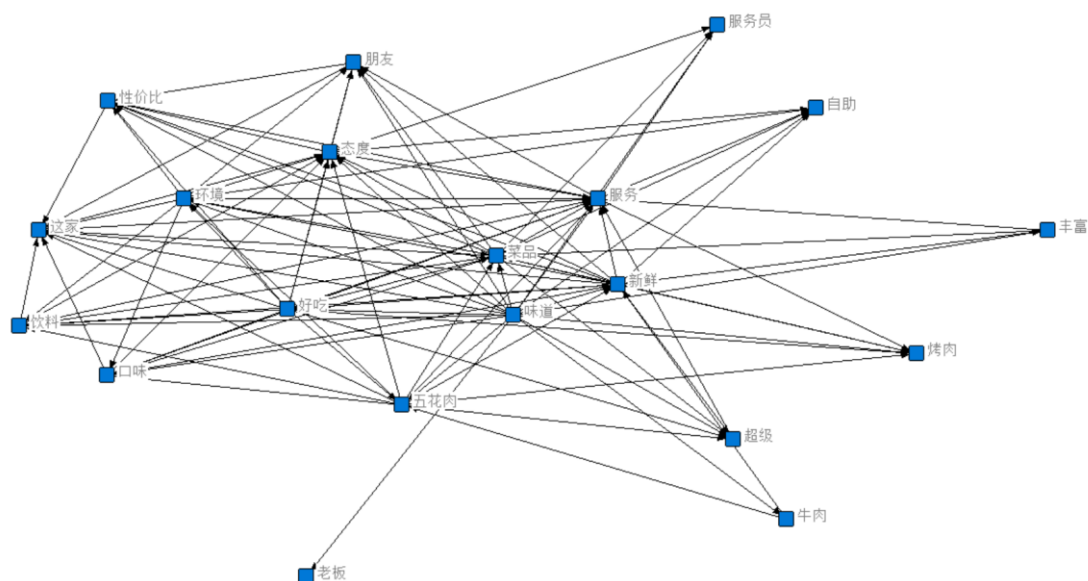


图 4-3 烧烤烤肉语义网络图

川湘菜高频词共现频数表和语义网络图分别如表 4-4 和图 4-4 所示,从表 4-4 和图 4-4 中可以看出,环境、服务、味道是关键节点,以服务节点为例,与之距离较近的节点是环境、味道、态度等,由此可以看出:川湘菜的环境、味道、态度,是影响消费者满意度的重要因素。

表 4-4 川湘菜高频词共现频数表

语义关系		频数	语义关系		频数
味道	好吃	27	服务	服务员	23
味道	服务	26	服务	分量	22
态度	服务	25	环境	口味	22
好吃	口味	25	味道	份量	22
好吃	服务	25	好吃	分量	22
味道	口味	24	菜品	态度	22
环境	服务	24	味道	分量	22
好吃	环境	24	份量	好吃	22
味道	态度	24	份量	环境	22
好吃	态度	24	份量	服务	22
味道	环境	24	味道	性价比	22
环境	态度	23	味道	服务员	22
味道	菜品	23	口味	态度	22
菜品	服务	23	好吃	服务员	22
口味	服务	23	环境	服务员	22

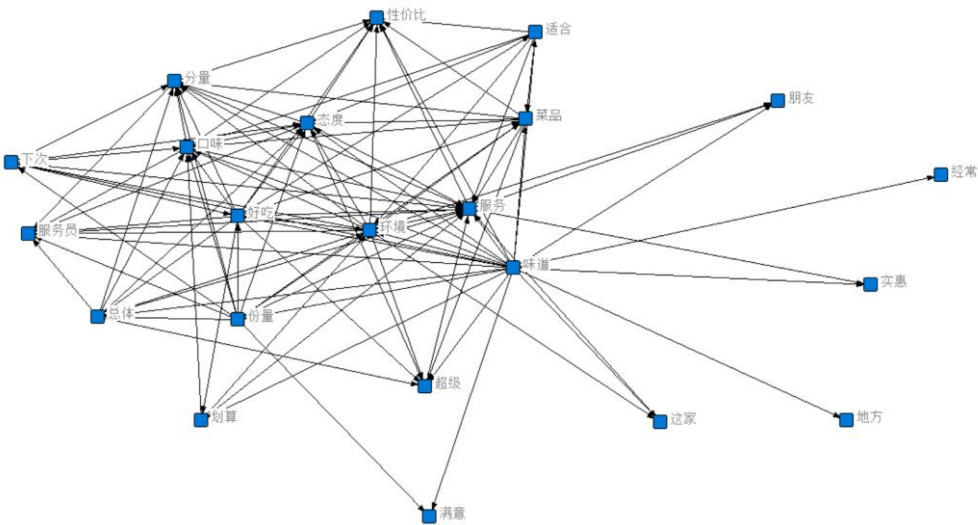


图 4-4 川湘菜语义网络图

汤锅高频词共现频数表和语义网络图分别如表 4-5 和图 4-5 所示，从表 4-5 和图 4-5 中可以看出，汤锅、养生、味道是关键节点，以汤锅节点为例，与之距离较近即共现频数较高的是养生、味道、玉米等关键词，由此可以看出：对于汤锅来说，养生、味道、玉米是影响消费者满意度的重要因素。

表 4-5 汤锅高频词共现频数表

语义关系		频数	语义关系		频数
好吃	玉米	12	汤锅	新鲜	11
味道	养生	11	汤锅	菜品	11
味道	好喝	11	汤锅	玉米	11
味道	好吃	11	养生	好喝	11
味道	新鲜	11	养生	好吃	11
味道	菜品	11	养生	新鲜	11
味道	口味	11	养生	菜品	11
味道	玉米	11	养生	玉米	11
味道	朋友	11	菜品	玉米	11
好吃	新鲜	11	好喝	好吃	11
好吃	菜品	11	好喝	新鲜	11
味道	汤锅	11	好喝	菜品	11
汤锅	养生	11	好喝	玉米	11
汤锅	好喝	11	新鲜	菜品	11
汤锅	好吃	11	新鲜	玉米	11

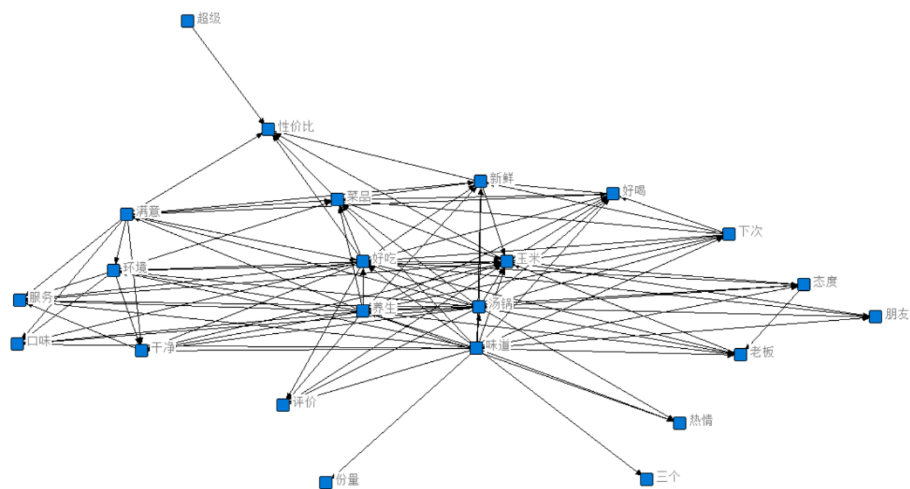


图 4-5 汤锅语义网络图

火锅高频词共现频数表和语义网络图分别如表 4-6 和图 4-6 所示，从表 4-6 和图 4-6 中可以看出，味道、环境、菜品是关键节点，以味道节点为例，与之距离较近即共现频数较高的是服务、环境、菜品等关键词，由此可以看出：消费者对于火锅的味道、服务、菜品、环境更为关注。

表 4-6 火锅高频词共现频数表

语义关系		频数	语义关系		频数
味道	口味	157	态度	服务	138
味道	服务	156	好吃	环境	138
新鲜	菜品	154	菜品	口味	138
味道	菜品	152	好吃	服务	137
环境	服务	151	味道	火锅	137
菜品	服务	149	环境	口味	137
味道	新鲜	149	好吃	菜品	137
味道	环境	148	好吃	口味	136
热情	服务	147	环境	服务员	135
环境	菜品	146	好吃	新鲜	134
口味	服务	145	新鲜	口味	134
味道	好吃	144	味道	服务员	134
新鲜	服务	144	火锅	服务	133
新鲜	环境	142	味道	锅底	133
服务	服务员	141	菜品	服务员	132

