

学校代码 10125

专业代码 120801



山西财经大学
SHANXI UNIVERSITY OF FINANCE AND ECONOMICS

本科毕业论文（设计）

中文题目 基于 NLP 的天猫用户评论情感分析及
其在商品推荐中的应用研究

英文题目 *Sentiment analysis of Tmall user reviews
based on NLP and its application in
product recommendation*

姓名 张宇萌

学号 202107030149

班级 电子商务班

专业 电子商务

学院 信息学院

指导教师 张瑞 讲师

完成时间 2025 年 4 月 25 日

摘 要

在激烈的电商平台竞争环境中，用户评论在影响消费者决策和商品声誉方面具有至关重要的作用。目前传统的基于热度或销量的推荐机制往往过于重视热度，而忽视评论蕴含的情感信息，导致部分口碑欠佳的商品因前期营销手段积累的高热度频繁出现在推荐列表中，潜在优质但曝光度较低的新品则被埋没。为解决这一行业痛点，本研究的核心目标是在推荐流程中融入用户评论的情感倾向，以提升推荐质量和用户满意度。

本研究以 5000 条天猫评论数据为研究对象，通过情感分析（TF-IDF+SVM）量化用户情感为好评率差评率，并基于这些情感度量，设计并实现了多种侧重不同口碑目标的推荐策略，并将这些情感度量融入推荐策略中。最后对这些推荐策略进行实证分析。经过实证分析，本研究发现与传统推荐策略相比，情感驱动的推荐策略显著提高了推荐结果的整体口碑水平。本研究也深入探讨了不同策略的优劣势及适用场景，力求为不同电商平台的不同运营目标提供灵活的推荐方案。本研究证实，将用户情感分析纳入推荐系统，并根据经营战略选择适合自己的推荐系统，是改善电商平台推荐效果、增强用户信任的有效途径。

关键词: 用户评论；电子商务；商品推荐

Abstract

In the highly competitive e-commerce platform environment, user reviews play a crucial role in influencing consumer decisions and product reputation. Currently, traditional recommendation mechanisms based on popularity or sales volume often overemphasize popularity while neglecting the emotional information contained in reviews. This leads to subpar products with high popularity due to early marketing efforts frequently appearing in recommendation lists, while potentially high-quality but less exposed new products are buried. To address this industry pain point, the core objective of this study is to integrate the emotional tendencies of user reviews into the recommendation process to enhance recommendation quality and user satisfaction.

This study takes 5,000 Tmall review data as the research object and quantifies user sentiment into positive and negative review rates through sentiment analysis (TF-IDF SVM). Based on these sentiment measurements, various recommendation strategies focused on different word-of-mouth goals are designed and implemented, incorporating these sentiment metrics into the recommendation strategies. Finally, an empirical analysis of these recommendation strategies is conducted. The empirical analysis found that compared to traditional recommendation strategies, sentiment-driven recommendation strategies significantly improve the overall word-of-mouth level of the recommendation results. This study also explores the strengths and weaknesses of different strategies and their applicable scenarios in depth, aiming to provide flexible recommendation solutions for different operational goals of various e-commerce platforms. This study confirms that incorporating user sentiment analysis into recommendation systems and selecting suitable recommendation systems based on business strategies is an effective way to improve the recommendation effectiveness of e-commerce platforms and enhance user trust.

Key words: User Reviews; E-commerce; Product Recommendations

目 录

1 导 论..... 1

1.1 选题背景与意义..... 1

1.2 文献综述..... 2

1.3 研究目标与内容..... 4

1.4 论文的结构及主要内容..... 6

2 商品推荐系统概述..... 8

2.1 主流推荐算法介绍..... 8

2.2 主流推荐算法局限性..... 8

2.3 情感信息在推荐系统中的应用与价值..... 9

3 数据与方法..... 10

3.1 数据来源与基本情况..... 10

3.2 商品情感画像构建..... 12

4 情感驱动的推荐策略设计与实证研究..... 18

4.1 情感驱动推荐策略设计..... 18

4.2 实证分析..... 19

4.3 三种情感驱动策略比较与选用场景分析..... 22

5 结论与展望..... 24

5.1 研究结论..... 24

5.2 管理建议..... 24

5.3 研究局限与未来工作..... 25

参考文献..... 26

致 谢..... 28

1 导 论

1.1 选题背景与意义

电子商务已成为现代生活的重要组成部分，其规模随着科技的发展持续扩张。2024 年中国电子商务交易额达到 155,255.0 亿元，较 2023 年的 154,264 亿元同比增长 7.2%，展现出稳健增长态势（如图 1-1）¹。在此背景下，以天猫为典型代表的头部电商平台凭借强大的资源整合能力，汇聚了海量商品与庞大用户群体，使得电商市场竞争态势愈发白热化。消费者面对如此信息爆炸式增长，对用户生成内容（UGC）的依赖程度与日俱增。于电商平台而言，用户评论不仅在很大程度上左右潜在消费者的购买意愿，其背后蕴藏的口碑效应，更会对商品转化率、用户黏性以及品牌形象产生深远影响^[3]。因此，有效解读和利用用户评论已成为电商平台及商家提升核心竞争力的战略要务。

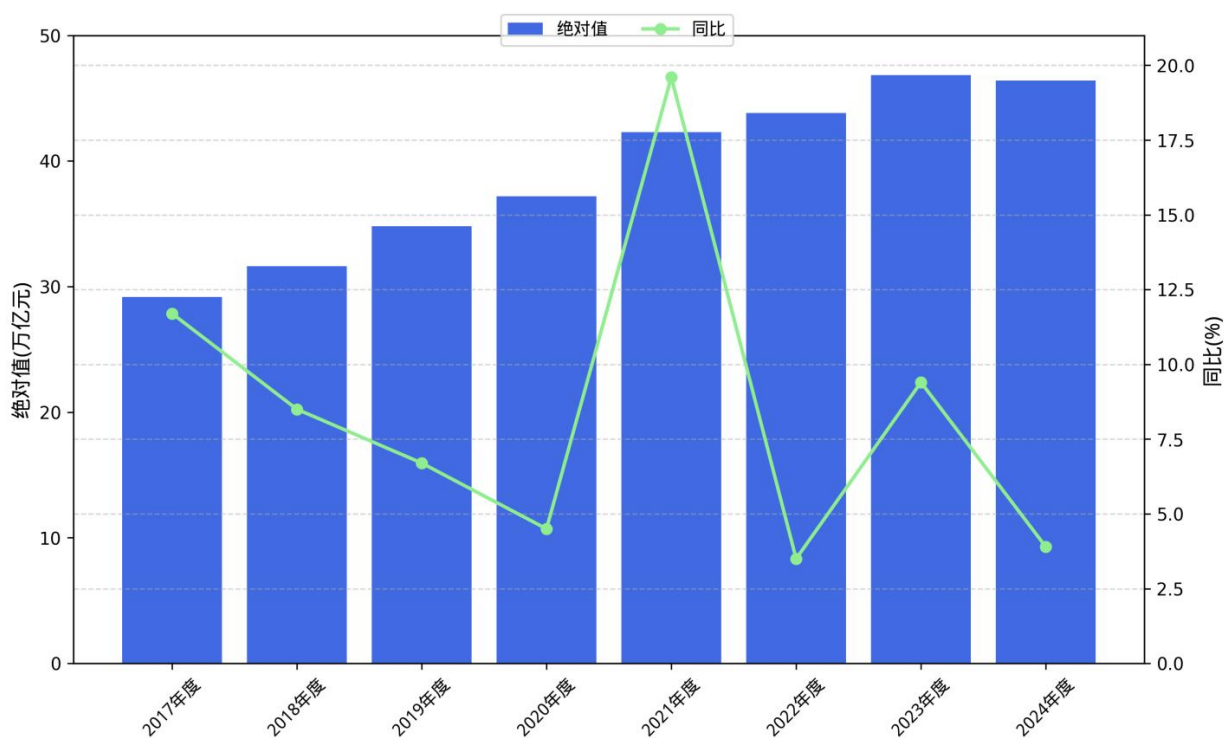


图 1-1 2017-2024 年电商交易量及增长率

然而，当前主流的推荐系统主要依赖用户的历史交互行为进行偏好预测，却普遍忽视了用户评论文本中蕴含的宝贵情感信息^[5]。某些商品可能因早期营销积累了高销量，但在用户评论中却充斥着对质量、服务或描述不符的负面评价，但传统推荐系统无法有效识别这些负面信号，仍会将其推荐给用户，极大地损害了用户体验和满意度。更严重的是，这可能引发平台内“劣币驱逐良币”的恶性循环——口碑优良但曝光不足的商品被埋没，低质商品反而占据推荐位，长此以往将破坏平台的健康生态与用户信任根基^[11]。

¹ 数据及图片来源:全国电子商务公共服务网

鉴于此，深入挖掘用户评论文本，运用情感分析（Sentiment Analysis）技术来精准识别并量化用户的情感倾向显得尤为迫切，且具有重要的现实意义。通过对海量评论的情感剖析，我们不仅能洞察消费者的真实感受与满意度，为消费者提供更全面的商品信息以辅助决策，帮助消费者做出更明智的购买选择。同时也为电商平台优化推荐策略，为商家改进产品与服务提供直接洞察。本研究正是基于这一认识，致力于探索如何将 NLP 情感分析技术有效应用于天猫用户评论数据，并将情感量化结果融入商品推荐框架，旨在显著提升推荐质量与用户满意度。

1.2 文献综述

1.2.1 情感分析研究现状

情感分析作为自然语言处理领域的一个重要分支，近年来受到了国内外学术界和工业界的广泛关注，围绕着情感分析的方法、应用场景和评价指标等方面，涌现了大量的研究成果。

1.2.1.1 国外研究现状

国外在情感分析领域的研究起步较早，积累了丰富的理论基础和实践经验。早期的研究主要集中在基于情感词典的方法上，Turney（2002）利用情感词典对电影评论进行情感分类，取得了较好的效果^[25]。随着机器学习技术的发展，基于机器学习的情感分析方法逐渐成为主流。Pang 等人（2002）使用朴素贝叶斯、支持向量机等算法对电影评论进行情感分类，验证了机器学习方法在情感分析中的有效性^[21]。

近年来，情感分析领域的研究现状呈现出模型与方法持续创新和应用场景深度拓展两大核心趋势。

在模型与方法创新层面，研究者们致力于多模态与深度学习的深度集成。例如，Singh et al. (2025) 提出的 TextAI 3.0 框架，通过结合多模态数据、注意力机制及超球空间中的深度学习，显著提升了情感分析的性能^[24]。同时，Miriya et al. (2025) 的研究表明，利用思维链提示等策略能够大幅提高 LLMs 在复杂情感分类任务上的准确率^[20]。

在应用拓展层面，情感分析技术已从传统的社交媒体舆情监控和电商评论分析，广泛渗透到更多元、更专业的领域。在金融领域，情感驱动的预测模型日益受到重视，Batoool et al. (2025) 的工作揭示了新闻情感在市场波动预测中的重要性^[16]。此外，情感分析还在文化研究（Peak & Lee, 2025）、社会工作以及人机交互等多个交叉领域展现出其独特的价值和潜力^[22]。

