

基于电影消费者在线评论的情感分析及网络营销策略研究

曾强青

贵州大学管理学院, 贵州 贵阳

收稿日期: 2025年6月5日; 录用日期: 2025年6月23日; 发布日期: 2025年7月24日

摘要

在互联网时代, 电影评论成为消费者表达观影感受的重要途径, 也为电影行业的创作及网络营销提供了丰富的数据资源。随着电影市场的竞争加剧, 如何有效利用消费者在线评论中的情感信息, 制定精准的网络营销策略成为电影行业关注的焦点。本文以豆瓣网站Top 250榜单中前五部电影的评论文本为研究对象, 运用Python网络爬虫技术收集数据, 并结合TF-IDF算法、词云图和LDA主题模型等文本挖掘方法, 深入分析电影消费者的情感倾向和兴趣偏好, 挖掘评论中隐含的主题信息, 发现观众对电影多数持积极态度并比较关注电影的剧情, 演员、演技等属性特征。基于此, 本文提出针对性的网络营销策略, 优化电影行业的营销手段, 让用户的评论文本数据发挥出最大的价值。

关键词

豆瓣网站, 在线评论, 文本挖掘, 情感分析, 网络营销

Sentiment Analysis and Network Marketing Strategy Research Based on Online Reviews of Movie Consumers

Qiangqing Zeng

School of Management, Guizhou University, Guiyang Guizhou

Received: Jun. 5th, 2025; accepted: Jun. 23rd, 2025; published: Jul. 24th, 2025

Abstract

In the current Internet age, movie reviews have emerged as a crucial platform for audiences to

文章引用: 曾强青. 基于电影消费者在线评论的情感分析及网络营销策略研究[J]. 电子商务评论, 2025, 14(7): 2049-2057. DOI: 10.12677/eci.2025.1472406

articulate their viewing experiences and emotions. Meanwhile, they also serve as a vast treasure trove of data, holding significant value for film creation and online marketing. As competition in the film market grows increasingly fierce, the question of how to effectively harness the emotional cues embedded in online reviews and develop precise online marketing strategies has garnered widespread attention within the film industry. This study focuses on the review texts of the top five movies from the Top 250 list on Douban website. By employing Python web-crawling technology to collect relevant data, and integrating multiple text-mining techniques including the TF-IDF algorithm, word-cloud visualization, and the LDA topic model, we delve into the emotional tendencies and interest preferences of film audiences. Furthermore, we explore the thematic information underlying the reviews. The findings reveal that audiences generally hold a positive attitude toward these movies and show a higher level of concern for certain attribute characteristics of the films, such as plot, actors, and acting skills. Based on these insights, this paper puts forward targeted network marketing strategies aimed at optimizing the marketing approaches in the film industry. The ultimate goal is to maximize the value derived from user-generated comment text data.

Keywords

Douban Website, Online Reviews, Text Mining, Sentiment Analysis, Network Marketing

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

党的二十大报告指出,要坚持以人民为中心的创作导向,打造增强人民精神力量的优秀作品,讲好中国故事,展现中国形象,推进文化自信自强,为电影行业发展指明方向。电影作为重要的文化载体,在文化传播和价值观塑造方面承担着关键使命。在当今互联网蓬勃发展的背景下,电影评论作为观众观影后的情感表达和观点分享,已成为电影文化传播的重要组成部分,同时也为电影网络营销提供了丰富的数据资源。电影评论平台,如豆瓣网,积累了海量的用户评论文本。这些评论文本不仅揭示了用户的情感倾向,还可用于深入分析其购买行为,辅助观众做出更明智的观影选择。同时,可为电影制作方和发行方提供正负面反馈,为电影行业制定精准的网络营销策略,实现对目标受众的精准定位和个性化推荐,增强电影市场的活力和竞争力,以发挥电影评论文本价值的最大化。因此,深入研究电影评论的特性和功能,探索其在网络营销中的应用路径,对于推动电影产业的持续健康发展具有重要的现实意义。

