Artificial Intelligence and Robotics Research 人工智能与机器人研究, 2023, 12(4), 361-372
Published Online November 2023 in Hans. https://www.hanspub.org/journal/airr
https://doi.org/10.12677/airr.2023.124039

Hans汉斯

弹幕文本挖掘与情感分析

刘宇婷*，杨　燕#

南宁师范大学数学与统计学院，广西　南宁

收稿日期：2023年9月25日；录用日期：2023年11月15日；发布日期：2023年11月29日

摘　要

弹幕文本数据的流行，为短文本处理和实时数据处理提供了大量新的文本数据。本文首先对近年来关于弹幕文本的研究进行了系统性梳理归纳，然后基于文本挖掘技术对节目视频弹幕进行深层数据分析，围绕弹幕文本情感分析的关键技术和基本流程进行重点阐述，主要包括通过Python进行文本获取、文本预处理、高频词与词云图可视化、弹幕文本主题词分析、弹幕文本情感分析等多个模块，完成弹幕情感倾向分析，探究弹幕文本数据结构及文本特征，提高弹幕文本情感分析准确度。

关键词

文本挖掘，弹幕情感分析，WordCloud，LDA，SnowNLP

Danmu Text Mining and Emotion Analysis

Yuting Liu*, Yan Yang#

College of Mathematics and Statistics, Nanning Normal University, Nanning Guangxi

Received: Sep. 25th, 2023; accepted: Nov. 15th, 2023; published: Nov. 29th, 2023

Abstract

The popularity of bullet screen text data provides a lot of new text data for short text processing and real-time data processing. Firstly, this paper systematically summarizes the researches on bullet screen text in recent years, and then analyzes the deep data of program video bullet screen based on text mining technology. This paper focuses on the key technologies and basic processes of bullet-screen text emotion analysis, including text acquisition, text preprocessing, visualization

*第一作者。
#通讯作者。

文章引用: 刘宇婷, 杨燕. 弹幕文本挖掘与情感分析[J]. 人工智能与机器人研究, 2023, 12(4): 361-372.
DOI: 10.12677/airr.2023.124039

of high-frequency words and word cloud images, bullet-screen text theme word analysis, bullet-screen text emotion analysis and other modules through Python to complete bullet-screen emotion tendency analysis and explore the data structure and text characteristics of bullet-screen text. Improve the accuracy of emotion analysis of bullet screen text.

Keywords

Text Mining, Barrage Emotion Analysis, WordCloud, LDA, SnowNLP

1. 引言

近年来由于网络视频日益扩张、娱乐内容愈加丰富，观众更加注重线上信息的交互，弹幕应运而生。作为一种新兴互动技术，弹幕在年轻群体中逐渐成为潮流，它被广泛应用于各大视频平台，不断发展、更新，成为了新的“网络舆情传播载体”，并开始对用户决策产生重大影响。相较于偏“理智”、“克制”、“守规矩”的传统式文本发言，弹幕作为一种新媒体盛行下的短文本表达，它以“随意”、“有感而发”的实时评论方式表达了大量的对于当前视频用户的即刻思维认知与即时情感倾向，具有更强情感色彩、时效性和研究价值。

弹幕文本出现初期，弹幕文本情感研究多利用短文本情感分析方法，其研究成果也广泛应用于视频推荐。弹幕文本情感分析是指对短小的弹幕文本进行情感判断，通常包括积极、消极和中性三种情感分类。实现短文本情感分析的方法有很多，其中最常见的是基于机器学习的方法，例如 SVM、朴素贝叶斯等模型方法。这些方法通常需要大量标注好的训练数据，用来训练模型，从而使其能够自动识别文本中的情感倾向。但基于机器学习的方法不但需要大量的标注好的训练数据，而且在数据质量和特征选择上也有一定的要求。同时，模型的性能和泛化能力也会受到数据分布和样本不平衡等因素的影响。因此，在实际应用中，需要不断优化算法和改进数据处理方法，以提高短文本情感分析的准确性和鲁棒性。随着深度学习的发展，一部分学者将神经网络引入到弹幕的研究中。

另外，还有一些基于规则的方法，例如基于情感词典、否定词和程度副词等。这些方法通常需要手工制定一些规则，来判断文本的情感倾向。基于规则的方法具有可解释性和灵活性的优势，但基于规则的方法需要人工定义和调整规则，受到复杂的语义和上下文信息处理限制，因此对领域和语言的适应性相对较低。

