

基于文本挖掘的财经视频评论分析与热度研究

——以哔哩哔哩财经类视频在线评论为例

钱立辉, 宋瑾钰*

浙江理工大学计算机科学与技术学院(人工智能学院), 浙江 杭州

收稿日期: 2025年3月14日; 录用日期: 2025年3月27日; 发布日期: 2025年4月29日

摘 要

随着数字经济与新媒体的深度融合, 短视频平台已成为公众获取财经资讯和参与投资讨论的重要渠道。本研究以哔哩哔哩(Bilibili)财经类视频的在线评论为研究对象, 旨在通过文本挖掘技术分析财经视频评论的内容特征和用户情感倾向, 并构建视频热度评估模型。研究方法包括LDA主题模型、SnowNLP情感分析和TF-IDF特征提取技术。通过对高质量评论数据的系统分析, 本研究揭示了财经内容传播的影响机制, 并为创作者优化内容生产、平台完善推荐算法、投资者把握市场情绪提供了实证依据。研究结果表明, 评论中积极情绪占主导, 用户对财经领域整体持乐观态度, 但对具体问题和潜在风险保持警惕。不同主题下的情感倾向差异显著, 高频词分析揭示了用户讨论财经话题时的双重心理。

关键词

文本挖掘, LDA主题模型, 情感分析, 财经视频评论

Analysis and Research on the Popularity of Financial Video Comments Based on Text Mining

—Taking Online Comments on Financial Videos on Bilibili as an Example

Lihui Qian, Jinyu Song*

School of Computer Science and Technology (School of Artificial Intelligence), Zhejiang Sci-Tech University, Hangzhou Zhejiang

Received: Mar. 14th, 2025; accepted: Mar. 27th, 2025; published: Apr. 29th, 2025

*通讯作者。

文章引用: 钱立辉, 宋瑾钰. 基于文本挖掘的财经视频评论分析与热度研究[J]. 电子商务评论, 2025, 14(4): 2483-2495.
DOI: 10.12677/ecl.2025.1441158

Abstract

With the deep integration of the digital economy and new media, short video platforms have become an important channel for the public to obtain financial information and participate in investment discussions. This study focuses on the online comments of Bilibili financial videos, aiming to analyze the content characteristics and user emotional tendencies of financial video comments through text mining techniques, and construct a video popularity evaluation model. The research methods include the LDA topic model, SnowNLP sentiment analysis, and TF-IDF feature extraction technology. Through systematic analysis of high-quality review data, this study reveals the impact mechanism of financial content dissemination and provides empirical evidence for creators to optimize content production, platforms to improve recommendation algorithms, and investors to grasp market sentiment. The research results indicate that positive emotions dominate in comments, and users hold an optimistic attitude towards the overall financial field, but remain vigilant about specific issues and potential risks. There are significant differences in emotional tendencies under different themes, and high-frequency word analysis reveals the dual psychology of users when discussing financial topics.

Keywords

Text Mining, LDA Topic Model, Sentiment Analysis, Financial Video Comments

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

在数字经济与新媒体深度融合的背景下,短视频平台已成为公众获取财经资讯、参与投资讨论的重要渠道。哔哩哔哩(Bilibili)作为以“Z世代”(1995~2009年出生群体)为核心用户群体的综合性视频平台,其财经类视频内容呈现爆发式增长。通过用户发送的弹幕、评论等形成的实时互动数据,不仅反映了受众对经济热点的认知动态,更蕴含着市场情绪传播与内容传播规律的深层信息。如何从海量非结构化评论中提取有价值的信息,已成为优化内容生产、提升平台运营效率的关键。

传统财经内容分析多聚焦于专业媒体报道或机构研究报告,对用户生成内容(UGC)的挖掘尚存空白。相较于结构化金融数据,财经视频评论兼具专业性与大众化特征:一方面包含股票走势、宏观经济等专业话题讨论,另一方面则融合了碎片化表达、网络流行语甚至情绪化观点。这种双重属性使得传统统计分析方法难以有效捕捉语义信息,亟需引入文本挖掘技术实现多维度的内容解构。

现有研究表明,文本挖掘技术在社交媒体分析领域已形成成熟的方法论体系。例如,叶贵等[1]基于新浪微博中的低碳出行博文数据,采用BERT-BiLSTM模型和LDA(Latent Dirichlet Allocation)主题模型对居民的低碳出行行为意向和关注主题进行分析。杜恒波等[2]通过Python语言工具对美国亚马逊网站上的中国图书评论数据进行分析,构建LDA主题模型以探讨中国图书的出口效果。罗向东等[3]以“京东商城”为例,使用LDA主题模型对10万条跑鞋在线评论进行文本挖掘,结合词频共现分析、主题聚类 and 情感分析来研究消费者购买跑鞋的关注点。郭羽婷和姚宣合[4]提出了一种在线健康社区用户评论分析模型,运用LDA主题模型挖掘患者评论的主题,并使用分类模型对患者评论进行主题分类,通过词频筛选、TF-IDF关键词提取以及SO-PMI方法构建领域情感词典,进而计算各个主题的患者评