近年来,随着人们日益增长的文化需求和观影热情,电影评论文本的分析研究逐渐成为学术界的热点领域。众多学者运用不同的方法对电影相关数据展开深入探究。涂小琴(2017)使用 PMI 算法对电影的影评进行对比分析[1],徐圣方等人(2024)对影评数据实现了可视化[2],汪韬等人(2018)通过构建情感词典进行文本分析[3],但这些研究主要停留在数据描述和可视化层面。姜霖等人(2017)通过对豆瓣电影评论信息进行细粒度的情感分析,发掘用户偏好,但该研究对评论中隐含主题信息未能充分识别[4]。任力等人(2022)通过分析手工收集整理的 949 部国产电影的微观数据,探讨了主演流量和电影评分对电影票房粘度的影响[5],但该研究未能充分揭示观众情感倾向与票房表现之间的微妙关系。尹梦瑶(2023)基于网络口碑特点,以“锚定效应”为理论视角,通过对《你好,李焕英》进行案例分析,提出网络口碑的锚定效应优化方案[6],但该研究的案例分析可能存在样本规模较小、结论推广性有限的问题。任力等人(2024)运用多元线性回归模型与 Heckman 两阶段方法对电子口碑以及质量感知进行实证分析[7]。然而,该研究

主要关注宏观层面的质量感知,对个体观众的情感倾向和深层次需求挖掘不足。谢治海等人(2020)提出一种基于影评情感类型与强度的自回归票房预测模型,帮助电影行业分析者对一部电影在上映前后的影评情感进行多角度探索与分析,但对如何依据情感分析结果制定具体营销策略指导不足[8]。黄小燕(2025)精准分析目标观众、强化数据收集与分析、打造个性化精准化营销,将成为提升电影票房、推动电影产业发展的重要方向,但该研究对于未充分考量不同电影类型在营销实践中的差异性[9]。

上述研究为电影评论文本的深入挖掘提供了方法支持,也为电影行业的数据分析奠定了基础。但这些研究多在挖掘评论文本的深层次主题信息方面存在一定的局限性,同时对于电影评论与电影营销的关系分析尚不够全面。因此,为深入分析电影评论文本中蕴含的丰富信息,挖掘评论背后的深层次主题和情感倾向,本研究以豆瓣网站为数据来源,采用了 Python 网络爬虫技术来获取豆瓣网站上的电影评论文本。基于此,综合运用了多种文本分析技术,包括 TF-IDF 算法来提取文本特征,通过词云图直观展示评论中的关键词,以及利用 LDA (Latent Dirichlet Allocation, 潜在狄利克雷分配)主题模型进行文本聚类分析。本研究的创新之处在于将 LDA 这一无监督的机器学习算法应用于豆瓣电影评论的文本集合中,以发现和识别评论中隐含的主题信息,为电影产业的网络营销策略和内容创作提供了数据支持和洞察。

2. 数据收集

豆瓣,一个集结了丰富用户生成内容(UGC)的社区平台。起初,豆瓣专注于书籍、电影和音乐领域的内容分享,仅提供与这些作品相关的信息,而所有描述和评论均出自用户之手,并非由网站创始人编纂,这一特色使其在 Web2.0 时代独树一帜,形成了一个独特的创新网络服务体系。鉴于豆瓣拥有庞大的电影评论数据库,本研究选取了豆瓣 Top 250 榜单中的前五部电影的评论文本作为分析样本(如图 1)。这一选择旨在确保数据集既具有代表性,又能满足研究的可行性,从而为论文提供丰富的训练材料。本研究利用 Python 对数据进行重复值、去除无用评论,最终采集到有效评论信息 5328 条。

3. 实证分析

3.1. 基于 TF-IDF 的文本特征提取

TF-IDF (Term Frequency-Inverse Document Frequency)是一种常用的文本特征提取方法,用于评估一个词对于一个文档集或语料库中某个文档的重要性[10]。TF (Term Frequency),即词频,是衡量一个词汇在特定文档中出现次数的指标,它通过将一个词的出现次数除以该文档中的词汇总数来计算得出。一个词汇如果在文档中反复出现,其 TF 值也随之升高,表明该词在文档中占有较高的比重。IDF (Inverse Document Frequency),即逆文档频率,用于评估一个词汇在整个文档集合中的显著性。IDF 的计算方法是对包含该词的文档数量取倒数,然后取对数。一个词如果在少数文档中出现,它的 IDF 值会较高,意味着这个词汇在文档集中较为稀有,对于区分不同文档具有较高的重要性。通过这种方式,IDF 有助于突出在特定文档集合中具有辨识力的词汇。IDF 计算的是词在文档集中出现的文档数的倒数的对数值。对于常见的词,其在文档集中出现的文档数会较大,其 IDF 值较低;而对于罕见的词,其在文档集中出现的文档数较少,其 IDF 值较高。TF-IDF 的计算公式为: $TF-IDF = TF * IDF$ 。通过将词频与逆文档频率相乘,TF-IDF 能够量化一个词对某个文档的重要程度。如果一个词在某个文档中频繁出现(TF 高),并且在整个文档集合中罕见(IDF 高),那么它的 TF-IDF 值就会很高,表明该词对于该文档的重要性较高。