1.2.1.2 国内研究现状

国内情感分析研究起步稍晚，但发展迅速，在方法和应用方面都取得了显著进展。在研究方法方面，国内学者在借鉴国外研究成果的基础上，结合中文语言的特点，提出了一些改进的方法。针对中文分词的难点，研究者提出了多种改进的分词算法，提高了分词的准确率（袁里驰，2023）^[12]。针对中文情感表达的复杂性，研究者构建了一些专门针对中文的情感词典，伍蕾等人（2024）提出较好地反映数字沉浸式场景特色词汇的词典，能够明晰数字沉浸式场景下旅游者情感体验特征^[9]。

此外，国内学者也积极探索深度学习技术在情感分析中的应用，并取得了一些创新性的成果。例如，黄丽凤等人（2025）提出了一种基于循环神经网络的情感分类模型，能够有效地处理长文本的情感分析问题^[2]。郑志建等（2024）提出的 LSTM 模型在用户情感分类中达到 89.6% 的准确率，验证了时序建模的有效性^[14]。

在情感分析的应用方面，国内研究者将其广泛应用于电商评论分析、金融风险评估、在线教育等领域。例如，崔婷等人（2024）利用情感分析技术对金融新闻进行分析，预测股票市场的走势^[1]。王新宇（2024）基于游客评论情感分析构建的景区满意度评估体系，为管理决策提供量化支撑^[8]。

1.2.2 情感分析技术在电商的应用

本部分旨在梳理国内外情感分析技术在电商的应用研究现状，为本研究的开展提供理论基础与参考。

1.2.2.1 国外推荐系统研究现状

国外在此领域的研究持续引领创新，尤其在融合先进 NLP 技术以提升情感感知推荐系统的性能和可解释性方面取得了显著进展。

早期研究已认识到用户评论中情感信息的价值，传统推荐系统主要依赖用户的显式评分，如评星级。或隐式行为，如点击、购买。但这些数据往往难以全面反映用户的真实偏好和对物品具体属性的看法。Agarwal et al. (2015) 尝试将用户情感信息其作为用户评分的补充或一种隐式反馈，以缓解数据稀疏性或修正评分偏差，并将其应用到推特数据中进行测试，获得良好成效^[15]。McAuley & Leskovec (2013) 等工作进一步揭示了评论文本在解释评分、理解用户深层偏好方面的潜力。他们提出一种结合评分数据与用户评论文本的推荐模型，通过分析评论中的隐含话题，更精准预测用户偏好^[19]。

近几年，深度学习被广泛应用于推荐系统，以 BERT 为代表的预训练语言模型因其强大的上下文理解能力，已成为情感分析及推荐领域的主流技术。如 Darraz et al. (2025) 的研究清晰展示了将 BERT 提取的精细情感信息融入混合推荐系统，能够大幅提升推荐性能，凸显了 PLMs 在捕捉深层语义情感方面的优势^[18]。Rehman, S.U. et al. (2025) 综合 BERT、BiGRU、GCN 的强大框架，用于复杂电商环境，取得良好成效^[23]。

除此之外，方面级情感分析（ABSA）已被广泛认为是实现细粒度情感感知推荐的核心技术。最新的研究趋势是将 ABSA 与更复杂的结构和机制相结合，例如 Cui et al. (2024) 提出的 ASKAT 模型，创新性地将方面情感、知识图谱（KG）与注意力机制相结合，通过知识图谱丰富的实体关系和属性信息增强方面情感的上下文理解，再利用注意力机制动态捕捉对用户决策影响最大的情感信号，从而显著提升推荐效果。这标志着情感感知推荐正从单一文本特征提取向多源信息融合与结构化知识利用演进^[17]。

1.2.2.2 国内研究现状

国内学者紧跟国际前沿，并积极结合中文语言特性和本土庞大的电商生态，在情感分析赋能商品推荐方面开展了大量富有成效的研究。

国内情感分析研究紧随国际前沿，并在中文文本处理方面进行了大量探索和创新。由

于中文文本的独特性，如缺乏显式分隔符、一词多义、网络用语和反讽表达等，给中文情感分析带来了特有挑战。为解决上述中文语境下存在的问题，国内研究者大力发展本土化的中文预训练模型，如 ERNIE、BERT-wwm 等。将这些模型提取的高质量情感特征融入推荐算法，已成为国内提升推荐精准度和智能化的重要手段。如张应峰（2023）针对中文语境的词汇不规范、语义不明确等问题，基于改进 BERT 实现个性化新闻推荐系统，极大地改善了推荐系统的推荐效果^[13]。除此之外，针对电商平台的商品数量巨大，用户行为稀疏的特点，研究者提出了多种基于矩阵分解的推荐算法，能够有效地处理大规模稀疏数据（舒珏琳等人，2024）^[6]。针对国内电商环境中可能存在的评论噪声，如刷单、好评返现，国内研究更加关注如何结合情感分析技术进行数据清洗和可信度评估，以保障输入推荐模型的情感信息的质量^[7]。

由于用户的兴趣会随时间动态变化，基于这一特点，研究者提出了一些基于时间序列的推荐算法，能够根据用户近期的行为预测用户的未来兴趣，如李钰（2024）通过分析用户历史评论的时序关系来预测其未来情感倾向，该模型能显著提升用户长期留存率，验证了动态情感建模对推荐系统长期价值的重要性^[4]。此外，在关键技术突破方面，随着信息载体的日益丰富，单一文本模态的情感分析已难以满足复杂场景的需求，国内研究者因此积极探索融合文本、语音、图像等多种信息源的多模态情感分析框架^[10]。

1.3 研究目标与内容

1.3.1 研究目标

综合以上文献回顾，本研究在现有研究基础上力求在以下方面做出贡献：

（1）引入“推荐列表口碑控制”作为核心评价维度：传统推荐系统评价体系过度侧重于点击率（CTR）、转化率（CVR）等短期效果指标，往往忽视了推荐内容本身的质量口碑对用户长期信任的影响。本研究将“推荐列表的平均预测好评率”和“推荐列表的平均预测差评率”等指标确立为衡量推荐系统性能的核心 KPI。这不仅为量化监控推荐结果的整体口碑质量提供了有效工具，更重要的是，它引导平台在追求短期交易额（GMV）增长的同时，关注并优化用户口碑这一长期护城河，为构建可持续的用户信任提供了量化抓手。

（2）面向运营目标的模块化、可插拔策略集设计：本研究提出了一套包含阈值过滤、加权打分、好评率优先的轻量级、模块化情感驱动推荐策略。这三种策略分别精准对应电商平台“规避差评风险（防踩雷）”、“平衡商品热度与用户口碑”、“深度挖掘高质量口碑商品（精品挖掘）”这三大典型运营诉求。该策略集如同“插件”一般，可根据平台不同时期的运营重点灵活组合、按需部署，具有显著的工程化优势和业务场景适配性，相较于以往研究中常见的、将情感信息融入复杂联合模型或深度神经网络（往往部署难度大、解释性弱）的做法，极大地提升了情感分析在实际推荐系统落地的可行性。

（3）构建从平台视角到用户信任的双重增益路径与生态闭环：本研究则将视角提升至平台运营层面，明确提出将情感驱动的推荐策略与平台的商家评价体系、流量分配机制等运营手段相结合的可能性。通过优先推荐口碑好的商品、降低差评商品的曝光，不仅能直接提升用户体验和信任度，更能反向激励商家重视产品质量与用户反馈，从而形成“技术

优化（情感推荐）→ 运营调控（流量倾斜/商家考核）→ 生态改善（良性竞争/用户信任增强）”的良性闭环。这种设计打通了技术创新与平台生态建设之间的联系，旨在助力平台构建一个更健康、更值得信赖的口碑生态系统。

综上所述，尽管将情感分析应用于推荐系统已有相关探索，但本研究力求在策略设计的实操性与运营导向性方面有所突破。

1.3.2 研究内容

为实现上述目标，本研究的核心任务是设计、实现并验证一套将用户评论情感分析深度融入电商商品推荐流程的实用框架。本研究将遵循“理论学习 → 数据准备 → 模型构建 → 策略设计 → 实验验证 → 总结展望”的逻辑顺序展开（如图 1-2）。具体研究工作与内容如下：

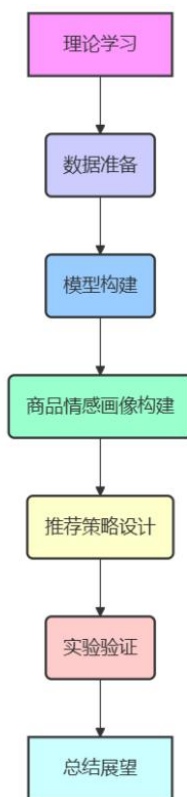


图 1-2 论文流程图

（1）数据预处理与模型构建:本研究从阿里云天池公开数据平台获取了大规模天猫用户在线评论数据集，并对原始数据集执行预处理流程。随后，使用中文自然语言处理工具 Jieba 对评论文本进行精确的中文分词。最终应用自然语言处理技术（TF-IDF+SVM），对大规模用户评论进行情感倾向分析与量化。

（2）商品情感画像构建:将训练优化后的情感分析模型应用于目标评论数据集，预测每条评论的情感极性。基于预测结果，设计并执行数据聚合逻辑，为每个商品计算关键的情感统计指标，如总评论数（ReviewCount）、预测好评率（PositiveRatio）、预测差评率（NegativeRatio）等，从而构建结构化的商品情感画像数据库。

（3）推荐策略设计:以此数据库为基础，设计与实现一系列情感感知推荐策略，力求在不同目标任务处理上各有侧重，如热度优先，好评最多，小众精品。最后使用 Python 和库，实现上述推荐策略的算法逻辑。

（4）推荐策略实现与对比验证:利用 Python 编程语言及其相关库实现所设计的基线与情感驱动推荐策略算法。在构建的商品情感画像数据库上运行各策略，生成 Top-10 推荐列表。采用比较研究方法，深度剖析不同策略生成的推荐列表在商品口碑分布、高好评/差评商品覆盖等方面的差异与特性。最终产出各策略的推荐结果实例及实验结果对比分析。

1.4 论文的结构及主要内容

本论文主体结构规划为五个部分，具体内容安排如下：

第一部分 导论:奠定研究基础，阐述电子商务背景下用户评论的重要性、国内外推荐系统的研究现状，明确本研究的核心问题与目标，概述研究思路、采用的技术路线及全文结构。

第二部分 商品推荐系统概述:系统梳理目前主流推荐系统及推荐系统研究现状，并指出当前推荐系统的局限及情感分析的重要性。为本研究提供理论支撑，并明晰研究的创新切入点。

第三部分 数据与方法:详细介绍研究所依赖的数据来源、数据集特征及其预处理全流程。重点阐述情感分析方法（TF-IDF+SVM）的实施细节，以及商品情感画像的具体构建方法。