然而，无论是基于机器学习的方法、基于深度学习的方法还是基于规则的方法，弹幕文本情感分析的关键在于特征提取。特征提取是指将文本转化为可供机器学习算法或规则引擎使用的特征向量。常用的特征包括词袋模型、TF-IDF、词向量等，这些特征可以帮助机器学习算法或规则引擎更好地理解文本，并进行情感分类。本文对节目视频的弹幕文本展开深度研究，探究弹幕文本数据结构及文本特征，寻求最优的特征提取模型，提高弹幕文本情感分析准确度。

弹幕具有话题开放性、多元化和情绪化等特点，弹幕文本的情感分析，对于多学科、多领域均具有十分重要的研究价值。海量的弹幕数据中蕴含着用户潜意识的行为认知和丰富的情绪价值，通过对弹幕进行文本挖掘、数据可视化、主题分类以及情感分析，不仅有助于视频作品传播、优化节目设置，而且

有助于为网络空间管理提供一些指导方向，例如探究弹幕背后所隐藏的情感倾向和舆情热点，分析追踪网民关注热点，管控热点事件，动态把握网络舆情态势走向，为防范化解网络舆情风险，完善舆情分析机制，构建和谐稳定网络空间做出贡献。

2. 文献回顾

2.1. 弹幕研究

弹幕文本数据的流行，为短文本处理和实时数据处理提供了大量新的文本数据。由于弹幕数据中透露出大量的用户行为认知和情绪价值，众多学者围绕弹幕文本展开相关研究。在研究领域方面，国内有关关研究从传播学领域逐渐发展到教育学、社会学和计算机学科等领域，并朝着跨学科、多学科融合的方向迈进。裴淑红等[1]采用文献研究法和案例分析法，从哔哩哔哩弹幕平台的实际经营视角出发，分析五个盈利模式的共性要素，发现经营问题并提出有效建议；贺思萱[2]则从利用传播学共情理论，分弹幕表现风格、语言结构及表达方式、传播场景三个维度进行深入探究，从而总结出弹幕共情现象中的独特传播特征；刘梦梦[3]以语言模因论为理论依据，通过研究 B 站弹幕的语言结构、传播途径、流行动因三个方面，分析解读 B 站弹幕的流行与传播并寻找语言复制与传播的一般规律；过山[4]通过深度剖析弹幕文本，发现弹幕技术与动漫产业发展之间的相互融合有助于动漫产业的升级与发展；高百慧[5]通过引入弹幕功能，为大学语文课程的在线教学注入新的活力。

2.2. 弹幕文本情感分析

弹幕文本出现初期，弹幕文本情感研究多利用短文本情感分析方法，其研究成果也广泛应用于视频推荐。早期的弹幕文本情感分析方法主要有基于情感词典的方法和基于机器学习的方法。

情感词典的评分赋值是指通过对弹幕文本进行情感词汇的匹配，汇总出情感词条并进行每个文本的分值，就可以得出该文本中的情感倾向。情感词典可通过人工编写、启发式算法来构建，且不同领域的情感词典对文本情感分析结果也有较大的影响。洪庆[6]建立了网络弹幕常用词词典；金丹丹[7]等人使用融合改进的词林构建多维情感词典和改进情感值计算方法对弹幕内容进行情感分析；王文韬[8]融合多种通用词典后构建情感词典交集，对清洗后弹幕文本进行情感分析；付兵[9]通过合并现阶段通用情感词典，进行求并集、分词、词性标注、整合、删选情感词等操作完成了弹幕情感词典的创建；郑飏飏[10]建立了基于情感词典的分析模型。

基于机器学习的方法需要一定数量的训练数据，并且需要对一定规模的数据进行人工标注、标签化。辛雨璇[11]利用贝叶斯分类器将电影短评进行二分类；马梦曦[12]针对弹幕文本碎片化、口语化的特征，构建了基于词频－逆文本频率指数(TF-IDF)与支持向量机(SVM)的情感极性分析模型，可达到使用少量有标签样本即可对大量弹幕评论样本进行情感极性分类效果。

