

分类号 C8/375
U D C 0005666

密级 公开
编号 10741

兰州财经大学

LANZHOU UNIVERSITY OF FINANCE AND ECONOMICS

硕士学位论文

论文题目 基于评论文本挖掘的电商助农水果
产品运营策略改进研究

研究生姓名: 余佳豪

指导教师姓名、职称: 牛胜强 教授

学科、专业名称: 应用经济学 统计学

研究方向: 社会经济统计

提交日期: 2024年6月5日

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。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果。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。

学位论文作者签名： 余佳豪 签字日期： 2024年6月3日

导师签名： 牛胜强 签字日期： 2024年6月3日

关于论文使用授权的说明

本人完全了解学校关于保留、使用学位论文的各项规定， 同意（选择“同意”/“不同意”）以下事项：

1. 学校有权保留本论文的复印件和磁盘，允许论文被查阅和借阅，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文；

2. 学校有权将本人的学位论文提交至清华大学“中国学术期刊（光盘版）电子杂志社”用于出版和编入 CNKI《中国知识资源总库》或其他同类数据库，传播本学位论文的全部或部分内容。

学位论文作者签名： 余佳豪 签字日期： 2024年6月3日

导师签名： 牛胜强 签字日期： 2024年6月3日

Research on Improving the Operation Strategy of E-commerce Assisted Agricultural Fruit Products Based on Comment Text Mining

Candidate: Yu Jiahao

Supervisor: Niu Shengqiang

摘 要

农村电子商务作为“三农”发展的重要抓手，既可以促进农产品上行，提升农产品流通效率，又可以帮助农村地区优化产业结构，推动农业升级、农村发展、农民增收，是我国全面推进乡村振兴战略的重要实践选择。农村电子商务持续助力“三农”发展，主要表现为电商助农模式的运行。在电商助农模式下，通过电子商务平台，将农产品与消费者直接连接起来，有效解决了传统农产品销售的中间环节问题，同时为农产品提供了更多元化的销售渠道，使得农产品能够更广泛地进入市场，得到消费者的认可和购买。但是目前电商助农的发展过程中出现了一些关键问题，如服务质量低、消费者需求未得到满足等导致农产品缺乏产品竞争力，使得其销售滞后。而产品竞争力是企业或商家发展的核心，率先满足消费者的需求是产品竞争力最直接的价值体现，与电商助农产业的可持续发展息息相关。因此，需要从消费者需求出发，对电商助农产品运营策略进行改进。

本文以京东平台电商助农水果产品为研究对象，设计了电商助农水果产品运营策略改进方案，为电商助农水果产品商家提供产品运营策略改进决策支持。首先，通过八爪鱼采集器在京东平台爬取了多家助农水果产品商家和一般水果产品商家消费者在线评论数据，对其进行评论去重、文本去噪等文本数据预处理操作；其次，构建水果产品领域属性词典，确定产品评价维度即产品属性，基于 PaddleNLP 搭建 SKEP 模型进行属性级情感分析，从在线评论数据中提取出相应的属性词、观点词及情感倾向，再利用基于情感词典的方法计算水果产品各项属性的情感得分；之后，利用属性级情感得分结果计算助农水果产品商家和一般水果产品商家水果产品的属性绩效表现、属性重要性值，得出电商助农水果产品整体的市场表现；确定电商助农水果产品目标研究商家，通过竞争者识别模型识别其竞争对手，再通过 IPCA 和 DIPCA 模型分析得出目标商家水果产品的优劣属性；最后，针对目标商家第一优先级改进劣势属性，通过文本挖掘技术 TF-IDF 对此属性的原始负面评论提取消费者不满意因素的热词，利用 K-Means 聚类算法对消费者不满意因素的热词进行聚类分析，最终以热图形式将运营策略改进方案形成依据进行可视化展示，提出具体的改进方案。

研究得出目标商家的“果面”属性为第一优先级改进劣势属性，应该采取的产品改进方案为：一方面，重点改进“果面”属性，即严控水果采摘保鲜环节、严控筛检环节、

改善包装以及选择可靠的物流公司；另一方面，还需要根据属性改进收益合理分配资源投入其他属性改进。通过对全文工作和结果的总结，得出了以下研究结论：第一，电商助农水果产品的消费者满意度较低且市场表现相对较差，在电商水果市场上的竞争力较弱；第二，消费者的偏好和需求可能会随着时间的推移而变化，电商助农产品商家应及时关注消费者的偏好和需求，并适时调整运营策略和资源配置；第三，消费者满意度越高的产品属性其竞争力越强，且竞争对手对产品属性的改进可能会削弱自身产品的竞争力，因此电商助农产品商家还应不断完善自己的产品，避免陷入“不进则退”的陷阱。

根据研究结论，对电商助农不同主体提出了不同的建议：对消费者来说，一是积极对商品进行真实的评价，二是京东平台是水果等生鲜产品和支持消费助农的不错选择，三是购买水果等生鲜产品要了解清楚退换赔付情况。对电商助农水果商家来说，一是严格控制产品质量，二是完善自身服务，三是合理分配资源投入，四是紧跟消费需求、把握消费趋势。对于平台来说，一是优化物流体系建设，二是完善包装管理，三是健全售后服务体系。

关键词：电商助农 评论文本挖掘 产品运营策略 属性级情感分析 产品竞争力

Abstract

As an important grasp of the development of agriculture, rural areas and farmers, rural e-commerce can not only promote the upward of agricultural products and improve the circulation efficiency of agricultural products, but also help rural areas optimize the industrial structure, promote agricultural upgrading, rural development and farmers' income increase. It is an important practical choice for China to comprehensively promote the rural revitalization strategy. Rural e-commerce continues to help the development of agriculture, rural areas and farmers, mainly reflected in the operation of the e-commerce helping agricultural mode. Under the e-commerce helping agricultural mode, through the e-commerce platform, agricultural products are directly connected with consumers, which effectively solves the problem of the intermediate link of traditional agricultural product sales, and provides more diversified sales channels for agricultural products, so that agricultural products can enter the market more widely and be recognized and purchased by consumers. However, there are some key problems in the development process of e-commerce helping agriculture, such as low service quality and unmet consumer demand, which lead to the lack of product competitiveness of agricultural products, making their sales lag behind. The product competitiveness is the core of the development of enterprises or businesses. To meet the needs of consumers first is the most direct embodiment of the value of product competitiveness, which is closely related to the sustainable development of e-commerce helping agriculture industry.

Therefore, it is necessary to improve the operation strategy of e-commerce helping agricultural products from the perspective of consumer demand.

Taking the e-commerce fruit products on JD platform as the research object, this paper designs the improvement plan of e-commerce fruit product operation strategy, and provides decision support for the improvement of product operation strategy for e-commerce fruit product merchants. Firstly, the online review data of several e-commerce fruit product merchants and general fruit product merchants are crawled on Jingdong platform through octopus collector, and the text data preprocessing operations such as deduplication and text denoising are performed on the review. Secondly, the attribute dictionary of fruit product field is constructed to determine the product evaluation dimension, namely product attributes. The SKEP model is built based on PaddleNLP for Aspect(Attribute)-based Sentiment Analysis, and the corresponding attribute words, view words and sentiment orientation are extracted from the online review data. Then the attribute-based sentiment score of fruit product attributes is calculated using the attribute-based sentiment score method. Then, the attribute performance and attribute importance value of fruit products of e-commerce fruit product merchants and general fruit product merchants are calculated using the attribute-level sentiment score results, and the overall market performance of e-commerce fruit products is obtained. The target merchants of e-commerce fruit products are determined, and their competitors are identified through the competitor identification model. Then the advantages and disadvantages of fruit products of target merchants are analyzed through

IPCA and DIPCA models. Finally, aiming at the first priority of the target merchants to improve the disadvantage attributes, the hot words of consumer dissatisfaction factors are extracted from the original negative comments of this attribute by the text mining technology TF-IDF, and the hot words of consumer dissatisfaction factors are clustered by the K-Means clustering algorithm. Finally, the basis of the formation of the operation strategy improvement plan is visualized in the form of heat map, and the specific improvement plan is proposed.

The study concludes that the "fruit surface" attribute of the target merchants is the first priority of the disadvantage attributes, and the product improvement plan should be adopted: on the one hand, the focus is to improve the "fruit surface" attribute, that is, strictly control the fruit picking and fresh-keeping link, strictly control the screening link, improve the packaging and select reliable logistics companies; on the other hand, it is also necessary to reasonably allocate resources to other attribute improvement according to the income of attribute improvement. Through the summary of the work and results of the full text, the following research conclusions are drawn: first, the consumer satisfaction of e-commerce agricultural fruit products is low and the market performance is relatively poor, and the competitiveness of e-commerce fruit market is weak; second, consumer preferences and needs may change over time, and e-commerce agricultural product merchants should pay timely attention to consumer preferences and needs, and timely adjust the operation strategy and resource allocation; Third, the improvement of competitors' product attributes

may weaken the competitiveness of their own products, so e-commerce agricultural products merchants should also constantly improve their products to avoid falling into the trap of "not to advance is to retreat".

According to the conclusions of the study, different suggestions have been put forward for different subjects of e-commerce to help agriculture: for consumers, first, you can actively make a true evaluation of the goods, second, the Jingdong platform is a good choice for fresh products such as fruit and support for consumption to help agriculture, and third, the purchase of fruits and other fresh products should be clear about the return of the compensation situation. For e-commerce to help farmers fruit merchants, one is to strictly control product quality, the second is to improve their own services, the third is to reasonably allocate resources to invest, the fourth is to keep up with consumer demand, grasp the consumer trend. For the platform, one is to optimize the construction of the logistics system, the second is to improve packaging management, and the third is to improve the after-sales service system.

Keywords: E-commerce Helps Farmers; Comment Text Mining; Product Operation Strategy; Attribute-based Sentiment Analysis; Product Competitiveness

目 录

1 绪论	1
1.1 研究背景与研究意义	1
1.1.1 研究背景.....	1
1.1.2 研究意义.....	2
1.2 国内外研究现状	3
1.2.1 电商助农研究现状.....	3
1.2.2 情感分析研究现状.....	4
1.2.3 基于在线评论的产品竞争力分析研究现状	5
1.2.4 文献述评.....	6
1.3 研究内容与技术路线	7
1.3.1 研究内容.....	7
1.3.2 技术路线.....	9
1.4 研究方法	10
1.5 研究创新点	11
2 相关概念界定及理论基础	12
2.1 相关概念界定	12
2.1.1 文本挖掘.....	12
2.1.2 电商助农.....	13
2.1.3 产品竞争力.....	14
2.2 相关理论基础	14
2.2.1 消费情感理论.....	14
2.2.2 属性级情感分析理论.....	15
2.2.3 重要度—绩效分析理论.....	18
3 电商助农水果产品运营策略改进方案设计	19
3.1 电商助农水果产品运营策略改进方案设计原则	19
3.2 电商助农水果产品运营策略改进方案设计	20

3.3 产品竞争力模型	22
3.3.1 产品属性绩效表现模型.....	22
3.3.2 产品属性重要性模型.....	22
3.4 产品竞争力分析模型	23
3.4.1 竞争者识别模型.....	23
3.4.2 IPCA 模型.....	24
3.4.3 DIPCA 模型.....	25
4 电商助农水果产品评论数据采集与预处理	27
4.1 数据来源与数据采集	27
4.2 评论文本数据预处理	29
4.2.1 评论去重.....	30
4.2.2 文本去噪.....	30
4.2.3 短文本删除.....	30
4.2.4 中文分词.....	31
4.2.5 去除停用词.....	32
5 电商助农水果产品运营策略改进方案形成	33
5.1 电商助农水果产品属性识别及领域属性词典构建	33
5.2 电商助农水果产品属性级情感分析	33
5.3 电商助农水果产品竞争力计算	36
5.3.1 产品属性绩效表现.....	36
5.3.2 产品属性重要性.....	36
5.4 电商助农水果产品竞争力分析	37
5.4.1 竞争者识别.....	37
5.4.2 静态竞争力分析.....	38
5.4.3 动态竞争力分析.....	41
5.5 电商助农水果产品运营策略改进方案形成	42
5.5.1 消费者不满意因素 TF-IDF 热词提取	42
5.5.2 基于 K-M e a n s 聚类算法的不满意因素分析	43
5.5.3 运营策略改进方案形成.....	45

6 总结、建议与展望	47
6.1 总结	47
6.2 建议	48
6.2.1 对消费者的建议.....	48
6.2.2 对电商助农水果产品商家的建议.....	49
6.2.3 对平台的建议.....	50
6.3 展望	51
参考文献	53
攻读硕士学位期间承担的科研任务及主要成果	57
致谢	58

1 绪论

1.1 研究背景与研究意义

1.1.1 研究背景

2018年1月，中共中央、国务院发表了《关于实施乡村振兴战略的意见》，强调要在党中央领导下，坚持解决好“三农”问题，持续加大加强乡村惠农政策力度，稳步推进镇域电商助农，全面实现乡村振兴战略。当前，我国正值“十四五”时期，随着数字化生活消费方式的变革，我国农村大市场将重塑，电子商务将成为数字乡村最大的推动力和最坚实的发展基础。在此背景下，农村电子商务生态要素将加速整合，电子商务作用于农业生产和农村消费的巨大潜能将加速释放，未来其必将成为推动乡村振兴取得新进展、农业农村现代化迈出新步伐的巨大引擎。

电子商务与农村产业相融合助力乡村振兴和农业农村现代化发展主要表现之一为电商助农模式。电商助农是结合了电子商务、物联网、大数据等技术，为传统农业提供了全新的转型升级之路，有助于提高农业生产效率、降低农产品销售成本以及增强农产品市场竞争力，同时也有助于提升农民收入水平、改善农村地区经济状况。在电商助农模式下，农户得以参与电子商务生态系统，并从在线市场交易中获益，实现“工业品下乡，农产品进城”流通新业态。近年来，在电子商务助力乡村振兴背景下，我国农产品生产能力有了质的飞跃，农业供给侧结构性改革跨上新台阶，农民收入持续稳步增长，农村生态文明建设显著提高，农业农村现代化发展取得了重大成就。据商务部统计显示，在电商助农效应的辐射下，2022年我国农产品网络零售额突破了5000亿元，在助农增收、促进就业以及增效经营等方面发挥巨大效力，电子商务的助农价值愈加凸显。但是，我国现阶段的电商助农模式仍存在一系列发展问题。对此，中共中央提出，我国要继续创新发展电子商务以此助力农产品销售、农民增收，达到乡村全面振兴的目标。

电商助农的主要途径在于平台销售助农，电商平台为助农产品搭建了一个良好的流通网络，为其提供了大量搜索曝光、流量推荐的机会，使农业生产与市场之间形成一种高效的衔接。然而，电商助农的发展过程中产品运营方面出现了一些关键问题，如产品质量难以保障、消费者需求难以满足、可市场化程度低等导致助农产品缺乏产品竞争力，使得其销售滞后。产品竞争力是企业或商家发展的核心，率先满足消费者的需求是产品

竞争力最直接的价值体现，与电商助农产业的可持续发展息息相关。由此，需要从消费者需求出发，从产品层面探讨如何提高电商助农产品的竞争力，为商家提供切实可行的具体对策来改进他们的产品。目前各大电商平台都设置了用户评价模块，在线评论中消费者对产品使用体验的表述和情感倾向的表现，都是基于消费者角度来反映他们的需求和满意度，相对于传统的消费者调查访谈、问卷等渠道获取消费者需求更有效便捷。因此，本文设计了产品运营策略改进方案，基于京东平台电商助农产品的在线评论文本，结合属性级情感分析计算产品的属性绩效表现及重要性，利用竞争力分析方法得出目标产品的市场表现及优劣属性，适应消费者对助农产品动态的感知和表现，针对产品的劣势属性特征，为电商助农产品运营提供相应改进策略决策支持。

1.1.2 研究意义

（1）理论意义

本文所做工作对电商助农和竞争力分析的学术研究都有一定程度的贡献：

第一，虽然目前已有研究认为电商助农产品的市场竞争力需要提高，以获得电商助农模式的持续成功，但很少有研究真正对电商助农产品的竞争力进行分析，本文一定程度上弥补了这一空白，也回应了电商助农模式长效运行的需求。

第二，在电商交易中，商品评价对于大众消费者而言，是最为关键的影响因子。消费者在购物后分享的产品使用体验，会传染到其他消费者，影响他人的消费决策，从而形成以口碑传播为核心的网络购物消费者决策的循环闭口模式。因此，本文通过挖掘电商平台在线评论文本，在了解消费者真实的情感偏好基础上，对电商助农产品的竞争力进行分析，为商家改进产品提供以消费者需求为导向的策略引导，丰富了电商助农领域的研究。

（2）现实意义

第一，本文是基于属性级情感分析进行情感计算的，在识别消费者关注的产品属性特征及其情感倾向的基础上，结合竞争力分析方法，能针对性地对电商助农产品的劣势属性进行提取、分析，挖掘其竞争优势与不足，找出当前的改进焦点，为商家改进产品、优化运营、建立品牌口碑提供切实可行的决策支持。

第二，由于产品竞争力是动态变化的，因此需要商家及时跟进消费者情感偏好以获得长效竞争力，而长效竞争力的基础来源于产品的持续改进。本文所提出的运营策略改进方案能及时响应内外部因素的变化，以满足消费者变化的偏好和行为为基础，针对竞

争产品优化带来的市场挑战性，对自身产品的劣势属性进行及时改进。

第三，由于本文的核心方法（如属性级情感分析、IPCA、DIPCA 及消费者不满意因素分析）独立于特定产品，因此本文所提出的运营策略改进方案也可以扩展到其他电商助农产品领域。

1.2 国内外研究现状

1.2.1 电商助农研究现状

乡村振兴是实现中华民族伟大复兴的一项重大任务，而电子商务作为新兴业态，既可以促进农产品“走出去”、拓宽农民增收致富渠道，又可以帮助农村地区优化产业结构，从而实现乡村振兴赋能助力。近年来，国家高度重视农产品电商的发展，电子商务成为促进农产品上行和工业品下行的重要流通渠道（宋瑛等，2021）。电子商务的技术支持能拓展农产品市场连接网络，实现农产品流通的规模化和组织化，减少流通的中介环节，降低交易成本（Lin 等，2020）。王瑞峰（2023）指出，电子商务不仅改变了农民传统观念、提升了生活品质，在农民增收、增能提质方面发挥了重要作用，更是成为了强化农村主体地位的重要引擎。目前，电子商务已经成为了提高农业效率和支持农村繁荣的重要推动力。在电子商务赋能效应下，很多农村地区经济方式的转变和助农增收方面都取得了显著成效（王翠翠等，2022）。