论文本的情感得分。

基于此, 本文以哔哩哔哩财经类视频评论为研究对象, 构建融合多模态分析的文本挖掘框架: 首先通过 LDA 主题模型识别评论核心主题, 运用 SnowNLP 实现情感倾向量化; 继而采用 TF-IDF 算法提取关键特征词, 结合用户互动行为数据构建视频热度评估模型。通过系统分析 4304 条高质量评论数据, 旨在揭示财经内容传播的影响机制, 为创作者优化内容生产、平台完善推荐算法、投资者把握市场情绪提供实证依据。相较于既有研究, 本文的创新点在于建立“主题-情感-热度”三维分析模型, 突破单一维度文本分析的局限性, 更精准地捕捉财经 UGC 的复杂信息特征。

2. 数据收集以及研究设计

2.1. 数据收集

哔哩哔哩(Bilibili)成立于 2009 年, 是中国领先的综合性视频分享平台, 以“年轻人的文化交流社区”为定位, 尤其受到“Z 世代”的青睐。这一群体具有较高的网络活跃度和消费潜力, 对财经类内容的关注度也日益增加。哔哩哔哩的财经类视频涵盖了宏观经济、股票投资、金融知识等多个领域, 吸引了大量用户的参与和讨论。此外, 哔哩哔哩的评论区互动性强, 用户可以通过弹幕和评论实时表达观点, 这些评论数据蕴含着丰富的信息, 能够反映用户对财经热点的认知动态和情感倾向。因此, 选择哔哩哔哩财经类视频的在线评论作为研究对象, 具有重要的研究价值和现实意义。

通过 Python 语言与 Selenium 工具等爬虫工具的结合, 本文实现了对哔哩哔哩财经类视频评论的自动化采集。共收集到 4304 条高质量的评论数据, 涵盖不同类型的财经视频, 具有较高的代表性, 适用于深入的文本分析研究。

2.2. 研究方法与可行性分析

本研究聚焦于哔哩哔哩财经类视频评论, 运用文本挖掘技术展开深度剖析, 采用 LDA 主题模型、SnowNLP 情感分析以及 TF-IDF 特征提取技术, 旨在从多维度揭示评论中的潜在信息。

隐含狄利克雷分布(LDA)主题模型是 Blei 提出的一种无监督的三层贝叶斯主题模型[5], 作为基于贝叶斯统计的生成式模型, 在文本数据分析领域具有独特优势。其核心假设为每篇文档是多个主题的混合体, 而每个主题又由一系列词汇所构成。通过迭代推断过程, 该模型能够精准识别文档集合中潜藏的主题及其分布态势。LDA 主题模型以多主题建模、无监督学习以及卓越的可扩展性为显著特征, 无需预先进行数据标注, 即可从大规模文本数据中解析出多个主题, 并给出各主题的概率分布。在财经视频评论分析情境下, 该模型可有效识别评论中涉及的主要财经话题, 诸如宏观经济、股票投资、民生经济等, 助力研究者深入洞察用户对不同财经热点的关注程度。

SnowNLP 情感分析作为一款专为中文文本设计的 Python 文本处理库, 在中文文本情感分析方面展现出强大效能[6]。其通过对文本中的情感词汇与语义结构进行深入分析, 进而计算出文本的情感极性分数, 该分数取值范围在 0 至 1 之间, 数值越趋近于 1, 表明文本情感越趋向于正面; 数值越趋近于 0, 则表明情感越趋向于负面。凭借对中文语境的良好适应性、操作的便捷性以及分析结果的高准确性, SnowNLP 情感分析能够有效应对中文文本情感分析的复杂需求[7], 其简洁高效的 API 接口为快速实现情感分析功能提供了便利。该技术可量化评论的情感倾向, 有助于研究者了解用户对财经内容的态度, 进而评估视频对用户情绪的影响。

TF-IDF 特征提取是一种在信息检索和文本挖掘领域得到广泛应用的加权技术。该技术通过计算词语在文档中的频率(TF)以及逆文档频率(IDF), 来评估词语在文档集合中的重要程度[8]。

TF 表示词条(关键字)在文本中出现的频率。

$$tf_{ij} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad (1)$$

其中 n_{ij} 是该词在文件 d_j 中出现的次数, 分母则是文件 d_j 中所有词汇出现的次数总和。为了防止偏向长文件, 通常会将数字归一(一般用词频率除以文章总词数)。

IDF 可以由总文件数目除以包含该词语的文件数目, 再将得到的商取对数得到。如果包含词条 t 的文档越少, IDF 越大, 则说明词条具有很好的类别区分能力。

$$idf_i = \log \frac{|D|}{|\{j: t_i \in d_j\}|} \quad (2)$$

其中, $|D|$ 是语料库中的文件总数。 $|\{j: t_i \in d_j\}|$ 表示包含词语 t_i 的文件数目(即 $n_{ij} \neq 0$ 的文件数目)。如果该词语不在语料库中, 就会导致分母为零, 因此一般情况下使用 $1+|\{j: t_i \in d_j\}|$ 。