随着深度学习的发展，一部分学者将神经网络引入到弹幕的研究中。陈霞[13]利用神经网络算法及文本挖掘技术对弹幕文本进行多角度情感分析；陈志刚[14]等人针对弹幕文本口语化、网络化、一词多义等特点，引入 BERT-wwm 预训练模型，利用 BiLSTM 提取特征，构建了 BERT-wwm-BiLSTM 模型，有效提高情感分类准确率；曾诚[15]提出一种结合 ALBERT 预训练语言模型与卷积循环神经网络(CRNN)的弹幕文本情感分析模型 ALBERT-CRNN；周卫桐等[16]对直播平台获取的弹幕文本数据进行清洗后，使用 Word2vec 训练得到词向量，并构建了 CNN、BiLSTM 以及改进的 BiLSTM-Attention 三种深度学习的模型进行实验，对比和评估三种模型得到最优的情感分析模型。

由于弹幕文本具有实时性、灵活性、匿名性、交互性等多种鲜明特征，对弹幕领域的情感分析研究仍有待发展。

3. 数据处理与模型构建

3.1. 研究设计

针对传统评论方式相对延迟、偏“理性”的问题，以实时、灵活的弹幕文本为研究对象，通过文本挖掘、情感分析等多种方式探究弹幕文本与网络舆情之间的潜在关联。首先利用网络爬虫技术搜集网络舆情相关弹幕数据，使用 Jieba 库实现分词，去除停用词、去除重复项、无效符号、表情、文本符号、机械降重及高频词统计，基于 WordCloud 库实现词频可视化，并通过 SnowNLP 库计算网络舆情中弹幕的情感得分，运用 LDA 模型进行主题词聚类，实现对弹幕的情感分类和主题分析。本研究利用 Python 编程爬取芒果 TV《声生不息 • 港乐季》节目的视频弹幕信息，对弹幕内容进行文本挖掘、热点话题挖掘及情感分析，以探究这类节目的舆情传播规律和用户情感特征。

3.2. 数据采集

本文的数据来源于《声生不息 • 港乐季》视频的弹幕文本，首先对芒果 TV 网页进行页面分析，找到网页发送弹幕的异步请求包，并分析目标网页的 URL 变化，通过观察分析发现页面遵循的规律，利用变化规律就可以快速实现数据的分段爬取处理。其次对目标网页结构进行分析之后，找到数据的接口，由于网页返回的数据是 JSON 格式，我们可以利用 json.loads 对数据进行直接解析，最后进行数据的存储。其中存储的数据内容包含用户名、评论内容等字段。本文通过 Python 获取了芒果 TV 网页上《声生不息 • 港乐季》全十二期的弹幕，得到弹幕数据共 499,552 条，其中第一期 58,624 条，第二期 32,854 条，第三期 31,867 条，第四期 29,911 条，第五期 33,714 条，第六期 43,984 条，第七期 43,015 条，第八期 36,424 条，第九期 40,763 条，第十期 32,429 条，第十一期 34,885 条，第十二期 56,698 条。

3.3. 数据清洗

Table 1. Raw danmu data
表 1. 原始弹幕数据

	stype	id	uname	content	time	v2_up_count
0	16060732	7090015395617111103		笔笔我来了！	849	46
1	16060732	7090488499150211585		呜呜呜梅艳芳和哥哥	5	
2	16060732	7102294921877124693		何老师的旁白太有感觉了	0	2
3	16060732	7100506720831353348		新生代一个都唔识，哈哈哈	2	
4	16060732	7090144944715893193		安崎我来了	3	
5	16060732	7117112348614071712		何老师的旁白太有感觉了	0	1
6	16060732	7090887879569445480		荣迷来了！！！！！(gorgor 要还在就好了	1000	41
7	16060732	7090148049977253921		毛不易毛不易	1960	174
8	16060732	7090015958257827958		来了！！！我的杨千嬅李健！！！！	1000	30
9	16060732	7094194437454640394		荣迷来了！！！！！(gorgor 要还在就好了	1000	9

通过 Python 爬取获得大量弹幕文本数据，形成初始的弹幕数据集。首先对收集到的弹幕数据进行预处理，清洗空数据项，去除重复项、无效符号、表情、文本符号以及机械降重。然后，利用哈工大停用词表，同时自设停用词表，形成较为完善的停用词表。最后，根据清洗后的弹幕数据集进行分词。原始弹幕见表 1，清洗处理后弹幕见表 2。