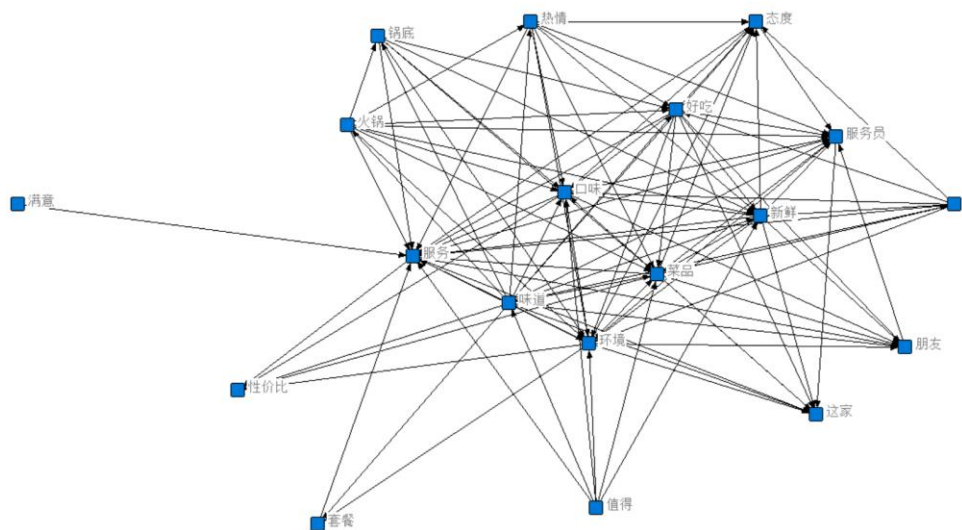


图 4-6 火锅语义网络图

除此之外，蛋糕甜品、东南亚菜、京菜鲁菜、聚餐宴请、日韩料理、甜点饮品、西餐、香锅烤鱼、小吃快餐、自助餐等高频词共现频数表及语义网络图如附件所示，其中：蛋糕甜品的环境、服务、美味是影响消费者满意度的重要因素；东南亚菜的菜品、味道、环境是影响消费者满意度的重要因素；京菜鲁菜的服务、菜品、态度是影响消费者满意度的重要因素；聚餐宴请的味道、热情、服务是影响消费者满意度的重要因素；日韩料理的火锅、环境、套餐是影响消费者满意度的重要因素；甜点饮品的好吃、水果、态度是影响消费者满意度的重要因素；西餐的好吃、服务、环境是影响消费者满意度的重要因素；香锅烤鱼的分量、味道、服务是影响消费者满意度的重要因素；小吃快餐的服务、味道、环境是影响消费者满意度的重要因素；自助餐的好吃、味道、环境是影响消费者满意度的重要因素。

4.5 结果分析——情感分析及关键词提取算法（以江南道为例）

4.5.1 各类词频与观点分析

首先对江南道所有的评论数据进行关键词抽取，绘制成词云图。



图 4-7 江南道词云图

江南道作为一家餐饮行业的商店，客户最关心的当然是其菜品，这是在评论中被提到最多的。除此之外，客户对菜品关心的方面也不尽相同，大部分人关心食物的味道，一些人关心实不实惠，一些人关心是否新鲜。除了对菜品的关心外，大部分人都关心在就餐过程中江南道提供的服务，客户很重视服务员的态度是否热情。在对这些方面的点评之外，有很多菜品的名字也出现在了词云图中，这是因为客户在评论时会选取一些推荐菜品，这给了我们后续分析推荐菜的想法。

为了进一步分析客户关心的内容，把江南道的评论分成积极情感的评论和消极情感的评论，分别对其进行关键词抽取与词云图的绘制。由于两方面数据前两个高频词都为菜品与味道，为了更好地展现其不同，去掉这两个词再绘图。

积极评论的词云图如下:



图 4-8 积极评论词云图

通过这张图，我们可以看出江南道能成为这 68 家店中最火爆的秘诀，首先作为餐饮店，它的事物口味大受好评，尤其是五花肉、鸡翅、肥牛这些经典菜式。江南道的服务态度也受到了大部分人的肯定。学生、室友、聚餐等词的出现，说明江南道的客户大部分都是在校学生。

消极评论的词云图如下:



图 4-9 消极评论词云图

在消极评论中，江南道的环境是最让人不满意的。座位、爆满、位置等词汇的出现可以看出江南道的生意很火爆，但却没有处理好排队等位等问题。服务不仅在积极评论中被大量提到，在消极评论中也一样。说明江南道的服务并不是百分之百做的很好，需要进一步改进。一个很严重的问题是很多消极评论都是针对肉类、海鲜这样的菜品，甚至出现了臭味这样的词汇。这需要江南道把控好肉类菜品的质量，因为菜品才是其立身之本。

4.5.2 推荐菜分析

在对数据进行探索性分析时发现很多评论中都会出现客户单独点评的菜品，并且有固定的格式。

表 4-7 评论格式展示

店铺名称	用户名	评论
江南道自助烤肉海鲜涮锅（柳浪湾店）	KqN8873 12316	「#套餐：江南道单人自助」「#五花肉」「#炸鸡翅」不错的烤肉，好评（。ò ∨ ó。）
江南道自助烤肉海鲜涮锅（柳浪湾店）	Tbi40840 4305	「#套餐：江南道单人自助」多吃水果哈哈哈
江南道自助烤肉海鲜涮锅（柳浪湾店）	姓 116	「#套餐：江南道单人自助」「#五花肉」「#扇贝」「#水果蔬菜」「#西瓜」都是随便吃 【环境】自助餐环境基本上都属于一般 最主要是看菜品新鲜不 这家店在大学附近 周边学生聚会的好地方 【服务】老板还是多热情 根据自己需求选室内或者室外

因此采用正则表达式匹配的方法对推荐菜品进行提取。在对全部评论、积极评论、消极评论中的出现次数最多的菜品进行分析后，发现排在前面的菜品相差不大，因此选取全部评论中出现次数最多的 5 种人气菜品进行好评与差评的比例可视化。