因此, TF-IDF 的计算公式为:

$$TF\text{-}IDF = TF * IDF \quad (3)$$

TF-IDF 值越高, 意味着该词语在文档中的重要性越大。TF-IDF 特征提取技术具备精准的权重计算能力、高效的特征提取能力以及良好的可扩展性, 能够有效区分词语的重要性, 避免常见词汇对分析结果的干扰, 适用于大规模文本数据的特征提取任务, 可快速识别出关键特征词。同时, TF-IDF 特征提取技术能够提取评论中的关键特征词, 识别出对财经内容传播具有重要影响的词汇, 为深入理解用户关注的核心内容提供有力支持。

将 LDA 主题模型、SnowNLP 情感分析以及 TF-IDF 特征提取技术应用于哔哩哔哩财经类视频评论分析, 具有高度的可行性与显著的研究价值。从技术层面考量, 这三种技术在各自领域均已得到广泛应用且技术成熟度高, 拥有大量成功实践案例可供参考借鉴。从数据层面来看, 哔哩哔哩平台财经类视频评论数据量庞大且公开可得, 为这三种技术的应用提供了充足的数据基础。LDA 主题模型可有效定位热点话题, SnowNLP 情感分析能够量化用户态度, TF-IDF 特征提取则能精准抓取核心词汇, 三种技术相互补充、协同作用, 可全方位揭示评论中的主题分布、情感倾向以及关键特征词, 为财经内容的传播与优化提供坚实的理论与实践支撑。

2.3. 研究流程

通过爬虫技术从哔哩哔哩平台获取财经视频的用户评论数据是第一位的。利用 Python 编写爬虫程序, 以“财经”为关键词搜索相关视频, 并提取其评论内容。在收集资料的过程中, 对复核资料进行了初步筛选, 剔除了重复的、无效的复核, 最终得出了 4304 份高质量的复核材料。

接下来, 资料预处理收集到的解说文字。通过建立自定义词典, 确保专业术语的准确分词并在金融领域加入专门的词汇。同时, 剔除停用词, 减少无关词汇对分析结果的干扰。利用吉巴分词工具进行精确的分词, 将文字转换成与解析相适应的格式, 用 Jieba 分词工具进行精确的分词。

在数据分析阶段, 使用 TF-IDF 算法对评论中的关键特征词进行提取, 通过词频统计对高频词进行识别。画好词云图, 直观展示复习中的重要词汇分布。进一步, 利用 LDA 主题模型进行主题分析预处理后的数据, 识别主要财经话题及评论。

最后, 利用 SnowNLP 工具对评论进行情感倾向分析, 统计情感分布, 评估用户对财经内容的正面或负面态度。结合 LDA 主题模型分析结果, 确定情感特征词的权重, 深入探讨用户关注的核心问题及其情感倾向, 为财经内容的传播和优化提供实证依据。

流程图如图 1。

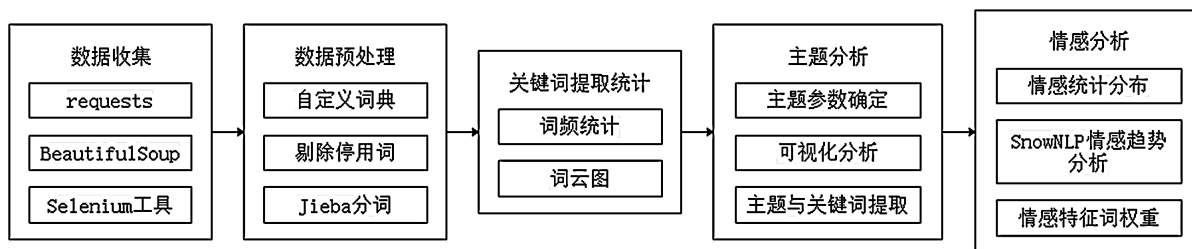


Figure 1. Flow chart of data crawling and text mining

图 1. 数据爬取与文本挖掘流程图

3. 数据分析与结果

3.1. 词频分析

词频分析是有效揭示文中核心信息的语篇分析的关键步骤。本文利用 python 中的 Jieba 分词工具，做分词处理收集到的文字素材。通过自定义词典，扩充分词词库，确保专业术语和特定词汇的准确识别。同时，剔除副词、无意义词语及停用词，并合并同类词，以提高词频统计的准确性。如许多用户评论习惯用表情等非规范性符号，例如“doge”、“金箍”等，需要对这些内容进行删除。

分词后，对文中的词频进行统计，得出每个字的频率。高频词的提取，对课文的主要内容、重点都能有直观的体现。例如，本次分析中，“投资”“市场”“中国”等词汇的高频率出现，突显了文本的核心主题。