Table 2. After cleaning the danmu data
表 2. 清洗后弹幕数据

	id	content	time	count	人物提及
1	7090015395617111103	笔我来了	849	46	周笔畅
2	7090488499150211585	呜梅艳芳和哥	5	0	其他
3	7102294921877124693	何老师的旁白太有感觉了	0	2	其他
4	7100506720831353348	新生代一个都唔识	2	0	其他
5	7090144944715893193	安崎我来了	3	0	安崎
6	7117112348614071712	何老师的旁白太有感觉了	0	1	其他
7	7090887879569445480	荣迷来了	1000	41	其他
8	7090148049977253921	毛不易	1960	174	毛不易
9	7090015958257827958	来了	1000	30	其他
10	7094194437454640394	荣迷来了	1000	9	其他

对清洗后的弹幕进行分类，加入人物提及标签并进行人物提及次数统计，见图 1：

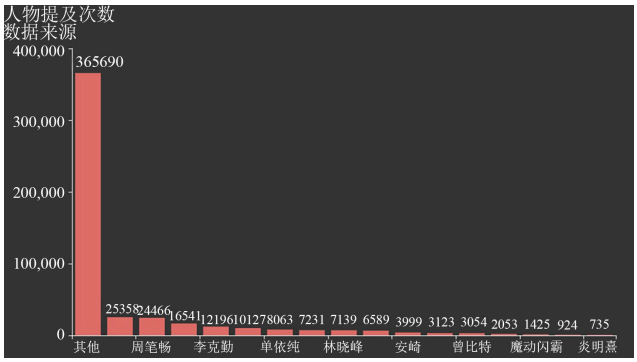

Figure 1. Character mentions
图 1. 人物提及次数

下图图 2 为与“叶倩文”相关的弹幕，共 3817 条：

	id	content	time	count	人物提及
20	7090602397388075478	叶倩文好漂亮哦	3818	51	叶倩文
42	7096513913070007994	叶倩文怎么保养的	6095	1	叶倩文
361	7090453576771098000	林子祥和叶倩文	60151	4	叶倩文
466	7092211073095024661	啊叶倩文	77135	1	叶倩文
496	7092017391544647142	叶倩文	82602	3	叶倩文
...	...	...	...	...	...
497862	7132344677796987859	叶倩文林子祥再见	4163069	1.0	叶倩文
497874	7118774646989120557	叶倩文林子祥再见	4165236	2.0	叶倩文
497969	7118625903681406914	叶倩文真好看	4181607	17.0	叶倩文
498010	7118737839119313168	叶倩文真好看	4188186	3.0	叶倩文
498524	7118937933055865576	再见倩文姐夫妇	4279000	1.0	叶倩文

3817 rows × 5 columns

Figure 2. Ye Qianwen related danmu data
图 2. 与“叶倩文相关弹幕数据”

3.4. 高频词统计

词频，是指一个词重复出现的次数。TF-IDF 是一种常用于词语检索与文本挖掘的加权技术，其优点是能过滤掉一些常见的却无关紧要的词语，同时保留影响整个文本的重要字词。TF-IDF 值越大则重要程度越高。

本研究通过 TF-IDF 得出高频词，表 3 为排名前十的高频词。

Table 3. Top 10 most frequent words
表 3. 排名前十高频词

Word	Count	Frequency	Word	Count	Frequency
好听	256662	0.026019851	可爱	10273	0.010416255
真的	16302	0.016529328	刘惜	10108	0.010248954
李健	15501	0.015717158	健哥	7685	0.007792166
周笔畅	14702	0.014907016	声音	7123	0.007222329
喜欢	13901	0.014094846	粤语	6396	0.006485191

3.5. WordCloud 词云图绘制

本研究通过 Python 绘制词云图，实现高频词的可视化。分别将每一期、全十二期的弹幕分别绘制成词云图，如图 3、图 4 所示。词云图可以直观地体现出弹幕中的焦点和主题，例如“好听”“厉害”“可爱”“喜欢”“感动”等词语。在弹幕词云图中，还可以发现对节目嘉宾关注的相关词汇，例如“李健”“林子祥”“杨千嬅”等。