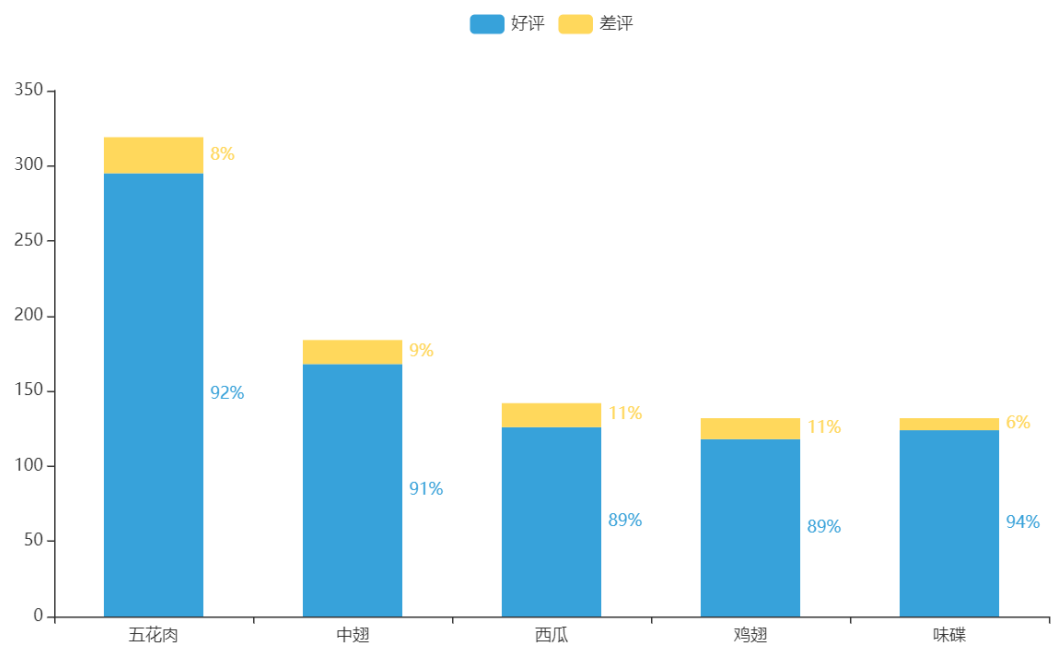


图 4-10 人气菜品好评差评比例可视化

从比例可以看出，五花肉当之无愧是江南道的招牌菜，不仅好评条数多，好评比例也大。在前五名菜品中仅有味碟的好评率比五花肉高。

4.5.3 情感分析

江南道的情感值变化如下：

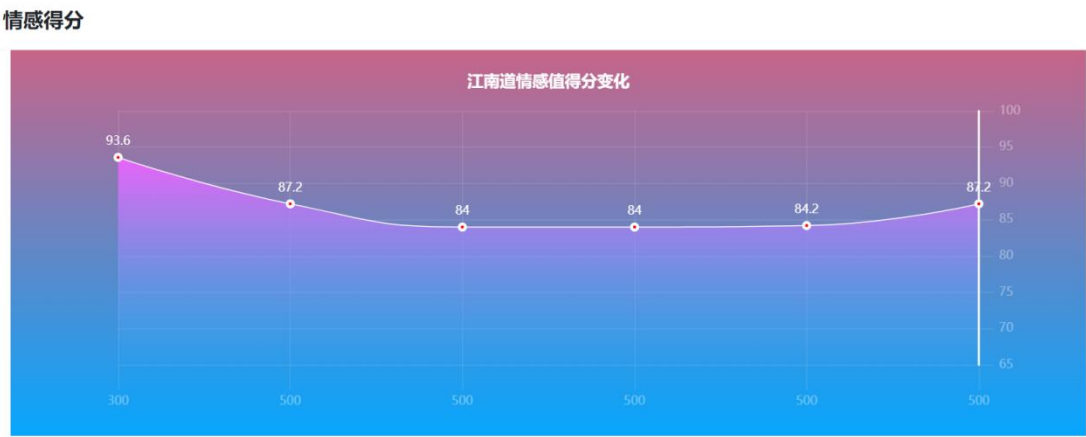


图 4-11 江南道情感值得分变化

近一段时间，江南道的情感值先下降，再上升，不太稳定，并且比最开始低了很多。

虽然爬取的数据没有标签，但我们可以用关键词匹配的方法对情感极性做一个简单的分析。我们把客户点评的方面分成了五个：味道、环境、服务、性价比、等位。采用相关关键词在评论中进行匹配，然后分析五个方面的情感值倾向。匹配规则如下：

味道：'好吃','味道','难吃','新鲜','不好吃','发苦','口味','酸'

环境：'环境','挤','卫生','吵','脏','乱','差'

服务：'服务','态度'

性价比：'性价比','便宜','贵','价格','物美价廉'

等位：'等位','排队','慢','等','爆满','预约'

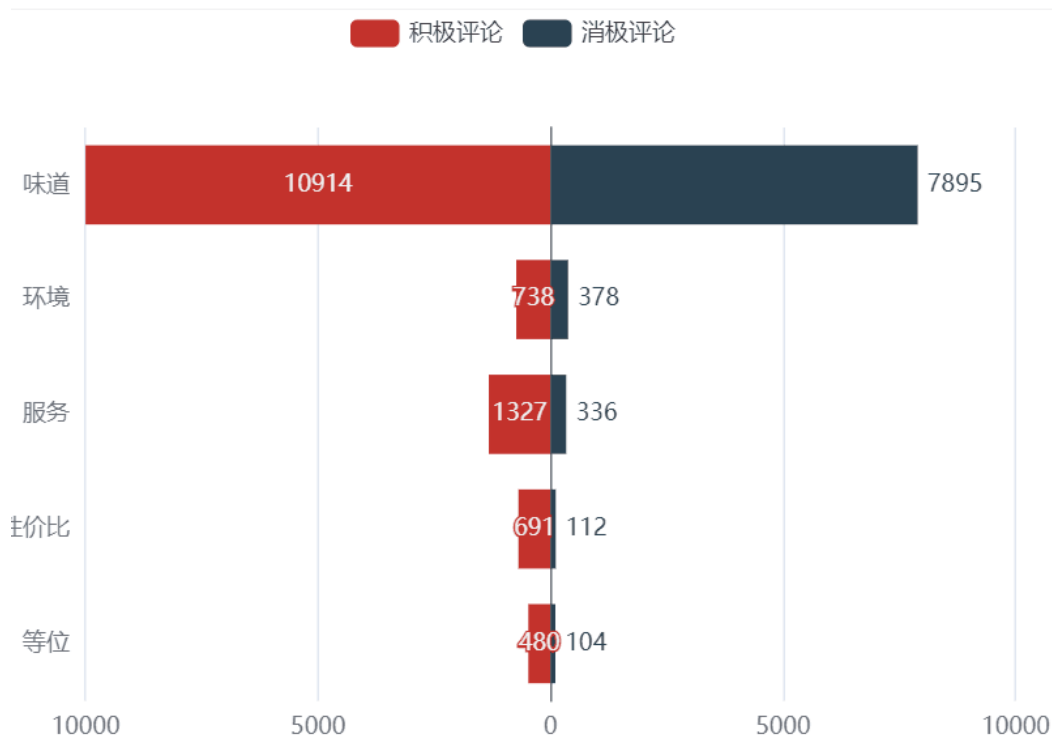


图 4-12 聚类——情感极性分析

味道的情感得分为 58，环境的情感得分为 66.2，服务的情感得分为 79.8，性价比的情感得分为 86，等位的情感得分为 82.2。

4.5.4 竞品店分析

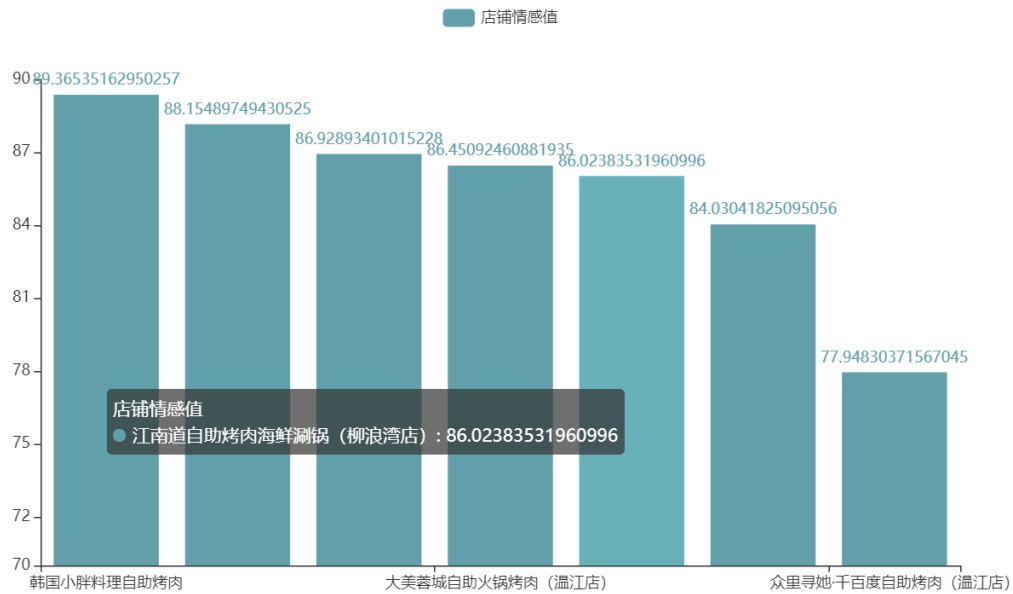


图 4-13 情感排行

在同类餐厅的情感值得分中，江南道排名倒数第三，不是一个很好的成绩。最受欢迎的烤肉店是韩国小胖料理自助烤肉。

4.5.5 结论与建议

对江南道的评论数据进行文本分析的结果如下：

（1）菜品方面：首先，江南道作为餐饮行业的店铺，把控好菜品的质量和口味是立身之本。当前好评多集中于对菜品味道的点评上，也解释了江南道能将生意做大的原因。以五花肉为代表的人气菜品也广受好评。但同时也要注意，差评的点评也多集中于菜品的味道上。虽然江南道生意火爆，好评很多，但好评比例在整个温江区内的餐饮店铺中，甚至仅仅是烤肉这一类餐饮店铺中都不占优势。除此之外，在差评中经常提到肉类产品的不新鲜问题。这是一个很严重的缺陷。