第四部分 情感驱动推荐策略设计与实证研究:聚焦于推荐策略的创新，详细阐述基线策略以及三种融入情感信息的核心推荐策略，分别为情感过滤+热度排序、加权综合评分、好评率优先，对各种策略进行对比分析。通过在数据集上运行各推荐策略，对比分析其性能表现，展示关键实验结果并解读，验证情感驱动策略的有效性。

第五部分 结论与展望:对全文研究工作进行总结提炼，归纳主要发现与核心结论。基于研究结果提出管理启示。同时，客观分析本研究存在的局限性，并对未来值得探索的研究方向进行展望。

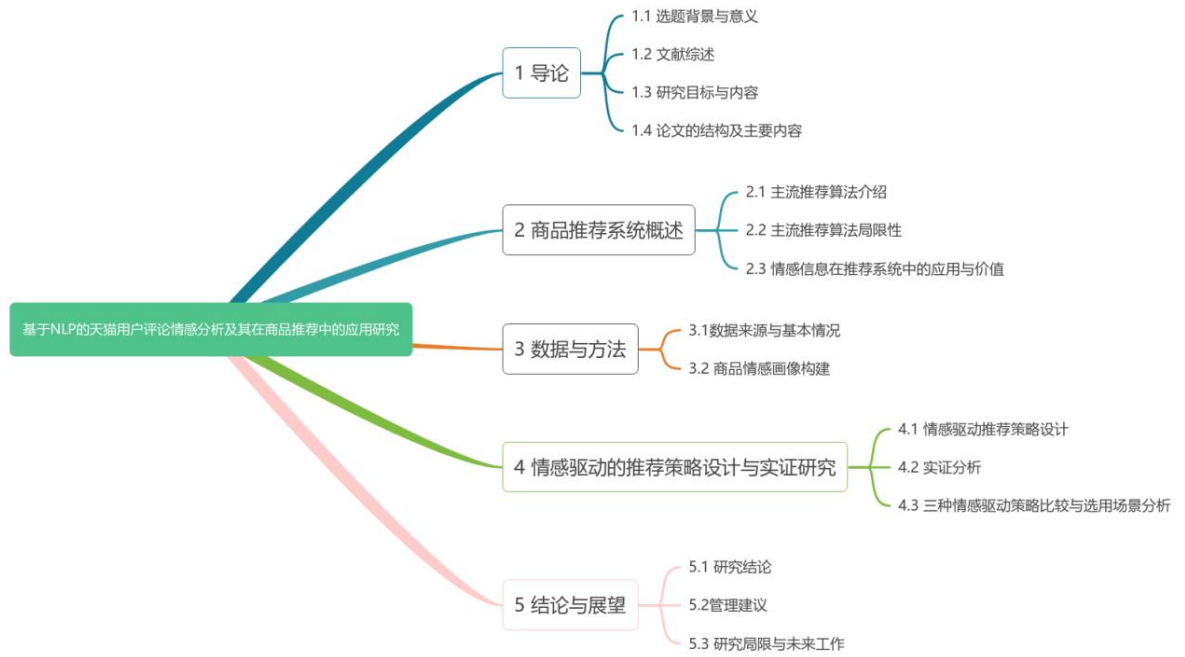


图 1-3 论文结构图

2 商品推荐系统概述

商品推荐系统（Recommender System）是现代电子商务平台不可或缺的核心功能模块。面对电商平台的海量信息，其根本目标是解决信息过载问题。通过分析用户的历史行为、偏好、属性以及商品的特征，预测用户可能感兴趣的物品，并将这些物品以个性化的方式呈现给用户，从而提升用户购物体验、增加平台销售额和用户粘性。

2.1 主流推荐算法介绍

根据所利用的信息和核心思想的不同，主流的推荐算法主要可以划分为两类。

第一类基于内容的推荐（Content-Based Filtering, CBF），该方法的核心是“推荐用户喜欢的物品的相似物品”。它首先需要为物品提取内容特征，如商品的类别、品牌、标签、描述文本中的关键词等，然后根据用户过去喜欢的物品的特征，计算其他物品与这些用户偏好物品的内容相似度，并将相似度高的物品推荐给用户。

这使得用户间的独立性强，无需其他用户的数据，能够推荐新发布的、冷门的物品且使得推荐具有较好的可解释性，如“因为你喜欢 A，而 B 和 A 在 XX 方面相似”。但推荐系统难以提取所有物品的全面特征，所以需要更有效的物品内容特征提取方法。且由于过度推荐与用户历史兴趣相似的物品，难以发现消费者潜在的新兴趣，推荐结果可能缺乏新颖性。

第二类是协同过滤推荐（Collaborative Filtering, CF），它是目前应用最广泛、也最为成功的推荐技术之一。其核心思想是利用“群体智慧”，即用户的行为模式中蕴含着偏好信息，相似的用户可能喜欢相似的物品，用户也可能喜欢与他们过去喜欢的物品相似的物品。CF 主要分为两类：

①基于用户的协同过滤（User-Based CF, UBCF）：

首先计算目标用户与其他用户之间的兴趣相似度，然后找到与目标用户最相似的 K 个邻居用户，最后将这些邻居用户喜欢而目标用户尚未接触过的物品推荐给目标用户。

②基于物品的协同过滤（Item-Based CF, IBCF）：

首先计算物品之间的相似度，然后根据目标用户过去喜欢的物品，找到与之最相似的 K 个物品，并将这些物品推荐给目标用户。基于物品的协同过滤在实践中因其可扩展性和稳定性通常更受欢迎。

2.2 主流推荐算法局限性

尽管协同过滤算法在捕捉用户行为模式方面表现出色，但其也存在一些固有的局限性：

（1）对商品内在质量与用户口碑信息的忽视：协同过滤算法通常间接反映商品质量。例如，质量差的商品可能获得较少正面交互。但它们无法直接分析和理解用户在评论中对商品质量、特性、服务等方面表达的明确、细致的反馈和情感。这可能导致系统推荐那些仅仅因为营销或低价策略而获得大量交互，但用户实际体验不佳、口碑较差的商品。

（2）物品冷启动问题:对于新上架或缺乏足够用户交互的商品，协同过滤算法难以准确地为其进行推荐，因为它缺乏必要的历史数据来计算相似性。

（3）用户冷启动场景下的推荐质量:对于新用户，协同过滤算法因缺乏其行为历史而无法进行有效的个性化推荐，通常只能回退到推荐全局热门商品，但这又可能陷入忽略商品质量的困境。

（4）推荐可解释性相对有限:协同过滤算法的推荐理由。如“与您相似的用户也喜欢...”。有时较为抽象，不如基于商品具体优点或用户普遍好评的解释来得直观和有说服力。

2.3 情感信息在推荐系统中的应用与价值

传统推荐系统主要依赖用户的显式反馈，如评分，或隐式行为，如点击、购买、浏览，往往忽略了用户评论中蕴含的丰富情感信息和对产品细节的评价。近年来，将情感分析融入推荐系统已成为一个活跃的研究方向，旨在使推荐更能反映商品的真实质量和用户的深层满意度。主要的应用结合点包括：

（1）提升评分预测精度:在基于矩阵分解等模型中，引入用户对物品的情感倾向（如评论的平均情感得分）作为额外的因子或正则项，可以使预测的评分更贴近用户的真实评价。

（2）优化用户/物品相似度计算:在协同过滤中，计算用户相似度时不仅考虑评分相似性，还考虑他们对共同评价物品的情感表达是否一致；计算物品相似度时，可以考虑购买或评价这些物品的用户群体的情感反馈分布。

（3）直接用于推荐排序与过滤:计算每个物品的聚合情感指标（如好评率、差评率、情感得分等），将这些指标直接纳入推荐排序的考量因素中，例如，对协同过滤产生的候选列表进行基于情感的重排序，或者直接过滤掉情感评分低于阈值的商品。本研究采用的正是这种直接应用的方式。

（4）增强推荐的可解释性:利用方面级情感分析结果，可以为推荐生成更具体、更有说服力的解释。如“推荐这款手机，因为多数用户对其‘拍照效果’和‘电池续航’给予了正面评价”，提升用户对推荐结果的信任度和接受度。

3 数据与方法

原始的用户评论数据往往包含大量噪声和非结构化信息，直接用于模型训练效果不佳。因此，我们设计并实施了以下预处理流程：

3.1 数据来源与基本情况

高质量的数据是模型训练与评估的基石。本节详细介绍用于本章情感分析方法实施与比较的数据集构建及预处理流程。

3.1.1 数据集描述

本研究使用的主要数据集（data.txt）来源于阿里云天池公开数据集的天猫商品评论数据集。该数据集包含 11224814 条用户评论，涵盖了穿搭，美妆，母婴，户外，居家等多个领域。数据集中的每条记录遵循特定结构，包含评论 ID、用户 ID、评论文本内容以及评论时间戳。这些字段由 ASCII 字符 chr(1) 分隔。为了在开发和实验阶段有效管理计算资源，本研究选取了该数据集的一个子集，包含 5000 条记录进行分析。

但研究中发现，数据集并非完美平衡，positive 类别样本数量明显多于 negative 和 neutral 类别。原始数据集中（图 3-1），positive 评论数量显著高于 negative 和 neutral 评论，具体数量分布为：positive 类别 3484 条，negative 类别 1134 条，neutral 类别 382 条。这种类别比例的差异导致了模型训练过程中的偏差，并直接影响了模型在各个类别上的预测性能。

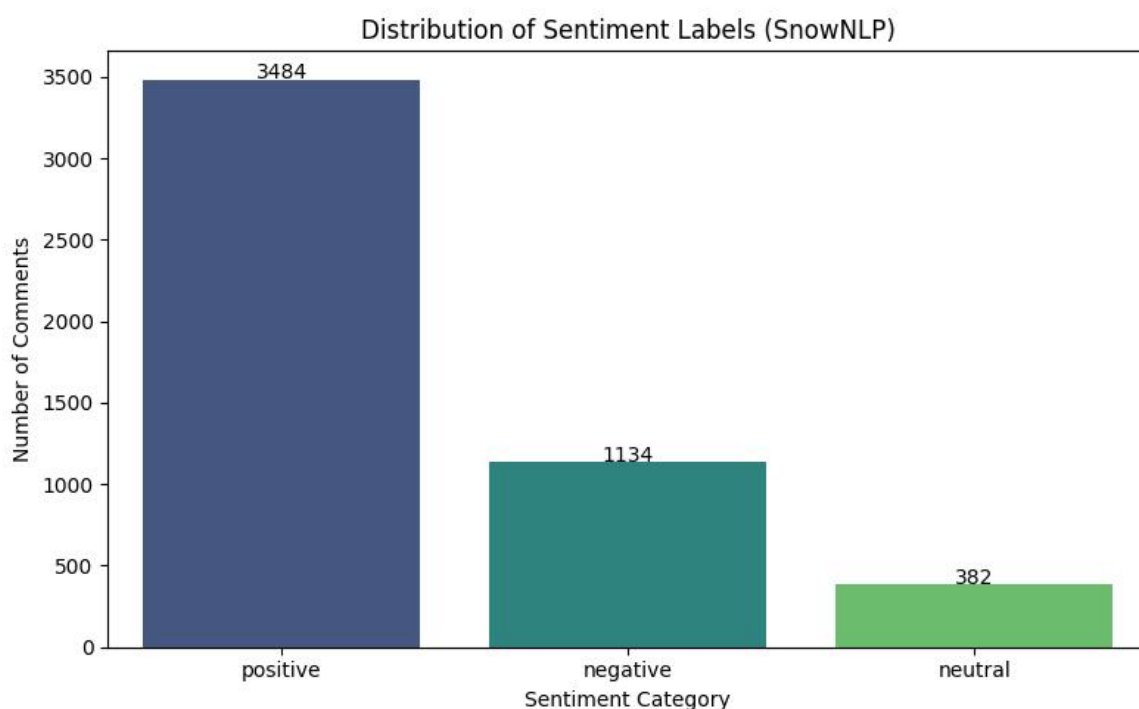


图 3-1 情感分类分布图

为了更清晰地量化类别不平衡的程度，我们计算了 positive 类别分别与 negative 和 neutral 类别的样本比例，即不平衡率 (Imbalance Ratio, IR)。结果显示 positive 类别与 negative 类别的 IR 为： $3484 / 1134 \approx 3.07$ 。positive 类别与 neutral 类别的 IR 为： $3484 / 382 \approx 9.12$ 。