Figure 3. Word cloud map of the whole twelve phase
图 3. 全十二期词云图

Figure 4. Word cloud map of the eighth phase
图 4. 第八期词云图

对个别嘉宾个人弹幕进行词云图绘制，见图 5：

Figure 5. Lin Zixiang's personal word cloud map
图 5. 林子祥个人词云图

剔除嘉宾名称后的词云图，见图 6：

Figure 6. Word cloud map after removing guest names
图 6. 剔除嘉宾名称后的词云图

4. 弹幕文本分析

4.1. LDA 主题模型

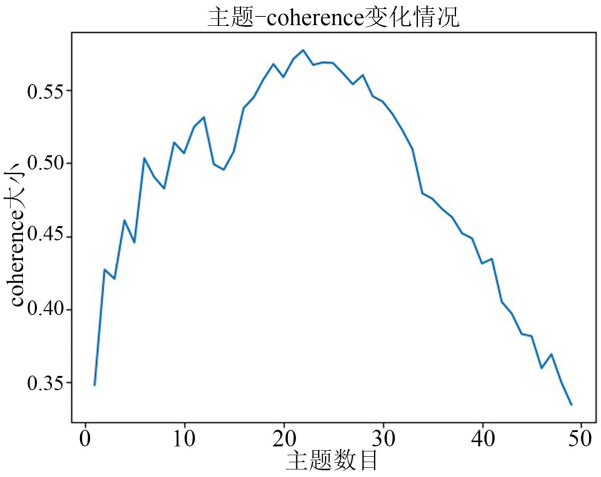

Figure 7. Topic consistency curve
图 7. 主题一致性曲线图

LDA 主题模型是一种文档主题生成模型，该模型假设每篇文章都是以一定概率选择某个主题，然后从这个主题中以一定概率选择一个词语，最后由若干个选出的词语构成。LDA 主题模型在概率潜在语义分析的基础上利用狄利克雷分布得到文档主题和词语的先验分布，并通过 Gibbs 采样来得到文档中的文档－主题分布和主题－词语分布。

利用 LDA 模型对清洗后弹幕文本进行主题聚类，这里的主题分析实验将全部数据进行清洗、分词，然后建立词典、建立语料库，接着进行 LDA 模型训练，最后得到结论，输出结果。

本文使用主题一致性 coherence 作为 LDA 模型评价标准，主题一致性越高，说明模型效果越好。通过绘制主题数目与主题一致性曲线图(见图 7)，选择最佳主题数目。

根据观察图 7 的曲线，可以选择 20 作为最佳主题数，重新设置主题数运行并得出结果见表。获得 20 组主题关键词，选取前五组如下表表 4：

Table 4. Five groups of subject keywords
表 4. 五组主题关键词

第一组	第二组	第三组	第四组	第五组
粤语	节目	健哥	李健	梅姐
有点	广东	比特	标准	闪霸
芒果	孩子	不错	适合	魔动
千嬅	湖南	唱功	看到	这首
普通话	期待值	加油	经典	歌曲
张学友	广州	很多	等到	更好
最好	第一	以为	好多	年轻
综艺	李建	依纯	郑秀文	后面
不会	打卡	可能	完全	认识
大湾	骄傲	居然	前辈	当年

同时运用 pyLDAvis 对 LDA 模型进行可视化得到各主题中词语的贡献度分布情况，见下图图 8、图 9。

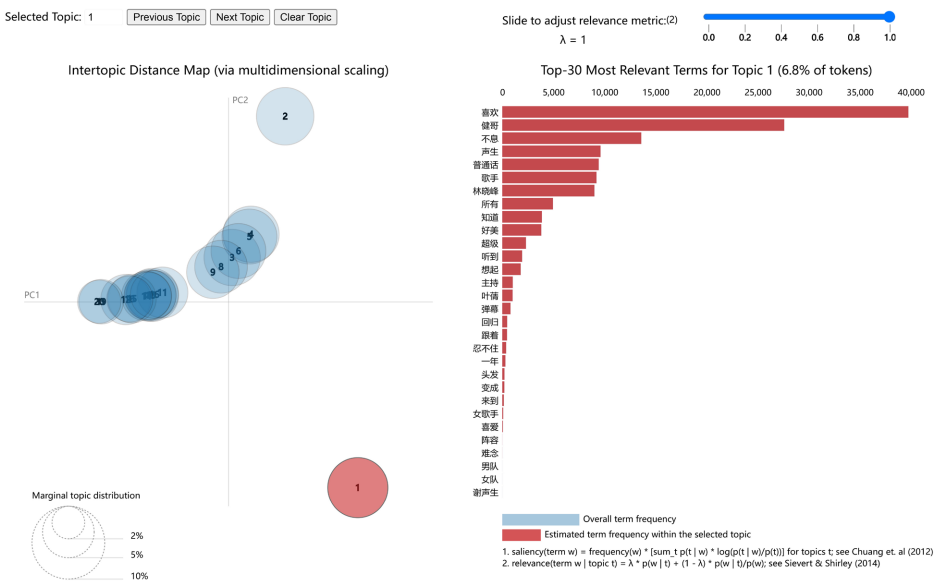

Figure 8. Distribution of contribution degree of the first group of topic words
图 8. 第一组主题词语贡献度分布情况