在国家积极提倡和政策大力支持下，各大电商平台纷纷开展电商助农活动，逐渐呈现助农特色，不断涌现电商助农帮扶的新模式。2019年，淘宝启动“村播计划”，培育农民主播，深入农产品原产地，助力农村地区打开新市场，截至2020年7月，该计划带动了农产品销售80亿元（陈光，2022）。2020年6月京东的“京心助农”项目全面升级，通过整合供应链、物流、人才、流量、直播等全平台和全渠道资源，在3年内带动项目覆盖的农民收入增长50%以上。拼多多率先打造“政企合作，直播助农”系列活动，探索“市县长当主播，农户多卖货”的助农电商新模式（郭红东等，2020）。目前直播带货成为了电商助农的新模式，在解决农产品销售难方面发挥了重要作用。拼多多、淘宝、京东以及抖音等电商平台利用直播引流等方式，借助网红、流量效应，助力破解滞销农产品的销售困境。

鉴于电子商务助农的优势和效益，许多学者致力于探索如何建立一个更完善、更健

全的电商助农运行机制。电商助农本质上就是通过线上网络平台推广并销售与“三农”相关的产品，帮助农民增收致富。因此，孙文婷等（2022）建议应充分考虑电商平台和消费者的助农偏好，以此为基础定位生产与定价策略。胡珺祎等（2022）指出小农户的生产经营具有盲目性、分散性和信息滞后性，如果随波逐流加入平台电商浪潮，极易陷入低价竞争怪圈，在市场上仍旧会处于弱势地位，面临“边缘化”风险，因此电子商务与小农户的衔接是一个需要探索的关键问题。王瑞峰（2022）研究得出要构建电商助农增收的长效机制，就必须加大政府扶持力度，有效挖掘整合乡村资源，重视农民的创新赋能工作，让农民真正具备从事高质量产业的能力。对于以上研究，都是针对电商助农效应、模式以及运行机制进行的，但目前为止，具体的电商助农产品如何改进仍有待研究，这是电商助农模式长效运行的根本所在，需要更多的学术研究来构建评估和提高电商助农产品竞争力的方法体系。

1.2.2 情感分析研究现状

国外学者 Nasukawa 最早提出了情感分析方法（Agarwal, 2015），此后随着情感分析研究的逐渐成熟，其开始广泛应用于各类领域。情感分析，即意见挖掘，旨在对带有情感色彩的主观性文本进行分析、处理、归纳和推理，从中提取出信息表达者所表达情感状态的过程（Pang, 2002）。国外的情感分析研究起步早，且在英文文本评论数据分析方面有了较成熟的算法体系，研究的英文文本主要是基于 Twitter 数据、亚马逊数据等（史振杰, 2020）。国内的研究起步相对较晚，主要是对中文评论类数据进行分析，如新闻、微博、微信、电商平台、短视频平台等。对于文本情感分析的应用，也主要集中在网络舆情监控、商品评论分析、个性化推荐等领域。

国外对于情感分析的研究范围广泛，主要分为两类，一是基于情感词典的文本情感分析，二是基于机器学习的文本情感分析。Riloff 和 Shepherd（1997）构建了相应的特定语义词典，建立了一个基于语料数据的语义分析系统，可以通过此系统快速匹配类别相关的单词。Wilson 等（2005）在大规模语料的基础上验证了连词对文本正负面的情感倾向的影响，结果发现每个连词被认为是独立的情况下准确率较高。Abbasi 等（2008）通过细粒度语义差异，对网络评论进行了语义情感分类。使用机器学习进行文本情感分析，主要方法有支持向量机（SVM）、K 最近邻（KNN）和最大熵（ME）等。Pang（2002）最早将机器学习的方法应用于电影评论中，比较了 Naive Bayes 等三种机器学习算法对电影评论的分类效果。Whitelaw 等（2005）提出文本中抽取和分析评价词组，利用半自

动方法建立词组，然后将标准词袋和评价词组作为支持向量机的特征输入进行情感分类模型的训练。Socher 等（2011）提出一个距离监督学习算法来实现 Twitter 的自动情感分类。Denil 等人（2014）提出了一种多层次 CNN 模型，该模型在经典分类 CNN 模型中加入多个卷积层和池化层，增强了对文本局部特征的识别能力，在文本特征提取效果上得到了更好的优化。

随着情感分析应用价值的逐渐凸显，国内也逐渐开展了对情感分析的全面研究。李寿山（2013）等利用英文种子词典，通过机器翻译系统构建了中文情感词典。但这只是最基础的情感词典，情感词覆盖率低，后来研究学者对情感词典进一步进行了完善。阳爱民（2013）利用情感种子词计算了基础情感词的倾向值，借助搜索引擎返回的共现数构建了情感词典；王志涛（2015）通过 40 万余条微博数据构建了新词词典，以表情符号作为附近信息提供辅助，对已有情感资源进行了拓展；张克亮（2015）等融合了情感词典资源和概念层次网络语境框架的优势，将文本情感分析分为特征词、句子和句群判定阶段。后来学者们逐渐意识到传统情感词典的局限性，转而采用基于机器学习的方法对文本进行情感分析。唐晓波等（2013）为实现文本情感强度阈值的可视化，提出了一种基于旋进原则和 AdaBoost 集成技术的回归 SVM 情感分类模型。韩开旭（2014）等根据潜在语义分析方法（LSA）改进了概率潜在语义分析（PLSA）的参数初始方法。之后学者们发现深度学习方法在对文本逐步分析、特征抽取且自动学习优化模型输出方法更具优势，开始研究弱监督的深度学习方法提高文本分类的准确性。朱少杰（2014）引入了基于深度学习的半监督 RAE 方法进行文本情感分析。梁军等（2015）对 LSTM 进行了扩展，以应用于基于树结构的递归神经网络，用于捕获文本的深层次信息。谢铁等人（2017）运用深度递归神经网络算法来获悉句子的语义信息，并引入中文“情感训练树图资料库”作为训练数据，以识别词语的情感信息。

1.2.3 基于在线评论的产品竞争力分析研究现状

产品竞争力分析通常是指对目标产品与竞争产品的各项属性表现进行比较分析（Liu 等，2020），来帮助目标企业或商家提高自身产品的竞争力。近年来，随着网络的普及和电子商务的蓬勃发展，越来越多的人养成网上购物的习惯，每天都有大量的用户在网上发布产品使用体验和意见的评论信息。这些评论文本为消费者和商家都提供了重要的指导，消费者可以根据已有评论调整决策，商家也可以通过分析消费者的在线评论帮助了解自身产品的优劣特征。因此，如何基于在线评论分析产品竞争力已成为了一个

受业界高度关注的问题。

基于在线评论的产品竞争力分析方法认为，产品在某一特定属性上的竞争力是通过积极或消极的情感来体现的（Bi 等，2018），因此，竞争力可以通过计算消费者对产品的不同属性的情感值来衡量。例如，Wang 等（2014）通过半监督机器学习方法提取到产品特征维度，在综合考虑比较和非比较意见的基础上计算产品各特征维度情感得分，由此得到产品自身缺陷和比较缺陷。Feng 等（2018）基于 LDA 主题模型对产品在线评论文本进行主体提取，通过正负面主题之间的比较，确定了目标产品的竞争优势。Liu 等（2019）利用特定的产品特征维度，基于情感词典测量社交媒体在线评论数据中的消费者情感，之后通过与竞争对手的比较来挖掘目标产品的竞争优势和劣势。

然而，消费者对产品的看法是在动态变化的，企业或商家面临的竞争地位也可能发生变化（Hemphill 等，2004）。因此，竞争力的动态性驱使管理者要能够捕捉动态并有效应对（Cracolici 等，2009）。不过，现有的竞争力分析方法通常是从静态视角来衡量给定时间点的产品特征，这与现实需求不符。因此，需要开发一种能够进行动态分析的竞争力方法，以帮助商家了解产品属性特征的动态变化，从而制定匹配策略。此外，虽然在各种研究中使用的情感分析揭示了产品属性特征的表现水平，但不能揭示其背后的潜在和具体原因（Kim 等，2017），即现有的竞争力分析方法不能帮助商家理解消费者满意或不满意的原因，也导致了无法产生有效和具体的产品改进策略。因此，在情感分析的基础上，还需要探索更多的有效的文本挖掘技术。

1.2.4 文献述评

通过以上的文献梳理可以得到，学者们分析了电子商务的助农效应，对于电商助农模式的长效运行机制和电商助农发展过程中存在的问题进行了有益探索。另外，当前国内外学者们在利用文本挖掘技术基于在线评论进行文本情感分析方面，已经取得了诸多成果，应用于产品竞争力分析领域的研究也逐渐深入。但通过总结文献资料，本文也发现了当前研究存在的一些不足，有待进一步探究：

（1）当前对于电商助农领域的研究，主要在于宏观、中观层面，即对电商助农的政策体系、影响效应及运行机制的研究，而较少涉及到电商助农产品这一微观层面。

（2）当前的细粒度情感分析能揭示产品属性特征的表现水平，但不能进一步帮助商家掌握消费者对产品具体属性情感表现的原因。

（3）现有的竞争力方法通常忽略了竞争力的动态性，不能帮助商家及时跟进产品

属性特征的动态变化，制定相应的匹配策略。

根据以上的文献研究总结，本文吸取前人研究成果，结合文本挖掘技术、属性级情感分析方法和改进的竞争力分析方法对京东电商助农水果产品进行研究，深入挖掘消费者不满意因素的原因，为商家提供具体的助农水果产品运营策略改进方案，以促进电商助农水果产品商家提升消费者满意度和市场竞争力，助力电商助农模式的长效运行。

1.3 研究内容与技术路线

1.3.1 研究内容

本文是基于在线评论文本挖掘，对京东平台助农水果产品进行属性级情感分析、竞争力分析，以明确助农水果产品相对于竞争产品的比较优劣属性。针对劣势属性，提取消费者对此属性特征的不满意因素，了解消费者需求和偏好，将结果可视化，最终形成产品运营策略改进方案。

本文研究内容从以下六个方面展开：

第一章 绪论。绪论部分首先系统论述了本文的研究背景、研究意义，国内外文献综述部分主要从电商助农、情感分析及基于在线评论的竞争力分析这三个方面进行梳理归纳和述评，之后对本文的研究内容与技术路线、研究方法进行阐述，最后提出本文可能的创新点。

第二章 相关概念界定及理论基础。本章对于电商助农、评论文本挖掘、产品竞争力相关概念进行界定，并概述了消费情感理论、属性级情感分析理论及重要度-绩效分析理论，为后续研究提供理论支撑。

第三章 电商助农水果产品运营策略改进方案设计。本章设计了电商助农水果产品的运营策略改进方案，构建了运营策略改进方案形成的核心分析模型——产品竞争力模型和产品竞争力分析模型，目的是为了明确目标商家水果产品的优劣势属性。产品竞争力模型分为产品属性绩效表现模型、产品属性重要性模型。产品属性绩效表现是根据属性级情感分析结果得出，产品属性重要性是利用多元线性回归来计算。产品竞争力分析模型包括了竞争者识别模型、IPCA 和 DIPCA 模型。其中竞争者识别模型是首先需要为产品定义一个竞争力指数 ($Comp_i$)，然后计算每个产品与目标产品之间的不相似指数 ($Dissim_i$)，之后定义一个竞争对手指数 ($Competitor-index$) 识别目标商家产品的竞争

对手产品。IPCA 模型（重要性-绩效竞争对手分析模型）和 DIPCA 模型（动态重要性-绩效竞争对手分析模型）是基于 IPA 模型（重要性-绩效表现模型）构建，目的是为了适应动态的市场竞争环境，了解目标商家水果产品的属性变化特征。

第四章 电商助农水果产品评论数据采集与预处理。本章主要是采用八爪鱼采集器获取京东平台电商助农水果产品和一般水果产品的评论文本数据。在爬取到相应的评论文本后，对文本数据进行预处理，包括评论去重、文本去噪、短文本删除、中文分词与去除停用词等操作，以此为后续研究做准备。

第五章 电商助农水果产品运营策略改进方案形成。本章可以划归为三部分内容——属性级情感分析、竞争力分析及运营策略改进方案形成，主要目的是通过应用本文所提出的“电商助农产品运营策略改进方案”找出目标商家水果产品的劣势属性，以此进行针对性的运营策略改进方案制定。首先构建相应的水果产品领域的属性词典，进行属性级情感分析提取出在线评论中的属性词、观点词及情感倾向，并通过基于情感词典的方法计算出水果产品相应的属性情感得分；之后基于水果产品属性情感得分计算得到属性绩效表现、属性重要性，再进行竞争力分析以找出目标商家产品与竞争商家产品的比较优劣属性；最后针对需重点改进的目标产品劣势属性，通过文本挖掘技术 TF-IDF 提取消费者对其不满意的因素，分析不满意原因，对其进行 K-Means 聚类并以热图呈现可视化结果，最终形成具体的产品运营策略改进方案。

第六章 总结、建议与展望。根据实证分析的结果，对本文提出的电商助农水果产品运营策略改进方案进行探讨总结并得出相关结论，根据研究结果对电商助农不同主体提出不同的建议，并对本文后续研究作出展望，以期未来在该领域取得更好的研究成果。

1.3.2 技术路线

在研究主要内容的基础上做出相应的技术路线图，如图 1.1 所示。

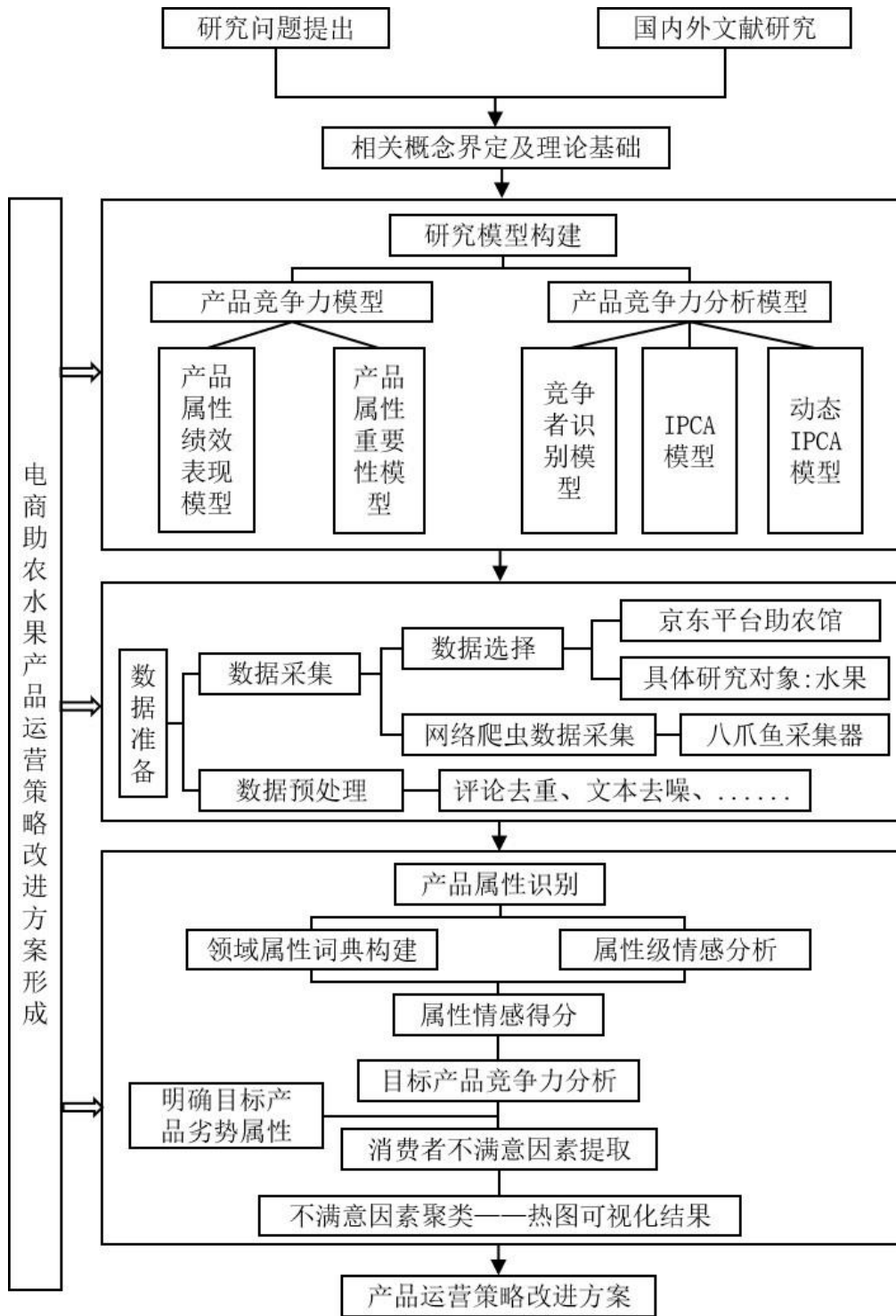


图 1.1 技术路线图

1.4 研究方法

(1) 文献资料分析

主要通过收集、整理和评析与电商助农、情感分析以及基于在线评论的竞争力分析相关的文献资料，了解已有文献的研究现状与存在的不足，结合电商助农模式长久运行的现实需求，从中找到可创新之处，在此基础上确定研究方向，保证研究的合理性。

(2) 文本挖掘技术和工具

在数据的采集、处理和实证分析中，本文用到了八爪鱼采集器以及自然语言处理(jieba分词等)、属性级情感分析、TF-IDF算法等方法。八爪鱼采集器应用于在线评论数据的爬取，自然语言处理主要用于对于评论文本进行预处理，属性级情感分析则基于PaddleNLP搭建SKEP模型提取消费者在线评论中的属性词、观点词及情感倾向，TF-IDF算法用于提取产品劣势属性对应评论中消费者不满意的因素。

(3) 竞争力分析

竞争力分析包括了产品竞争力模型和竞争力分析模型。本文产品竞争力模型，即重要性—绩效表现模型(IPA)，包括了对产品属性绩效表现和属性重要性的计算。其中产品属性绩效表现是根据属性情感得分得到的，产品属性重要性是由产品属性绩效表现构建的计量模型，由多元线性回归得到。本文竞争力分析模型是产品运营策略改进方案形成的核心模型，包括了竞争者识别模型、IPCA模型(重要性-绩效竞争对手分析模型)和DIPCA模型(动态重要性-绩效竞争对手模型)，目的是为了明确电商助农目标产品相对于竞争产品的比较优劣属性。其中竞争者识别模型是首先需要为产品定义一个竞争力指数($Comp_i$ ， $Comp_i$ 取决于属性的绩效表现和重要性)，然后计算每个产品与目标产品之间的不相似指数($Dissim_i$ ， $Dissim_i$ 反映目标商家产品和竞争商家产品之间属性绩效表现的差异程度，使用欧几里德距离计算)，之后通过定义竞争对手指数 $Competitor_index$ 来识别潜在的竞争对手。而IPCA模型和DIPCA模型是基于IPA模型构建，主要通过GAP(属性表现与消费者期望之间的差距)和PD(目标产品与其竞争对手在某属性上的表现差距)两个指标表现。