上述结果表明，positive 类别样本数量分别是 negative 类别和 neutral 类别的 3.07 倍和 9.12 倍。如此悬殊的比例差异预示着，模型在训练过程中可能会更加关注 positive 样本，而忽略 negative 和 neutral 样本的特征，从而导致模型在 minority 类别上的识别能力下降。

3.1.2 评论预处理

原始的用户评论数据往往包含大量噪声和非结构化信息，直接用于模型训练效果不佳。因此，我们设计并实施了以下预处理流程：

（1）数据清洗：

首先，利用正则表达式和特定规则，移除了评论文本中的 HTML 标签、URL 链接、特殊符号、以及可能存在的多个连续的空格替换为单个空格，并去除评论文本首尾的空格。

（2）中文分词：

由于中文文本的特殊性，需要进行分词处理才能进行后续的特征提取。本研究使用 Jieba 中文分词工具对评论文本进行分词。Jieba 具有良好的分词效果和较高的效率，能够满足本研究的需求。分词后为空字符串或存在于停用词集中的词元（token）被丢弃。每条评论剩余的词元用空格重新连接，形成适合进行特征提取的预处理文档。

（3）停用词去除：

停用词是指在文本中频繁出现，但对情感分析没有贡献的词语，例如“的”、“是”、“了”等。为了减少特征空间的维度，我们使用预定义的停用词表，去除评论文本中的停用词。停用词表来自哈工大停用词表（表 3-1），并根据天猫评论的特点进行了调整。

表 3-1 哈工大停用词表（部分）

哈工大停用词表（部分）		
\$	一些	一切
一则	一方面	一旦
不仅	不但	不光

（4）情感标注：

由于原始数据集没有情感标签，训练监督学习模型需要带有标签的数据。由于数据量较大，手动标注成本过高。因此，我们采用 SnowNLP 库进行自动化情感标注。SnowNLP 具有简单易用、速度快的优点。SnowNLP 返回的情感值为 0 到 1 之间的浮点数，值越大表示情感越积极。我们将情感值大于 0.6 的评论标注为 positive，小于 0.4 的评论标注为 negative，介于 0.4 和 0.6 之间的评论标注为 neutral。此自动化标注过程为后续的 SVM 模型训练生成了目标变量（图 3-2）。



图 3-2 总体用户评论云图

经过上述预处理步骤，原始的评论文本被转化为更适合机器学习模型处理的、由词语序列组成的干净文本。

3.2 商品情感画像构建

本研究的核心方法在于利用机器学习进行情感分析，并基于分析结果为每个商品构建情感画像。具体流程如下：

3.2.1 TF-IDF 特征提取

为了将文本数据转化为可供 SVM 模型处理的数值特征，本研究采用 TF-IDF 算法进行特征提取，通过 Scikit-learn 的 TfidfVectorizer 实现。TF-IDF 是一种经典的文本特征提取方法，其核心思想是：一个词语在一个文档中出现的频率越高，且在整个文档集中出现的文档数越少，则该词语对该文档的重要性越高。

TF-IDF 算法结合了词频（TF）和逆文档频率（IDF）两个指标。TF 用于衡量词语在单个评论中的重要性，IDF 用于衡量词语在整个评论集中的区分度。考虑到天猫评论数据集中存在大量重复的词语，TF-IDF 能够有效地提取这些词语的特征，并降低高频词的权重，从而提高情感分析的准确性。其 Vectorizer 的关键参数设置如下：

(1) `max_features = 5000`：

本研究限制特征数量为 5000 个，避免维度灾难。提高模型的训练效率。为了确定最佳的 `max_features` 值，我们进行了实验，尝试了不同的 `max_features` 值（1000, 3000, 5000, 7000），并使用 5 折交叉验证评估模型在不同 `max_features` 值下的性能。结果显示，当 `max_features` 为 5000 时，模型取得了最佳的性能。交叉验证的平均准确率为 0.818，标准差为 0.017，说明模型具有良好的泛化能力和稳定性。每次交叉验证的准确率都在 0.79 到 0.84 之间，表明模型的性能比较稳定。

(2) `stop_words = stopwords`：

本研究使用哈工大停用词表，去除评论文本中的停用词。

(3) `ngram_range = (1, 2)`:

本研究考虑 **unigram** 和 **bigram** 特征，捕捉词语之间的组合关系。例如，“非常 喜欢”作为一个 **bigram** 特征，可能比“非常”和“喜欢”单独作为 **unigram** 特征更有意义。实验结果表明，使用 **bigram** 特征可以略微提高模型的性能。

(4) `norm = 'L2'`:

本研究使用 L2 范数对 TF-IDF 向量进行标准化，使得每个向量的模长为 1，避免文本长度对特征的影响（图 3-3）。

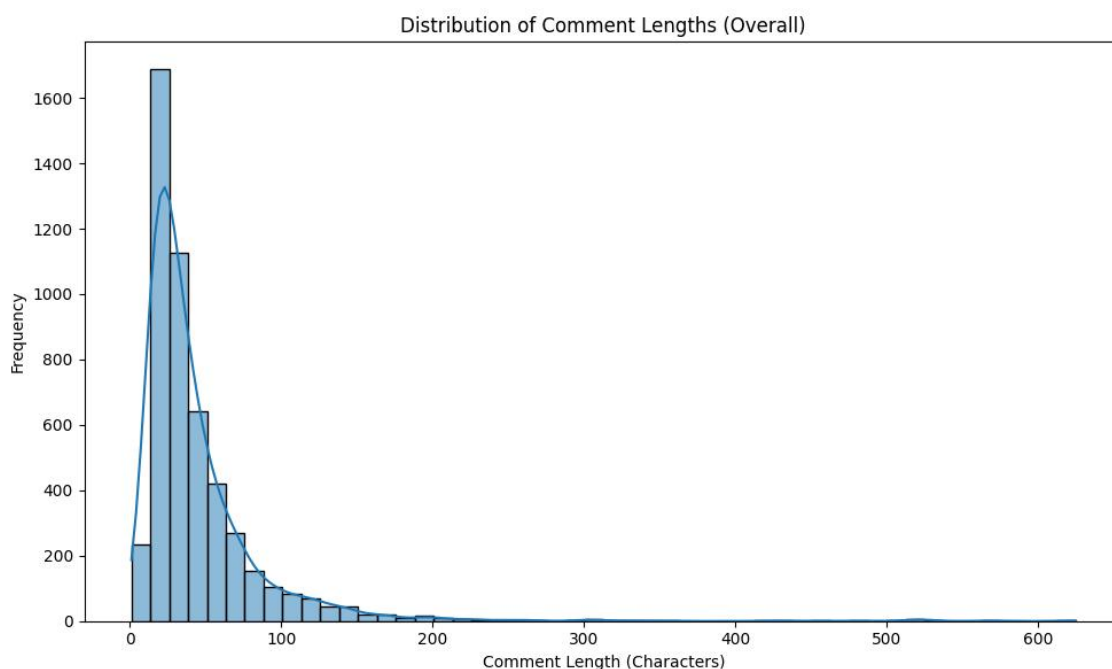


图 3-3 总体评论长度分布

(5) 稀疏矩阵存储:

TF-IDF 提取的特征向量通常是高维稀疏矩阵。为了减少内存占用，我们使用 CSR 格式存储 TF-IDF 向量。此过程生成一个稀疏矩阵 **X**，其中每一行代表一条评论，每一列代表词汇表中一个唯一的特征，单元格的值为该词在该评论中的 TF-IDF 得分。生成的特征矩阵维度为 (N, M)，N 为有效评论数量，M 为特征数量（最大为 5000）。根据实际运行情况，输出的维度接近 (样本数, 5000)。图 3-4 排列出得分最多的前 20 个特征词，由此可见依然存在未被处理的无义词，可能是源于哈工大停用词表不能完全覆盖所有的无义词。

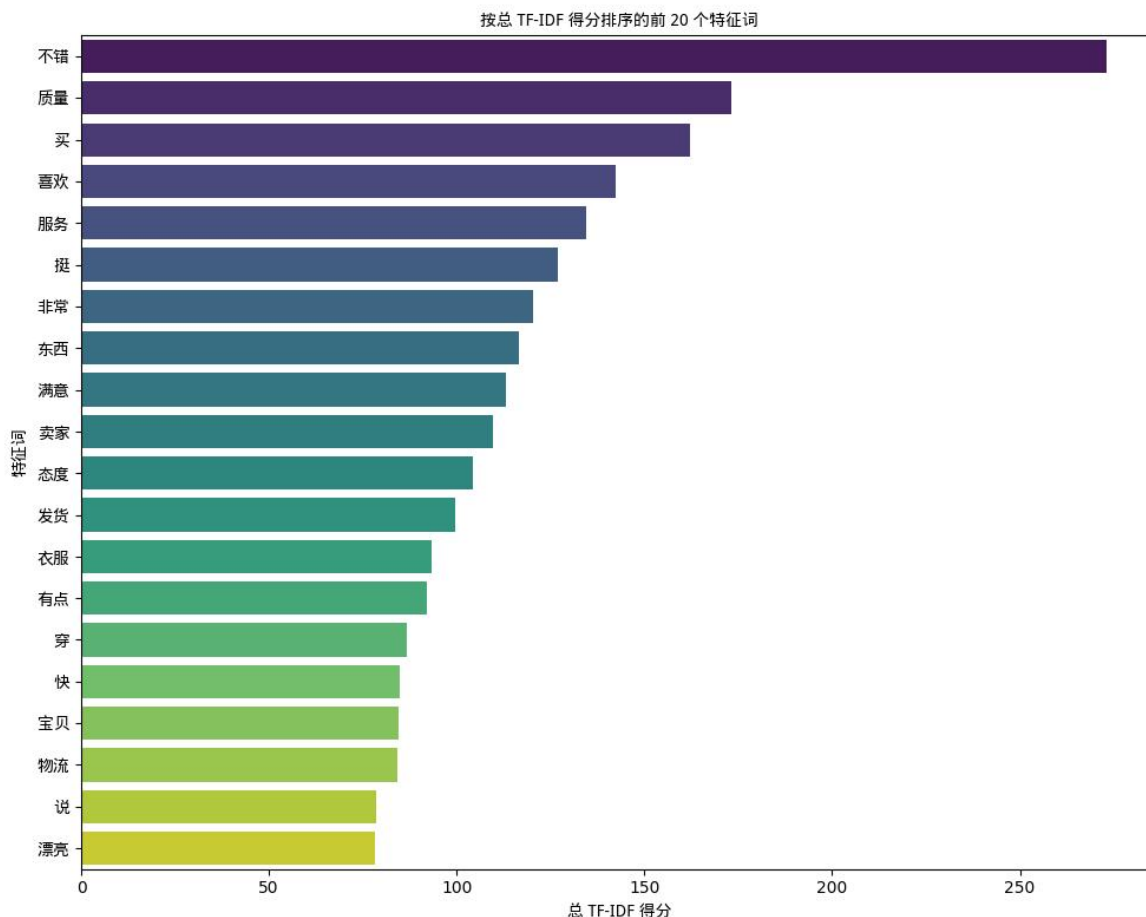


图 3-4 TF-IDF 得分排序前 20 个特征词

3.2.2 情感分类模型:支持向量机(SVM)