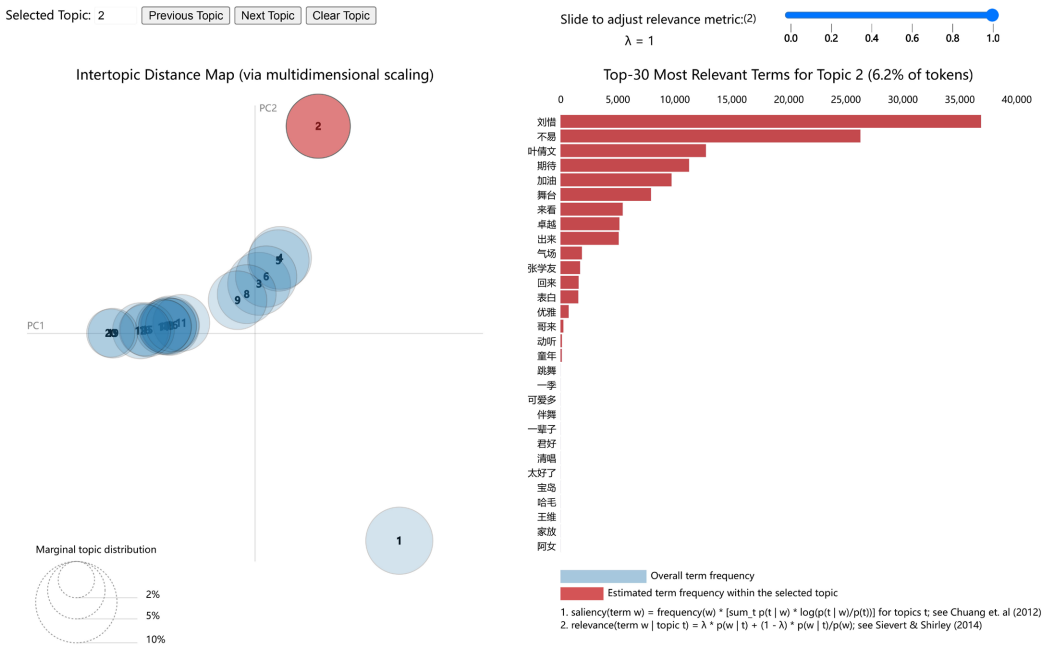

Figure 9. Distribution of contribution degree of the second group of topic words
图 9. 第二组主题词语贡献度分布情况

4.2. 基于情感词典 SnowNLP 情感分析

文本情感分析常用技术常见的文本情感分类方法有两种：一种是基于情感词典的方法，一种是基于机器学习的方法。基于情感词典的方法通常需要建立情感词典，词典里面的词汇越全面准确，情感分析才能越准确。大部分的研究者选择整合多个情感词典来获得更加完备的情感词典，然后结合文本的句法结构对文本进行情感值的计算，根据情感分值判定文本表达的情感倾向。

	id	content	time	count	人物提及	score	情感极性
211203	7103002152668220297	第一	2017	3.0	其他	0.550351	正面
25671	7091256542382538695	其实他不一定要唱粤语	2734026	3	其他	0.750436	正面
387247	7113557777840887477	又看见家驹了	2191000	4.0	其他	0.481961	负面
364648	7118333528077280258	啊	5228183	1.0	其他	0.526233	正面
33448	7090535730905626323	我估	4034425	4	其他	0.500000	中立
199953	7104473939177723066	晓峰独唱真厉害	4426896	1.0	林晓峰	0.581475	正面
215278	7107578225579559228	毛筆組合也很期待	704093	3.0	毛不易	0.733247	正面
31974	7101593820004874084	我觉得这样的节目还是要请有一点有实力的	3787682	1	其他	0.444676	负面
352453	7118356553397003312	输了	3146894	2.0	其他	0.823529	正面
210205	7102803583445014512	广东人民表示很有共鸣	6239000	1.0	其他	0.856198	正面