（2）服务方面：服务态度同样是一个广受大家注意的方面，无论是好评还是差评，服务都是被提到最多的方面之一。首先是服务员的服务态度，从差评中服务也出现了很多就可以得知，江南道的服务员并不是一直做的很好。有时人们即使与服务人员没有过多接触，也会顺手进行好评，但如果是差评，应该重点注意。第二，江南道作为一家生意火爆的餐厅，等位是不可避免的。等位、排队的问题经常出现在差评中，这是一个需要江南道去着力解决的问题。

（3）同类竞品：在温江区较大的 68 家店铺中，江南道的情感得分已经排在了倒数第十五位。在最相近的七家店中，排在倒数第三位。这是一个不太好的成绩。

总的来说，从情感值的变化也可以看出，江南道的好评率是有下降的。即使江南道作为温江区一家较大的餐饮店铺，拥有无数好评，但这也许仅仅是由地理位置的优势带来的。要想长久在这片竞争激烈的区域中立足，不能被大量好评迷惑，要注意到自己逐渐下滑的口碑与暴露出来的问题。

改进建议如下：

（1）严格把控好菜品的质量，尤其是肉类制品。做好经典菜，也要进一步探索客户对其他菜品的意见。

（2）优化排队等候机制。可以实行电话预约、小程序预约等功能。也要加强在客户排队等候过程中的服务，如提供小零食、根据等候时长提供优惠等。

（3）加强对服务员的培训，改进服务态度问题。

（4）抓住主要客户（学生）的特点，进一步探究其需求，并有针对性地做出改进。

五、可视化设计

5.1 网页设计

为了方便直观地进行数据展示，本团队自行开发了可视化网站，并以“江南道自助烤肉海鲜涮锅(柳浪湾店)”作为商家样例进行分析结果展现。网站开发基于 bootstrap4 响应式布局 WEB 项目框架，利用 CSS/HTML + JavaScript/jQuery 编写，大量数据可视化图表调用 ECharts 开源可视化图库和百度地图 API 进行展现。网站总体框架如下图所示：

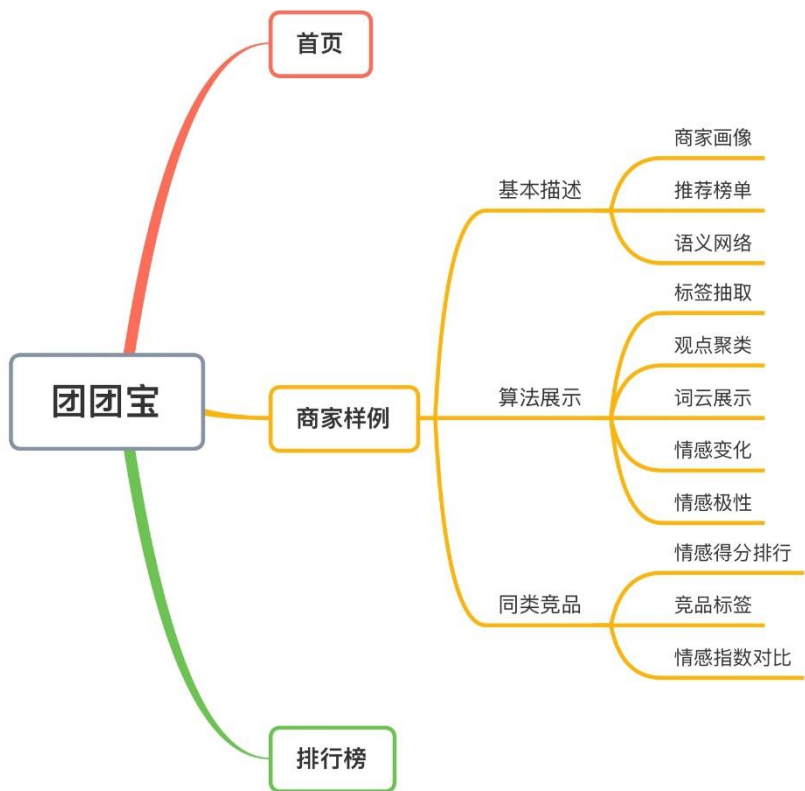


图 5-1 网页框架

5.1.1 网站首页

在网站首页，向用户进行产品总体功能的简要说明。本产品通过挖掘消费者评论中的核心观点和情感，帮助餐饮商家了解消费者，洞察市场，改善经营策略。用户点击“了解更多”，可以查看商家样例中的详细功能。

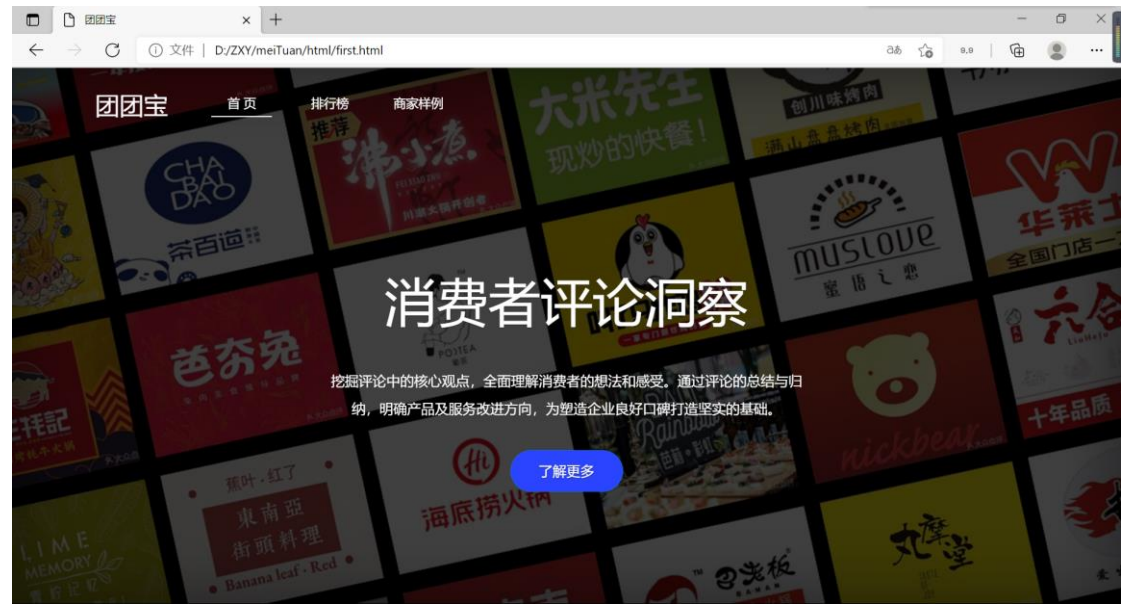


图 5-2 网站首页

5.1.2 基本描述

在商家样例中，主要包括基本描述、算法展示、同类竞品三个功能模块，下面针对不同模块进行详细介绍。

在页面上方首先展示商家的名称（江南道自助烤肉海鲜涮锅(柳浪湾店)）、所属餐饮品类（自助餐、烤肉）、该商家在大众点评上的评论数量、好评率、差评率、人均消费价格等基本数据。在基本描述模块，我们主要反映商家自身的特质信息，通过商家特质评分、推荐菜榜单和评论语义网络进行商家画像。