表 1 列出了 TF-IDF 权重值排名前二十的特征词。从表 1 可以看出,“电影”一词权重最高,为 0.253,表明其是电影领域文本的核心词汇,反映了文本主要聚焦电影本身。其后“国产”“观众”等词与电影产业紧密相关,从创作、市场、受众等角度反映文本信息。



图片来源：豆瓣网站。

Figure 1. The top five movies on Douban’s Top 250 list
图 1. 豆瓣 Top 250 榜单中的前五部电影

Table 1. The top 20 text feature words with TF-IDF values
表 1. TF-IDF 值排名前 20 的文本特征词

排名	TF-IDF	权重	排名	TF-IDF	权重
1	电影	0.253	11	喜欢	0.031
2	国产	0.138	12	现实	0.03
3	观众	0.082	13	孩子	0.029
4	题材	0.071	14	不错	0.029
5	流浪	0.063	15	反派	0.029
6	人物	0.051	16	表演	0.028
7	男主	0.046	17	看过	0.028
8	喜欢	0.043	18	女主	0.026
9	原著	0.041	19	不错	0.024
10	感觉	0.037	20	影片	0.023

3.2. 词云图

词云可视化是一种基于语言分析的方法,对大量的文本数据进行词频分析,形成一幅可视化的图像,其中的文字颜色和尺寸代表着该词汇在影视作品中的出现频率,从而将文本数据中蕴含的信息以图片的方式呈现出来。本文通过构建词汇云图来实现词汇的可视化,见图2所示。



Figure 2. Douban movie user review word cloud map
图2. 豆瓣电影用户评论词云图

“电影”在词云图中占据较大面积,凸显了电影本身作为焦点话题的地位,观众围绕电影展开多维度的讨论。从整体来看,其中“真的”“不错”“喜欢”等词汇以较为显著的字体呈现,反映出观众对该电影持有一定程度的正面评价倾向,表明影片在整体上获得了观众的认可。具体到电影的各个构成要素,“演员”“剧情”“特效”等词较为突出,这说明观众对影片中的演员表演给予了较多的关注,并且演员的表现可能是影响观众评价的关键因素之一;同时,“剧情”关注度较高,意味着电影的叙事内容、情节设置等对观众产生了较强的吸引力,能够引发观众的讨论与评价;“特效”作为现代电影技术的重要组成部分,其在词云图中的突出地位也显示出影片的视觉效果对观众有着较好的呈现,进而成为观众评价中的一个重点方面。

3.3. 基于 LDA 主题模型的特征分析

LDA (Latent Dirichlet Allocation)主题模型是一个由文档、主题和词语三层结构组成的复杂系统。该模型基于这样的假设:每篇文档都是由多个主题以不同比例混合而成的,而每个主题则由一系列特定的词语集合构成。在文档生成的过程中,首先是文档基于一定的概率分布选择一个主题,随后该主题根据其词语分布选择一个特定的词,这样反复进行,直至文档构建完成。LDA 模型的应用实际上是上述过程的逆转。它通过分析已有文档的内容,识别出文档中所隐含的主题及其相应的关键词。在这个过程中,文档中的主题分布遵循狄利克雷分布,这是一种多变量概率分布,常用于描述一组随机变量的分布情况。而每个主题中的词语分布则遵循多项式分布,这种分布是离散概率分布的一种,适用于描述一系列可能结果的概率。

文本采用的是基于词语语义特征提取文章主题的 LDA 主题模型,由文档,主题和词语三个层次构成。我们将每一条电影评论看作一个文档,由于电影评论多为短文本,内容有限,每条评论所能传递的主题有限,如果主题个数设置过多,会导致信息分散,所以使用 pyldavis 将困惑度较小的主题数即主题 $k = 4$ 均可可视化(如图3)。