为了直观的展示词频分布，使用 python 的 wordcloud 库进行词云图的生成。词云图通过字体大小和颜色，直观呈现高频词的分布，帮助快速识别文本的核心内容。词频统计结果(前 23 名)及词云图见表 1、图 2。

Table 1. Word frequency statistics

表 1. 词频统计

高频词	频数
问题	238
中国	232
经济	225
美国	206
投资	157
市场	156
国家	147
公司	136
消费	128
企业	116
工作	113
赚钱	108
美元	108
金融	105

续表

东西	96
能力	93
世界	90
国内	89
发展	88
社会	84
资本	84
影响	82
股票	82



Figure 2. Bilibili finance video comment high frequency word cloud map
图 2. 哔哩哔哩财经视频评论高频词词云图

通过对文本数据进行词频统计与词云图可视化分析，可以发现高频词主要集中在“问题”“中国”“经济”“美国”“投资”“市场”等词汇上。这些高频词的出现频数显著高于其他词汇，表明文本内容主要围绕全球经济问题、中美经济关系、市场投资动态等方面展开。统计显示，“问题”(238次)、“中国”(232次)及“经济”(225次)权重突出，表明用户对经济议题的聚焦及对国家经济地位的关注，具体关联性需结合主题模型进一步阐释。此外，“美国”“投资”“市场”等词汇的高频出现，进一步印证了文本对国际经济形势、投资机会与市场变化的关注。这些高频词不仅揭示了文本的核心主题，也为后续的深入分析提供了重要依据。

3.2. 基于 LDA 主题模型的特征分析

LDA 主题模型的核心目标是从大量文本数据中自动发现潜在的主题结构。在 LDA 中, k 值起着至关重要的作用, 它决定了模型将把数据分割成多少主题类别, 其大小直接关系到模型的性能和主题的质量。合适的 k 值可以使 LDA 主题模型在复杂性和表现能力之间取得平衡, 从而更准确地挖掘文档中的潜在主题结构, 实现对文档内容的有效聚类 and 主题提取。在评估 LDA 主题模型的性能时, 困惑度和主题连贯性

是两个关键指标, 它们分别从不同的角度衡量模型的质量和有效性。

困惑度是衡量语言模型对数据的拟合程度的指标, 表示模型对未见数据的预测能力。其通过公式(4)计算:

$$\text{Perplexity} = \exp \left(- \frac{\sum_{d=1}^D \log P(\omega_d | \text{model})}{\sum_{d=1}^D N_d} \right) \quad (4)$$

其中 D 是文档数量, N_d 是文档 d 中的单词数量, $P(\omega_d | \text{model})$ 是模型对文档 d 中单词的预测概率。较低的困惑度表明模型对数据的拟合较好, 预测能力较强[9]。

主题连贯性衡量的是主题中单词之间的语义相关性。连贯性越高, 表示主题中的单词在语义上越相关, 主题质量越好。常见的计算方法包括 U-Mass、U-Score 和 NPMI 等, 其中 U-Mass 方法的公式为:

$$\text{Coherence} = \sum_{i=1}^{V-1} \sum_{j=i+1}^V \log \frac{C(\omega_i, \omega_j)}{C(\omega_i) + C(\omega_j)} \quad (5)$$

其中 V 是主题中的单词数量, $C(\omega_i, \omega_j)$ 是单词 ω_i 和 ω_j 的共现次数, $C(\omega_i)$ 和 $C(\omega_j)$ 是单词的出现次数。较高的主题连贯性表明主题中的单词在语义上更相关, 主题更易于理解和解释。

通过代码进行模拟测试后, 可以通过建立图表来综合判断不同主题数量 k 下的模型性能。

Table 2. Model evaluation metrics under different numbers of topics

表 2. 不同主题数量下的模型评估指标

主题数量 k	困惑度	主题连贯性
2	855	0.39
3	720	0.43
4	738	0.37
5	750	0.39
6	774	0.40

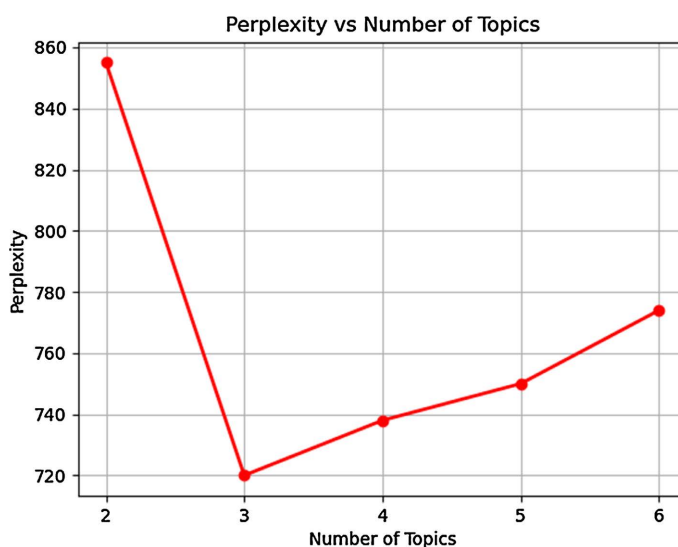


Figure 3. Line chart of perplexity under different numbers of themes

图 3. 不同主题数下的困惑度折线图

LDA 主题模型的最优主题数 k 通过交叉验证确定。如表 2 所示, 当 $k = 3$ 时, 困惑度(720) (图 3)最低, 主题连贯性(0.43)最高, 且主题间语义区分度最大(图 4)。