图 5-3 商家基本描述可视化

1) 商家画像

商家画像对店铺进行口味、服务、环境、品种和星级方面的评分，用于反映消费者对商家的总体印象和评价。



图 5-4 商家画像可视化

2) 推荐榜单

推荐榜单通过提取评论中消费者对菜品的推荐信息统计得出。可以看出，江南道最受欢迎的菜品是五花肉，其他肉类，蔬果海鲜也得到了较多好评。

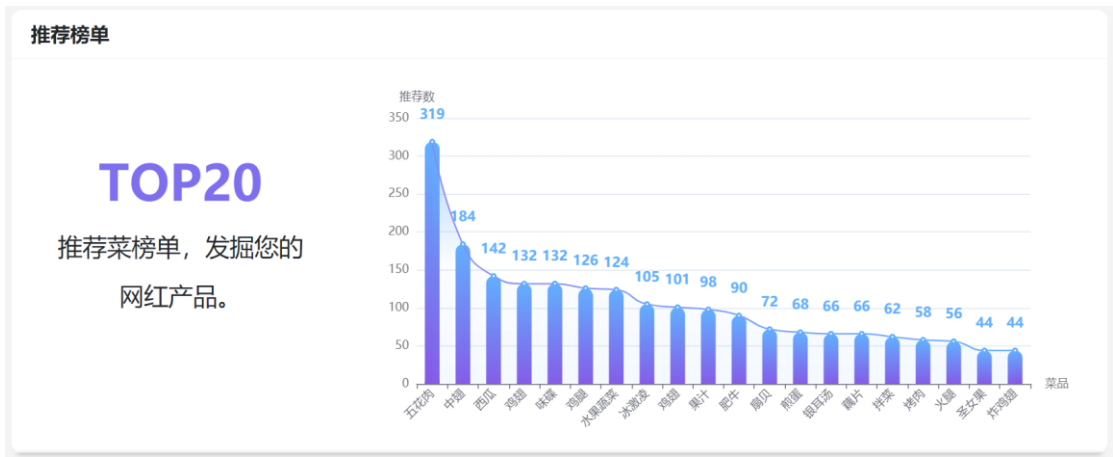


图 5-5 推荐榜单可视化

3) 语义网络

为了更深入的了解评论信息，我们制作了评论语义网络，发掘评论中的实体节点及其关系，直观地反映出评估对象与评论之间的联系，挖掘用户对商家服务的关注点。

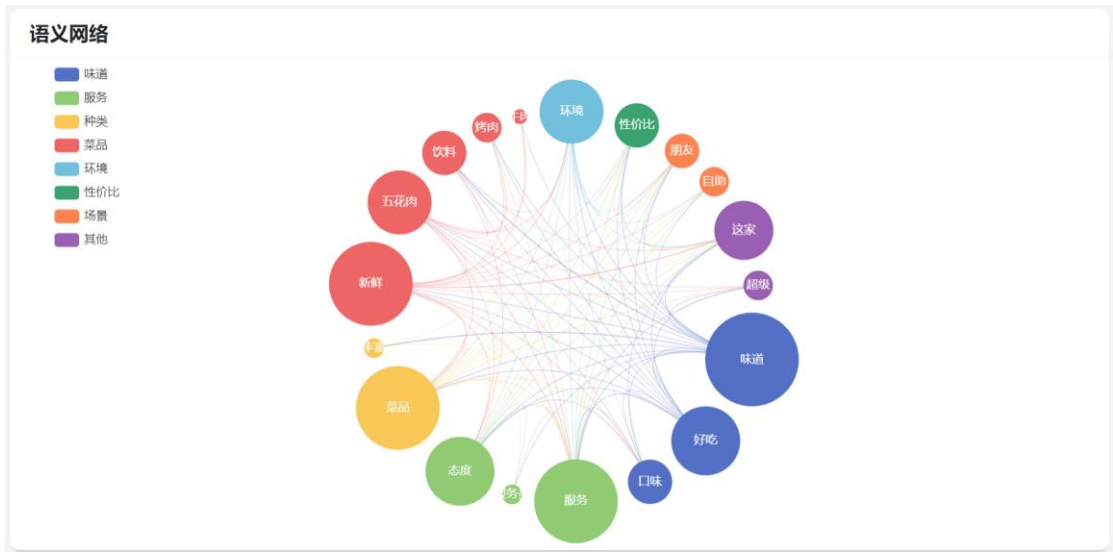


图 5-6 语义网络可视化

5.1.3 算法展示

1) 标签抽取



图 5-7 标签抽取可视化

这一部分展示评论观点标签抽取结果。同一条评论中可能包含积极和消极两方面的情感，也包括口味、服务、环境等多方面的评价，因此可能会有不同的标签展示。

2) 观点聚类

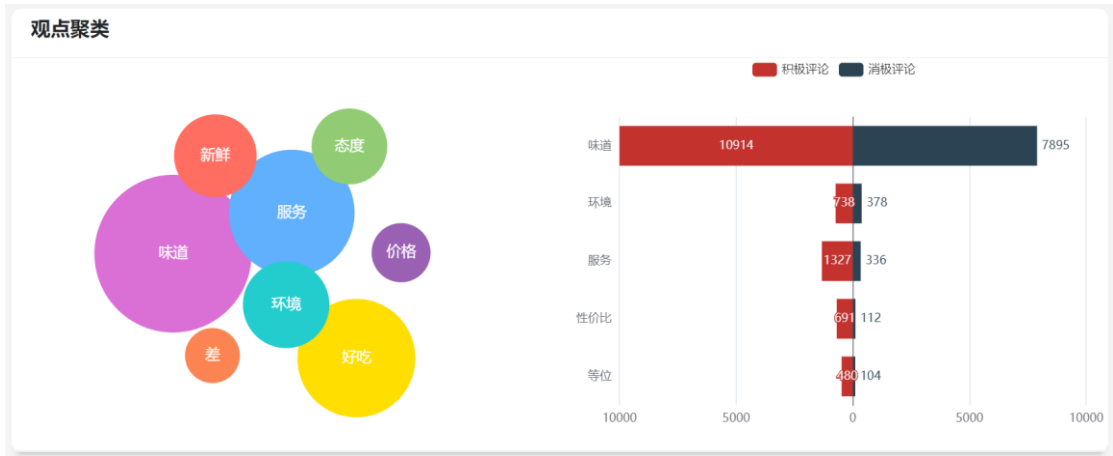


图 5-8 观点聚类可视化

抽取观点标签后，依据标签进行观点聚类，可以看出，关于“味道”的讨论和评价数量最多，其次是关于“服务”、环境、菜品等。可见江南道的消费者更加重视烤肉的口味是否符合心意，以及商家服务是否及时，服务态度是否良好。

3) 词云展示

利用 Echarts 的 WordCloud 模块将评论信息词云化，鼠标悬浮可查看每个词语在评论中出现的频次。



图 5-9 词云展示

4) 情感变化



图 5-10 情感变化可视化

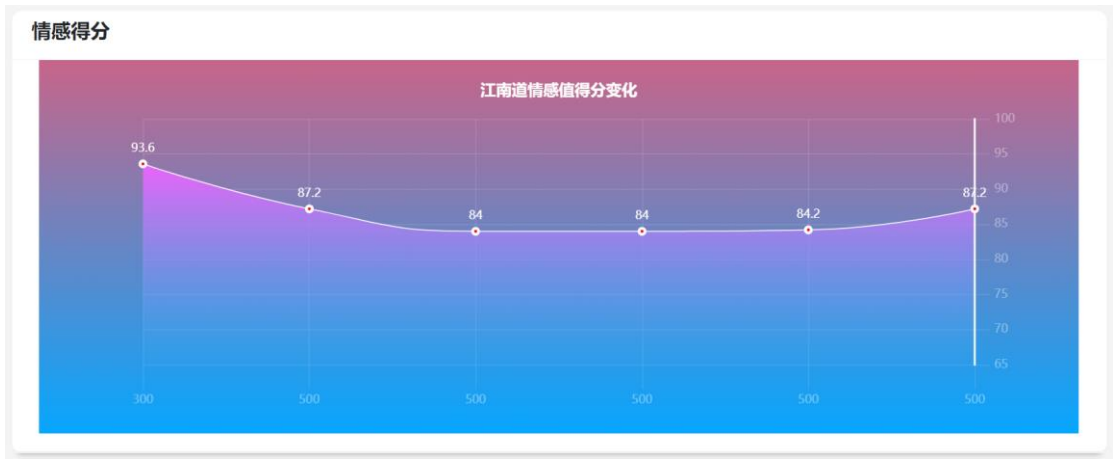


图 5-11 情感得分可视化

我们分析了评论的情感变化，包括：积极评论和消极评论分别随时间和数据量变化的改变；针对不同菜品，好评和差评的改变；江南道总体情感得分的变化。可以看出江南道的情感得分略有降低，但近期有回升趋势。