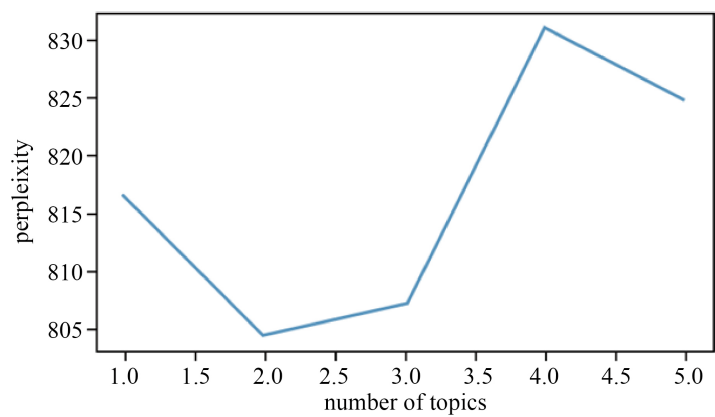


Figure 3. The topic is 4 levels of confusion
图 3. 主题为 4 的困惑度

本文使用 LDA 主题聚类模型得到 4 个主题，并选取了每个主题排名前 8 的特征词及对应权重，结果如表 2 所示。

Table 2. LDA topic classification and subject word probability
表 2. LDA 主题分类及主题词概率

主题 1		主题 2		主题 3		主题 4	
特征词	权重	特征词	权重	特征词	权重	特征词	权重
战争	0.292	演员	0.173	煽情	0.113	剧情	0.108
历史	0.239	角色	0.135	教育	0.095	特效	0.089
社会	0.067	人物	0.097	国产	0.092	不错	0.078
喜欢	0.048	男主	0.078	观众	0.068	感觉	0.072
现实	0.039	女主	0.056	孩子	0.053	原著	0.06
题材	0.024	表演	0.043	流浪	0.047	看过	0.052
英雄	0.020	反派	0.029	导演	0.045	喜欢	0.043
可爱	0.018	配角	0.019	影片	0.045	好看	0.04

文章 LDA 模型通过 jieba 库进行分词，利用中文自然语言处理库 snownlp 进行情感分析，建立词典对文本数据进行训练，最后输出主题。通过对电影评论进行 LDA 主题模型分析，可将观众对电影的评价归纳为四类主题，分别反映出观众对不同维度的关注与思考。主题 1 以“战争”(0.292)、“历史”(0.239)等词为核心特征，表明观众对电影所承载的历史背景、战争题材较为关注，这类电影可能通过展现战争场景、历史事件来传递深刻的社会意义与人文价值，引发观众对过去时代的反思与共鸣，其中“社会”“现实”等词也显示出观众会结合当下社会现实去解读电影中的历史战争元素。主题 2 紧扣“演员”(0.173)、“角色”(0.135)、“人物”(0.097)等关键词，凸显出演员的表演以及对角色的塑造是观众评价的关键维度，演员能否精准诠释角色的内心世界与个性特点，直接影响观众对电影人物的认同感与代入感，同时“男主”“女主”“配角”等词显示观众对不同角色的细致考量。主题 3 围绕“煽情”(0.113)、“教育”(0.095)等词展开，说明部分电影凭借煽情元素触动观众情感，且可能蕴含教育意义，引发观众对成长、人生等命题的思考。主题 4 则聚焦于“剧情”(0.108)、“特效”(0.089)、“不错”(0.078)等特征词，反映出观众

对电影情节设置、特效制作等技术与叙事层面的重视，精彩的剧情推进、精良的特效呈现能提升观众的观影体验，获得观众的认可与好评，而“不错”“好看”等积极词汇也从侧面说明该主题下的电影整体上能够满足观众的期待。综上，此次 LDA 主题模型分析揭示出观众对电影的评价是一个多维度的综合考量过程，既包含对历史题材、演员表演等传统电影要素的关注，也涵盖对情感教育、技术特效等现代电影特质的审视，为电影创作者、研究者以及相关从业者提供了从不同维度洞察观众反馈的视角，有助于深入理解观众的观影偏好与评价逻辑，进而推动电影创作更好地回应观众需求，提升电影作品的品质与影响力。

3.4. 情感倾向分析

Snownlp 是一个基于 Python 的中文自然语言处理库，用于情感分析、文本分类等任务。它的情感分析功能通过对文本进行情感倾向性分析，判断文本是正面情绪还是负面情绪。情感分析原理主要基于机器学习和自然语言处理技术。Snownlp 使用 jieba 库进行中文分词，通过切分文本为独立的词语，为后续步骤提供基本的文本单位。其内置了一个情感词典，其中包含了大量的情感词汇。情感词典中的词汇被标记为正面情感或负面情感，例如，开心、喜欢被标记为正面情感，而悲伤、厌恶被标记为负面情感。情感词典是情感分析的基础，用于识别文本中的情感倾向。在情感词典的基础上，Snownlp 计算文本中情感词汇的情感强度。它通过查找文本中的情感词汇，并考虑它们的上下文环境、程度副词等因素，对情感词进行加权计算。这样可以更准确地评估文本中的情感倾向。基于情感强度计算的结果，Snownlp 对文本进行情感极性判断。如果文本中的正面情感强度大于负面情感强度，那么文本被归类为正面情感；反之，如果负面情感强度大于正面情感强度，则文本被归类为负面情感。