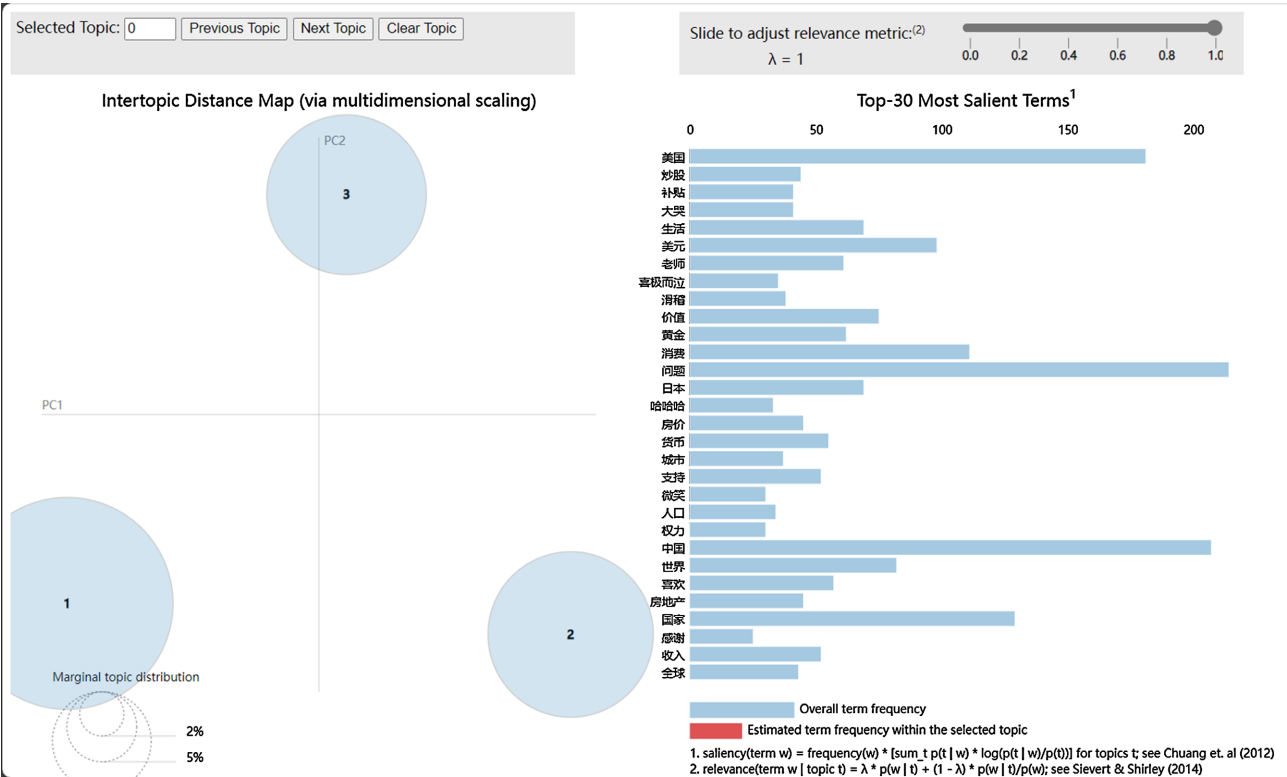


Figure 4. Visualization results of LDA model with theme number $k = 3$
图 4. LDA 模型主题数 $k = 3$ 时的可视化结果

第一类主题聚焦宏观经济, 特征词权重显示讨论重点为国家经济治理。特征词权重显示讨论聚焦国家经济治理。如表 3, “美国” (0.0202)与“国家” (0.0137)权重较高, 反映用户关注中美经贸关系对国内政策的影响, 凸显宏观经济决策的国际关联性; “市场” (0.0115)与“金融” (0.0112)权重接近, 表明用户重视市场机制完善与金融风险防范, 期待“有效市场与有为政府”的协同作用; “问题” (0.0217)权重最高, 指向用户对债务风险、产业升级等问题的担忧, 体现对系统性风险的敏感及政策解决的迫切期待, 视频创作者可据此创作。

第二类主题聚焦国际经济, 特征词呈现多边互动。“中国” (0.0206)与“美元” (0.0202)权重领先, 体现用户对“一带一路”战略效果及美元霸权的双重关注; “消费” (0.0115)与“数据” (0.0090)组合反映数字经济时代用户对全球消费模式及新兴规则(如数字服务税)的兴趣; “世界” (0.0099)与“日本” (0.0080)权重差异显示用户既关注全球化议题(如 WTO 改革), 也聚焦区域合作(如 RCEP); “资产” (0.0070)与“炒股” (0.0056)低权重关联, 表明跨境资本流动议题讨论不足, 或与用户认知局限性相关, 反映用户在该领域认知不足及相对忽视, 创作者可深入探讨创作以提升用户认知。