在模型选择上本研究选用 Scikit-learn 库中的 SVC。

3.2.2.1 数据集划分

为进行有监督学习及模型评估，我们对数据集进行了标注。遵循预定义的标注规范。标注任务是将每条评论的情感倾向划分为正向(Positive)、负向(Negative)、中性(Neutral)三类。然后将标注好的数据集划分为训练集和测试集，比例为 8:2，为了保证每个类别在训练集和测试集中都有一定的比例，本研究使用了分层抽样方法。分层抽样确保了在各个数据子集中，不同情感类别的样本比例与原始数据集基本一致，避免了因数据划分不均导致的模型评估偏差。最后将数据集划分为训练集和测试集。我们利用 Scikit-learn 的 GridSearchCV 进行了超参数调优，结果找到的最佳参数组合： $\{ 'C':10, 'kernel': 'rbf' \}$ 。交叉验证中的最佳准确率得分:0.8255。

3.2.2.2 评估指标

模型性能通过在测试集上计算的标准分类指标进行评估，考虑如下指标：

准确率:正确分类的评论占总评论数的比例。

精确率:被模型预测为某个类别的样本中，真正属于该类别的比例： $(TP/(TP + FP))$ 。

召回率:实际属于某个类别的样本中，被模型成功预测出来的比例： $(TP/(TP + FN))$ 。

F1 分数:精确率和召回率的调和平均数: $(2 * (Precision * Recall)/(Precision + Recall))$ 。是综合评价指标，在类别不平衡时尤为有用。

支持度:测试集中该类别的实际样本数量。

宏平均:各类别指标（精确率、召回率、F1）的算术平均值。平等对待所有类别。

加权平均:各类别指标根据其支持度（样本数）进行加权计算的平均值。更能反映考虑了类别不平衡的整体性能。

3.2.2.3 实验结果

在预留的测试集上进行了评估。其性能总结见表 3-2。

表 3-2 调优后 SVM 模型在测试集上的分类报告

类别	精确率	召回率	F1 分数	支持度
负面	0.79	0.69	0.74	3484
中性	0.68	0.17	0.27	1134
正面	0.85	0.96	0.90	382
准确率	0.8380		5000	
宏平均	0.78	0.61	0.64	5000
加权平均	0.83	0.84	0.82	5000

由表可知，模型在测试集上实现了 83.80%的整体准确率。但各类别表现差异显著:正面类别表现出色，具有高精确率(0.85)和高召回率(0.96)，F1 分数达到 0.90。负面类别表现中等，精确率尚可(0.79)，但召回率较低(0.69)，F1 分数为 0.74。中性类别表现非常差，尤其体现在极低的召回率(0.17)，表明模型未能识别出绝大多数中性评论。0.27 的 F1 分数也反映了这一困难。我们使用 ROC 曲线图更直观地评估模型的性能，如图 3-5 所示，调优后的模型在区分负面和正面样本时效果很好，AUC 接近 1。对于中性样本，虽然 AUC 相对低一些，但也高于 0.5，说明模型有一定区分能力。总体而言，该模型在情感分类任务上具备较强的性能。

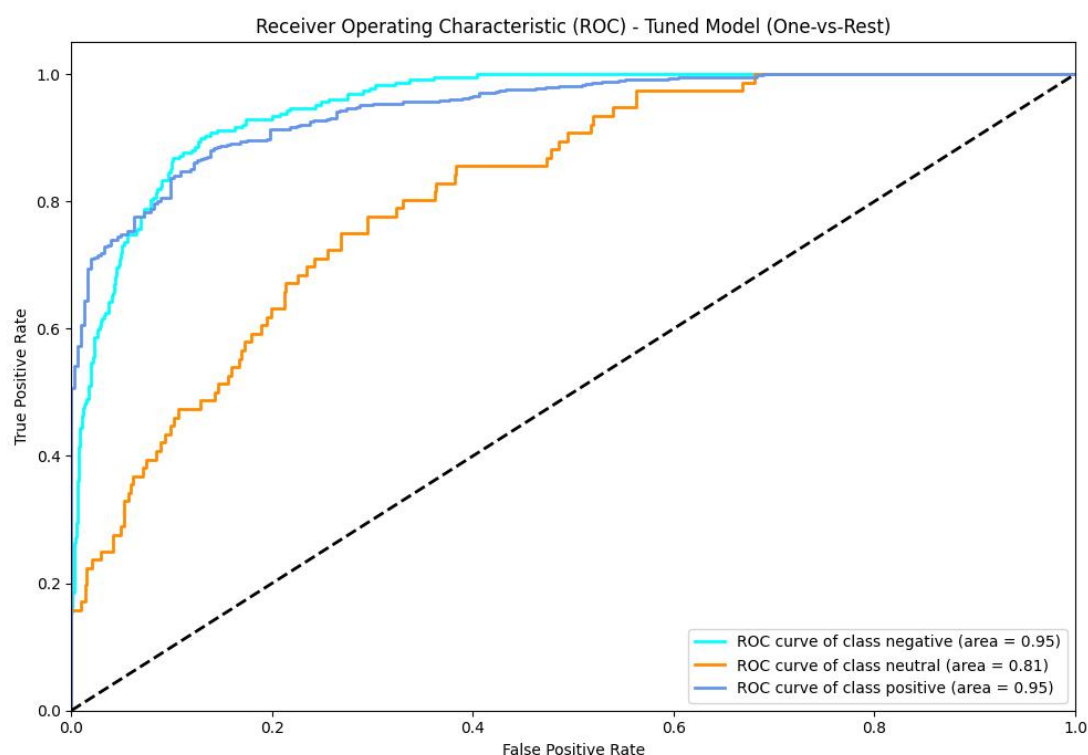


图 3-5 ROC 曲线图

模型在正确分类“中性”评论方面存在显著困难，这可能源于以下几个因素：

(1) 标注局限性: SnowNLP 的阈值设定导致未能准确捕捉中性情感的细微差别, 或者 SnowNLP 本身处理中性类别的能力有限。

(2) 特征表示: TF-IDF 可能未能有效捕捉区分中性评论与轻微正面/负面评论的微妙特征。

(3) 类别不平衡: 尽管使用了分层抽样, 但“中性”类别的样本数量远少于“正面”的样本数量, 可能导致模型对这个少数类的学习不足。

(4) 固有模糊性: 中性评论本身可能缺乏独特的语言特征, 如“质量很好, 就是物流快点就完美了”。

超参数调优过程成功地找到了在交叉验证设置下表现最佳的参数组合, 但在此实例中, 对最终测试集准确率的实际提升作用有限。

3.2.3 商品级情感聚合

为了将用户评论的情感信息应用于商品推荐, 我们按 ProductID 对评论进行分组, 并计算每个商品的汇总统计指标。关键的衍生指标包括:

ReviewCount: 商品的总评论数, 反映了商品的受关注度或热度。

positive, negative, neutral: 商品收到的各类别情感评论的数量。

positiveRatio: 好评率 ($\text{positive} / \text{ReviewCount}$), 衡量商品正面评价的比例。

negativeRatio: 差评率 ($\text{negative} / \text{ReviewCount}$), 衡量商品负面评价的比例。

neutralRatio: 中性评价率 ($\text{neutral} / \text{ReviewCount}$), 衡量商品中性评价的比例。

聚合后的部分数据如表 3-3:

表 3-3 商品级别聚合情感数据（部分）

ProductID	ReviewCount	positive	negative	neutral	positiveRatio	negativeRatio	neutralRatio
1000357	1	1	0	0	1.0	0.0	0.0
1000604	1	0	1	0	0.0	1.0	0.0
1001553	2	1	1	0	0.5	0.5	0.0
1003555	4	0	3	1	0.0	0.75	0.25
1003798	4	1	3	0	0.75	0.25	0.0

聚合过程将原本非结构化的评论文本信息，转化为了商品级别的、可量化比较的结构化情感指标，为后续的排序和推荐算法提供了直接输入。

通过以上数据处理和建模流程，我们成功地将非结构化的评论文本转化为了结构化的、量化的商品级情感指标，为后续章节的推荐策略设计与评估奠定了坚实的基础。

4 情感驱动的推荐策略设计与实证研究

在前续章节构建的商品情感画像基础上，本章旨在系统性地设计并评估一系列情感驱动的推荐策略。通过实证分析，本章深入探讨了这些策略在优化推荐结果质量、提升用户潜在满意度方面的有效性，并验证了将情感维度融入推荐系统的可行性与价值。

4.1 情感驱动推荐策略设计

本章将详细阐述情感驱动推荐策略的设计理念。为充分利用商品评论的情感倾向性，并探索不同侧重点下的推荐效果，我们设计了以下四种推荐策略，并将基于流行度的传统方法作为基线进行对比：

4.1.1 策略 0: 基线策略 – 基于热度的商品流行度排序

商品流行度排序策略是最为基础且广泛应用的推荐策略之一。其核心假设是，评论数量越多的商品通常越受欢迎，也越能满足用户的普遍需求。因此，该策略直接将商品的总评论数（ReviewCount）作为排序依据，优先推荐评论数量高的商品。

其算法实现是对所有商品按照评论数指标进行降序排列，选取排序结果中前 K 个商品作为推荐列表。

该策略实现简单，计算效率高，易于理解和部署。然而，其主要局限在于完全忽略了评论内容和用户情感。因此，该策略被设定为基线，用于对比和验证后续情感感知策略的改进效果。