Figure 10. Emotional tendency of 10 random bullet screens
图 10. 随机 10 条弹幕的情感倾向

本文主要运用 Python 的第三方库 SnowNLP 库对弹幕文本进行情感计算。情感 score 表示了弹幕文本积极的概率，越接近 0 情感越消极，反之，越接近 1 情感越积极。通过 score 将弹幕的情感倾向进行正面

情感、中立情感和负面情感分类，部分弹幕情感倾向见图 10。

绘制弹幕整体情感倾向图，直观了解弹幕的情感变化、走向趋势，见图 11。

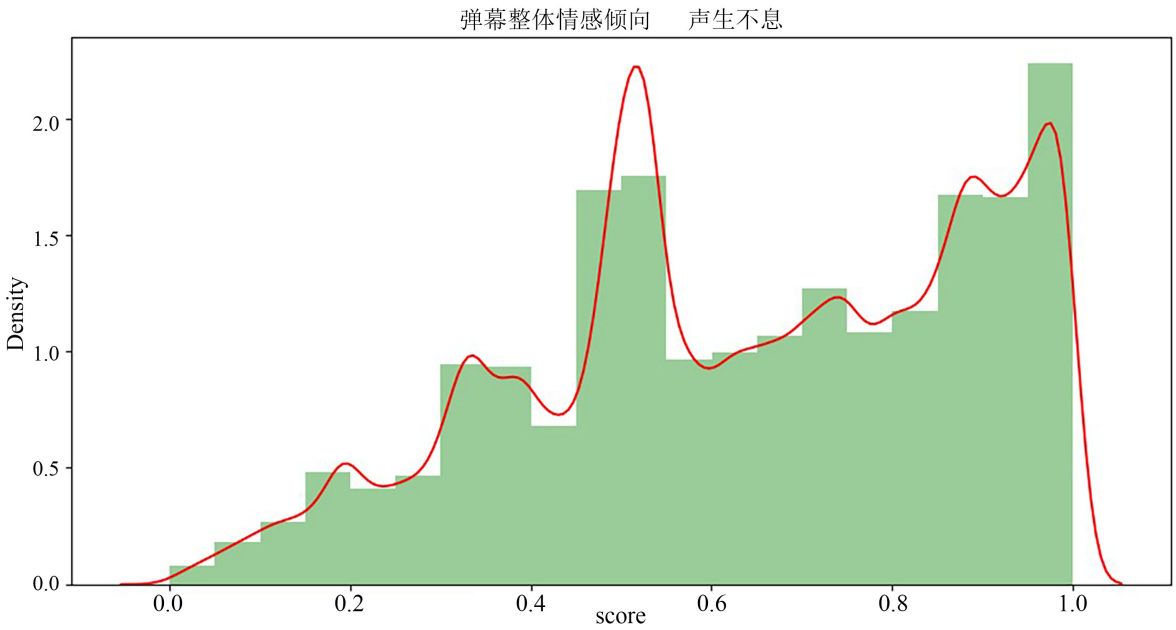

Figure 11. Overall emotional tendency of bullet screen
图 11. 弹幕整体情感倾向图

同时对各个嘉宾个人的弹幕进行情感计算，得出各个嘉宾的情感得分与情感分布，可以更加清晰观众对于嘉宾的喜爱程度，见表 5 与图 12。

Table 5. Emotional values of guests
表 5. 嘉宾情感值

嘉宾	情感得分	嘉宾	情感得分
马赛克乐队	0.297218	毛不易	0.764539
魔动闪霸	0.531607	炎明熹	0.777243
李玟	0.534056	叶倩文	0.807578
李克勤	0.535515	安崎	0.822832
周笔畅	0.544661	林子祥	0.839573
曾比特	0.627316	刘惜君	0.869717
单依纯	0.632644	林晓峰	0.889022
李健	0.683366	杨千嬅	0.918188

将弹幕情感得分划分为两类，分别为积极类(得分在 0.8 以上)和消极类(得分在 0.3 以下)，筛选出两大类分别进行分词。同样通过建立两大类别词典，建立两大类别语料库，其次，分别对两大类别弹幕进行 LDA 模型训练。与上一 LDA 实验结果进行对比讨论，有助于挖掘出观众情感产生的原因。