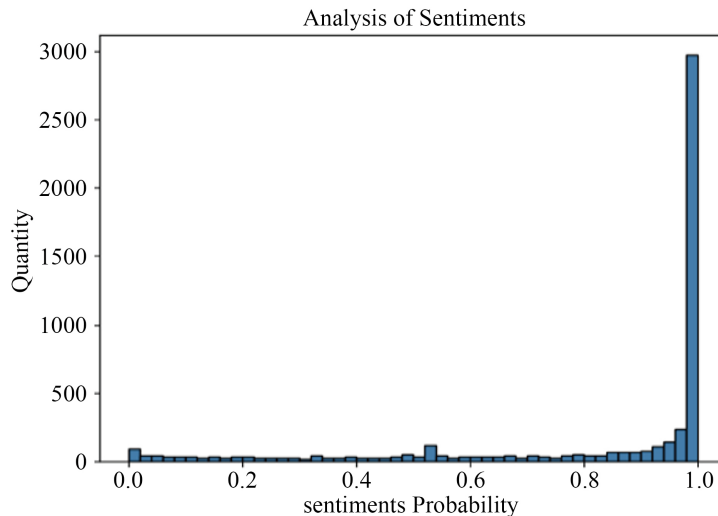


Figure 4. Affective propensity probability distribution
图 4. 情感倾向概率分布

图 4 展示了电影评论的情感倾向概率分布情况，从中可以洞察观众对电影的整体情感态度。从图中可见，情感倾向概率呈现出右偏态分布，大部分数据集中在接近 1.0 的位置，且在 1.0 处形成了一个较为显著的峰值，数量接近 3000。这表明观众对电影的评价普遍偏向积极，对电影整体较为满意，认为其在制作、表演、剧情等多个方面达到了较高的水准，能够满足甚至超出观众的观影期望。在情感倾向概率较低的区域(0.0 到 0.8 左右)，数据分布较为分散且数量较少，这代表了一部分观众对电影存在不同程度

的负面评价，但其影响力相对较弱。这种情感分布特点反映出电影在创作与呈现上取得了较好的成效，其积极正面的情感反馈可以为电影的后续传播与发展提供有力的支持。

4. 基于情感分析的电影网络营销策略

基于上述主题提取及情感倾向分析结果，为强化电影行业的数据利用能力与营销效果，本研究基于互联网的广泛覆盖性、即时互动性以及大数据分析的便利性等特点，提出相应的网络营销策略，并针对不同类型电影的特点进行细化，以提高策略的针对性和有效性。

4.1. 优化宣传推广

电影制作方和发行方可以根据评论中观众对电影剧情、人物、特效等方面的反馈，优化电影的宣传推广内容。对于动作冒险电影，应着重彰显其动作场景的震撼性与冒险情节的扣人心弦，突出影片的视觉冲击力和惊险刺激，吸引观众的目光。剧情片则需深度解析剧情架构与人物情感脉络，引发观众的情感共鸣，使他们对影片产生更深层次的兴趣。喜剧电影宜精选其中的爆笑桥段、幕后花絮以及主演的趣味互动视频进行推广，使观众预先领略影片的诙谐氛围，从而激发他们的观影欲望。而对于科幻电影，宣传重点应放在其独特的科幻世界观、尖端的特效制作工艺以及对未来的前瞻思考上，以满足观众对未知世界的好奇与探索欲。这些策略不仅能够精准地触及目标受众，还能显著提升电影宣传的效率和效果。

4.2. 精准定位目标受众

精准定位目标受众是电影营销的关键环节。电影行业可依据观众的年龄、性别、地域等特征，实施个性化推荐策略。动作冒险电影主要吸引青少年和年轻观众，推广途径包括游戏社区、体育赛事周边等，以实现精准触达。剧情片受众广泛，但可根据主题细分，如青春成长、爱情类剧情片，宜通过社交媒体、校园周边平台推广，吸引年轻观众。喜剧片可依地域特色或幽默元素普适性选择推广渠道，科幻片则针对科幻爱好者，借助特定媒体及科技品牌合作推广。通过这些精准定位策略，电影行业能够更有效地触达目标观众群体，提升营销效果与影响力。