5) 情感极性



图 5-12 情感极性可视化

针对不同情感类型的评论进行深入洞察。右侧玫瑰图显示江南道积极评论和消极评论的占比，点击图中区域可以查看对应的评论示例集。

5.1.4 同类竞品

除商家自身的评论分析之外，我们还向用户提供市场上同类餐饮商家的竞争情报，便于商家进一步了解竞争对手，发现自己的优劣势，寻求自身改进。

1) 情感得分排行榜

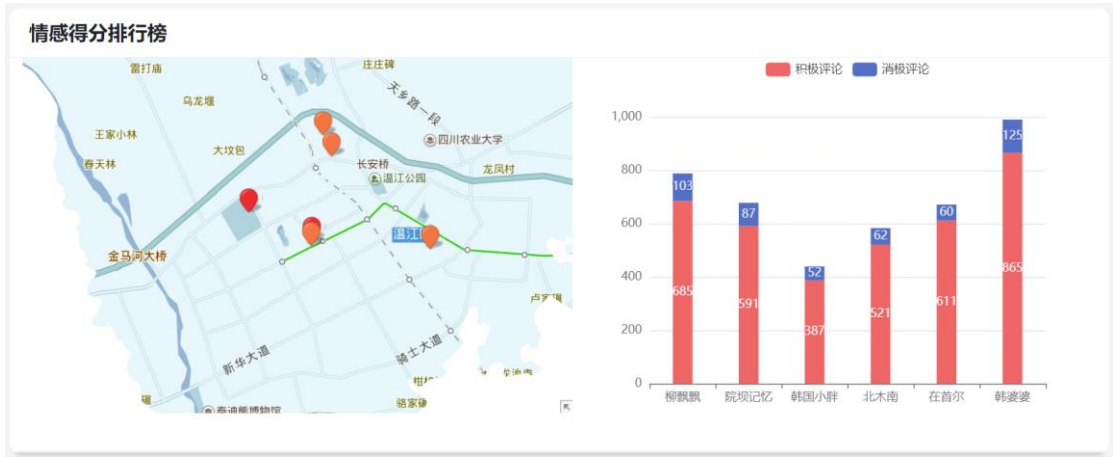


图 5-13 情感得分排行榜

我们选取了大众点评上品类为“烧烤烤肉”的餐饮商家，包括：柳飘飘亚洲烤肉集合店（温江店）、院坝记忆牦牛串串·龙虾·烧烤（鱼鳧路总店）、韩国小胖料理自助烤肉、北木南烤肉和酒（西南财大店）、在首尔韩国料理·自助烤肉（温江店）、韩婆婆小院烧烤 6 家店铺，分别统计各自的积极评论与消极评论数量，并调用百

度地图 API 接口展示各店铺的地理位置。分析结果可得，7 家店铺的情感倾向均为积极，地图中红色标注表示该店铺的情感得分在 88-100 分之间，橙色标注表示该店铺的情感得分在 80-88 之间。

2) 竞品标签



图 5-14 同类竞品标签展示

竞品标签展示各竞争商家的评论观点标签，分为积极向和消极向分别进行展示。用户可点击卡片下方的 tab 切换查看。还可以很直观地进行横向对比，比如在关于“味道”的积极评论中，大芙蓉城位居榜首；在“服务”方面，北木南烤肉相对而言店面较小，客流量较少，卖家服务相对周到，因此关于“服务”和“态度”的差评较少，而相比之下，大芙蓉城和众里寻她千百度烤肉在人员服务方面则需要改善。该功能可以帮助商家与众多竞争者及时对比，了解自身优势和不足，从而调整竞争策略。

3) 情感指数对比

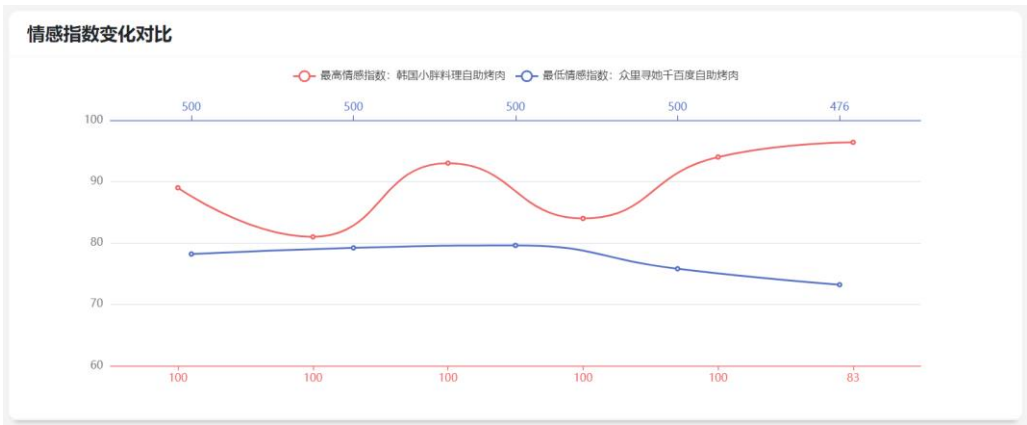


图 5-15 情感指数变化对比可视化

在这一部分，对比展示“烧烤烤肉”品类的最高情感得分和最低情感得分商家的分数变化情况。可以看出该品类下的商家情感得分良好，都在 70 分以上，众里寻她千百度烤肉的情感得分正在下降，韩国小胖料理自助烤肉的情感得分前期波动较大，目前已有明显提升。

5.1.5 排行榜

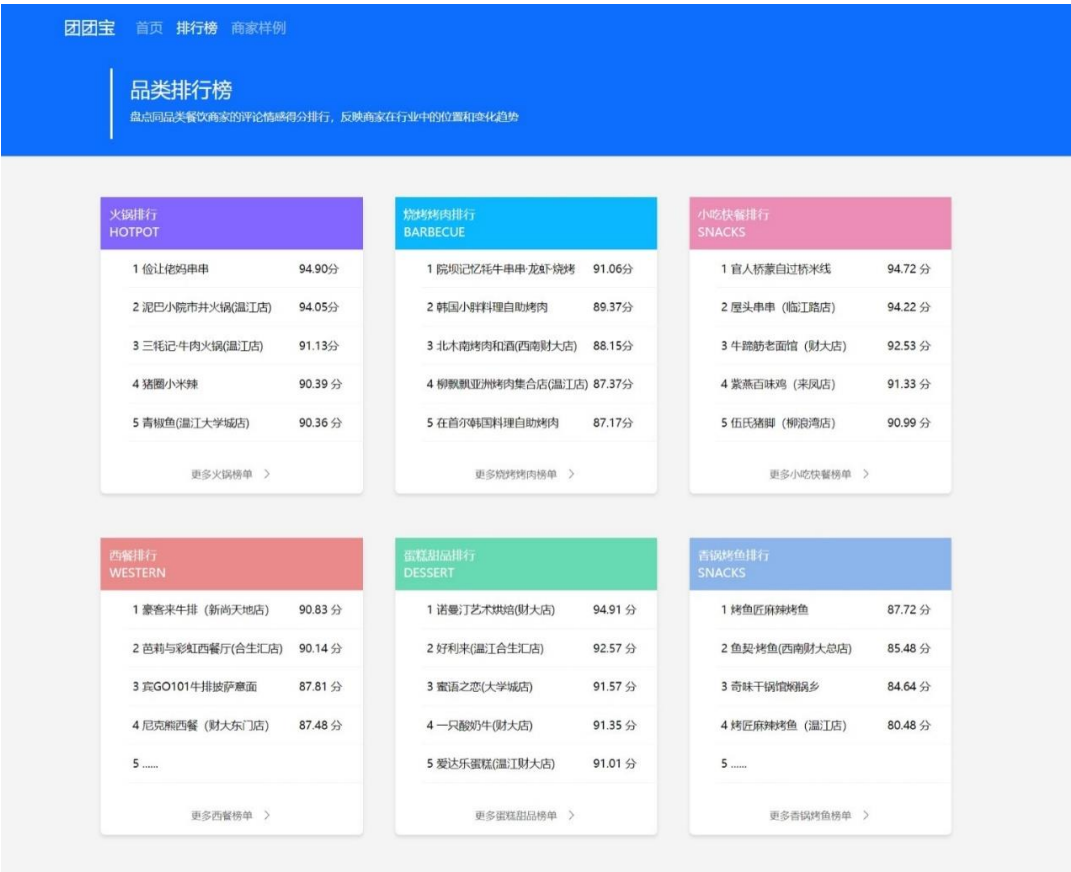


图 5-16 排行榜

我们对火锅、烧烤烤肉、小吃快餐、西餐、蛋糕甜品、香锅烤鱼 6 个品类分别统计，针对不同品类的商家进行评论情感得分排行，商家可以查看自己的情感得分在同类竞品中的次序，了解自己的竞争实力。

5.2 年度报告

为了丰富可视化及完善网站功能实现，我们基于 H5，为商家生成具有特色的

年度报告，基本内容包括：评论情况、年度热评、年度热评菜品、年度死忠粉、年度差评关键词、提升指南，以“江南道自助烤肉海鲜涮锅(柳浪湾店)年度报告”作为样例进行分析，结果展现如下：