第三类主题聚焦金融民生, 特征词显示居民经济行为的双重性。“投资” (0.0130)与“赚钱” (0.0112)高相关性凸显用户理财的功利性导向, 与普惠金融发展现实相呼应; “基金” (0.0076)与“黄金” (0.0067)权重差异反映风险偏好分层, 部分用户尝试新理财方式, 部分则偏好资产安全; “企业” (0.0101)与“收益” (0.0052)关联性弱, 可能源于用户对中小企业融资与居民收入传导机制的理解不足; “房子” (0.0063)

权重指向房产政策调控效果关注，体现用户对资产配置角色转变的敏感，创作者可围绕这些金融民生话题制作贴近生活的视频并普及知识以获取流量。

Table 3. LDA topic categories and feature word weights
表 3. LDA 主题类别及特征词权重

宏观经济		国际经济		金融民生	
特征词	权重	特征词	权重	特征词	权重
问题	0.0217	美元	0.0202	收益	0.0052
经济	0.0177	消费	0.0115	投资	0.0130
市场	0.0115	工作	0.0098	赚钱	0.0112
美国	0.0202	数据	0.0090	基金	0.0076
国家	0.0137	资产	0.0070	企业	0.0101
社会	0.0066	炒股	0.0056	美国	0.0202
发展	0.0069	中国	0.0206	中国	0.0206
公司	0.0124	日本	0.0080	黄金	0.0067
金融	0.0112	世界	0.0099	房子	0.0063

3.3. 情感分析

3.3.1. 情感统计分布

情感分析技术通过识别文本中的情感倾向，能够帮助我们洞察用户对财经话题的态度和看法。运用自然语言处理技术进行情感分类和结果分析，揭示用户情绪分布和变化趋势，从而为相关决策提供数据支持，以了解用户对金融政策、市场动态和企业行为的情绪反应。具体评论情感分析如表 4 所示。

Table 4. Statistical results of sentiment distribution in comments
表 4. 评论情感分布统计结果

情绪类型		数量(条)	所占百分比(%)
积极情绪(0.6, 1.0)		2824	65.61%
中性情绪[0.4, 0.6]		371	8.62%
消极情绪(0, 0.4)		1109	25.77%
积极情绪分段统计	一般(0.5, 0.6]	258	5.99%
	中度(0.6, 0.8]	301	6.99%
	高度(0.8, 1.0)	2265	52.63%
消极情绪分段统计	一般[0.4, 0.5)	619	14.38%
	中度[0.2, 0.4)	214	4.97%
	高度(0, 0.2)	647	15.03%

评论数据中，积极情绪占据了主导地位，占比高达 65.61%，其中高度积极情绪(评分 > 0.8)占 52.63%，反映出用户对财经话题的强烈乐观预期。这种积极情绪主要体现在三个方面：其一，用户普遍对经济复苏抱有较高期待，尤其是对亚洲和中国在全球经济中的引领作用持正面看法。例如，有用户评论称：“2024

年经济复苏还得靠亚洲, 东南亚和中国拉了全球一把!” 这表明用户对区域经济协同效应的信心。其二, 用户对特定投资领域表现出显著兴趣与信心, 如新能源、科技和低空经济等。例如, 有用户提到基金经理的专业布局, 并表达了对相关领域的积极预期。其三, 部分用户对政府经济政策表示支持, 认为政策能够有效推动经济发展, 如“国家正在出台政策, 希望把产业链和工作机会留在西部。” 总体来看, 积极情绪的广泛存在反映了用户对经济复苏、投资机会以及政策效果的高度信任。

尽管积极情绪占据主导地位, 消极情绪仍占有一定比例(25.77%), 其中一般消极情绪($0.4 \leq \text{评分} < 0.5$)占比 14.38%, 高度消极情绪(评分 < 0.2)占比 15.03%。这些消极情绪主要集中于以下方面: 其一, 用户对市场波动的担忧较为突出, 特别是对股市和房地产市场的不确定性表示忧虑。例如, 有用户提到: “产业链出海, 意味着资本抛弃了中国, 也意味着中国失业率的不可遏制, 如同美国的铁锈带。” 这种担忧可能源于对全球经济格局变化的敏感性。其二, 部分用户对政策实施效果提出质疑, 认为政策可能难以应对复杂的经济形势。其三, 一些用户对个人财务状况表现出焦虑, 尤其是在房价和投资收益方面的不稳定性感到不安。例如, 有用户分享自身经历, 提到购房后房价飙升带来的压力。这些消极情绪的存在表明, 尽管用户对整体经济前景持乐观态度, 但对具体问题和潜在风险仍保持警惕。