4.3. 开发衍生产品

电影评论能为衍生产品开发提供参考，助力电影的营销。动作冒险电影可以开发与冒险元素相关的模型和装备套装，让观众能够亲身体验影片中的冒险精神。剧情片则可以通过经典台词设计文化衫、徽章等周边产品，满足观众对影片情感和文化内涵的认同需求。喜剧片凭借其幽默元素，适合打造搞笑表情包和玩偶，进一步传播影片的欢乐氛围。科幻片的未来感和科技感可以通过开发具有未来科技元素的模型和智能设备周边来体现，满足观众对未来世界的想象。通过网络营销渠道推广销售这些衍生产品，能够有效拓展电影的商业价值和品牌影响力。

4.4. 与观众建立互动关系

与观众建立互动关系也是提升电影营销效果的关键策略。针对不同类型电影的特点，创新互动方式能够有效吸引目标受众，增强观众的参与感与忠诚度。动作冒险电影可通过线下主题活动、比赛及特效讲座，让观众体验冒险精神与特效魅力；剧情片借助深度剧情讨论、角色扮演和征文活动，激发观众对故事和人物的深入思考；喜剧片利用演员见面会、短视频大赛和即兴表演，营造欢乐氛围并激发创作热情；科幻片则通过主题交流、创作大赛和虚拟现实体验，吸引爱好者参与互动，构建粉丝社区。这些互动举措不仅能够深化观众对电影的情感连接，更能通过口碑传播和二次创作，拓宽电影的市场覆盖面，实现电影营销从单向推广到双向互动的转变，推动电影产业的可持续发展。

4.5. 评估与优化网络营销活动效果

电影行业可以利用评论数据对网络营销活动的效果进行评估和优化。通过对评论中观众对营销活动的正负反馈进行分析,了解观众对不同营销渠道、营销内容和营销方式的接受程度,及时调整宣传内容和推广方式,提高营销资源的利用效率,增强观众的观影体验和满意度,实现电影营销效果的最大化。

5. 结语

豆瓣网凭借其独特的设计理念和良好的用户体验等因素脱颖而出,在电影评论网站中占据了相当大的市场份额。多年来,豆瓣电影积累了大量的电影评论数据,这些数据蕴含着观众对电影的丰富反馈和情感倾向。通过深入挖掘和分析这些评论文本,电影行业可以更好地理解观众需求,从而优化营销策略,提升电影的市场竞争力。未来,随着互联网和数据分析技术的不断发展,基于电影评论文本的网络营销策略研究将具有更为广阔的应用前景。

参考文献

- [1] 涂小琴. 基于 Python 爬虫的电影评论情感倾向性分析[J]. 现代计算机(专业版), 2017(35): 52-55.
- [2] 徐圣方, 王金阳. Python 爬虫获取豆瓣观众影评数据及可视化分析[J]. 网络安全技术与应用, 2024(4): 59-62.
- [3] 汪韬, 张再跃. 面向电影评论的情感词典构建方法研究[J]. 计算机与数字工程, 2022, 50(4): 843-848.
- [4] 姜霖, 张麒麟. 基于评论情感分析的个性化推荐策略研究——以豆瓣影评为例[J]. 情报理论与实践, 2017, 40(8): 99-104.
- [5] 任力, 脱虹. 电子口碑与消费者质量感知——基于中国院线电影市场的分析[J]. 新闻与传播研究, 2024, 31(6): 74-86, 127-128.
- [6] 尹梦瑶. 网络口碑对电影票房的锚定效应分析[J]. 传媒, 2023(21): 44-46.
- [7] 任力, 章阳. 主演流量、电影评分与票房粘性——基于国产电影数据的经验分析[J]. 新闻与传播研究, 2022, 29(4): 93-109, 128.
- [8] 谢治海, 朱敏, 牛红宇, 等. 面向票房预测的影评情感可视分析[J]. 计算机应用研究, 2020, 37(10): 2945-2950.
- [9] 黄小燕. 电影营销策略对票房销售影响的定量研究[J]. 电影文学, 2025(8): 42-46.
- [10] Qaiser, S. and Ali, R. (2018) Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents. *International Journal of Computer Applications*, **181**, 25-29. <https://doi.org/10.5120/ijca2018917395>