1) 首页



图 5-17 年度报告首页

2) 评论情况

评论情况页面向商家直观地展示了过去一年消费者在线评论的总数量，及根据情感分析所得正向评论数量与负向评论数量。如江南道在 2020 年共有 5538 位顾客进行评论，其中正向评论 4764 条，负向评论 774 条：

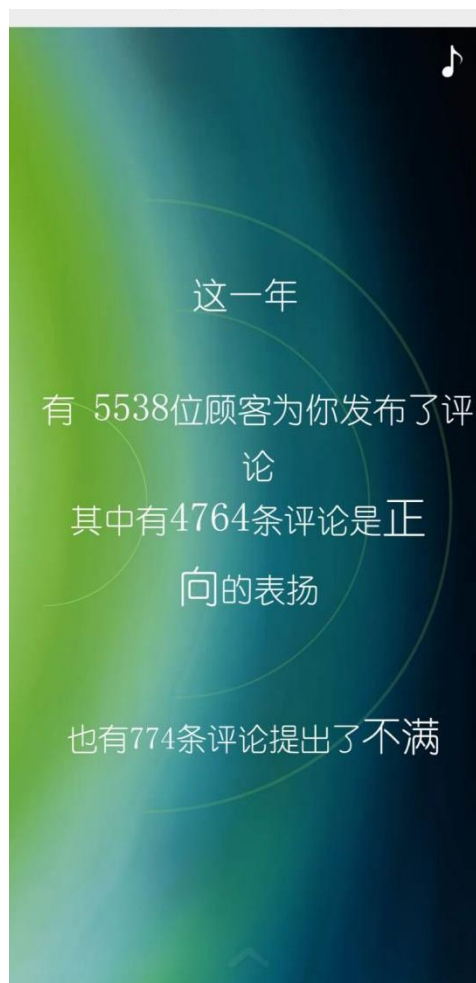


图 5-18 评论情况

3) 年度热评

年度热评主要通过词云图的形式向商家展示过去一年，消费者评论中的关键词，可以直观地向商家展示消费者对本店服务的关注点：



图 5-19 年度热评

4) 年度热评菜品

年度热评菜品主要是向商家展示一年内在消费者评论中出现的高频菜品，在一定程度上可以展现消费者关注的菜品。如在江南道 2020 年的评论中，“五花肉”共出现 319 次，“中翅” 184 次，“鸡翅” 164 次，“西瓜” 142 次，说明消费者对这些菜品关注更高：



图 5-20 年度热评菜品

5) 年度死忠粉

该页面主要展示在过去一年中，在美团商家店铺评论最多的用户及其评论集：

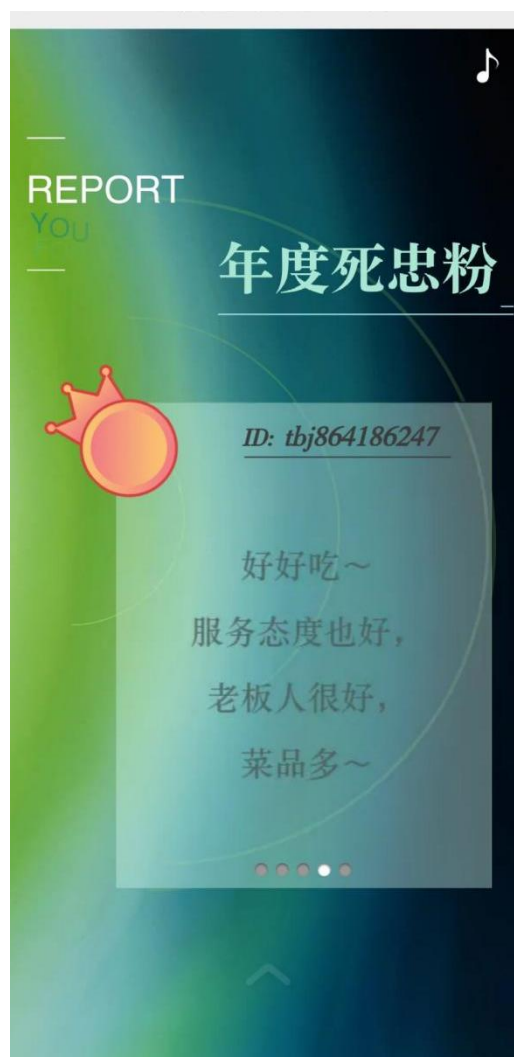


图 5-21 年度死忠粉

6) 年度差评关键词

该页面主要展示负面评论中出现的关键词，在一定程度上可以反映商家在服务中待提升的指标。如江南道在 2020 年差评关键词是菜品、味道、环境、服务、态度，商家可以根据这些指标，在新的一年里重点改善。



图 5-22 年度差评关键词

7) 提升指南

提升指南主要针对商家在过去一年的评论情况，为商家生成的改善建议。

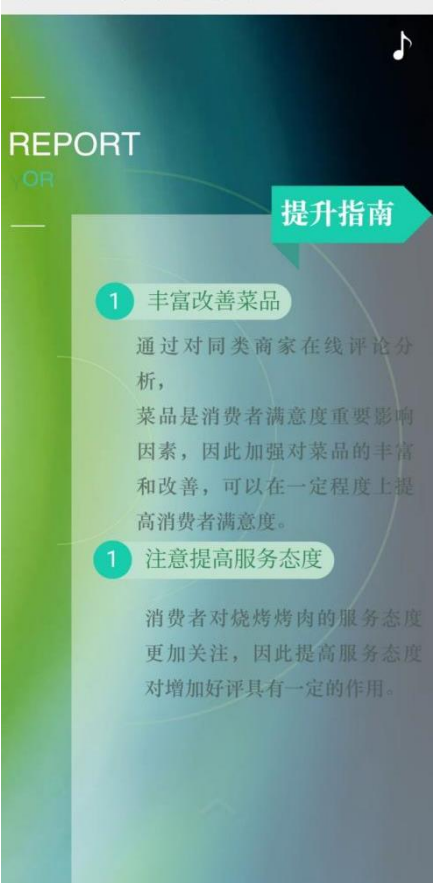


图 5-23 提升指南

8) 尾页



图 5-24 尾页

六、总结

随着互联网技术的发展，数据呈几何爆炸式增长。如何在海量数据中提取关键词，获得有用信息，如何将数据信息转化为经济效益，是一个值得思考并具研究潜力的方向。对评论数据进行情感分析是一个发展迅速，应用面较广的领域。美团平台作为一个被商家和消费者广泛使用的应用，每天接收大量订单，产生海量评论信息。本文的研究思路是利用问卷数据和结构方程模型探究什么在影响消费者的用餐选择，再根据餐厅的用户评论进行文本挖掘分析，从评论中探究顾客满意度因素，

并将两种不同研究方法所得出的结果进行比对，从而为商家提供消费者需求分析和改进建议。

6.1 主要研究工作及创新点

针对上述问题现将本文的主要研究工作及创新点总结如下：

（1）使用八爪鱼爬虫软件爬取了美团平台的评论数据，并对爬取的数据进行按照制定的数据清洗规则进行数据预处理。并将对收集到的商家数据进行分类处理，分别为川湘菜、蛋糕甜品、东南亚菜、火锅、京菜鲁菜、聚餐宴请、日韩料理、烧烤烤肉、汤锅、甜点饮品、西餐、香锅烤鱼、小吃快餐、自助餐共十四个类别。

（2）将统计因素分析与自然语言处理相结合，互相印证。目前消费者选择影响因素分析主要采取结构方程模型等统计方法，而将统计方法与情感分析相结合进行研究的很少。本文首先基于收集到的问卷数据建立结构方程模型，对消费者选择餐饮商家影响因素进行初步探索。然后，基于美团消费者在线评论，进行情感分析、观点挖掘、关键词抽取等自然语言处理方法，针对烧烤烤肉、火锅等十四类餐饮探究用户满意度影响因素，二者相互补充，互为印证，结构更为完整。