中性情绪在评论中占比相对较低, 仅为 8.62%, 但其客观性和理性特征值得关注。这类评论多为事实陈述或中立观点, 缺乏明显的情感倾向。例如, 有用户针对财经话题进行技术性解释: “农信社和农行不是一家子的, 早些年是农行管理, 后来分出来了。” 这类评论通常以信息传递为主, 较少涉及主观评价。中性情绪的存在表明, 部分用户在讨论财经话题时能够保持理性和客观态度, 避免过度情绪化表达。这一现象不仅为研究提供了更为平衡的数据参考, 也为平台内容生态注入了更多理性声音。

综合分析表明, 用户对财经话题的态度整体乐观, 但消极情绪仍需关注。基于此, 本研究提出以下思考: 首先, 建议动态监测用户情感分布, 特别是在政策调整或市场波动时期, 为决策提供数据支持。其次, 针对政策效果的质疑, 可通过透明化解读与沟通机制, 增强公众理解与信任。最后, 针对个人财务焦虑, 可通过金融知识普及和投资教育, 引导用户理性应对不确定性。这不仅有助于提升用户素养, 也为内容创作者和平台优化提供了参考, 未来可结合更多案例深入验证。

3.3.2. 情感特征分析

为深入挖掘数据, 使用 SnowNLP 模型量化评估各评论的情绪倾向。按标准, 评论分为积极(≥ 0.5)和消极(< 0.5)两类, 分别做词频分析, 揭示不同情感用户表达动机与心理特征, 为财经视频内容优化提供数据支撑。结合 LDA 主题模型提炼的 3 类主题, 用 SnowNLP 分析情感, 结果见表 5。可见, 三类题材情感倾向差异明显。“金融民生”主题情感得分最高(0.9318), 用户对该领域讨论态度积极, 可能因期待政策支持与民生改善; “国际经济”主题次之(0.8365), 体现用户对全球经济动态的关注和信心; “宏观经济”主题得分最低(0.5664), 用户对国内经济发展有担忧或不确定性。这表明财经话题情绪表达具领域依赖性, 不同主题下用户情绪分化。

表 6 结果显示, 积极情绪特征词多集中于国家政策、国际关系及经济领域(如“中国”“美国”“经济”), 反映了用户对宏观层面议题的高度关注; 而消极情绪特征词则更多聚焦于个人财务压力与市场

Table 5. Sentiment orientation of three topic categories
表 5. 三类主题情感倾向

主题	SnowNLP 情感倾向
宏观经济	0.5664
国际经济	0.8365
金融民生	0.9318

Table 6. Consumer sentiment words and TF-IDF weights (top 15)
表 6. 消费者情绪特征词及 TF-IDF 权重(前 15)

积极情绪		消极情绪	
特征词	TF-IDF 权重	特征词	TF-IDF 权重
中国	0.0274	消费	0.0186
美国	0.0261	赚钱	0.0150
问题	0.0241	问题	0.0143
经济	0.0215	工资	0.0136
国家	0.0187	炒股	0.0116
投资	0.0178	东西	0.0113
金融	0.0162	公司	0.0111
市场	0.0152	补贴	0.0110
老师	0.0149	之前	0.0099
世界	0.0139	不行	0.0098
公司	0.0128	企业	0.0094
能力	0.0125	地方	0.0092
工作	0.0123	银行	0.0090
国内	0.0116	成本	0.0088
美元	0.0114	分配	0.0083

风险(如“消费”“工资”“炒股”)。这一对比揭示了用户在讨论财经话题时的双重心理：一方面对国家经济发展抱有信心，另一方面对个体经济状况存在隐忧。

上述情感分布现象不仅揭示了用户对不同财经议题的心理倾向差异，更映射出内容传播与用户情绪的动态关联。用户对不同财经议题的情感差异，既反映了其对民生政策、个人财富等切身议题的显著关注，也表明内容传播中风险信息呈现(如宏观经济风险讨论)显著影响用户情绪倾向。

4. 结语

4.1. 结论

本研究运用 LDA 主题模型从评论中提取出宏观经济、国际经济、金融民生三大核心主题，并借助 SnowNLP 情感分析量化各主题情感倾向。结果显示，评论中积极情绪占主导(65.61%)，其中高度积极情绪占比 52.63%，表明用户对财经领域整体持乐观态度。该积极情绪主要体现为：对亚洲及中国在全球经济引领作用的信任，对新能源、科技等新兴投资领域的兴趣，以及对政府经济政策的支持。然而，消极情绪仍占 25.77%(高度消极情绪占比 15.03%)，主要集中在市场波动、个人财务压力及对政策实施效果的质疑，反映用户对具体问题和潜在风险保持警惕。