(4) K-Means 聚类算法

K-Means 聚类算法是一种基于数据划分的无监督聚类算法，主要用于将相似的对象自动归到一个类别中。K-Means 聚类算法核心思想为：首先随机选取 K 个对象作为初始的聚类中心，然后计算每个对象与各个聚类中心之间的距离，把每个对象分配给距离它

最近的聚类中心，聚类中心以及分配给它们的对象就组成一个种类，每分配一个对象，已经聚类的聚类中心会根据聚类中现有的对象被重新计算距离，这个过程将不断重复直到满足某个终止条件——没有对象被重新分配给不同的聚类、没有聚类中心再发生变化或者误差平方和局部最小。本文利用 K-Means 聚类算法对消费者不满意因素热词的特征进行聚类分析，目的是了解消费者对产品某属性不满意的主要原因在于哪些方面。

1.5 研究创新点

本文的创新之处在于以下两点：

(1) 电商助农领域研究视角的创新。将电商助农领域的研究扩展到了电商助农产品这一微观层面，以电商助农水果产品为代表性研究对象进行分析，了解电商助农水果产品的市场表现和竞争力水平，通过构建的产品运营策略改进方案为电商助农水果产品商家提供具体的改进策略决策支持，为解决其他电商助农产品“滞销”问题提供借鉴。

(2) 研究方法的创新。一方面是情感分析方法的创新，本文爬取电商平台在线评论文本作为研究数据，以消费者的角度出发，将情感分析与竞争力分析两种方法结合起来对产品属性特征进行分析，能帮助电商助农产品商家掌握消费者对产品某项属性情感表现的具体原因，得到针对消费者情感倾向和偏好的产品运营策略改进方案。另一方面基于现有竞争力分析方法进行改进，得到 IPCA 和 DIPCA 模型，该模型能直观展示产品各项属性竞争力的动态变化轨迹，应用该模型能使得最终形成的产品运营策略改进方案更好适用于动态变化的市场竞争环境，更具实际效用。

2 相关概念界定及理论基础

2.1 相关概念界定

2.1.1 文本挖掘

文本挖掘（Text Mining）是属于数据挖掘的一个分支，是指从文本数据中获取有价值的信息和知识的一种方法。文本挖掘在信息检索、情感分析、知识发现和舆情监测等方面都具有重要意义。在现代社会中，我们面临着海量的文本数据，如新闻、微博、社交媒体、商品评论等，通过文本挖掘技术，我们不但能迅速提取到大量文本中的关键价值信息，还可以发掘出隐藏在文字背后的知识和规律，为决策和研究提供支持，并应用于实际生活中指导实践。

文本挖掘的流程大致分为五个步骤，主要包括了文本数据采集、文本预处理、文本特征提取、文本数据分析和结果可视化等，如图 2.1 所示。

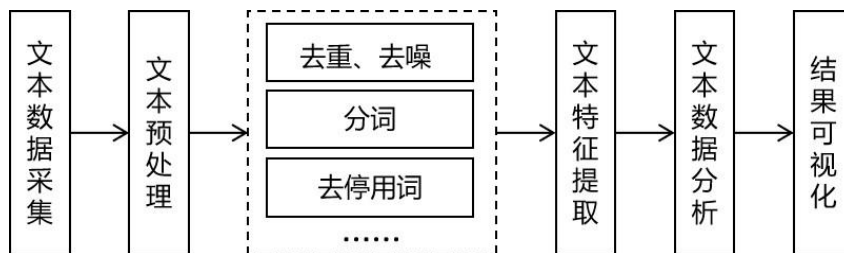


图 2.1 文本挖掘一般流程

(1) 文本数据采集。文本数据获取是对文本数据源的确定，文本挖掘所需数据需要从公开的网页利用 Python 等相关工具进行爬取，或使用现有提供的数据库中的文本数据。

(2) 文本预处理。由于从网页中爬取到的文本数据长短不一，且含有大量噪音，例如符号、表情、注释等，为了确保挖掘结果的精确度，必须对其进行预处理操作。文本预处理一般包括了去重、去噪、删除短句、中文分词、去除停用词等，以便于后续计算机能准确识别非结构化数据。

(3) 文本特征提取。针对预处理后的本文利用向量化方法将自然语言文本转化为向量表示，对数据进行降维，再采用相应的方法对文本数据进行特征选择，筛选出重要性较大的特征，从中提取文本中的关键信息。

(4) 文本数据分析。通过机器学习、数据挖掘、神经网络等相关技术，实现对文本数据的分类、聚类、情感分析、主题挖掘等。

(5) 结果可视化。这一步需要把挖掘到的有用信息转变为易于大众理解的视觉信息，主要是利用计算机技术和图像处理技术把结果以图形或界面等形式展示出来，其中最常用的是词云图或热图。

2.1.2 电商助农

广义上的电商助农指扎根于农村，围绕农产品出村进城和消费品下乡发展的电子商务，其将生产、销售和消费有机结合起来，实现农产品的产销无缝对接，包括利用互联网助力贫困地区实现特色农产品网上销售，实现特色产业、旅游业带动农村创业就业，壮大集体经济。狭义的电商助农指随经济社会发展而逐渐显现的独具特色的农村电商发展之路，主要是大小电商下沉农村，采用流量倾斜的方式，以平台销售、直播带货、商城货架、微商促销等形式为农特产品开辟上行新通道，拓宽农产品销售规模^[48]。

电商助农的实现过程主要分为三个环节：第一，生产环节。在生产环节，通过电商助农不仅可以帮助农民实现生产的规模化和标准化，还可以促进农产品供需平衡。通过线上线下融合的方式，电商企业可指派专业人员为农户提供种子、化肥等生产所需的物资支持以及栽培技术指导培训等相关技术支持。在此过程中，企业可以通过搭建数字虚拟监测平台，对农产品生产情况进行实时监测，确保产品的质量和安全，之后反馈到农户和自身，也可以根据市场需求动向及时调整生产。第二，流通环节。在流通环节，电子商务冷链物流等物流体系的优化完善，使得农户能通过与物流公司合作获得快速、便捷和安全的运输服务，解决农产品销售中的物流难问题。同时，电商平台也可以为农产品进行加工和包装，提高农产品附加值。第三，销售环节。在销售环节，电子商务为农户提供了广阔的销售渠道，农户通过入驻电商平台或与企业合作参与，能在网上实时与消费者沟通产品使用情况，得到消费者反馈和评价，进而反推生产端优化。此外，电商助农的服务模式也加强了农产品的品牌建设，通过电商平台对农产品进行品牌宣传和推广，提升其知名度与美誉度。通过建立品牌形象，农产品可以获得更高的溢价能力，提高农户收入。

2.1.3 产品竞争力

产品竞争力这一概念，至今尚未形成统一、标准的定义。目前对于产品竞争力的概念主要有以下观点：产品竞争力是指产品符合市场要求的程度，这种要求具体体现在消费者对产品各种竞争力要素的考虑和要求上，即消费者对于产品的需求和满意度越高，该产品的竞争力就越强^[49]。另外也有学者认为，产品竞争力是主体（公司、行业等）竞争力在市场中的表现，是在同行业中以产品为中心进行综合比较的重要指标，是通过产品各要素反映出来的公司或行业的整体竞争力水平。由于本文更多倾向于考虑产品竞争力的显示指标，即产品的市场地位和销售情况，没有将设计、研发、生产等环节中的潜在竞争力考虑在内，且是以在线评论作为研究基础，从消费者角度出发，因此本文对产品竞争力的定义与前者相符。

影响产品竞争力的要素可以从三个角度进行分析：基于市场视角的产品竞争力要素、基于竞争者视角的产品竞争力要素和基于消费者视角的竞争力要素。基于市场视角的产品竞争力要素主要在于市场营销因素。市场营销因素对产品销售有着关键影响，决定了产品的市场占有率，需要企业采取有力的市场营销策略来促进产品扩销，包括品牌声誉、广告流量、销量等。基于竞争者视角的产品竞争力要素是建立在了解竞争对手策略、产品、市场和客户等基础上的，包括市场趋势、价格变化、需求动态等，通过竞争对手了解市场趋势和客户需求，发现自身优劣势，以此不断创新优化产品设计。基于消费者视角的产品竞争力要素是要以消费者为中心，对影响消费者满意度的竞争力要素进行分析，主要包括产品特性、服务水平、综合质量、设计水平等，是在实现产品核心功能的基础上以满意消费者需求、提升其使用感为最终目的。

2.2 相关理论基础

2.2.1 消费情感理论

对情感的理解，最早来源于心理学^[50]。情感属于态度的一部分，是态度在生理上一种较复杂而又稳定的生理评价和体验。近年来，情感理论被引入到产品运营管理中，从不同视角理解情感对消费者行为的影响，形成了对消费情感的研究。在前期研究中，学者将情感引入消费者的认知和行为过程中，将其视为认知过程中的附属因素，认为消费情感是消费者在消费经历中对产品或服务的体验而产生的心理反应^[9]。随着研究的深入，

消费情感则被认为是消费者在产品或服务消费过程中产生的一系列情感反应，它是消费者对产品或服务属性感知与最终价值获得之间的匹配和情感性反应^[5]。消费情感具有动态变化的特征，当消费者面对差异性的产品或服务时，会产生不同的情感，从而影响消费者对商品或服务的认知和购买行为。

情感是分维度和极性的，需要对消费情感进行分类与衡量。通过建立在体验和面部表情变化来识别情感的基础上，Lzard 认为人们通常会有正面、负面和中性三种情感表达^[12]。后来部分学者通过双因素的形式构建测量量表，典型的双因素构建是将情感分为正面情感和负面情感，但在实证研究中，经常发现很难区分消费者的情感是正面还是负面。为了使消费情感的划分和衡量变得更加清晰，Mehrabian 和 Russell 提出“愉快—唤醒—支配”量表来衡量消费情感^[17]。在后续的研究中，对于消费者情感极性的划分也采用了这种划分方式，但对于中性情感的存在，仍然需要结合实际应用情况考虑。

消费者情感对消费者的认知和行为产生指导性影响，正面情感会使消费者对产品或服务产生良性认知，促使购买行为发生，而负面情感会抑制或减少购买行为。因此，影响消费情感的因素就成了关键。关于影响消费情感的变量研究，主要包括了消费前情感、服务态度和行为、产品特性、服务环境和感知服务质量等，所得到的结果则表现为消费者满意度和购买行为意向。在电商营销领域中，消费者情感结果还受到了心理期望、物流服务、售后服务等的影响，最为关键的便是在线评价，即受他人消费情感所影响的群体情感感染。在线评价不仅是消费者情感结果的直接体现，也是其他消费者预先了解该产品或服务的主要来源，是网上消费者情感形成的重要影响因素。因此，对于电子商务产品或服务的研究，基于在线评价展开，能更好地了解消费者情感的内在特征。

2.2.2 属性级情感分析理论

(1) 属性级情感分析

属性级情感分析 (Aspect-based Sentiment Analysis, ABSA) 是情感分析领域内的一个细分任务。普通的句子级情感分析是针对一个句子分析其整体的情感极性，然而这一方式很难满足一些特定的应用需求，例如针对某店铺水果的某条评论：“苹果个头很大，没有坏果，物流快，口感细腻”，商家或消费者更想知道该条评论针对苹果的哪些具体属性分别做出了什么样的极性评价。一般而言，这一任务首先要自动从评论文本中提取所有的相关属性，如“个头”“物流”“口感”等，然后分析这一评论对这些属性所展现出来的具体情感极性，因此属性级情感分析实际上主要包含了两个子任务，即属性抽取

和基于该属性的情感极性分析。

属性级情感分析目前主要的关注领域集中在产品评论上，其研究方法和模型与细粒度实体级别的情感分析非常类似。早期研究主要从关于某一特定主题的大规模评论语料中根据词频等相关的统计信息，自动抽取当前指定主题的属性，且在情感极性分类方面，一般采用基于情感词典的方式对每个属性进行情感分类。这一类方法一般都属于无监督的方法，除此之外，后续还有不少更为有效的无监督模型也被应用于属性级别情感分析的这两个子任务中，包括使用 LDA 主题模型用于属性识别，以及采用标签传播算法进行情感分类。为了提升这一类方法的实用价值，词性标注、句法语块以及依存句法等信息被融入到相关规则和统计信息中，使得最终系统能取得一定的效果。在中期阶段，基于有监督学习的统计机器学习模型占据了主流地位。对于产品属性提取，一般采用序列标注的方式实现。通过人工特征构建，结合条件随机场解码，这一子任务能取得比较好的性能。对于第二个子任务，即指定属性条件下的情感分类，一般采用最大熵模型或者支持向量机（SVM）进行分类，其中特征抽取需同时考虑输入评论语句和属性，它们之间的充分结合能够更有效的提升最终分析性能。为了很好的开展此任务的相关研究，不少研究者贡献了标准数据集用于模型训练和性能评价，方便对不同模型的性能进行公平对比。近年来，神经网络由于其强大的自主表示学习能力在自然语言处理领域受到了广泛的关注，在属性级别情感分析也不例外。首先，在词表示方面，主要的阶段性进步包括普通预训练词向量、面向情感的预训练词向量以及基于上下文的预训练词向量如 ELMO、BERT 等；其次，在特征自动抽取和表示学习方面，LSTM 神经网络、记忆神经网络、注意力机制、Transformer 神经网络，以及这些神经网络在融合语法语义等结构信息的加强版本均取得了极大的成功。

随着神经网络模型在属性级别情感分析上取得成功，部分研究者开始关注端到端的方法，采用一个单一的模型架构同时解决属性抽取和情感分类两个子任务，在避免了错误传播问题的同时又能使得这两个子任务充分的进行交互融合。关于属性级别的三元组抽取任务（Aspect Sentiment Triplet Extraction, ASTE）最为典型，它是将属性抽取、情感描述以及情感极性分类充分的结合在一起进行属性级情感分析。鉴于本文研究，要进行竞争力分析以找出电商助农产品目标商家的比较优劣属性，前提是需要划归属性集并计算属性情感得分，因此在属性级情感分析环节选择三元组抽取任务，即抽取结果应为：属性词-观点词-情感极性。

(2) 基于 PaddleNLP 的属性级情感分析

PaddleNLP 是百度飞桨自然语言处理开发库，具备易用的文本领域 API、多场景的应用示例和高性能分布式训练三大特点。PaddleNLP 集成了百度自研的情感知识增强预训练模型 SKEP (Sentiment Knowledge Enhanced Pre-training for Sentiment Analysis)，此模型在 14 项中英文情感分析典型任务上全面超越了 RoBERTa 和 SOTA，受到了业界的青睐。SKEP 算法采用了无监督方法自动挖掘情感知识，然后利用情感知识构建预训练目标，从而让机器学会理解情感语义。SKEP 模型示意图如图 2.2 所示。

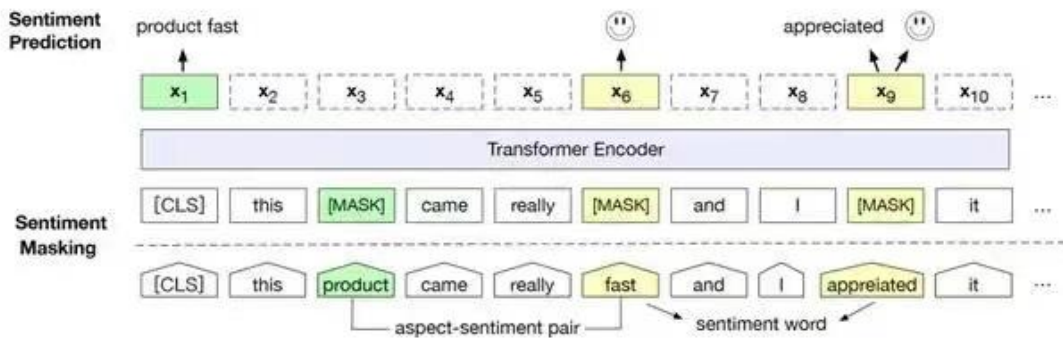


图 2.2 情感预训练 SKEP 模型示意图

具体来说，SKEP 首先基于统计方法从大量无标记数据中自动挖掘情感知识，包括情感词（如图中情感词 fast、appreciated）、情感词极性（如图中 fast 对应的情感为正向）以及观点搭配（如图中 <product, fast> 构成的二元组）。然后，基于自动挖掘的情感知识，SKEP 对原始输入句子中的部分词语进行屏蔽即替换为特殊字符 MASK。最后，SKEP 设计了三个情感优化目标，要求模型复原被屏蔽的情感信息，包括：基于多标签优化的观点搭配预测，如图中 x_1 位置预测 <product, fast> 情感搭配；情感词预测，如图 x_6 位置预测 fast；情感极性分类，如图 x_6 、 x_9 预测的该位置情感极性。这样，通过面向情感的优化目标进行预训练，自动挖掘的情感知识就被有效地嵌入到模型的语义表示中，最终形成面向情感的语义表示。

百度研究团队在句子级情感分析、属性级情感分析和观点抽取这三个典型情感分析任务，共计 14 个中英文数据上验证了 SKEP 模型的效果，结果表明该模型比 SOTA 等更具优势。因此，本文基于 PaddleNLP 搭建 SKEP 模型进行属性级情感分析。

2.2.3 重要度—绩效分析理论

在在线评论研究兴起之时，一些研究采用调查问卷等市场调查方法来获取顾客的需求，据此发现产品需要改进的地方，重要度—绩效分析(Importance Performance Analysis, IPA)便是其中一种行之有效的方法。IPA方法最早是1977年Martilla和James提出的，是一种简单有效的客户满意度评价方法^[18]。传统IPA方法的实施，是通过顾客对产品的各个属性的绩效和重要度进行打分，并计算单个属性的所有顾客评分的均值，将所得均值在分别代表重要度和绩效的双轴坐标系中表示出来，从而形成IPA图，如图2.3所示。

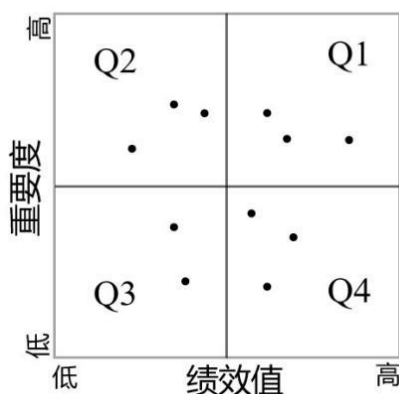


图 2.3 IPA 模型示意图

图 2.3 中：Q1 是优势区，表示对消费者重要的产品该属性绩效值较高，落在该象限中的属性应继续保持当前优势；Q2 是集中改善区，表示对消费者重要的产品该属性绩效表现却不理想，商家应考虑消费者期望并重视的因素；Q3 是消费者重视度低的区域，表示对消费者不重要的产品属性绩效表现也低；Q4 是过度重视区，表示商家对不重要的产品属性投入过度。