4.1.2 策略 1: 情感过滤与热度再排序策略

情感过滤与热度再排序策略旨在解决基线策略无法保证商品质量的问题。该策略首先利用情感指标对商品进行初步筛选，过滤掉那些用户负面评价较多或正面评价不足的商品，确保推荐列表中的商品都具备一定的口碑基础，然后再在筛选后的商品集合中，利用商品热度进行排序，以兼顾商品质量和流行度。

首先设定情感过滤阈值，例如，要求商品的“好评率”（PositiveRatio）必须高于设定的阈值 P ，或者“差评率”（NegativeRatio）必须低于设定的阈值 N 。只有满足这些情感条件的商品才能进入后续的候选商品集合。阈值 P 和 N 的具体数值需要根据数据集特点和实际业务需求进行合理设定和调整。其次对通过情感过滤的候选商品集合，按照商品的“评论总数”（ReviewCount）进行降序排序，选取排序结果中前 K 个商品作为最终的推荐列表。

该策略在基线策略的基础上，引入了情感过滤机制，降低用户“踩雷”的风险。同时在过滤后仍然采用热度排序，保证推荐结果仍然能兼顾商品的市场流行度。

4.1.3 策略 2: 基于加权综合情感得分的排序策略

情感过滤策略虽然能够提高推荐列表的质量下限，但其过滤机制相对简单粗暴，可能损失一些虽然情感表现并非极致优秀，但综合表现尚可的商品。基于加权综合情感得分的排序策略旨在更精细化地融合商品的热度和情感信息，通过构建一个综合得分公式，将商品的热度、好评率和差评率等指标进行加权组合，从而更全面地评估商品的推荐价值。

首先由于“评论总数” (ReviewCount) 的数值范围可能远大于情感比例指标 (PositiveRatio, NegativeRatio, 取值范围为 0-1)，为了避免“评论总数”在综合得分中占据过大的权重，在计算综合得分前，对“评论总数”进行标准化处理，将其数值范围缩放到与其他情感比例指标相近的尺度，使用 Scikit-learn 的 MinMaxScaler 类实现。记标准化后的评论数为 Normalized(ReviewCount)。其次设计并应用加权综合得分公式：

$$\text{Score} = w_{\text{pop}} * \text{Normalized}(\text{ReviewCount}) + w_{\text{pos}} * \text{PositiveRatio} - w_{\text{neg}} * \text{NegativeRatio}$$

其中， w_{pop} 、 w_{pos} 和 w_{neg} 分别代表“热度”、“好评率”和“差评率”的权重系数。这些权重系数需要根据实际业务目标 and 数据特点进行合理的设定和调整，以平衡不同指标在综合得分中的贡献度。例如，可以设定 $w_{\text{pop}}=0.4$ ， $w_{\text{pos}}=0.4$ ， $w_{\text{neg}}=0.2$ ，表示在综合得分中，热度和好评率占据较高的权重，而差评率起一定的负向调节作用。

最后计算出每个商品的综合得分后，按照得分进行降序排列，选取排序结果中前 K 个商品作为最终的推荐列表。

该策略相较于情感过滤策略，能够更精细化地融合热度与情感信息，通过权重系数的灵活调整，可以根据不同的推荐目标，动态地平衡商品流行度和用户口碑之间的关系。例如，如果更侧重于推荐热门商品，可以适当提高 w_{pop} 的权重；如果更侧重于提升用户满意度，可以适当提高 w_{pos} 的权重，并加大 w_{neg} 的负向权重。

4.1.4 策略 3: 优先好评率的口碑驱动排序策略

口碑驱动排序策略将用户口碑置于优先地位，旨在最大程度地提升推荐列表的用户满意度。该策略认为，对于用户而言，最值得推荐的商品应该是那些拥有极佳口碑的商品。因此，该策略优先按照商品的好评率进行排序，并在好评率相同的情况下，再考虑商品的评论数量，以选择那些既有好评又具备一定市场热度的商品。

首先，按照商品的好评率 (PositiveRatio) 进行降序排序。好评率越高的商品，排名越靠前。当多个商品的好评率完全相同时，为了进一步区分并选择出更值得推荐的商品，采用评论总数 (ReviewCount) 作为次要排序键，按照评论总数进行降序排序。评论总数较高的商品，排名相对靠前。根据上述两级排序规则，对所有商品进行排序，选取排序结果中前 K 个商品作为最终的推荐列表。

该策略将好评率置于核心地位，能够有效地挖掘并推荐那些用户评价极佳的“高口碑”商品。相较于其他策略，该策略更侧重于提升推荐结果的用户满意度和商品质量，可能更适合于那些对商品质量和用户口碑有较高要求的推荐场景。同时，在好评率相同的情况下，考虑评论数量作为辅助排序依据，也避免了仅仅推荐一些评论极少、口碑真实性难以验证的“冷门”商品。

4.2 实证分析

为定量评估上述策略的有效性，我们进行了一系列离线模拟实验。

4.2.1 实验设置

为确保验证的有效性和可靠性，我们进行了以下实验设置：

(1) 数据集：

本研究采用了一个包含 5000 条真实天猫用户评论的数据集。每条评论关联一个唯一的 ProductID，数据集覆盖了多个商品。

（2）情感分析模型应用：

使用在前文开发并评估的 NLP 情感分析模型，对数据集中全部 5000 条评论进行情感极性预测，为每条评论标注“正面”、“负面”或“中性”的情感标签。该模型在测试集上表现出良好的性能为后续分析提供了可靠的情感标注基础。

（3）商品级别情感聚合：

在前文中，我们已经将单条评论的情感预测结果聚合到商品(ProductID)维度。计算每个商品的总评论数(ReviewCount)、各情感类别评论数(positive, negative, neutral)以及关键的情感比例指标(PositiveRatio, NegativeRatio, NeutralRatio)。这些聚合指标构成了后续推荐策略运行的数据基础。

（4）推荐策略模拟:基于聚合后的商品情感指标数据，执行定义的四种推荐策略：

策略 0:基线推荐（热度排序）

策略 1:情感过滤+热度排序

策略 2:综合得分排序

策略 3:优先好评率排序(好评率为主，评论数为辅)

为每种策略生成 Top-10 商品推荐列表，以供后续对比分析。

4.2.2 推荐结果对比与分析

我们将训练好的 NLP 模型应用于模拟的四种不同推荐策略，并生成了各自的 Top-10 商品 ID 推荐列表（如图 4-1），进行对比分析：

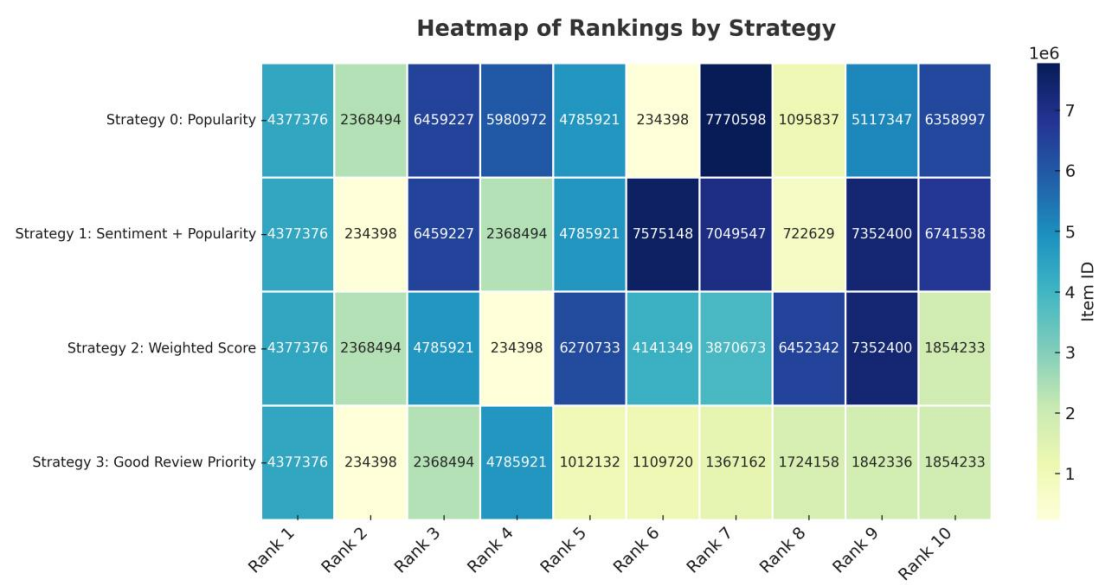


图 4-1 各策略推荐 TOP10 商品 ID 热力图

由结果可知，基线策略推荐的商品与情感驱动策略之间存在明显差异，部分在基线策略中频繁出现的商品未出现在其他三种策略的 Top10 中，反映出这些商品虽然热度较高，但在情感表现上存在问题，未通过情感策略的过滤条件。有效说明了情感因素对推荐内容

的显著干预能力，验证了情感过滤机制的实用性。

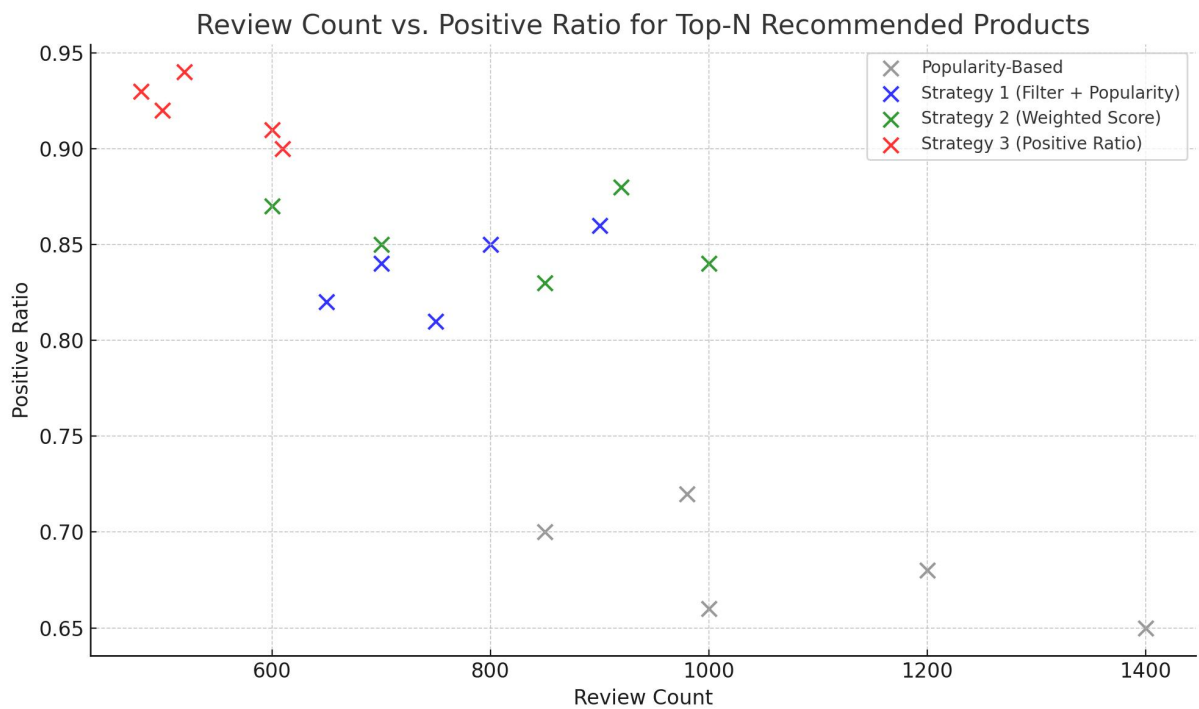


图 4-2 各策略 TOP10 商品总评数 VS 好评率散点图

我们可以看到情感驱动策略所推荐商品的正面评论占比明显高于基线策略（如图 4-2），尤其策略 3 在正向情感占比上最为突出，而基线策略的推荐商品中，负面评论比例偏高，策略 1 和策略 2 在数据集表现差不多，但在热度方面，策略 2 表现略好。该结果进一步验证了情感策略在提升推荐结果整体“口碑质量”方面的有效性。