（3）针对现有研究中大多采用情感分析模型或语义网络技术来挖掘评论文本的潜在信息的不足之处，本文将情感分析模型与语义网络分析相结合，运用到基于美团在线评论的消费者满意度影响因素研究中。本文通过对两种技术原理的研究，将两种技术相结合，运用语义网络分析，从整体评论中挖掘消费者对各类餐饮的关注点，具有一定的数据可视化功能，可以从整体上呈现出消费者评论状况，进而全面地获取消费者满意度影响因素。同时，通过情感分析和观点挖掘，对消费者评论进行正负情感分类，分别获取正负面评价消费者关注点，进而获取影响消费者满意度的主要因素。

（4）搭建面向商家的评论情感数据可视化平台，将研究更加直观地向商家进行回馈反映，提升研究效果。大多数基于在线评论的消费者满意度影响因素研究，仅停留在理论分析，不能将研究结果直接反映给商家，而本文在研究后，依托特色网站，辅以微信公众号平台，将研究数据可视化，为商家提供了具有项目特色的服

务平台。

6.2 工作成果

通过以上的的工作本文取得了相应的成果：

基于结构方程模型的分析，本文发现消费者选择餐厅时更加注重服务质量、菜品口味、平台评分和菜品价格。这四个因素对消费者选择行为有显著影响，餐厅可以结合自身经营情况，着重在上述四个方面进行改进和提高。

同时，本文通过自然语言处理技术针对商家的用户评论数据进行分析，得出的结果为商家提供更精准的顾客需求和满意度指导，帮助商家在激烈的市场竞争中抢夺先机。

本文根据 14 类商家分别做了语义网络分析（具体结果见附录），得出了 14 不同类别的商家的消费者的关注点，发现不同类别下消费者的关注点不同，比如消费者对于烧烤烤肉的菜品、味道、服务更为关注，但对川湘菜而言影响消费者满意度的重要因素则是环境、味道、态度。

此外，本文还运用了情感分析和观点挖掘技术，针对商家进行更深入的分析，分析内容及结果为：①各类词频及观点分析；②推荐菜分析；③情感分析；④竞品店分析⑤改进建议。

结合以上三种技术，本文以江南道（柳浪湾店）为例进行了具体分析。分析结果简要概括如下：

作为烧烤烤肉店，消费者对于它的菜品、味道、服务更为关注。

词频及观点分析结果：；积极：学生占顾客群体的一大部分；消极词频分析:环境和服务都有待改进，同时针对肉类、海鲜这样的菜品的消极评论出现频率较高。

推荐菜分析：江南道最受欢迎的菜是五花肉。

情感分析结果为：味道的情感得分为 58，环境的情感得分为 66.2，服务的情感得分为 79.8，性价比的情感得分为 86，等位的情感得分为 82.2。从情感变化结果看，江南道的情感值呈下降再上升的趋势，波动较大，且从数值上看并且比几年

前低了很多。

在同类餐厅的情感值得分中，江南道排名倒数第三，同类中最受欢迎的烤肉店为韩国小胖料理自助烤肉。

本文在通过情感分析、关键词提取、语义网络等技术对江南道用户评论进行挖掘后，提出如下建议：1. 严格把控好菜品的质量，尤其是肉类制品。做好经典菜，也要进一步探索客户对其他菜品的意见；2. 优化排队等候机制。可以实行电话预约、小程序预约等功能。也要加强在客户排队等候过程中的服务，如提供小零食、根据等候时长提供优惠等。3. 加强对服务员的培训，改进服务态度问题。

具体结果展现在可视化网站和微信公众号平台（团团宝儿）中。

6.3 改进

以上总结了本文所做的工作，但由于时间限制，针对海量评论数据的情感分析工作仍有很多研究工作亟待开展，根据所做的研究与了解，现对之后的工作进行展望，未来研究和改进的方面包括以下几个方面：

（1）数据收集方面，本文的数据利用八爪鱼的采集模板进行数据爬取，所能爬取到的数据有限，比如不能爬取用户评论的发布时间，使分析受限，在之后的研究中，可以考虑通过 `python` 进行爬虫，解决美团反爬问题，得到更多维度的数据。

（2）数据预处理本文中数据清洗只是使用了一些简单的方法，比如，空数据的剔除，重复数据的去重。未来还可以用更严格的数据剔除规则，去除垃圾评论和灌水评论。本文目前采用的只是简单的数据清洗，后续可以考虑更加复杂的数据预处理手段。高质量的数据集是算法精确度的保障。

（3）语义网络分析中本文利用全部的评论进行分析，而未能分别基于积极和消极的评论文本进行分析，探究人们对美团商家正向的满意点和负向的建议。在后续的研究中，可以细化评论集合，分别对积极和消极的评论进行语义网络分析。

（4）算法的改进：本文当前情感分析方法基于情感词典的统计学分析，此外，情感分析还可以使用效果更好的语义分析方法，通过使用预训练模型如 `BERT` 得到效果更好的初始表示，然后探索不同深度学习模型结构对文本数据的情感倾向

进行识别。除此之外，还可以使用模型进行更多层次的情感分析。本文使用的是文档级别的情感分析，后续还可对单句话进行分析，或是进行方面级别的细粒度情感分析，深度挖掘用户在不同观点上的情感倾向。

参考文献

- [1]曹伟,丁雪华.社区教育服务质量评价模型及指标体系研究[J].中国远程教育,2014(11):69-75+92+96.
- [2]贾薇,叶怀远,林迪芬.物业服务质量对顾客价值及顾客忠诚的影响:居住档次的调节作用[J].财经论丛,2017(03):92-102.
- [3]吴卿毅,黄斐.在线评论对消费者行为意愿的影响——基于数值型在线评论的效价[J].技术经济与管理研究,2019(11):60-65.
- [4]刘凤军,段坤,董泽瑞,孟陆.味觉感知与隐喻对于服务补救的影响研究[J/OL].南开管理理论:1-25[2021-06-23]. <http://kns.cnki.net/kcms/detail/12.1288.f.20210415.1843.020.html>.
- [5]胡少华,吴鹏.品牌服装店铺顾客动线的影响因素及权重研究[J].上海纺织科技,2019,47(10):18-20+56
- [6]高芸菲. 基于在线评论的用户痛点挖掘[D].暨南大学,2019.
- [7]张行健,焦政,钟瑞童.基于文本挖掘技术的社会网络舆情分析框架刍议——以某知名艺人偷税漏税事件为例[J].西部学刊,2019(10):134-138.
- [8]王泽云,刘子豪.数据挖掘在旅游电子商务中的应用与研究[J].中国管理信息化,2019(20):138-139.
- [9]封丽. 面向微博短文本的情感分析和特征抽取[J]. 农业图书情报学刊, 2018(9):56-60.
- [10]Xu G , Yu Z , Yao H , et al. Chinese Text Sentiment Analysis Based on Extended SentimentDictionary[J]. IEEE Access, 2019, 7:43749-43762.
- [11] Dos Santos C, Gatti M. Deep convolutional neural networks for sentiment analysis of shorttexts[C]//Proceedings of COLING 2014, the 25th International Conference on ComputationalLinguistics: Technical Papers. 2014: 69-78.
- [12] McDonald R, Hannan K, Neylon T, et al. Structured models for fine-to-coarse sentimentanalysis[C]//Proceedings of the 45th annual meeting of the association of

computational linguistics. 2007: 432-439.

[13] Lin Q , Zhu Y , Zhang S , et al. Lexical based automated teaching evaluation via students' short reviews[J]. Computer Applications in Engineering Education, 2018.

[14]岳子静,章成志,周清清.利用在线评论挖掘用户饮食偏好——以北京地区为例[J].图书馆论坛,2017,37(03):108-115.

[15]马松岳,许鑫.基于评论情感分析的用户在线评价研究——以豆瓣网电影为例[J].图书情报工作,2016,60(10):95-102.

[16]蒋超. 基于用户评论数据的线上大家电产品用户满意度影响因素研究[D]. 哈尔滨工程大学,2020.

[17]王银. 超短微博文本关联语义表示模型的构建与优化[D].安徽理工大学,2018.

[18]李丹丹. 面向在线评论的情感信息分类与挖掘[D].浙江理工大学,2018.

[19]郭美娟.基于语义网络分析的媒体报道中食品议题探究[J].新闻研究导刊,2017,8(20):139.

[20]夏琳康. 武汉市旅游服务质量评价研究[D].中南民族大学,2016.

[21]刘俊杰,张丽娟.飞行疲劳事件语义网络分析[J].中国安全科学学报,2016,26(01):34-39.

[22]Mihalcea, P. (2004). TextRank: Bringing Order into Text. In Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (pp. 404–411). Association for Computational Linguistics.