IPA 方法优点在于其模型直观形象、操作简单、实用性强等，在学术界已广泛应用。但有学者对其简易性提出了批评。例如，Mikulic 等认为，当不同的产品进行比较时，IPA 模型只能评估本商家产品属性的表现，无法考虑竞争产品的表现，最终得到的分析结果通常不准确，制定的产品改进策略也与实际的市场竞争环境不相符合。为解决该问题，本文研究考虑了竞争产品的信息，构建了适应消费者需求和市场竞争环境的更为全面的方法——IPCA 方法和 DIPCA 方法。

3 电商助农水果产品运营策略改进方案设计

电商助农水果作为农业领域水果市场的一个细分市场，是通过电子商务网络平台，将水果种植户农民与消费者直接连接起来，不仅能实现农户和消费者的直接交流和交易，还能将供应链、销售链以及物流链等环节进行整合和优化，提高了水果产品的销售效率和农户的收入水平，同时也为消费者提供了优质、安全的水果农产品。这种模式通过电商平台打通了水果生产者与消费者之间的直接联系，消除了中间环节，农户可以通过电商平台将自己农产品信息发布到平台上，消费者可以直接通过电商平台购买新鲜水果，实现水果产品的及时销售，很大程度上解决了水果产品销售渠道不畅、销售周期长导致“水果烂在树上、烂在地里无人问津”的问题，缓解了水果产品滞销和价格波动的难题。

虽然在电商助农模式效应的推动下，水果农产品得以加速流通实现农户群体收入的增加，但在电商领域市场，同质化等加剧的市场竞争也使得水果农产品销售再次受阻。因此，作为电商商家，需要持续了解消费者需求状况、市场竞争状况，不断改进产品以满足消费者需求，才能更好地促进产品销售。对此，本文设计了电商助农水果产品运营策略改进方案，为挖掘消费者需求和了解市场竞争状况、制定针对性的产品运营改进策略提供支持。

3.1 电商助农水果产品运营策略改进方案设计原则

本文电商助农水果产品运营策略改进方案的设计遵循了以下原则：

(1) 以消费者需求为核心。以消费者需求为核心和导向，能够有效抓住市场走向，尤其是电商领域，以满足消费者需求为出发点改进产品，能很大程度上提高自身产品的市场竞争力，促进销售。因此本文制定的电商助农水果产品运营策略改进方案是在了解消费者需求和痛点的基础上设计的，即基于在线评论展开研究。

(2) 匹配性。产品改进是一个持续的过程，商家应持续关注市场动态和消费者需求，不断优化产品以满足消费者需求，及时调整运营策略以应对动态市场环境的变化，才能提升产品竞争力，扩大销售。因此，最终得到的产品运营改进策略需要能适应动态变化的消费者需求和市场环境。

(3) 效益性。作为商家，资源投入分配是有限的，追求的是效益最大化，那么利

用有限的资源投入实现产品改进的最优化、最大效益化，则是本文设计电商助农水果产品运营策略改进方案的主要目标。因此，本文是通过找出目标商家产品的优劣属性，针对产品应重点改进的劣势属性展开分析并制定产品运营策略改进方案的。

3.2 电商助农水果产品运营策略改进方案设计

本文提出了一种基于评论文本挖掘的 IPCA 和 DIPCA 的电商助农水果产品运营策略改进方案，该方案分为 4 个部分，主要流程如图 3.1 所示。

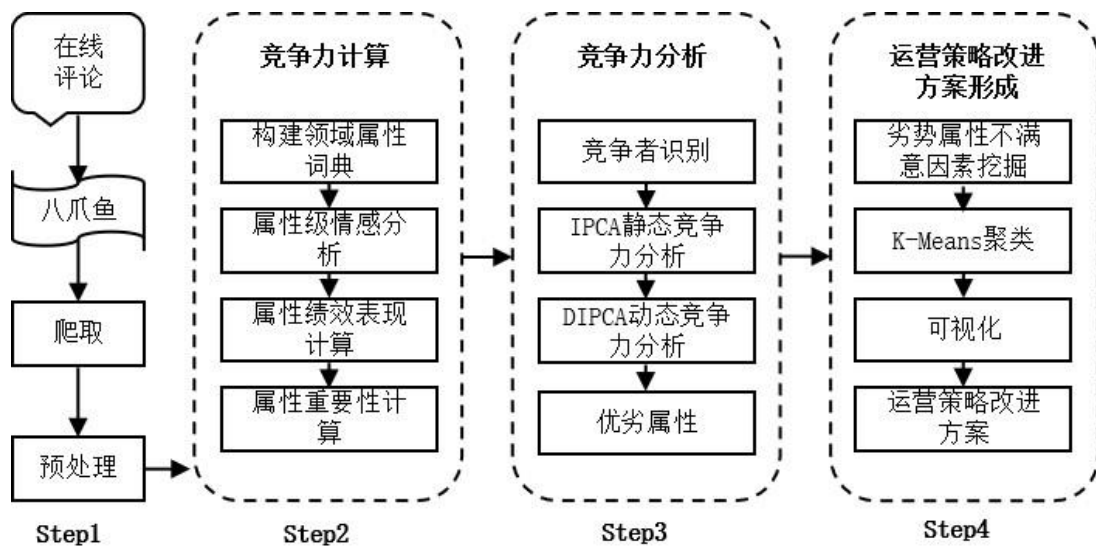


图 3.1 电商助农水果产品运营策略改进方案

本文所构建的电商助农水果产品运营策略改进方案的具体实现流程如下：

(1) 电商助农水果产品评论数据采集与预处理

使用八爪鱼采集器爬取电商平台水果店铺的在线评论文本数据，采集的评论数据集包括了助农水果产品评论数据和一般水果产品评论数据两部分。直接从电商平台上采集的评论数据中，不仅包含了产品属性词、观点词等，也包含了诸多与产品属性无关、对产品属性级情感分析无意义的噪音词语。为了得到有利于后续研究的数据，需要对原始评论数据进行相应的预处理，包括评论去重、文本去噪、短文本删除、中文分词、去除停用词等操作。

（2）属性级情感分析

要准确分析电商助农水果产品的优劣属性，就要先识别和确定产品属性，可以通过人工提取（结合产品自身特点和消费者在在线评论中对产品提及的不同方面等途径）、LDA 主题分析等相关操作来确定产品属性，再结合属性级情感分析所抽取的“属性词-观点词-情感极性”三元抽取任务结果对属性词进行完善，进而构建相应产品的领域属性词典，以此明确产品的评价维度。

在线评论中，情感得分反映了消费者对产品的情感强度即满意程度，情感得分越高，说明消费者对产品相应属性的表现越满意。因此，本文将消费者对产品属性的情感得分作为产品属性的实际绩效值。使用 PaddlePaddle 开源情感计算工具，基于 PaddleNLP 进行属性级情感分析，能从大量评论数据中提取出产品的各个属性词、观点词及情感极性。再利用基于情感词典的情感分析方法量化属性观点的情感强度，计算出产品各个属性对应的情感得分，通过产品竞争力计算模型转化为产品属性的绩效值。

（3）竞争力计算

在得到水果产品各个属性的情感得分后，利用本文所构建的产品竞争力模型，计算出电商助农水果产品和一般水果产品相应的属性绩效表现和属性重要性，依据计算结果可以获悉消费者对水果产品各项属性的满意度和重视程度，将计算结果进行比较还可以了解电商助农水果产品的市场表现如何。

（4）竞争力分析

首先需要确定电商助农水果产品目标研究商家，运用竞争者识别方法识别其电商助农水果产品市场竞争对手及一般水果产品市场竞争对手。之后运用本文所构建的 IPCA 模型进行静态竞争力分析，得出相应的 *GAP* 值（某属性的绩效表现与消费者期望之间的差距）和 *PD* 值（目标产品与竞争产品在某属性上的绩效表现差距），据此绘制出相应的 IPCA 分析图，可以得出目标商家与电商助农水果产品市场竞争对手及一般水果产品市场竞争对手的比较优劣属性，这是产品运营策略改进方案形成所需的核心所在。

之后，根据采集数据集的时间标签，将数据划分为 *T* 段，利用 IPCA 模型计算出每个时间段水果产品的 *GAP* 值和 *PD* 值进行动态竞争力分析，据此绘制出相应的 DIPCA 图，可以直观了解到水果产品各个属性在每个时间段的动态变化特征，能更进一步识别产品运营需改进的方向，明确接下来应及时、重点改善的属性。

（5）运营策略改进方案形成

根据 IPCA 和 DIPCA 分析结果，得到了目标商家水果产品的具体劣势属性及属性改

进优先级，识别了改进方向后，为了解产品属性的改进细节，通过文本挖掘技术 TF-IDF 进一步挖掘待重点改善属性的原始负面评论，提取更细粒度的消费者需求并发现主要的产品属性问题所在，以热词方式呈现消费者对产品属性不满意的高频词，再通过 K-Mans 聚类算法进一步了解热词特征，以热图方式可视化结果，进而提出具体的产品运营策略改进方案。

3.3 产品竞争力模型

电商助农产品运营策略改进方案形成的核心在于竞争力计算及分析部分，因此，需要结合当前基于在线评论的产品竞争力研究的不足，符合市场竞争环境的动态变化，构建切合电商助农产品发展需求的竞争力模型。

3.3.1 产品属性绩效表现模型

消费者满意度是产品竞争力的基础，电商平台评价模块中在线评论所表达的情感间接反映了消费者的满意度，也反映了产品属性绩效表现。因此，本研究利用了属性级情感分析来评估产品的属性绩效表现。具体而言，是通过开源情感计算工具 PaddlePaddle 搭建 SKEP 模型获悉每条在线评论对应属性、观点及相应的情感极性，再通过基于情感词典的情感分析方法计算情感得分，然后根据式(1)来计算各项属性 j 的属性绩效表现。其中 $senti_score_j^i$ 表示属性 j 下句子 i 的情感得分， N_j 表示属性 j 下的句子数。

$$Perf_j = \frac{\sum_i^{N_j} senti_score_j^i}{N_j} \quad (1)$$

3.3.2 产品属性重要性模型

产品竞争力分析的主要目标是识别产品的劣势属性，然后对其实施针对性的改进措施。由于资源有限，商家通常需要权衡产品的那些方面需要优先考虑，以获得最佳的改进效果。在此基础上，本文提出了考虑产品重要性的方法来确定产品属性改进的优先级，以此实现高效的产品竞争力分析。

消费者满意度对产品的市场竞争力起着至关重要的作用。因此，许多研究者从消费者的角度进行竞争力分析。同样，本文也基于消费者的角度来评估产品属性的重要性。参考 Phadernrod 等人^[21]的研究，本文构建了基于产品各项属性的绩效表现与整体消费

者满意度之间的模型，以此来评估产品属性的重要性，而并非要求消费者对属性进行排名。具体而言，将所得的各项属性的情感得分作为自变量，将整体消费者满意度作为因变量，通过多元线性回归（MLR）量化产品各项属性的重要性。

$$Satisfaction = \beta_0 + \sum_{j=1}^T \beta_j \times senti_score_j + \varepsilon \quad (2)$$

式（2）表示 MLR 模型。其中 *Satisfaction* 为整体消费者满意度（以消费者评价星级表示，包括 5 级，分别为 1、2、3、4、5，评价星级越大越满意），*senti_score* 为属性情感得分，*T* 为属性个数， ε 为误差项。将各项属性对应的情感得分和整体消费者满意度代入，可以得到回归系数 β_j 。这里 β_j 可以看作属性重要性，因为该系数反映了在其他变量保持不变的情况下，自变量 *senti_score* 每增加一个单位导致因变量 *Satisfaction* 增加或减少的程度。

3.4 产品竞争力分析模型

3.4.1 竞争者识别模型

在进行竞争力分析之前，需要确定电商助农产品的竞争对手，而在竞争激烈的市场环境中，从消费者的角度来识别竞争对手是相对合理的。参考 Gao 等人^[8]的研究，本文提出了以下方法。具体而言，首先需要定义一个竞争力指数 *Comp_i*，然后计算每位商家产品与目标商家产品之间的不相似指数 *Dissim_i*。

$$Comp_i = \sum Imp_{ij} * Perf_{ij} \quad (3)$$

Comp_i 取决于产品属性绩效表现和属性重要性，由式（3）计算得到。其中 *i* 为产品编号，*j* 为属性编号，*Imp_{ij}* 为 *i* 产品 *j* 属性的重要性，*Perf_{ij}* 为 *i* 产品 *j* 属性的绩效表现。

$$Dissim_i = ED(Perf_c, Perf_i) \quad (4)$$

式（4）中 *Dissim_i* 反映了两个产品之间属性绩效表现的差异程度，使用欧几里得距离计算。*Dissim_i* 值越大，两个产品之间的属性绩效表现差异也就越大。

由此，根据以上定义模型可以得到，当产品的 *Comp_i* 越大而 *Dissim_i* 越小时，则说明该商家产品是目标商家产品的潜在的强大竞争对手。对此，可以定义一个竞争对手指数 *Competitor_Index* 来识别潜在的强大竞争对手，如式（5）所示。其中，*Comp_{Normi}* 和 *Dissim_{Normi}* 分别表示标准化后的 *Comp_i* 和 *Dissim_i*。*Competitor_Index* 值越大的商家产品，

其相对于市场中的其他商家，越可能是目标商家产品的潜在的强竞争对手。

$$Competitor_Index_i = Comp_{Norm_i} - Dissim_{Norm_i} \quad (5)$$

3.4.2 IPCA 模型

根据产品属性绩效表现，可以了解目标产品的市场卖点，即属性优劣表现。属性重要性可以帮助商家了解产品的各项属性对消费者的重要性，即消费者对产品不同属性的重视程度。将这两个指标结合起来使用，可以更好地了解该商家产品的市场竞争力信息。为了更好地评估产品在市场中的竞争地位，需要有效可靠的竞争力分析方法和工具来对这些指标进行建模。

重要性-绩效分析（IPA）是一种常用的商业分析技术，用于了解产品的市场竞争力和消费者满意度并形成相应改进策略，与本文研究非常吻合。因此，本文对产品属性绩效表现和属性重要性综合建模，构建 IPCA（Importance-Performance Competitor Analysis）和 DIPCA（Dynamic Importance-Performance Competitor Analysis），以分析目标产品相对于竞争产品的市场竞争力。传统的 IPA 如图 3.2（a）所示，其中 x 轴为属性绩效表现，y 轴为属性重要性。

在 IPCA 建模过程中，需要计算 GAP 和 PD 两个指标。 GAP_j^f 表示属性 j 的绩效表现与消费者期望（消费者对属性重视程度，即属性重要性）之间的差距，可由式（6）计算所得，其中 \overline{Perf}_j^f 和 \overline{Imp}_j^f 分别表示标准化的 $Perf_j^f$ 和 Imp_j^f 。 PD_j^f 表示目标产品与竞争产品在属性 j 上的绩效表现差距，由式（7）计算得到。

$$GAP_j^f = \overline{Perf}_j^f - \overline{Imp}_j^f, \quad j = 1, 2, \dots, m \quad (6)$$

$$PD_j^f = \overline{Perf}_j^f - \overline{Perf}_j^c, \quad j = 1, 2, \dots, m \quad (7)$$

在得到相应的 GAP 和 PD 两个指标后，可以绘制相应的 IPCA 图，如图 3.2（b）所示，包括了 Q1、Q2、Q3 和 Q4 四个区域。图中：Q1 是“稳定的竞争优势”，表示了位于本象限的相应属性是目标产品相对于竞争产品的主要优势属性，应该继续保持；Q2 是“正面竞争”，表示了该象限的产品属性满足消费者期望，但落后于竞争对手，因此需要改进产品进而追赶竞争对手；Q3 是“改进重点”，位于该象限的属性是目标产品的主要劣势属性，因此需要及时采取行动措施对其进行改进，以提高产品竞争力；Q4 是“虚假优势”，目标产品位于该象限的属性虽然优于竞争对手，但没有达到消费者的期

望。因此，根据 IPCA 图四个区域所载属性的意义可以得到目标产品属性改进的优先顺序为：Q3>Q2>Q4>Q1。

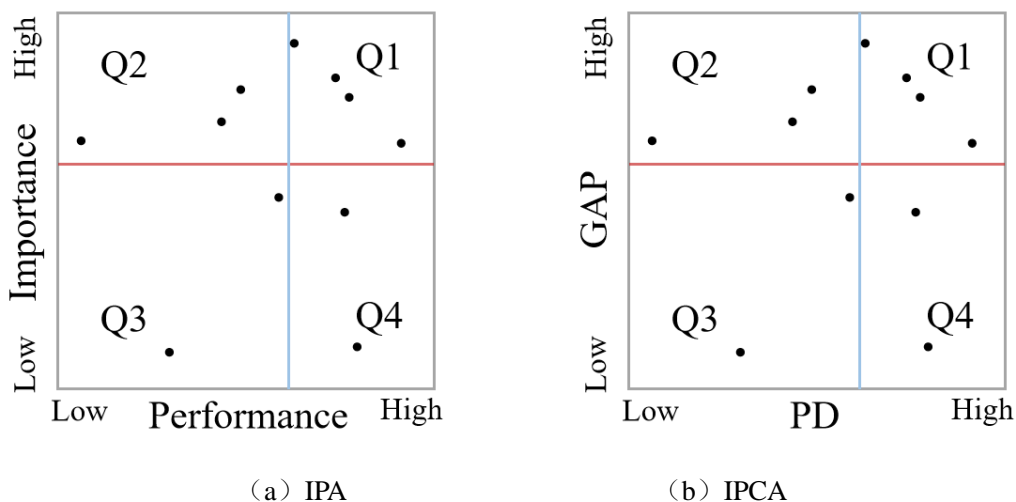


图 3.2 IPA 和 IPCA 图

3.4.3 DIPCA 模型

随着市场环境和产品技术的不断变化，商家的竞争战略也需要不断更新调整。在激烈的市场竞争环境下，商家需要时刻关注竞争对手的动态变化，以及市场的新趋势和新需求，找到应对策略。因此，为了分析电商助农产品竞争力的动态变化，有必要构建动态的 IPCA，即 DIPCA。

根据爬取在线评论数据的时间标签、数据集可以将其分为 T 个子集，反映 T 个时间段。然后通过竞争力模型计算出产品在各个时间段的属性绩效表现 $Perf_j^t$ 和属性重要性 $Im P_j^t$ 。之后由式 (6) 和式 (7) 分别计算出 GAP_j^f 和 PD_j^f ，据此绘制出 DIPCA 图。DIPCA 图如图 3.3 所示，箭头指向为时间推移顺序，该图直观展示了产品竞争力的动态变化轨迹。依据此图，我们可以了解到目标商家产品相对于竞争对手产品在 T 个时间段的优劣属性变化，以及随着时间推移其竞争力强弱的变化。