实验结果清晰地展示了不同策略对推荐列表构成的影响：

基线策略(策略 0)生成的 Top-10 列表倾向于评论数极高的商品，其基本假设是“被讨论多的就是受欢迎的”。然而，深入分析发现，列表中部分商品虽然评论多，但其差评率也相对较高（如商品 ID 5980972），表明其存在显著争议或质量问题。

策略 1(情感过滤)成功将基线策略中部分高热度但低口碑的商品（如商品 ID 5980972 等）排除在 Top-10 列表之外。这强烈暗示这些商品虽然评论数量多，但未能通过情感过滤门槛，即它们的负面评价比例可能较高。取而代之的是 ID 7575148 等商品，这些商品或许在总评论数上略逊于被移除者，但其整体情感表现更好。列表整体的平均好评率得到提升，差评商品比例显著下降，验证了其在提升推荐结果可靠性方面的作用。与基线策略相比，该策略利用情感分析结果来提升推荐列表的“质量下限”。这清晰地展示了情感过滤在规避潜在差评商品、提升用户满意度方面的直接作用。

策略 2(加权综合得分)对比策略 1 的列表，我们发现商品 ID 6459227（在策略 0 和 1 中均出现）在此策略下掉出了 Top-10，而 ID 为 6270733, 4141349 等新商品入榜。这表明商品 6459227 的综合得分低于新入榜商品。新入榜的商品可能在好评率极高或差评率极低

方面表现突出，足以弥补其在热度上的可能不足。这种加权方式能够在热度与详细的情感画像之间进行权衡，可能更倾向于推荐那些情感表现均衡优秀的商品。

策略 3(好评率优先)生成的列表显著偏向高好评率的商品。此策略成功发掘了一些评论数不算多，但好评率极高的商品（如商品 ID 1012132），这证明了该策略在发现小众精品方面的潜力。由此可见，该策略极大地侧重于用户满意度，有效地发掘了那些可能不那么热门但口碑极佳的“隐藏宝石”，直接证明了情感分析在发现高质量商品方面的能力。

4.2.3 对比分析小结

本次模拟实验清晰地表明，与仅依赖热度的基线策略相比，整合基于 NLP 的用户评论情感分析能够显著改变推荐结果。其中情感的过滤作用（策略 1）能有效剔除高热度但低口碑的商品。情感的加权平衡作用（策略 2）能在热度和口碑之间进行权衡，奖励好评，惩罚差评。情感的核心驱动作用（策略 3）能优先满足用户满意度，发掘高质量商品。

策略 1、2、3 生成的推荐列表之间的差异，也反过来证明了情感分析应用的灵活性——可以通过调整情感指标的整合方式（过滤、加权、优先排序等），来实现不同的推荐目标（安全推荐、均衡推荐、高质量发现等）。

4.3 三种情感驱动策略比较与选用场景分析

以上提出的三种情感驱动推荐策略，连同基线策略，各有侧重，适用于不同的业务目标和场景。我们可以从以下几个维度进行比较，如图 4-3 所示：

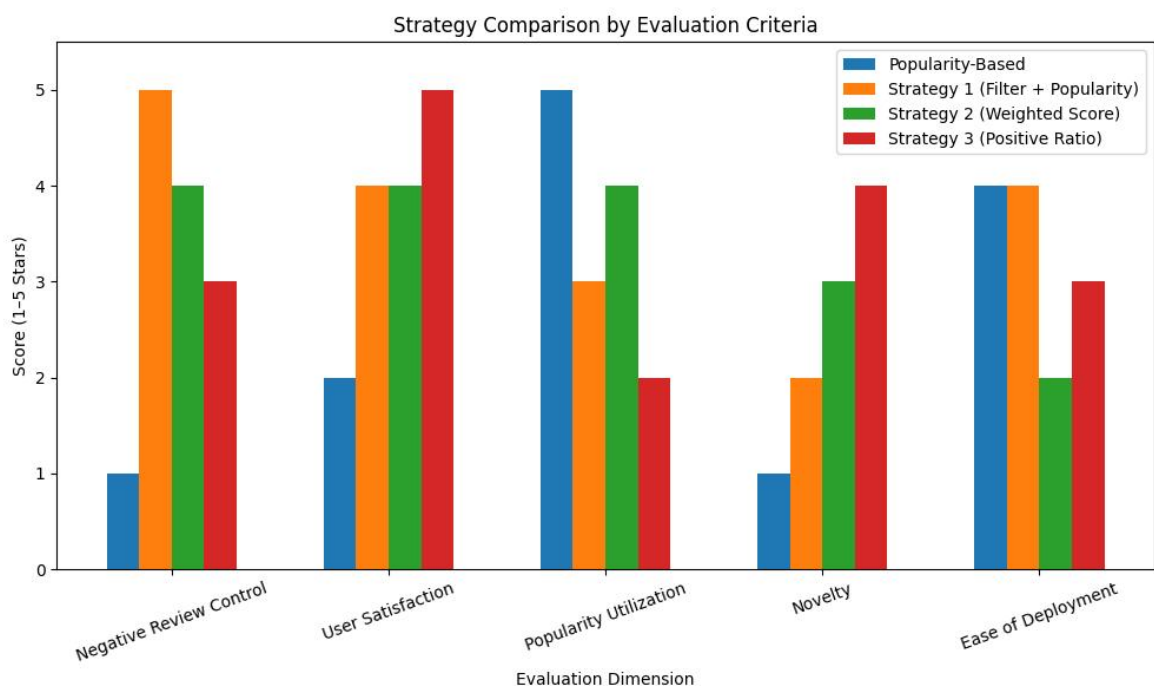


图 4-3 各策略横向对比图

负面评论控制方面（Negative Review Control）：策略 1 在控制负面评价方面表现最佳；基线策略在处理负面评价上效果不佳。用户满意度方面（User Satisfaction）：策略 3 在提升用户满意度上优势明显；基线策略表现较差。热度方面（Popularity Utilization）：基线策略

最优，体现其在利用流行度方面的强大能力。策略 3 最差，在流行度利用上存在不足。发现新商品方面（Novelty）：策略 3 在新颖性方面表现突出；基线最缺乏新颖性。部署难易度方面（Ease of Deployment）：四种策略得分较为接近，基线策略和策略 1 在部署上相对更容易；策略 2 部署难度相对较大。

综合用户体验、平台运营目标与推荐系统策略设计等多个维度，我们对比分析了三种情感驱动策略与基线策略的性能表现，并通过图表进一步揭示其优劣势与适用场景，如表 4-1 所示。

总体来看，不同策略在各评估维度上各有优劣，没有一种策略能在所有维度上都表现出色。在实际应用中，需根据具体需求和侧重点来选择合适的策略。例如，若重点关注用户满意度，可优先考虑策略 3；若看重流行度利用，则基线策略更为合适。在实际应用中，平台可以根据自身的业务目标、用户群特征以及对风险和收益的权衡，选择合适的策略或组合使用这些策略，在不同推荐模块或对不同用户群体采用不同策略。

表 4-1 各策略特点汇总

策略名称	主要机制	侧重点	优点	缺点	适用场景
基线策略	热度排序	热度	简单， 反应大众趋势	忽视口碑质 量，易推劣质 爆款	对比基准，特定 热门榜单场景
策略一	过滤差评， 再按评论 量降序	口碑+热度	有效规避差评 商品，提升用 户信任，实现 简单	阈值设定敏 感，可能过滤 掉小众商品	对平台声誉敏 感，需要保证推 荐结果质量
策略二	加权综合 评分	口碑热度 平衡	灵活度高，可 调整权重适合 不同目标	权重设定需人 为调整，计算 复杂	希望综合考虑质 量和热度时
策略三	优先好评 率	口碑优先	挖掘高口碑商 品，提升推荐 质量感知，增 强新颖发现	偏好评论少的 商品，对冷启 动商品不友好	精品推荐，高分 榜单，面向对质 量敏感的用户， 新品发现模块

5 结论与展望

5.1 研究结论

通过对天猫用户评论数据分析与实验，本研究取得以下主要成果：

（1）验证了情感驱动策略对推荐结果优化的提升

与传统仅基于热度的排序策略相比，情感驱动策略显示出明显优势。其中，策略一有效提升推荐结果的质量下限；策略二实现了热度与口碑的可调节平衡；策略三则显著增强了推荐的新颖性与高满意度。

（2）揭示了不同情感驱动策略的差异化效用与适用场景

本研究不仅证明了情感驱动的普遍有效性，还通过对比分析揭示了不同策略间的细微差别和侧重。这些差异化的表现说明，不存在唯一的“最优”情感策略，选择应依据具体的业务目标、用户群体特征和平台发展阶段而定。