进一步分析表明，不同主题下情感倾向差异显著。“金融民生”主题情感得分最高(0.9318)，与国家普惠金融政策及民生保障措施密切相关。这一现象符合行为经济学中的“前景理论”，即用户对直接关乎自身利益的政策(如居民收入、社会保障)更易产生积极预期，因其决策行为受潜在收益的显著影响[10]。相比之下，“宏观经济”主题得分较低(0.5664)，可能源于用户对系统性风险的敏感性；而“国际经济”主题得分次之(0.8365)，则与平台内容偏向全球化议题的议程设置效应相关。

高频词分析揭示用户讨论财经话题时的双重心理。积极情绪特征词多集中于国家政策和经济领域, 消极情绪特征词聚焦于个人财务压力与市场风险。“问题”一词同时出现在积极和消极情绪列表中, 体现用户对部分议题态度的两极化倾向。这反映用户讨论财经话题时的多层次心理, 也为内容创作者和平台运营者提供启示: 传递积极信息同时需正视用户现实关切, 助其理性应对不确定性。

4.2. 策略与建议

4.2.1. 创作者的内容优化建议

创作者可聚焦用户高度关注的宏观经济、国际经济和金融民生主题, 平衡各主题的情绪表达。在解读宏观经济内容时, 采用积极视角可缓解用户对系统性风险的担忧。在保证内容专业性的前提下, 融入趣味性元素, 例如运用案例故事或幽默表达, 这可能会增强内容的吸引力, 进而拓宽受众范围, 提升用户的观看体验。创作者通过制作专题视频、开展直播答疑等形式回应用户对市场波动、个人财务压力等问题的关切, 有助于用户理性应对不确定性, 进而增强用户粘性, 构建创作者与观众间的信任关系。

4.2.2. 平台推荐算法优化措施

平台可以动态监测用户情感分布变化趋势, 尤其在政策调整或市场波动时期, 及时捕捉情绪信号以优化推荐机制。可以优先推送能引发高度积极情绪的内容, 适度降低高度消极情绪内容曝光率, 营造健康社区氛围。在推荐算法中, 提高中性内容的权重, 鼓励理性讨论, 为平台注入客观声音, 这可能会提升内容生态质量, 吸引更多理性用户参与互动, 丰富平台内容的多样性。

4.2.3. 投资者市场情绪把握要点

投资者可以关注评论中积极与消极情绪的分布及变化趋势, 将其作为市场情绪的重要参考。当某类财经话题出现大量高度消极评论时, 可能预示相关领域潜在风险, 提醒投资者留意动态变化。积极情绪特征词如“新能源”“科技”等, 可作为挖掘投资机会的线索。投资者结合这些关键词, 深入分析相关领域市场动态与发展前景, 辅助决策制定, 从而精准把握市场情绪, 从用户讨论中挖掘潜在投资方向, 提升决策科学性与前瞻性。

4.3. 研究局限性

本研究数据来源集中于哔哩哔哩平台, 用户群体以“Z世代”为主, 其观点可能受年龄特征及平台文化影响。不同平台(如抖音、小红书)用户结构及内容生态差异显著, 结论外推需谨慎。未来研究可扩展至多平台、多年龄段样本, 结合纵向数据追踪情绪动态变化, 以增强结论的普适性。

参考文献

- [1] 叶贵, 李长帆, 李晋鹏, 等. 基于社交媒体文本挖掘的居民低碳出行意向分析[J]. 重庆大学学报, 2024, 47(11): 15-22.
- [2] 杜恒波, 王绍运, 罗润东. 基于 LDA 主题模型的中国图书出口效果研究——以亚马逊海外读者评论为例[J]. 商业经济研究, 2024, 45(2): 45-52.
- [3] 罗向东, 强威, 张希莹, 等. 基于文本挖掘的跑鞋用户评价及情感分析[J]. 陕西科技大学学报, 2024, 42(6): 112-118.
- [4] 郭羽婷, 姚宣合. 基于主题挖掘和情感分析的在线健康社区用户评论研究[J]. 东北师范大学学报, 2024, 62(12): 89-96.
- [5] 刘丹璇. 基于文本挖掘的建筑工程质量风险提取研究[J]. 建筑经济, 2023, 44(S2): 169-173.
- [6] 夏海峰, 陈军华. 基于文本挖掘的投诉热点智能分类[J]. 上海师范大学学报(自然科学版), 2013, 42(5): 470-475.
- [7] 王璐, 李诗轩, 陈烨. 基于主题-情感融合分析的新冠疫苗舆情演化研究[J/OL]. 情报科学, 1-27.

- <http://kns.cnki.net/kcms/detail/22.1264.G2.20240129.0935.004.html>, 2025-02-16.
- [8] 杜宇灏, 李环宇, 林晓霞. 基于 TF-IDF 算法的舆情分析研究——以日本排放核废水事件为例[J]. 现代计算机, 2024, 30(23): 108-112.
- [9] Blei, D.M., Ng, A.Y. and Jordan, M.I. (2003) Latent Dirichlet Allocation. *Journal of Machine Learning Research*, **3**, 993-1022.
- [10] Kahneman, D. and Tversky, A. (1979) Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, **47**, 263-292. <https://doi.org/10.2307/1914185>