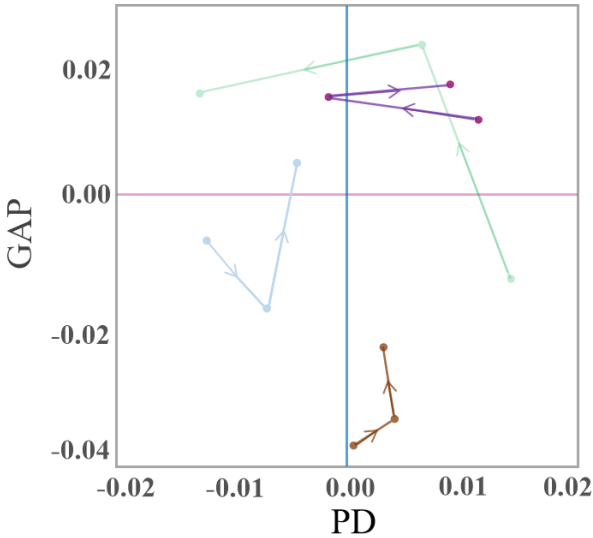


图 3.3 DIPCA 图

4 电商助农水果产品评论数据采集与预处理

4.1 数据来源与数据采集

近年来，随着网络零售市场的不断扩大，国内多家电商巨头纷纷涉足农村市场，电商平台掀起助农热，“助农式”网络销售农产品已成为助农好抓手和新风尚。而“水果”作为大多数农户土地种植经济作物的选择，是电商助农的主要代表产品之一，帮助他们解决“卖不出去”“卖不好”的难题，能有助于他们促销增收，且分析助农水果产品对于研究其他电商助农产品如何实施运营策略改进方案具有重要价值。

目前电商平台众多，且都具有自身助农产品经营模式特色，因此本文综合考虑了平台模式、消费者使用频率及平台评价结构，最终选取京东助农馆频道中的水果为研究对象，选取了苹果、橙子、梨、桃子四种水果品类，将这四类水果的在线评论作为本文产品改进策略形成的主要研究目标。由于消费者在电商平台购买产品时，往往会对多个产品进行比较，选择最满意的产品，因此本文为使最终的产品运营策略改进方案更符合实际，也为了更好地了解助农水果产品的市场表现，还选取了京东平台一般商家水果产品（即非助农商家水果产品）作为外部市场竞争对手与目标商家产品进行比较分析。为便于后续研究分析，剔除了在线评论数在 200 以下的商家后，确定最终研究对象为助农水果商家（S1-S28）28 家和一般水果商家（S29-S51）23 家，如表 4.1 所示。

表 4.1 店铺信息

店铺名	ID	店铺名	ID	店铺名	ID
中国特产·洛川助农馆	S1	中国特产·辛集馆	S18	农鲜淘京东自营旗舰店	S35
中国特产·洛川馆	S2	中国特产·临汾助农馆	S19	陇原红官方旗舰店	S36
中国特产·静宁助农馆	S3	中国特产·绍兴农特产馆	S20	播鲜匠京东自营旗舰店	S37
中国特产·静宁农特产馆	S4	中国特产·大荔助农馆	S21	堡鲜生旗舰店	S38
中国特产·庆阳合水助农馆	S5	中国特产·新疆助农馆	S22	绿水秦川生鲜旗舰店	S39
中国特产·旬阳助农馆	S6	中国特产·蒙自馆	S23	京鲜惠自营旗舰店	S40
中国特产·灵宝馆	S7	中国特产·眉县农特产馆	S24	众百亿生鲜专营店	S41

续表 4.1 店铺信息

店铺名	ID	店铺名	ID	店铺名	ID
中国特产·延安馆	S8	中国特产·翁源馆	S25	甫之牛生鲜专营店	S42
中国特产·渭南馆	S9	中国特产·秦安馆	S26	果果牛旗舰店	S43
中国特产·民勤助农馆	S10	中国特产·相城农特产馆	S27	硕懿果水果专营店	S44
中国特产·镇安助农馆	S11	中国特产·平遥助农馆	S28	骄瑰水果旗舰店	S45
中国特产·秭归助农馆	S12	自然故事自营旗舰店	S29	爱沃尔德果蔬专营店	S46
中国特产·利川助农馆	S13	自然搭档旗舰店	S30	阳山蜜语桃源专卖店	S47
中国特产·当阳馆	S14	云鲜风京东自营旗舰店	S31	荣鼎生鲜专营店	S48
中国特产·汶川馆	S15	曙光沙窝生鲜专营店	S32	蔓悦生鲜专营店	S49
中国特产·赵县农特产馆	S16	趣鲜语京东自营旗舰店	S33	农仙味生鲜旗舰店	S50
中国特产·洋符馆	S17	青沃生鲜专营店	S34	顶端果业官方旗舰店	S51

本文利用八爪鱼采集器（Octoparse）进行在线评论文本数据爬取，在京东平台上爬取了 28 家助农水果商家评论共计 20369 条，爬取了 23 家一般水果商家评论共计 25057 条，数据起止时间为 2022 年 1 月~2023 年 6 月，部分初始评论数据展示如图 4.1 所示。因京东平台上的在线评论只展示前 100 页，在数据采集的规定起止时间范围内，所爬取的每户商家评论数据最多只爬取到商品评价模块的前 100 页。

会员	级别	评价星级	评价内容	时间	追评时间	追评内容	商品属性
j***l	PLUS会员	star5	物流很快, 包装完整无损, 快递员送货上门到位	2023/6/27 13:45			80mm中果 (带箱5斤)
E***y	PLUS会员	star5	再次回购他家苹果, 苹果大小均匀, 果型端正, 脆爽多汁, 酸甜可口, 很好吃!	2023/6/27 12:57			80mm中果 (带箱5斤)
j***p		star5	包装很严实, 也没有任何损坏。苹果六个都很完	2023/6/23 20:07			6枚75mm 试吃装
晨***8	PLUS会员	star4	洛川馆的苹果?买过多次了, 商家发货及物流的速	2023/6/22 14:06			80mm中果 (带箱5斤)
y***6		star5	发货速度还可以, 吃起来也不错, 有点忘了一共	2023/6/20 11:45			80mm中果 (带箱5斤)
5***1	PLUS会员	star5	已经多次订购洛川助农馆提供的苹果了, 苹果新	2023/6/20 10:26			80mm中果 (带箱9斤)
吃***踢		star4	苹果很甜, 汁水非常多。虽然买的是最小个的试	2023/6/19 21:20			6枚70mm 试吃装
j***k		star5	苹果个头匀称, 买的小果给老爸打果汁喝, 收到	2023/6/15 12:56			12枚70mm
李***宝		star5	发货很快, 已经收到货了, 商品和描述的差不多	2023/6/15 18:10			12枚70mm
A***L		star3	买的小果, 发货速度很快, 店家服务态度也不错	2023/6/12 16:06			70-75mm 小果混装 (带箱9斤)
y***o	PLUS会员	star4	质量: 洛川苹果品质好, 脆甜口感, 75mm大小合适。还赠送羊角蜜, 羊角蜜也不错;	2023/6/12 7:41			24枚75mm 一级小果
l***7	PLUS会员	star1	口感硬的要命, 没吃一两口就腻了	2023/6/11 13:14	[购买50天, 放到烂了	看	18枚90mm 一级大果
残***尘	PLUS会员	star5	京东买水果的物流速度还是蛮快的, 苹果被包裹	2023/6/11 9:01			80mm中果 (带箱5斤)
j***8		star5	包装完好, 到货迅速。没有破损的地方, 苹果也	2023/6/9 15:14			70-75mm混装 (带箱9斤)
j***5	PLUS会员	star5	洛川苹果, 好吃不贵, 物流也很快。苹果味道酸	2023/6/9 13:22			70-75mm混装 (带箱9斤)
y***p	PLUS会员	star4	商品一收到坏了一个, 碰伤了一个, 与店家沟通	2023/6/7 13:43			70-75mm混装 (带箱9斤)
w***-		star5	苹果很好吃, 清脆酸甜可口, 但是箱子有点搞脏	2023/6/5 19:19			80mm中果 (净重9斤)
大***胆	PLUS会员	star5	还行吧看都是挺好看一个的, 个头也超级大。味	2023/6/3 13:20			80mm中果 (净重9斤)
j***1	PLUS会员	star5	选购了京东物流送货 就很方便快捷 省心省力了,	2023/6/2 13:39			70-75mm混装 (带箱9斤)
u***7		star5	苹果个头很大, 5斤才8个, 差点担心份量不足了	2023/6/2 12:31			80mm中果 (带箱5斤)
猫***分	PLUS会员	star5	我买的5斤80mm中果, 9个, 不大不小个头均匀	2023/6/2 10:59	[购买6天后吃了, 很好		80mm中果 (带箱5斤)
j***r	PLUS会员	star5	早市买的苹果都没有苹果的味道, 自从知道洛川	2023/5/27 22:39			12枚70mm
J***0		star5	从快递点拿回来我本来挺生气的, 因为箱子都破	2023/5/27 18:09			12枚75mm

图 4.1 部分评论数据展示

4.2 评论文本数据预处理

商品用户在线评论是用户在购买商品后根据对商家的商品、服务、物流等环节的实际体验来撰写评论, 是用户对商品使用的主观反馈, 能够在很大程度上影响其他用户的购买决策行为, 从而影响该商家商品的销售量。然而并非每一位用户都是认真给出评论的, 有的用户做出的评论会存在多余的标点符号、没有实际价值的英文字符等, 如“苹果已收到,, , 很新鲜, 很好吃! ?? &Sebu” 该条评论中的标点符号和最后的英文及符号, 有的用户评价过于简短, 如“好吃”、“非常好”, 有的用户直接复制其他用户评论或留空评论, 如“此用户未填写评价内容”。另外, 由于网络环境特性, 用户评论也并非都是标准化、规范化的, 而优质的文本规范化处理可以提高主题提取等文本分析任务或 NLP 项目的准确度, 因此有必要对已爬取到的用户评论进行文本清洗预处理。本文通过评论去重、文本去噪、短文本删除、中文分词与去除停用词一系列操作将上述所爬

取到的助农水果和一般水果评论文本进行规范化处理。

4.2.1 评论去重

评论去重，是指对用户评论中重复出现的文本数据进行归一处理，这是进行文本数据预处理最基本也是最重要的一步。由于电商平台为避免一些用户长时间不发表评论，往往都会设置一道自动回复程序，即如果用户购买商品后超过规定时间仍未做出评论，系统就会自动代替用户进行评论，表现为“此用户未填写评价内容”或“非常满意”等，往往这些评论都是好评。然而，这些评论显然没有任何分析价值，且存在大量重复出现，所以我们需要对其进行去除。

这里需要特别提及的是，本文是以“苹果”、“橙子”、“梨”、“桃子”四类水果为研究对象，所采集到的在线评论文本数据中“苹果”、“橙”、“梨”、“桃”必定为高频词汇，但考虑到如果去除掉这些词，许多文本会出现只有观点词的情况，在属性级情感分析环节提取属性词和观点词时会产生无法提取到的问题，例如评论“苹果可以，都是好果。”，去掉词“苹果”后，机器会认定此句无法进行属性词-观点词-情感极性分类三元组任务，最终会导致属性级情感分析结果缺失掉很多关键信息。因此，考虑到消费者对水果产品进行评论通常会以水果名作为产品的属性评价对象，本文不去除“苹果”、“橙子”、“梨”、“桃子”这类高频词。

本文使用 Python 代码进行文本去重，最终结果为助农水果评论共删除了 2376 条重复评论，一般水果评论共删除了 3902 条重复评论。使用 `value_counts` 函数统计重复评论，可以得到重复评论文本的重复数量，数量最大的评论文本可能是系统默认评论。

4.2.2 文本去噪

文本去噪，是指去除评论文本数据中英文字母、数字、符号等无效文本，这些不仅对评论文本整体意思的表达没有实际价值，而且还可能对后续文本分析产生影响，因此需要将其去除。利用自然语言处理技术进行文本去噪的方法主要包括正则表达式、规则匹配和机器学习等。其中正则表达式是一种定义了搜索模式的特征序列，主要用于字符串的模式匹配，本文使用 Python 中的 `re` 正则表达式模块进行文本去噪。

4.2.3 短文本删除

短文本，是指字数较少、语句较短的评论文本。根据语言的特点，如果一条评论所

包含的字数越少，那么这条评论所表达的含义越少，能挖掘的有用信息就越少。例如本文所爬取到的初始评论文本数据中出现的“很好”“不错”“差”等评论，只能确定该评论的情感极性，无法从中提取到相应的水果属性特征，即不能识别评价用户对产品的哪一个属性满意或不满意，对后续属性级情感分析没有什么实际价值。因此，为了提高文本预处理结果的精度，本文设置最短字符长度门槛为 3 进行短文本删除。

4.2.4 中文分词

由于所爬取到的在线评论为非结构化数据，计算机在自然语言处理过程中无法直接识别，因此需要将其转化为计算机可以理解和处理的结构化数据，即对中文文本评论数据进行中文分词处理。中文分词是将一个句子按照规范的准则拆分为相应的次序列，即将连续的中文文本切分为一个个有意义的词语过程，并且不改变原始语句的整体含义。在爬取到的在线评论中，词是可以反映用户对商品的使用体验、需求和期望的最小文本单元。不同于英文文本中单词之间是以空格作为自然分界符，中文只是字、句和段能通过明显的分界符来划界，在词项上没有一个形式上的分界符，在分词之前需要首先识别词与词之间的边界。对于中文而言，常见的分词算法有基于统计、基于规则和基于理解三类。随着自然语言处理技术的发展，基于这三类算法的较为成熟中文分词开源工具也越来越多，如 jieba、LTP-cloud、Jcseg、Stanford 汉语分词、庖丁解牛分词等。Jieba 分词提供了 Python 和 R 在内的多种编程语言实现，支持自定义领域词典使用以提高分词效果，提供多种分词模式，因此本文选择 jieba 进行中文分词。Jieba 分词目前支持四种分词模式，如表 4.2 所示。

表 4.2 jieba 分词模式

分词模式	特点
精确模式	试图将句子最精确的切开，适合文本分析
全模式	将句子中所有可能成词的词语都扫描出来，速度非常快，但不能解决歧义
搜索引擎模式	在精确模式的基础上对长词再次切分，提高召回率
paddle 模式	利用 PaddlePaddle 深度学习框架，训练序列标注（双向 GRU）网络模型实现分词

相对于传统的分词算法，paddle 模式采用了深度学习模型，能获得更高的分词准确度和更快的分词速度，因此本文选择 jieba 分词的 paddle 模式对本文的评论文本数据进行中文分词操作。

4.2.5 去除停用词

在对在线评论文本进行分词后，文本中还存在一些与数据挖掘分析无关的词，即停用词，这些词在句子中仅起到结构作用，对后续分析作用不大，还会增加词向量的维度，这些词主要是介词、连词、语气助词等，为了简化文本数据并提高机器处理效率和分析准确率，有必要采取去除停用词操作。

本文基于现有的停用词表，包括百度停用词表、哈工大停用词表、四川大学停用词表，并加上电商领域专有的词，以此构成新的停用词表，对在线评论文本进行停用词过滤，得到去停用词后的文本数据，部分评论文本数据如表 4.3 所示。

表 4.3 分词及去停用词后效果样例

序号	分词结果
1	包装/精美/大小/均匀/值得/购买
2	多次/回购/质量/一如既往/好/味道/不错/香甜/汁水/多
3	发货/太慢/一周/才/送到/打开/箱子/看/表皮/损坏/质量/糟糕
4	味道/好/虽然/个头/很小/但是/很甜/很脆/非常棒/一次/购物/体验/下次/还会/回购
5	品质/很好/口感/不错/发货/没/坏果/客服/态度/很好/口感/脆甜/放在/冰箱/口感/更好
6	优惠活动/购买/超级/便宜/苹果/很大/很甜

5 电商助农水果产品改进策略形成

5.1 电商助农水果产品属性识别及领域属性词典构建

对于如何有效制定电商助农水果产品的改进策略，其核心是找出其优劣属性，因此首先要确定产品的主要属性。本文参照苹果、橙子、梨、桃子四类水果的相关《中华人民共和国国家标准 GB》，并结合水果产品特点和消费者在线评论中提到的不同方面，确定了水果产品的综合属性，具体包括了“果面”“果形及大小”“颜色”“口感”“包装”“物流”“服务”“价格”“综合质量”“店铺”11个属性以及448个属性特征词，如表5.1所示。电商助农水果产品领域属性词典中的属性及特征词，其全面涵盖了消费者对水果产品关注的各个方面，能实现后续制定产品改进策略的针对性和具体化。

表 5.1 电商助农水果产品领域属性词典（部分）

属性	特征词
新鲜度	新鲜，新鲜度，生产日期，保质期，装箱日期
果面	果面，品相，果皮，外观，表面，表皮，烂，好果，坏果
果形及大小	果形，个头，大小，分量，重量，个数
颜色（色泽）	颜色，色泽，色相，诱人，食欲，红润
口感	口感，味道，口味，风味，果肉，肉质，细腻，甜度，可口，好吃，脆甜
包装	包装，外包装，密封，泡沫包装，纸箱
物流	物流，快递，发货，配送，速度，运输，时效
服务	服务，客服，态度，售后，售前，沟通，反馈
价格	价格，性价比，便宜，实惠
综合质量	质量，总体，整体，品质，回购
店铺	店铺，店家，卖家，商家，京东，品牌

5.2 电商助农水果产品属性级情感分析

按照本文所构建的电商助农水果产品运营策略改进方案，在明确水果产品的评价维度即产品属性后，基于 PaddleNLP 搭建 SKEP 模型对预处理后的在线评论数据集进行属性级情感分析，提取出相关的属性特征词、观点词及情感极性，计算出水果产品各个属性的情感得分，是进行竞争力分析以识别目标产品的优劣属性的前提。

首先，需要搭建评论观点抽取模型。模型采用序列标注的方式进行评论观点抽取，

即会抽取评论中的属性及属性对应的观点。本文在具体实践上基于 BIO 的序列标注体系进行了标签的扩展：B-Aspect, I-Aspect, B-Opinion, I-Opinion, O，其中前两者用于标注评论属性，后两者用于标注相应观点。如图 5.1 所示，首先将预处理好的在线评论文本数据输入 SKEP 中，SKEP 将会对文本数据的每个 token 进行编码，产生对应向量序列，然后基于向量序列进行预测每个位置上的输出标签。