（3）提出了轻量级情感推荐系统构建思路

本研究提出的策略模块简单清晰、部署成本低，工程可实现性和系统集成灵活性高，如图 5-1 所示。

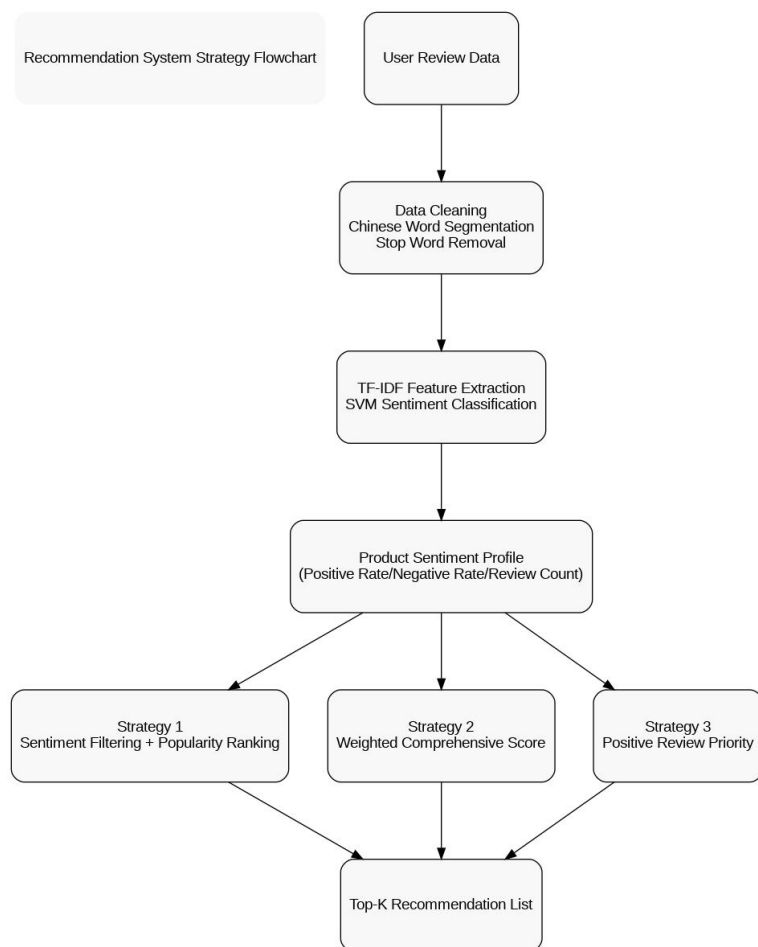


图 5-1 推荐系统策略流程图

5.2 管理建议

基于上述研究结论，我们向电商平台管理者及入驻商家提出以下具有实际操作价值的管理建议：

（1）对电商平台：

平台应开放给商家的“商品口碑仪表盘”，实时展示基于情感分析的各项指标，如好评/差评率趋势、用户关注点热词、与竞品对比等，使数据驱动的口碑管理成为可能，并通过算法设计，让获得持续正面评价的商品和服务能够获得更多的自然流量倾斜和曝光机会，以此正向激励商家专注于提升产品和服务质量，共同维护和提升平台的整体口碑形象。

（2）对入驻商家：

商家应主动利用平台提供的工具或第三方服务，将用户评论的情感分析纳入日常运营监控体系，如同关注销量、流量一样关注用户的情感反馈，及时发现问题和机会。然后深入分析负面评论指向的具体问题，将其作为产品迭代、服务流程优化的直接输入；同时，提炼正面评论中的亮点，强化自身的核心竞争优势，并在营销宣传中加以突出。通过对比分析自身与竞品的情感画像，了解各自的优劣势和用户评价焦点，为制定差异化的市场竞争策略和产品定位提供依据。

对用户的评论要进行及时、恰当的回应。对负面反馈积极沟通解决，对正面反馈表达感谢，有助于化解危机、增强用户粘性，将每一次用户反馈都视为一次建立和巩固客户关系的机会。

5.3 研究局限与未来工作

本研究虽取得一定进展，但仍存在一些局限性，这些局限性也为未来的研究指明了方向：

首先，本研究主要基于特定来源的公开数据集子集，其规模、覆盖品类的广度、时间跨度以及是否能完全代表当前天猫平台的评论生态都可能存在局限，结论的普适性有待在更大、更多样化、更新的数据集上进行验证。除此之外，本文采用的离线评估方式无法完全模拟真实用户在动态交互下的行为反应。

针对上述局限，未来的研究可以向以下几个方向深化和拓展：

首先，选择在更大规模、覆盖更多品类、包含更长时间跨度的真实数据集上进行研究。并尝试使用更高级的预训练语言模型（如 BERT、ERNIE 等）或图神经网络（GNN）等深度学习方法进行情感分析，以期获得更高的精度和更强的语义理解能力。针对离线实验的不足，设计并实施严格的在线 A/B 测试，直接观测情感驱动策略对核心业务指标的实际影响。

总之，将用户评论情感分析融入电子商务推荐系统是一个富有前景且持续发展的研究领域。本研究只是初步的探索和实证结果，未来的研究应在尚有不足之处取得更大的突破，最终实现更智能、更人性化、更值得信赖的商品推荐服务。

参考文献

- [1] 崔婷,黄斐然. 基于情感分析大模型的股票预测:结合 GRU 和 ALBERT 的预测模型[J]. 东岳论丛, 2024, 45(02): 113-123.
- [2] 黄丽凤, 吴宗波. 基于 LSTM 模型循环神经网络的情感分析研究与实现[J]. 普洱学院学报, 2023, 39(03): 34-36.
- [3] 雷克, 刘懿璇. 旅游虚拟社区中在线评论对消费者行为意向的影响研究[J/OL]. 广西职业师范学报, 1-15[2025-04-17].
- [4] 李钰. 融合时间信息的混合推荐算法研究[D]. 江西理工大学, 2024
- [5] 林廷昱. 基于 RFM 模型的协同过滤算法研究——以电商平台为例[J]. 产业创新研究, 2025, (01): 61-64.
- [6] 舒珏淋, 谢红韬, 袁公萍. 基于改进矩阵分解和谱聚类的协同过滤算法[J]. 现代信息技术, 2024, 8(09): 73-76.
- [7] 王魁. 基于隐式反馈的推荐系统数据去噪方法研究[D]. 武汉纺织大学, 2024.
- [8] 王新宇. 基于情感分析技术的景区游客网络评价研究[J]. 科技风, 2024, (26): 154-156.
- [9] 伍蕾, 胡佳, 伍蓓. 数字沉浸式场景下旅游者情感词典建构[J]. 旅游学刊, 2025, 40(02): 74-89.
- [10] 于玉海, 邢志琦, 孟佳娜, 等. 基于多模态桥连接的社交媒体多标签情感分析[J/OL]. 数据分析与知识发现, 1-16[2025-05-10].
- [11] 袁剑锋, 刘佳. 融合物品评价关键词的电商平台推荐算法设计[J]. 电脑知识与技术, 2024, 20(28): 48-51+55.
- [12] 袁里驰. 基于 BiLSTM-CRF 的中文分词和词性标注联合方法[J]. 中南大学学报(自然科学版), 2023, 54(08): 3145-3153.
- [13] 张应峰. 基于改进 BERT 和门控循环单元网络的新闻推荐系统研究和实现[D]. 南京邮电大学, 2023.
- [14] 郑志建, 林土水, 杨蕊平, 等. 基于 LSTM 模型的用户情感分析方法研究[J]. 计算机与网络, 2024, 50(04): 366-370.
- [15] Agarwal A, et al. *Sentiment analysis of Twitter data*[J]. Proceedings of the Workshop on Languages in Social Media, 2011(7): 30-38.
- [16] Batool U, Khattak A, Maqbool F, et al. *From traditional to explainable AI for financial market prediction: A survey*[J]. Journal of King Saud University - Computer and Information Sciences, 2025.
- [17] Cui Y, et al. *ASKAT: Aspect sentiment knowledge graph attention network for recommendation*[J]. Electronics, 2024, 13(1): 216.

- [18] Darraz N, et al. *Integrated sentiment analysis with BERT for enhanced hybrid recommendation systems*[J]. Expert Systems with Applications, 2025, 228: 120312.
- [19] McAuley J, Leskovec J. *Hidden factors and hidden topics: Understanding rating dimensions with review text*[J]. Proceedings of the 7th ACM Conference on Recommender Systems, 2013: 165-172.
- [20] Miriyala S S, Maharjan S, Solorio T. *Can Chain-of-Thought Improve Fine-Grained Sentiment Classification?*[J]. 2025.
- [21] Pang B, Lee L, Vaithyanathan S. *Thumbs up? Sentiment classification using machine learning techniques*[C]. In: Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing, Vol.10, 2002: 79-86.
- [22] Peak S A, Lee W. *Sentiment analysis in cultural studies: a systematic literature review*[J]. Cultural Trends, 2025.
- [23] Rana R R M, Nawaz A, Rehman U S, et al. *BERT-BiGRU-Senti-GCN: An Advanced NLP Framework for Analyzing Customer Sentiments in E-Commerce*[J]. International Journal of Computational Intelligence Systems, 2025, 18(1): 21-21.
- [24] Singh S, Bhatia K K, Kumar A. *TextAI 3.0: Combining Multimodal Sentiment Analysis, Attention Mechanisms, and Deep Learning in Hypersphere Space*[J]. IEEE Transactions on Affective Computing, 2025.
- [25] Turney P D. *Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews*[J]. arXiv preprint cs/0212030, 2002.

致 谢

行文至此，已到终点。原以为遥遥无期的日子，终于在无意间来临。回望来时路，筌路蓝缕，酸甜苦辣咸具在。

感恩母校山西财经大学，望不够，财大的一草一木，一花一叶都在我眼中，财大给了我自由成长的土壤，让我可以茁壮成长。我的青春在这里，我的爱在这里，我的回忆在这里，四年里多有遗憾，但更多的惊喜，是财大给我的安全感。

感恩所有信息学院的老师们，让我成长。想念夏季每一个汗津津的日子，也想念冬季穿着棉服拥挤的教室，想念信息学院每一位老师，没有他们就没有如今的我，我也无法完成此篇论文。曾以为青春难偷，师长教诲永在我心头。

感恩山西这个城市，四年前我带着憧憬和好奇来到这里，而今我深深为山西的文化底蕴着迷，山西人的真诚热情让我逐渐爱上这里。我的山西之旅从 2021 年的钟楼街开始，从 2025 年的壶口瀑布而止，但我希望未来有机会重回故土，以一个崭新的身份。

感恩同学，一路陪伴。夏来方知离别后，才感受立信楼屋顶住过的风，吹过青春，一去再难留。去年冬天，我在室友的欢声笑语中整理好围巾，原来是在整理 6 月的行李。

感恩亲人们，给我鼓励。东北的烈烈寒风吹不到山西，但家人的关心和爱能跨越山海让我感受得到。他们一直是我坚实的后盾，无论我遇到什么困难他们都会给予我无条件的支持和鼓励。四年间多有心情低落，也是他们给予我力量，让我不断前行。

感恩音乐艺术与大自然，感谢 Taylor Swift 和五月天。无数次我疲惫的时候，音乐和大自然是我的解药，这一路走来对我来讲很久，对艺术来讲不算久，对无止境的宇宙来说也不算久。“金融，工程，建筑，计算，固然很重要，那是我们生存的条件；但诗歌，浪漫，爱，是我们生而为人原因。”

感谢编程与代码，此刻我站在无数前辈们的肩膀上，我是多么幸运可以了解到曼妙的计算机世界，我敲下的每一个语句都是前辈们智慧的结晶，我惊叹于代码的奇妙，惊叹于科技的进步，梦想着有天我也能成为其做一点点的贡献，当然，那还要很久很久很久。

感谢自己，从未放弃。论文从初稿到终稿改了无数遍，我一直在想如何把一个包含很多内容的东西讲明白，无论是 NLP 还是推荐系统，我想要尽我最大的能力完成这篇论文。这篇论文就像是我的孩子，我用心血灌溉它，滋养它，期待有天它能成为我想要的样子。即便最后它并不如我所愿，它也是我的一部分，是我思想的诠释，我对其照单全收。

我的本科生涯结束了，但我的人生还在继续。路漫漫其修远兮，吾将上下而求索。我依然要继续为心中的理想而奋斗，这条路也许不都是鲜花和掌声，也会有荆棘和谩骂。路虽远，行则将至。愿心中宇宙，且让少年游。

This was the very first page, not where the story line end.



修德立信 博学求真