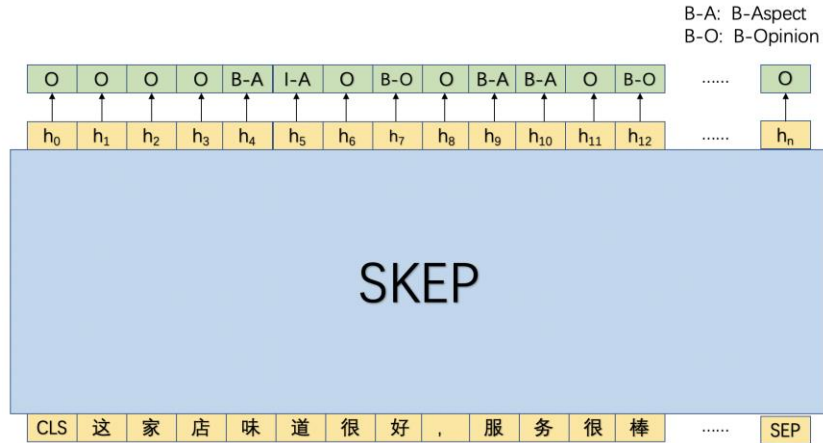


图 5.1 评论观点抽取模型示意图

在抽取完评论观点后，便可以有针对性地对各个属性进行评论，并进行情感分类，即搭建属性级情感分类模型。具体而言，是将抽取出的评论属性和评论观点进行拼接，然后和原始评论语句进行拼接作为一条独立的训练语句。如图 5.2 所示，首先将评论属性和观点词进行拼接为“味道好”，接着将“味道好”和原始评论进行拼接，再传入 SKEP 模型，SKEP 模型将会对评论文本的每个 token 进行编码产生对应向量序列，并使用“CLS”位置对应的输出向量进行细粒度情感极性分类。

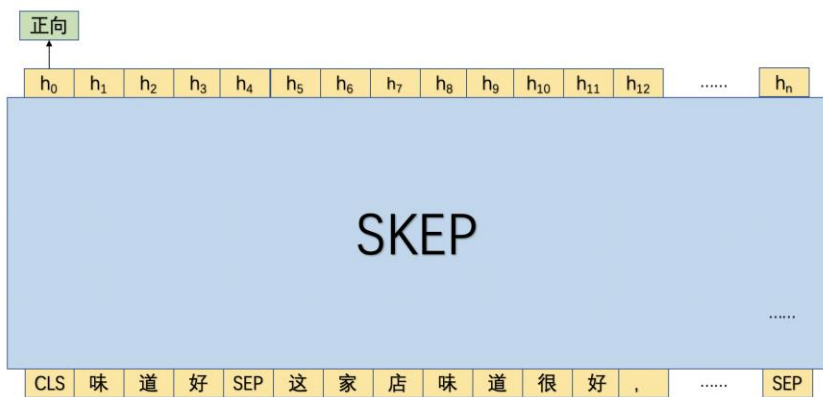


图 5.2 属性级情感分类模型示意图

基于 PaddleNLP 搭建 SKEP 模型，对预处理后的助农水果产品和一般水果产品在线评论数据集分别进行属性级情感分析，得到结果如下表 5.2 所示。

表 5.2 属性级情感分析结果示例

文本	属性级情感分析结果		
	属性	观点	情感倾向
包装精美，大小均匀，发货快，	包装	精美	正向
推荐购买	大小	均匀	正向
	发货	快	正向
多次回购，质量一如既往好。味	质量	好	正向
道不错，香甜，汁水多	味道	不错，香甜	正向
	汁水	多	正向
发货太慢，一周才送到，打开箱	发货	慢	负向
子看表皮损坏，质量糟糕	表皮	损坏	负向
	质量	糟糕	负向

属性级情感分析结果中，有不少文本会出现这样的情况，如“苹果可以，都是好果”“这家店梨好吃，脆脆甜甜”，得到的结果为“属性：‘苹果’，观点：‘可以，好果’，情感分类：‘正向’”和“属性：‘梨’，观点：‘好吃，脆脆甜甜’，情感分类：‘正向’”这样属性词为产品名时，这时候则需要以观点词为属性界定依据，根据领域属性词典将属性级情感分析结果划归到相应的属性类别，例如将“可以，好果”划归到“果面”属性，将“好吃，脆脆甜甜”划归到“口感”属性进行情感得分计算。

在完成属性级情感分析后，根据领域属性词典可以得到每个商家不同属性的观点词数据集，采用基于情感词典的情感分析方法对其进行量化，计算各个属性 j 的情感得分 $senti_score_j$ 。量化标准参考知网情感词典（HowNet）自带的分类标准。

5.3 电商助农水果产品竞争力计算

5.3.1 产品属性绩效表现

利用 3.3.1 节提出的式 (1)，可以计算出电商助农水果产品和一般水果产品的属性绩效表现，结果如图 5.3 所示。从图 5.3 中可以发现，电商助农水果产品各项属性绩效表现都明显低于一般电商水果产品，表明了电商助农水果产品在电商水果市场上的竞争力较弱，这一发现也验证了前人关于电商助农产品普遍市场化程度低、消费者满意度较低的研究结果。电商助农水果产品商家要提升消费者满意度并促进其市场发展，需要识别和明确其水果产品的劣势属性，找出导致消费者不满意的原因，并针对性地采取相应的改进方案。

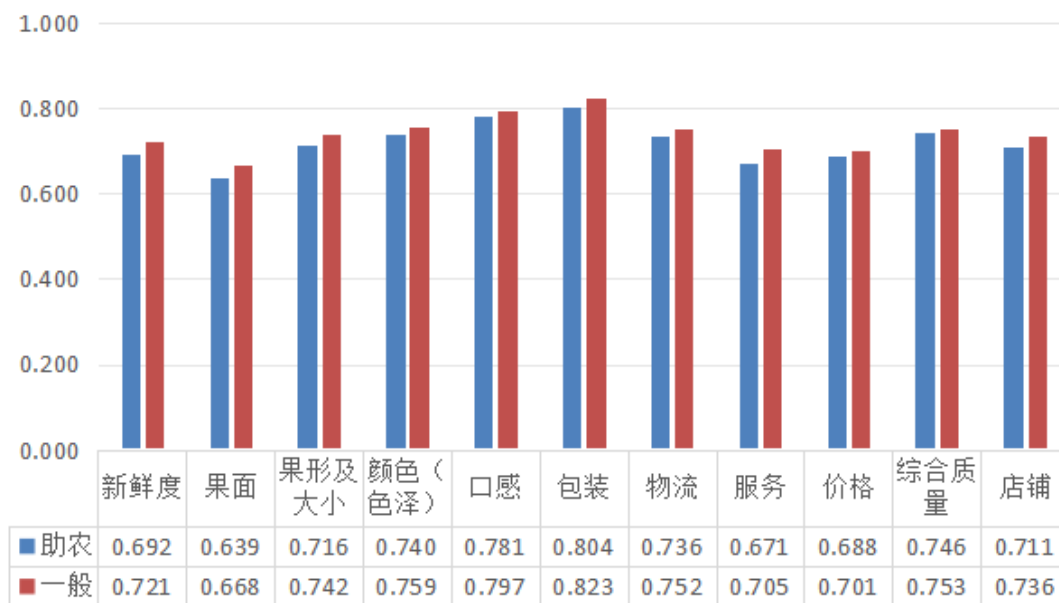


图 5.3 属性绩效表现

5.3.2 产品属性重要性

根据 3.3.2 节中的式 (2)，可以计算出电商助农水果产品和一般水果产品的属性重要性，结果如表 5.3 所示。所有属性的绩效表现对消费者满意度都有着显著的正向影响，表明不存在冗余属性。从表 5.3 中可以看出，最重要的三个属性是“新鲜度”“果面”“口感”，这表明消费者对水果产品最关心的是产品本身。由于生鲜水果产品本身具有储存

周期短、易腐坏等特性，消费者对“物流”属性也很重视。

表 5.3 属性重要性

属性	重要性 (β_j)	Standard error	VIF
新鲜度	0.445***	0.007	1.022
果面	0.428***	0.012	1.014
果形及大小	0.222***	0.016	1.024
颜色 (色泽)	0.191***	0.011	1.037
口感	0.398***	0.006	1.025
包装	0.235***	0.017	1.017
物流	0.382***	0.015	1.025
服务	0.157***	0.012	1.016
价格	0.379***	0.007	1.027
综合质量	0.214***	0.010	1.022
店铺	0.203***	0.015	1.011

注: *** $p < 0.01$

此外，表 5.3 中的 Standard error (标准误差) 和 VIF (方差膨胀因子) 值可以说明模型的效用和变量之间的相关性。具体而言，标准误差是一组值偏离平均值的程度的度量概念，可以通过计算方差的算术平方根得到。该值越大，说明回归模型误差越大；该值越小，说明回归模型误差越小。由表 5.3 可知，各属性的标准误差均小于 0.05，说明多元线性回归效果良好，即各属性重要性能反映不同属性对消费者满意度的影响程度。另外，由表 5.3 也可知，不同属性之间的 VIF 值都在 1.0 左右，说明属性之间不存在显著的相关性。

5.4 电商助农水果产品竞争力分析

5.4.1 竞争者识别

随机选择目标 S5 中国特产·庆阳合水助农馆作为本文电商助农水果产品目标研究商家，运用 3.4.1 节竞争者识别方法识别其竞争对手，即商家的竞争对手指数 (*Competitor_Index* 线) 越大，就越有可能是目标商家的强竞争对手。如图 5.4 和图 5.5 所示，S4、S37 的竞争对手指数相对于其他商家更大，可以确定 S4 为电商助农水果产品商家中的竞争对手，S37 为一般水果产品电商商家中的竞争对手。

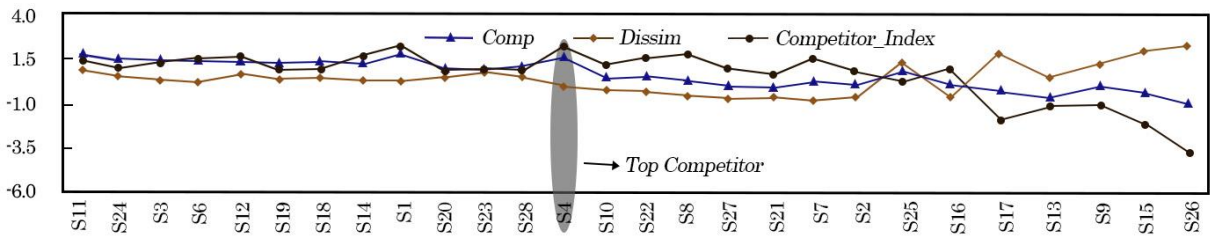


图 5.4 电商助农水果产品商家竞争力指数

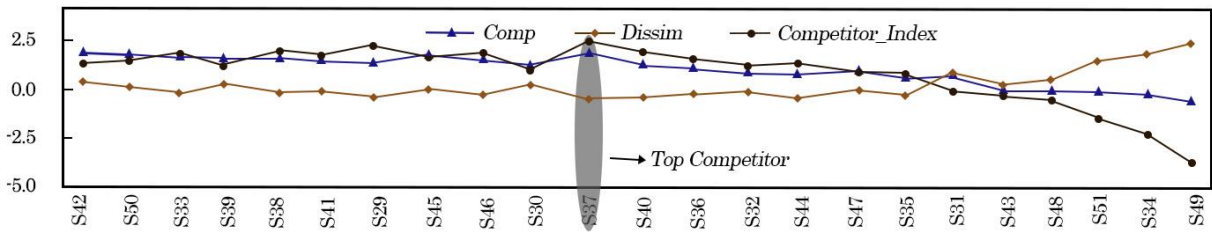


图 5.5 一般水果产品电商商家竞争力指数

5.4.2 静态竞争力分析

利用 3.4.2 节的方法，本文构建了如图 5.6 所示的 IPCA 图，其中图（a）为目标商家 S5 与电商助农水果产品商家竞争对手 S4 的比较，图（b）为目标商家 S5 水果与一般水果产品电商商家竞争对手 S37 的比较。

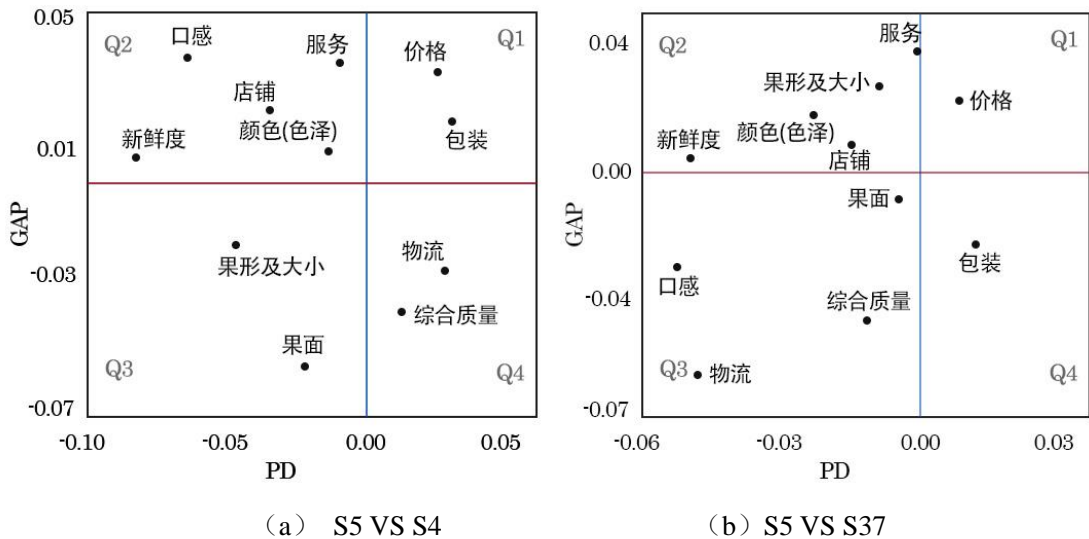


图 5.6 IPCA 结果

从图 5.6 (a) 中可以发现, 目标商家 S5 水果产品的“果面”和“果形及大小”属性均落在 Q3 (改进重点) 内, 即这些属性没有达到消费者预期, 均落后于竞争对手。Q2 (正面竞争) 中显示目标商家 S5 水果产品的“新鲜度”“口感”“颜色 (色泽)”“服务”“店铺”属性超出了消费者预期, 但仍比竞争对手表现差。此外, Q4 (虚假优势) 中的“物流”“综合质量”“服务”属性虽然表现优于竞争对手, 但并未满足消费者期望。从 Q1 (稳定的竞争优势) 中可以发现“包装”和“价格”属性, 这是目标商家 S4 水果产品在电商助农水果产品市场中的优势。

与电商助农水果产品市场表现不同, 在和一般水果产品电商商家 S37 的竞争中, 目标商家 S4 水果产品的更多属性处于劣势。如图 5.6 (b) 所示, 除了“果面”属性外, “口感”“物流”及“综合质量”属性也位于 Q3。在 Q1 只发现了“价格”属性, Q4 只有“包装”属性, 其余都位于 Q2, 即落后于竞争对手。

通过以上竞争力分析, 可以帮助目标商家获得消费者综合评价以及水果产品市场信息。根据 IPCA 结果, 目标商家 S5 最迫切需要改善的水果产品属性是“果面”、“果形及大小”, 这两个产品属性的提升对促进目标商家的市场发展有更大作用。第二优先级改善属性是“新鲜度”“颜色 (色泽)”“服务”“店铺”等。从长远来看, 与一般电商商家相比, “口感”“综合质量”属性也需要提升才能获得竞争优势, 还需要改善“物流”属性, 以提升目标商家在整个电商水果产品市场中的竞争力。

5.4.3 动态竞争力分析

对于商家而言, 与消费者建立长期的关系有利于其持续发展。由于产品竞争力是动态变化的, 这就需要商家针对竞争产品优化带来的市场挑战, 及时改进产品运营策略以满足消费者变化的偏好和需求, 与消费者建立长期良好的关系。因此, 对目标商家 S4 进行动态竞争力分析是有必要的, 能及时捕捉到自身水果产品优劣属性的动态变化特征, 对其进行适时的调整和改善。

具体来说, 根据时间标签将数据集分为三组, 即将 2022 年 1 月~2023 年 6 月数据分为 2022 年 1 月~2022 年 6 月、2022 年 7 月~2022 年 12 月、2023 年 1 月~2023 年 6 月三个时间段。同样, 我们将 S4 作为竞争对手。根据 3.4.3 节的方法, 可以计算出每个时间段的属性绩效表现和重要性, 具体结果见下表 5.4、表 5.5 和表 5.6。从表 5.4、表 5.5 和表 5.6 中可以看出, 水果产品属性绩效表现是不断变化的, 在属性重要性方面, 大多数

属性的重要性排序变化不大。例如，在第一时间段“服务”属性重要性排名第 11 位，在第三时间段升至第 8 位，表明消费者逐渐重视服务。总得来说，对于电商产品，尤其是水果这类特性不稳定的产品，消费者的偏好和需求可能会随着时间的推移而改变，因此从动态的角度理解产品属性变化至关重要。

表 5.4 2022 年 1 月~2022 年 6 月数据

属性	<i>Perfs₄</i>	<i>Perfs₅</i>	<i>Perfs₄- Norm</i>	<i>Perfs₅- Norm</i>	<i>Imp</i>	<i>Imp- Norm</i>	<i>Imp- Ranking</i>	<i>GAP</i>	<i>PD</i>
新鲜度	0.689	0.702	0.101	0.119	0.320	0.102	6	-0.001	-0.018
果面	0.633	0.650	0.082	0.088	0.424	0.107	3	-0.025	-0.006
果形及大小	0.642	0.725	0.081	0.097	0.365	0.103	5	-0.022	-0.016
颜色（色泽）	0.756	0.759	0.071	0.073	0.263	0.068	7	0.003	-0.002
口感	0.707	0.817	0.108	0.125	0.483	0.130	1	-0.012	-0.017
包装	0.671	0.682	0.089	0.085	0.197	0.059	9	0.030	0.004
物流	0.715	0.688	0.073	0.071	0.388	0.104	4	-0.031	0.002
服务	0.746	0.712	0.081	0.089	0.153	0.031	11	0.050	-0.008
价格	0.814	0.766	0.086	0.089	0.185	0.043	10	0.043	-0.003
综合质量	0.755	0.776	0.072	0.072	0.447	0.128	2	-0.056	0.000
店铺	0.769	0.794	0.080	0.097	0.241	0.059	8	0.021	-0.017

表 5.5 2022 年 7 月~2022 年 12 月数据

属性	<i>Perfs₄</i>	<i>Perfs₅</i>	<i>Perfs₄- Norm</i>	<i>Perfs₅- Norm</i>	<i>Imp</i>	<i>Imp- Norm</i>	<i>Imp- Ranking</i>	<i>GAP</i>	<i>PD</i>
新鲜度	0.759	0.797	0.102	0.117	0.433	0.084	6	0.018	-0.015
果面	0.598	0.632	0.097	0.107	0.584	0.141	1	-0.044	-0.010
果形及大小	0.651	0.799	0.092	0.103	0.459	0.103	5	-0.011	-0.011
颜色（色泽）	0.802	0.761	0.089	0.085	0.326	0.076	7	0.013	0.004
口感	0.744	0.805	0.138	0.152	0.461	0.106	4	0.032	-0.014
包装	0.739	0.784	0.083	0.072	0.228	0.048	11	0.035	0.011
物流	0.736	0.748	0.080	0.072	0.477	0.115	3	-0.035	0.008
服务	0.724	0.702	0.105	0.105	0.295	0.063	9	0.042	0.000
价格	0.826	0.787	0.107	0.102	0.274	0.053	10	0.054	0.005
综合质量	0.738	0.756	0.076	0.074	0.503	0.118	2	-0.042	0.002
店铺	0.776	0.798	0.099	0.104	0.318	0.067	8	0.032	-0.005

表 5.6 2023 年 1 月~2023 年 6 月

属性	$Perf_{S4}$	$Perf_{S5}$	$Perf_{S4-}$ <i>Norm</i>	$Perf_{S5-}$ <i>Norm</i>	Imp	$Imp-$ <i>Norm</i>	$Imp-$ <i>Ranking</i>	GAP	PD
新鲜度	0.735	0.742	0.103	0.113	0.336	0.082	6	0.021	-0.010
果面	0.607	0.669	0.088	0.098	0.506	0.147	1	-0.059	-0.010
果形及大小	0.687	0.782	0.099	0.103	0.349	0.104	5	-0.005	-0.004
颜色（色泽）	0.774	0.735	0.087	0.090	0.239	0.065	10	0.022	-0.003
口感	0.708	0.743	0.151	0.158	0.366	0.105	4	0.046	-0.007
包装	0.755	0.728	0.107	0.094	0.255	0.062	9	0.045	0.013
物流	0.689	0.712	0.082	0.077	0.421	0.120	2	-0.038	0.005
服务	0.715	0.696	0.120	0.118	0.275	0.073	8	0.047	0.002
价格	0.807	0.741	0.132	0.124	0.294	0.075	7	0.057	0.008
综合质量	0.741	0.772	0.084	0.083	0.387	0.114	3	-0.030	0.001
店铺	0.748	0.754	0.103	0.100	0.202	0.061	11	0.042	0.003

根据三组竞争力计算结果可以绘制出相应的 DIPCA 图，如下图 5.7 所示，图中有向箭头反映了时间推移顺序。从图 5.7 中，可以发现所有属性都在不断变化。图中显示目标商家 S4 水果产品的“果面”属性从 2022 年 1 月至 2023 年 6 月与竞争对手相比进一步处于不利地位，应迅速采取相应措施，防止局势进一步恶化；对于“价格”“服务”“店铺”属性相比于竞争对手逐渐获得了竞争优势，需要继续保持；对于“包装”属性，在 2022 年 1 月至 2023 年 6 月期间一直领先于竞争对手，也应继续保持；对于“新鲜度”“口感”属性在 2022 年 1 月至 2023 年 6 月期间逐渐缩小了与竞争对手的差距，但仍然落后于竞争对手，需要及时采取相应措施；对于“果形及大小”属性，虽然在不断提升，但仍然位于改进重点区，应重点关注，及时制定相应改进策略；对于“颜色（色泽）”“物流”“综合质量”属性，2022 年 1 月至 2022 年 12 月相比于竞争对手获得了领先优势，但随后又因竞争对手的改善削弱了目标商家 S4 的领先优势，这也证实了“不进则退”的商业规律。

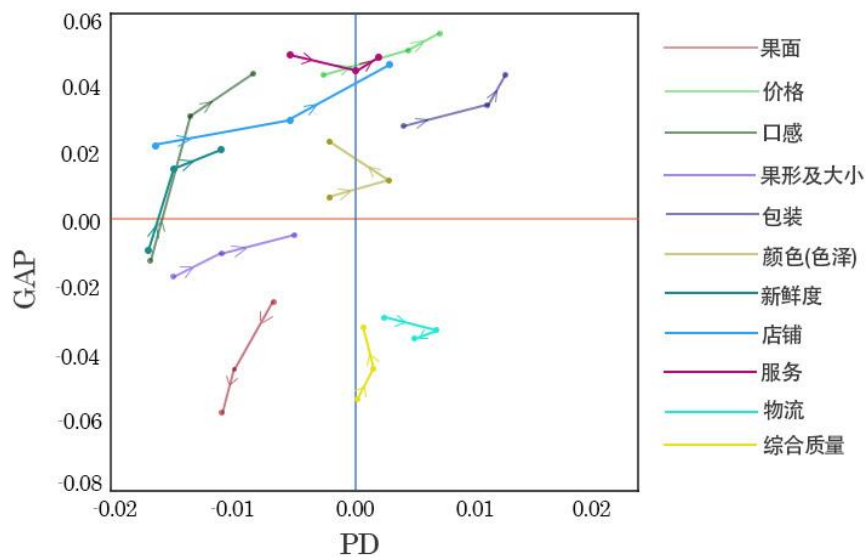


图 5.7 DIPCA 结果

5.5 电商助农水果产品改进策略形成

5.5.1 消费者不满意因素 TF-IDF 热词提取

根据 IPCA 和 DIPCA 分析结果显示，综合来看目标商家 S5 水果产品的“果面”属性不仅处于 Q3（改进重点）区，与竞争对手相比劣势局势还进一步恶化，因此需要优先考虑改进其水果产品的“果面”属性。由于 5.2 节属性级情感分析能提取出相应的属性词、观点词及情感倾向，因此更方便了各个属性与对应原始评论的划归。利用文本挖掘技术 TF-IDF 从“果面”的原始负面评论中挖掘导致消费者不满意的因素，如表 5.7 所示。其中 TF-IDF 值越大，热词排名越高。从表中可以发现，消费者对“果面”属性不满意的主要热词是“坏”“烂”“碰伤”等，这可能是由于水果属于易腐坏产品，果面若出现损伤会严重影响其整体口感，因此消费者对水果这类生鲜产品尤其在乎其果面是否完整。

表 5.7 “果面”属性消费者不满意热词提取（部分）

Rank	Word	TF-IDF Value	Rank	Word	TF-IDF Value	Rank	Word	TF-IDF Value
1	坏	0.245	9	坏果	0.099	17	摔坏	0.066
2	烂	0.237	10	挤压	0.099	18	伤疤	0.066
3	碰伤	0.234	11	烂坏	0.099	19	磕碰	0.047
4	差评	0.232	12	烂果	0.099	20	太差了	0.047
5	斑点	0.182	13	表皮	0.085	21	不会回购	0.047
6	烂果	0.151	14	虫眼	0.082	22	疤痕	0.034
7	破损	0.139	15	疤	0.082	23	腐烂	0.034
8	损伤	0.129	16	斑	0.082	24	摔坏	0.034

5.5.2 基于 K-M e a n s 聚类算法的不满意因素分析

为了了解消费者不满意因素热词的特征，在使用 TF-IDF 算法将热词进行矢量化后，接着利用 K-Means 聚类算法进行聚类分析。使用 K-Means 聚类时，可以根据“肘部法则”确定簇数。肘部法则的基本思想是通过绘制簇类平方和与聚类簇数的关系图找到一个拐点或肘部，拐点或肘部对应的聚类簇数即被认为是最佳的聚类簇数。如图 5.8 所示，该图在 K=4 时出现肘点，因此最佳聚类簇数为 4。

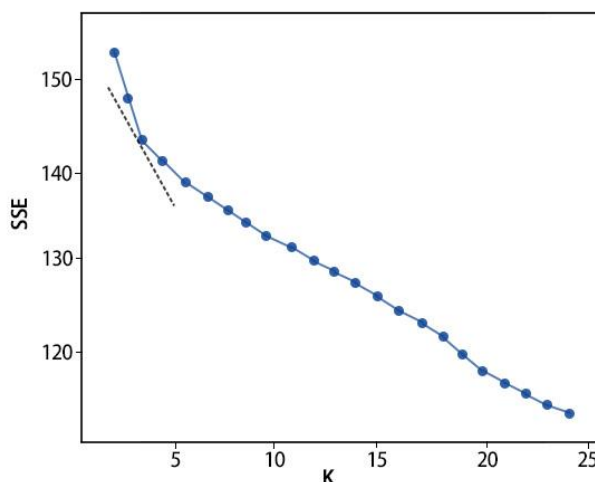


图 5.8 肘形图

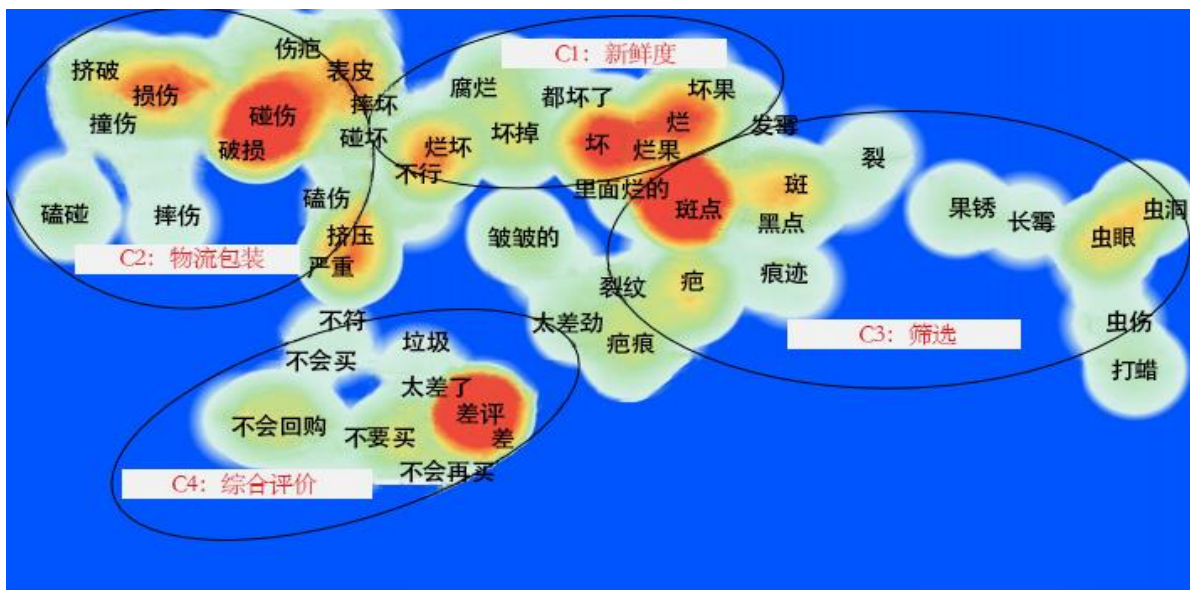


图 5.9 消费者不满意因素热图

进一步地，通过热图方式将消费者不满意因素聚类结果可视化，如图 5.9 所示，其中颜色越深表示对应热词的 TF-IDF 值越高，该词越重要。根据聚类结果，将消费者不满意因素划归为 4 个集群。

图 5.9 中的 Cluster-1，根据集群热词含义将其命名为“新鲜度”，包括“烂”“坏”等热词。根据这些消费者不满意因素，目标商家一方面应该重点排查供应端问题，确保水果是否是农户采摘的新鲜水果，是否为储存时间过长的水果，另一方面也需要考虑水果自身易腐不易储存的特性，可以通过提升冷藏保鲜和温控技术来解决采摘后的储存问题。除此之外，还需要选择物流配送效率高和保鲜冷藏技术有保障的第三方物流公司，保证流通配送环节中的水果产品的果面品质；

图 5.9 中的 Cluster-2，根据集群热词含义将其命名为“物流包装”，主要包括“碰伤”“损伤”等热词。该类热词说明目标商家在水果运输上出现了问题，可以通过改善内外包装、选择合适的包装材料来解决水果运输途中出现磕碰的问题，如考虑更换更为耐压防磕碰的包装、内包装使用加厚泡沫网套等，起到更多缓冲保护作用；

图 5.9 中的 Cluster-3，根据集群热词含义将其命名为“筛选”，包括“斑点”“虫眼”等热词。水果表皮上有斑点和虫眼等缺陷会严重影响部分消费者食用体验，目标商家需要注意发货时的水果挑拣筛选环节，确保果面干净完好，以满足消费者需求；

图 5.9 中的 Cluster-4，将其命名为“综合评价”，包括“差评”“不会回购”等热词。

对于这些评价，目标商家可以根据其他集群的具体原因对水果产品进行改进，以提高消费者满意度来提升整体综合评价水平。

5.5.3 运营策略改进方案形成

基于以上目标商家 S5 对于“电商助农水果产品运营策略改进方案”的应用与分析结果，提出具体的改进方案如下：

(1) 重点改进“果面”属性

依据 IPCA 分析结果显示，目标商家 S5 水果产品的“果面”属性，不论是与电商助农水果产品商家还是与一般电商水果产品商家竞争对手相比，都处于极端劣势，而且 DIPCA 分析结果更显示目标商家“果面”属性竞争力在不断下滑，因此“果面”属性为目标商家第一优先级改进的劣势属性，需要针对“果面”属性及时采取相应措施。基于 TF-IDF 热词提取和 K-Means 聚类分析结果，消费者对其不满意的主要原因在于水果出现果面“烂”“坏”等影响其新鲜度、因物流运输或包装不当导致水果表皮受损以及发货时对水果的筛选把控不严，可以考虑采用相应的改进方案：

第一，从源头保证水果新鲜度和质量达标，即严控水果采摘保鲜环节。采摘新鲜完整、符合标准的水果，采摘完成后到发货环节注意保鲜冷藏，发货前确保水果处于合适的温度、湿度和通风环境，以此延长水果保鲜期，减少目标商家水果出现烂、坏的现象；

第二，严控筛选环节。目标商家在此环节存在问题，需提高在发货前对水果挑选筛选的把控力度，避免给消费者的水果存在烂坏、表皮受损、斑点、虫眼等状况；

第三，改善包装。根据消费者不满意因素分析结果可以得出目标商家存在填充包装问题，为减少或避免在运输过程中水果相互碰撞和挤压，在内包装方面应选择更合适的耐压、耐摩擦的包装材料，也可以在包装盒中添加充分的填充材料，如加多气泡膜等，但也需避免过度包装；在外包装方面，用于物流运输的外包装盒上注明“生鲜水果”、“轻拿轻放”等标识，提醒货运人注意要货品的特殊性质，能够很大程度上减少或避免水果在运输途中出现损伤。

第四，选择可靠的物流公司。生鲜水果对时效和运输条件的要求都非常严格，如果物流运输条件不达标，会直接影响其果面品质。由于目标商家是由店铺发货、第三方物流提供配送服务，因此目标商家在选择第三方物流时，需要重点考虑其快递运输的时效、保鲜技术等方面能力水平，选择合适的水果冷链物流公司。对于京东物流，其提供一体化的智能供应链服务，具有高效优质的流通配送网络和货到付款、夜间配送等特色配送

服务，其京东冷链专注于生鲜食品等物流，在设备、设施、人员等方面具有一定技术优势，可以提供更有保障和专业化的物流配送，目标商家可以考虑与京东物流合作，以更好地保证产品的果面品质。

（2）根据属性改进收益合理分配资源

结合 IPCA 和 DIPCA 分析结果，目标商家 S5 的“果形及大小”属性虽然在不断提升但仍然位于重点改进区，要提高消费者满意度、促进自身的市场发展，在优先改进“果面”属性后，还需要合理分配资源投入“果形及大小”属性的改进中。另外，从长远来看，要提升目标商家的市场竞争优势，在“果面”、“果形及大小”属性得到改善后，“口感”“物流”以及“综合质量”属性也需要投入适当资源，有助于全面提升消费者满意度，满足消费者多元化需求。

6 总结、建议与展望

6.1 总结

在促进助农产品销售和提高助农产品市场竞争力目标下，本文以电商助农水果产品为例，设计了电商助农水果产品运营策略改进方案，并以消费者需求为基础对改进策略形成依据进行可视化展示，为改善电商助农水果产品目标商家当前问题提供具体的决策支持，也为解决当前电商助农产品市场化程度低、消费者满意度低问题提供参考。在电商助农水果产品在线评论数据的采集与预处理阶段，本文选取京东平台助农馆频道中的生鲜水果为研究对象，运用评论去重、文本去噪、短文本删除、中文分词与去除停用词等自然语言处理技术对非结构化的消费者在线评论数据进行文本规范化处理。在确定水果产品评价维度即产品属性阶段，本文结合与生鲜水果相关的《中华人民共和国国家标准 GB》内容以及水果产品特点和消费者在线评论中提到的不同方面，提取出了水果产品的综合属性及其特征词，具体包括了“果面”“果形及大小”“颜色”“口感”“包装”“物流”“服务”“价格”“综合质量”“店铺”11个属性以及448个属性特征词。之后，在属性级情感分析阶段，基于 PaddleNLP 搭建 SKEP 模型对预处理后的在线评论数据集进行属性级情感分析，提取出相关的属性特征词、观点词及情感极性，并采用基于情感词典的情感分析方法计算出水果产品各个属性的情感得分。在竞争力分析阶段，基于属性级情感分析结果计算出每个商家各项属性的绩效表现和重要性，了解电商助农水果产品的整体市场表现状况和消费者满意度情况，再以助农商家 S5 为目标研究对象，运用竞争者识别模型识别电商助农水果产品市场和一般电商水果产品市场中竞争者，通过本文构建的 IPCA 模型和 DIPCA 模型分析得出目标商家水果产品的优劣属性。最后，针对目标商家第一优先级改进劣势属性，本文通过文本挖掘技术 TF-IDF 对此属性的原始负面评论提取消费者不满意因素的热词，通过 K-Means 聚类算法对消费者不满意因素的热词进行聚类分析，最终以热图形式将改进方案形成依据进行可视化展示，提出具体的产品运营策略改进方案。

通过对全文工作和结果的总结，可以得出以下结论：

(1) 电商助农水果产品各项属性绩效表现都明显低于一般电商水果产品，表明了电商助农水果产品的消费者满意度较低且市场表现相对较差，在电商水果市场上的竞争力较弱。电商助农水果产品商家要提升消费者满意度并促进其市场发展，需要识别和明

确其水果产品的劣势属性，找出导致消费者不满意的原因，并针对性地采取相应的改进策略。

(2) 本文随机选择以中国特产·庆阳合水助农馆(S5)作为电商助农水果产品目标研究商家，展开应用于构建的电商助农水果产品运营策略改进方案，得出“果面”属性是优先重点改进的劣势属性，消费者对其不满意的主要原因在于水果出现果面“烂”“坏”等影响其新鲜度、因物流运输或包装不当导致水果表皮受损以及发货时对水果的筛选把控不严，可以通过采取严控水果采摘保鲜环节、严控筛选环节、改善包装以及选择可靠的物流公司等策略提升“果面”属性，进而提高消费者满意度，同时还需要根据属性改进收益合理分配资源，以促进市场发展并提高市场竞争力。

(3) 在研究分析过程中，属性重要性的动态计算结果也揭示了消费者的偏好和需求可能会随着时间的推移而变化，电商助农产品商家都应及时关注消费者的偏好和需求，并适时调整运营策略和资源配置。进一步的从动态竞争力分析结果发现，消费者满意度越高的产品属性，其竞争力越强，且竞争对手对产品属性的改进可能会削弱自身产品的竞争力，因此电商助农产品商家还应不断完善自己的产品，避免陷入“不进则退”的陷阱。

6.2 建议

6.2.1 对消费者的建议

基于本文研究，结合电商领域实际情况，对消费者提出以下建议：

(1) 消费者的评价对于商家来说至关重要，不仅会直接影响到商家销量、信誉等，还会对商家改进产品或服务、提升竞争力起到关键作用^[51]。因此，消费者在购买水果产品时可以积极对商品进行客观真实的评论，为其他消费者提供宝贵的购买意见和参考，这也是为商家改进产品和服务提供的宝贵反馈。

(2) 京东平台在生鲜领域，尤其是生鲜水果产品，在新鲜度、口感、质量以及物流速度等方面都得到了广大消费者的认可，在生鲜电商平台众多选择中，京东是一个不错的选项。另外，京东电商平台的助农模式不仅吸引了大量农户合作，包括京东百亿补贴、京东零售“乡村振兴·京东千县名品”、京东助农馆开设等方式来保障农产品的销售，其京东云助农“321模式”的推广应用更是逐步实现了把农产业高利润环节和利益

留在乡村，为丰盈农产业生态、实现乡村“自我造血”以破除传统电商助农困境提供了引导路径，对于支持助农产品的消费者来说，可以考虑在京东电商平台助农频道助力消费助农。

(3) 对于水果这类生鲜类产品，由于其保质期短，京东平台店铺大多不支持 7 天退换服务，并且大多数店铺也说明不会因消费者以快递慢、不新鲜、不想要类似理由单方面拒收而承担退换等责任，因此消费者在京东平台下单购买水果这类生鲜产品时，应注意发货货源地信息，对物流运输时间有较大影响。在收到产品后需要及时开箱检查，如有问题可以及时联系商家客服或平台提出理赔等诉求。

6.2.2 对电商助农水果产品商家的建议

基于本文研究，对电商助农水果产品商家提出以下建议：

(1) 严格控制产品质量，保证产品的新鲜完整度符合消费者期望

从本文的分析中可以得到消费者对水果产品的新鲜度、果面、口感等关注度较高，但一些水果到消费者手中却出现烂坏、腐烂、损伤等状况，严重影响了商家水果产品的竞争力。因此，商家对水果的新鲜完整问题应给予更多重视，完善水果产品的品控体系，确保农户提供的水果是新鲜完整、标准合格的；在水果发货前，对水果再次进行质检，将已经出现腐烂、损坏或表面有较大瑕疵的产品进行替换，完善供应链体系，以确保消费者收到新鲜完整、符合期望的水果。

(2) 完善自身服务

本文分析中主要是因水果产品本身问题使得消费者产生消极情感态度，商家应对于消费者收到损坏的水果产品给予一定合理的补偿，保证售后服务的质量，从而弥补消费者心理期望差距，增强消费者的购买信心和满意度。

(3) 合理分配资源投入，以实现收益最大化

商家在制定产品运营策略改进方案时，首先需要制定明确的目标，即确定产品属性改进优先级，权衡资源投入和收益大小之间的关系，优先满足最重要的需求，以确保最重要的利益得到保障，在满足第一优先级需求后根据实际情况灵活调整资源配置，适应不同的需求和环境变化，以实现收益最大化。

(4) 紧跟消费需求，把握消费趋势

从本文分析结果中可以发现，消费者满意度越高的水果产品属性其竞争力越强，与产品带来的收益成正比，这也进一步证实了市场是由需求驱动的，消费者的需求决定了

商家产品的销量和市场表现。因此，在制定产品运营策略和改进方案时，商家都需要以消费者需求为出发点和落脚点。而对于电商商家来说，获取消费者需求、把握消费趋势最有效的途径便是在线评论信息，这是消费者对产品和服务的意见和建议，对其进行深入分析和挖掘能得到消费者的直观感受和潜在需求，据此制定的产品运营策略和改进方案能更有力地提升产品质量和服务，提高消费者满意度，实现“产品对接需求”、“商家对接消费者”、“销售对接市场”的有利局面，达到提升市场竞争力并扩大销售的目标。

6.2.3 对平台的建议

（1）优化物流体系建设

京东电商平台的快递物流服务受到了广大消费者的认可，但由于水果这类生鲜农产品具有易腐变质、不易运输的特征，使得冷链物流体系的建设亟待进一步加强。目前京东生鲜建有自身的物流体系，并且具备标准化、智能化特点，这为解决水果产品保鲜问题提供了解决方案。第一，推进更为高效的仓储系统的建立和完善，根据助农水果产品特性，仓库选址需靠近消费者集中地、交通网络便利的区点建立仓库，合理规划仓储和运输是保证物流效率和质量的关键因素，为了满足用户的需求和生鲜产品的特性，应该根据客户订单制定科学的库存管理和物流配送计划；第二，完善配送网络，利用大数据、人工智能等技术优化配送路线和节点，结合物流智能化系统规划最优配送路线，提高配送效率；第三，加强物流信息共享，与商家、客户共享数据，建立完善的物流跟踪系统让客户实时了解订单状态和物流流程，并鼓励客户对物流服务进行评价和反馈，及时修正问题改进服务，以提高客户满意度；第四，提高保鲜温控水平，在生鲜的“冷”处理上，使用更有效的保鲜保温材料，以保障生鲜产品品质。

（2）完善包装管理

包装问题是消费者产生负面情感态度的重要原因之一，京东平台应当重视水果产品的包装管理。由于水果种类众多且产地不同导致运输距离不同，所以在包装上设计上可以差异化，此外通过技术手段控制包装内温度，保证水果产品的新鲜度。另外，还需采用有效防磕碰的包装材料，保证水果产品完整度。

（3）健全售后服务体系

一个高效的售后服务体系不仅能提升客户的购物体验，还能增强客户的忠诚度，促进平台回购率上升。京东平台应加强售后服务体系的建设，积极开展客服服务团队质检、专业培训等各项工作，明确客服人员的职责范围和工作流程，对客户需求和疑问及时处

理并解答，确保售后服务的高效运作。由于水果农产品的自身特性，在物流运输环节或包装问题可能会造成各种损失，消费者收到损坏的产品必然会找到售后客服提出退货、补偿等诉求，这就需要平台建立多渠道的售后服务渠道保障诉求畅通，且客服的有效处理也起到关键作用。平台需要建立高效的售后服务流程，备好处理预案，从而及时有效地解决消费者诉求，减少纠纷，增强消费者的信任度。在解决消费者诉求之后，平台还需重视售后服务的数据分析，如退货率、投诉次数等指标，对处理结果及时追踪跟进，发现遗留问题后及时采取措施改进售后服务体验。平台要想留住客户，提高市场竞争力，就要维护好与消费者之间的关系，重视客户服务，提高服务质量。

6.3 展望

在本文的主要研究工作中，提出了挖掘消费者在线评论来评估和提高电商助农产品的竞争力，根据竞争力分析得出目标产品的比较优劣属性，这些在线评论是否具有足够的代表性，可能存在争议。因此，本文提出这样代表性的讨论：据“电数宝”电商大数据库显示，2023年上半年我国网络购物用户规模已达8.84亿人，预计到2024年将超过9亿人，超过60%的网购用户会对购买的商品发表评论。与此同时，许多消费者在做出购买决定之前会参考产品评论。鉴于在线评论的影响，各个领域的研究人员已经开始从消费者在线评论这个数据源中探索商业机会，一些学者还利用在线评论数据来研究产品竞争力，如餐厅竞争力分析和酒店竞争力分析。

此外，本文选择水果作为研究电商助农产品，这也因为水果是一种高附加值的农产品，具有很高的经济收益，对资源投入和环境要求相对较低，从可持续性的角度来说，考虑到市场需求和销售情况，农民大多会选择种植水果作为经济收入的目标来源。当然，水果也是电商助农的主要代表产品之一，电商平台上对于生鲜水果也提供了大量的在线评论数据，这有助于展开属性级情感分析、竞争力分析、消费者不满意因素分析等相关研究。因此选择水果作为电商助农产品的具体研究对象，对于扩展其他电商助农产品的研究具有很好的代表性。

对于本文的研究，也存在一些不足，可以在未来不断改进。一方面，本文使用的是在线评论数据，属于文本数据，而消费者评论也包括了图片和视频，图片和视频也携带了更多的消费者感知和需求，对于图片和视频的数据挖掘可能有助于制定更实用的产品运营策略改进方案。另一方面，本文分析的数据来源于京东平台，还需要进一步将提出

的电商助农产品运营策略改进方案扩展到淘宝、拼多多等其他电子商务平台，以期在未来获得更为广义的结果。

参考文献

- [1] Abbasi A, Chen H, Salem A. Sentiment analysis in multiple languages: feature selection for opinion classification in Web forums[C]// ACM Transaction on Information Systems, 2008.
- [2] Agarwal B, Mittal N. Prominent feature extraction for sentiment analysis[J]. 2015.
- [3] Bi J-W, Liu Y, Fan Z-P, & Zahng J. Wisdom of crowds : Conducting importance-performance analysis (IPA) through online reviews[J]. Tourism Management, 2018(09):460-478.
- [4] Crocolici, M.F, & Nijkamp P. The attractiveness and competitiveness of tourist destinations: A study of Southern Italian regions[J]. Tourism Management, 2009, 30(03):336-344.
- [5] Dube L, Menon K. Multiple roles of consumption emotions in post-purchase satisfaction with extend service transactions[J]. International Journal of Service Industry Management, 2000, 11(3):287-304.
- [6] Denil M, Demiraj A, Kalchbrenner N, et al. Visualising and Summarising Documents with A Single Convolutional Neural Network[J]. Neural Computing & Applications, 2014(10):14-20.
- [7] Feng Y, Wang W.X, Dai W.Q. Topic analysis of online reviews for two competitive products using latent dirichlet allocation[J]. Electronic Commerce Research and Applications, 2018, 29:142-156.
- [8] Gao S, Tang O, Wang H, & Yin P. Identifying competitors through comparative relation mining of online reviews in the restaurant industry[J]. International Journal of Hospitality Management, 2017(71):19-32.
- [9] Holbrook M.B, Gardner M.P. Illustrating a Dynamic Model of the Mood-updating Process in Consumer Behavior[J]. Psychology and Marketing, 2000, 17(3):165-194.
- [10] Hemphill T.A. Antitrust, dynamic competition, and business ethics[J]. Journal of Business Ethics, 2004, 50(02):127-135.
- [11] Kim K, Park O, Yun S, & Yun H. What makes tourists feel negatively about tourism destination? Application of hybrid text mining methodology to smart destination

- management[J].*Technological Forecasting and Social Change*, 2017,123:362-369.
- [12]Lizard C.E. *Human emotions*[M].New York:Plenum,1977.
- [13]Liu Y, Jiang C, &Zhao H. Assessing product competitive advantages from the perspective of customers by mining user-generated content on social media[J].*Decision Support Systems*,2019,123:113079.
- [14]Luo Z.Y, Huang S.S, Zhu K.Q. Knowledge empowered prominent aspect extraction from product reviews[J].*Information Processing & Management*,2019,56(3):408-423.
- [15]Lin J, Li L, &Luo X.R. How do agribusinesses thrive through complexity?The pivotal role of e-commerce capability and business agility[J].*Decision Support Systems*,2020,135:113-124.
- [16]Liu Z, Qin C.X, &Zhang Y.J. Mining product competitiveness by fusing multisource online information[J].*Decision Support Systems*,2020,143:113477.
- [17]Mehrabian A, Russell J.A. *An Approach to Environmental Psychology*[M].Cambridge, MA:MIT Press,1974.
- [18]Martilla J.A, &James J.C. Importance-performance analysis[J].*Journal of Marketing*,1977,41(1):77-79.
- [19]Pang B, Lee L, &Vaithyanathan S. Thumbs up: Sentiment classification using machine learning techniques[R].*Empirical Methods Natural Language Processa*,Stroudsburg,PA,USA,2002.
- [20]Pang B,Lee L. A sentiment education: sentiment analysis using subjectivity summarization based on minimum cuts[C].*ACL*,2004.
- [21]Phadermrod B, Crowder R.M, Wills G.B. Importance-performance analysis based SWOT analysis[J].*International Journal of Informatuon Management*,2019,44:194-203.
- [22]Riloff E.M, Shepherd J. A corpus-based approach for building semantic Lexicons[J].1997.
- [23]Socher J, Pennington E.H,Huang A.Y.Ng, &C.D.Manning. Semi-supervised re-cursive autoencoders for predicting sentiment distributions[C]// In Proceedings of the Conference on Empirical Methods in Natural Language Processing,EMNLP'11,2011:151-161.
- [24]Whitelaw C,Garg N,Argamon S. Using appraisal groups for sentiment analysis[C]// *Acm*

- International Conference on Information & Knowledge Management,ACM,2005:625.
- [25] Wilson T,Wiebe J,&Hoffmann P. Recognizing contextual polarity in phrase-level sentiment analysis[C].HLT-EMNLP, 2005.
- [26] Wang H.W, Wang W. Product weakness finder: An opinion are system through sentiment analysis[J].Industrial Management & Data Systems,2014,114(08):1301-1320.
- [27] 宋瑛,蔡芳芳,张驰.乡村振兴战略背景下农户参与农产品电商的满意度研究——基于目的和过程双重视角[J].重庆社会科学,2021(08):104-119.
- [28] 王瑞峰.电商赋能农民共同富裕的演进逻辑、理论动因与实现机制[J].当代经济管理, 2023,45(07):8-17.
- [29] 王翠翠,夏春萍,蔡轶.农业电商扶贫可以提升农户的可持续生计吗? ——基于农产品上行视角[J].浙江农业学报,2022,34(03):636-651.
- [30] 陈光.淘宝“村播计划”如何助力乡村经济振兴[J].传媒,2022(01):61-64.
- [31] 郭红东,曲江.直播带货助农的可持续发展研究[J].人民论坛,2022(20):74-76.
- [32] 孙文婷,彭红军.电商助农背景下农产品生产与销售策略研究[J].中国管理科学,2022,7(20):13.
- [33] 胡珺祎,肖煜珠,覃丹妮,刘春辰,姚佳琪.微商如何助小农? ——基于砀山县的案例研究[J].中国农业大学学报(社会科学版),2022,39(05):93-113.
- [34] 王瑞峰.相对贫困视阈下农村电商助农增收的中介效应研究[J].湖南师范大学社会科学学报,2022,51(02):55-66.
- [35] 史振杰.基于深度学习的电商评论情感分析[D].河北经贸大学,2020.
- [36] 李寿山,李逸薇,黄居仁.基于双语信息和标签传播算法的中文情感词典构建方法[J].中文信息学报,2013,27(06):75-81.
- [37] 阳爱民,林江豪,周咏梅.中文文本情感词典构建方法[J].计算机科学与探索, 2013,7(11):1033-1039.
- [38] 王志涛,於志文,郭赋.基于词典和规则集中的中文微博情感分析[J].计算机工程与应用,2015,51(08):218-225.
- [39] 张克亮,黄金柱,曹蓉.基于 HNC 语境框架和情感词典的文本情感倾向分析[J].山东大学学报(理学版),2016,51(07):51-58.
- [40] 唐晓波,严承希.基于旋进原则和支持向量机的文本情感分析[J].情报理论与实践, 2013,36(01):98-103.

- [41]韩开旭.基于支持向量机的文本情感分析[D].东北石油大学,2014.
- [42]朱少杰.基于深度学习的文本情感分类[D].哈尔滨工业大学,2014.
- [43]梁军,柴玉梅,原慧斌.基于极性转移和 LSTM 递归网络的的情感分析[J].中文信息学报,2015,29(05):152-159.
- [44]谢铁,郑啸,张雷.基于并行化递归神经网络的中文短文本情感分类[J].计算机应用与软件,2017,34(03):205-211.
- [45]韩广华.企业产品市场竞争力研究[J].生产力研究,2007(17):111-112.
- [46]田金霓,尤天慧,袁媛.基于在线评论的产品竞争力分析方法[J].东北大学学报(自然科学版),2021,42(10):1498-1505.
- [47]王忠群,王族,钱寅亮,陈云霞.基于超网络和在线评论的产品竞争力分析[J].情报理论与实践,2022,45(04):113-119.
- [48]齐备.电商助农视角下基层政府职能研究[D].华北水利水电大学, 2021
- [49]邵研,阮平南.产品竞争力分析及营销预警[J].北京工业大学学报,2000(51):102-106
- [50]孟昭兰.情绪心理学[M].北京:北京大学出版社,2005:35-47.
- [51]朱丽叶,袁登,张静宜.在线用户评论质量与评论者等级对消费者购买意愿的影响——产品卷入度的调节作用[J].管理评论,2017,29(02):87-96.

攻读硕士学位期间承担的科研任务及主要成果

在学期间发表的学术论文

- [1] 余佳豪. 农村电商流通效率优化赋能甘肃乡村振兴——基于 DEA-Multiple Regression 模型[J]. 物流科技, 2023, 46(03):121-125.
- [2] 余佳豪. 农村电商发展对数字乡村建设的影响效应[J]. 河北农业大学学报(社会科学版), 2023, 25(04):118-126.

在学期间参与的科研项目

项目名称：破解深度贫困地区乡村振兴难题的长效机制研究

致谢

回首过往，点滴三年，兰州财经大学已经慢慢陪我走过了人生中意气风发、挥斥方遒的硕士时代。感谢这三年中与我相伴的同学，帮助过我的老师以及朋友，你们是我前进路上的动力，最好的提携者与引路人。每每回顾，不能自己。学校的每一片土地，每一棵花木，均见证了我的成长，心怀感恩，我将勇往直前。

首先，感谢我的导师牛胜强教授，牛老师根据我的专业背景，鼓励我以自身兴趣及专长为出发点，结合统计学知识技能，确立研究方向。在我的学术论文及毕业论文设计写作中，牛老师严格要求又耐心的指导，遇到问题时与牛老师讨论总能得到启发，让我开拓了思路，受益匪浅。在撰写论文过程中，牛老师给出的修改意见，我不仅仅学习到了论文的写作技巧，更得到了逻辑思维的锻炼。牛老师的言传身教，让我懂得了读研究生不仅仅只是学习知识，而是学习研究方法，学习解决问题的能力，更是对自身意志力的锻炼。相信在之后的学习路途上，不管遇到任何困难，我一定牢记教诲，努力做出一番成绩。感谢统计与数据科学学院的老师们，为我打开了一扇统计学的大门，你们教授的数据挖掘、非参数统计、多元统计分析等课程，让我学习到了相关的理论知识和前沿统计分析技术，收获了毕业设计的灵感，为学位论文的开展奠定了坚实的基础。毕业在即，由衷的对老师们表达深深的感谢，因为有你们，才让我这三年的时光更加充实向前。

同时感谢相互交流写作经验的师姐、同门和师弟师妹们，你们的建议让我获益不少。感谢朝夕相处三年的宿舍兄弟们，给予我生活和学习上的帮助。感谢给予我技术和精神上支持的路林，为我驱散了背井离乡的孤独和不安。最后，特别感谢我最爱的爸爸妈妈，感谢你们的养育之恩、你们的理解和包容，在我迷茫和面对挫折时，你们始终都是我心灵的港湾，也是我最坚强的后盾，祝福你们幸福安康！

未来可期，感恩兰州。