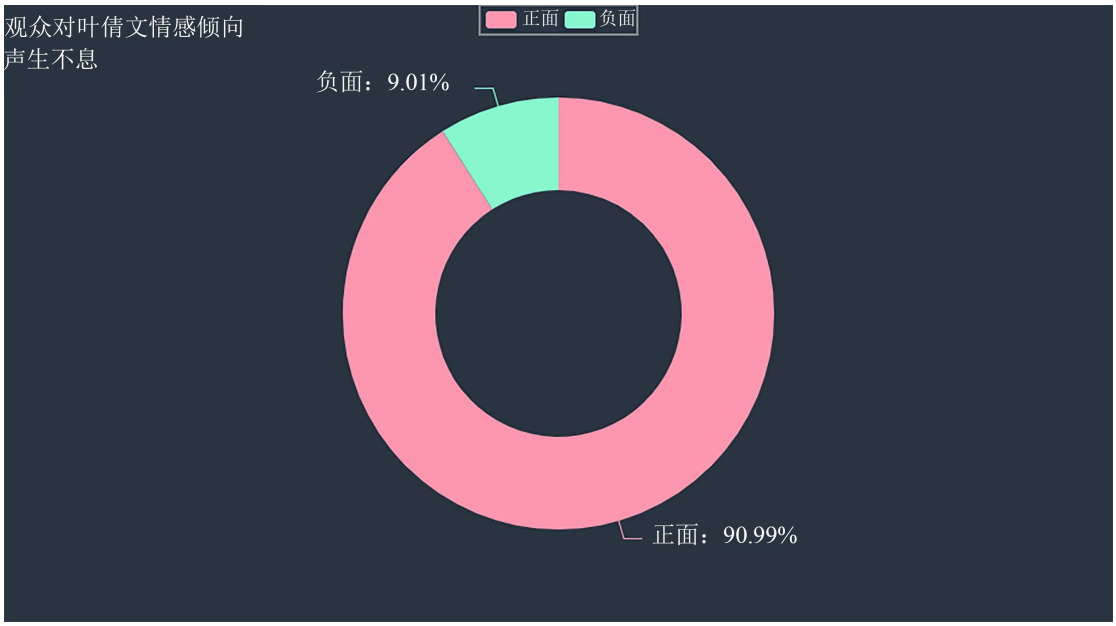

Figure 12. Audience's emotional tendency distribution towards Ye Qianwen
图 12. 观众对叶倩文的情感倾向分布

5. 总结

本文对弹幕文本研究进行了梳理，基于文本挖掘技术对节目视频弹幕进行深层数据分析，爬取弹幕，利用 TF-IDF、WordCloud 等工具进行高频词、词云图可视化，通过 LDA 模型对弹幕进行主题词分析以及利用进行 SnowNLP 完成弹幕情感倾向分析，深入探究弹幕文本大数据背后隐含的深层信息。

如今，国内关于弹幕文本的研究逐渐成为热门趋势，有关弹幕的情感分析、关键词识别等成为研究重点。弹幕中若出现低俗语言、敏感词汇、负面舆情信息，将导致弹幕氛围乌烟瘴气、负面情绪渲染，这不仅仅不利于弹幕文化发展、影响用户体验，而且有碍于弹幕用户实现文化认同，更有碍于动态把握网络舆情态势走向。通过对弹幕中的某些关键词进行适当屏蔽，建立敏感词识别系统监测，可以有效避免过多极端弹幕评论信息瀑布现象，可以有效减少用户发布负面情感弹幕的羊群效应，可以极大降低网络暴力发生概率，可以规范新媒体背景下的网络信息管理，这也是接下来弹幕研究的一大方向。

参考文献

[1] 裴淑红, 余舒琪, 戚少丽. 哔哩哔哩弹幕视频网盈利模式分析[J]. 会计师, 2022(19): 26-29.

[2] 贺思萱. 传播学视域下网络视频的共情传播价值——以四大名著弹幕为例[J]. 新闻前哨, 2022(14): 14-16.

[3] 刘梦梦, 范丽群. 从语言模因视角看 B 站弹幕的流行与传播[J]. 新媒体研究, 2021, 7(20): 86-89.

[4] 过山, 韩存玲. 从弹幕剖析互联网技术创新带动动漫产业升级大数据[J]. 新美域, 2022(11): 76-78.

[5] 高百慧. 大学语文中基于弹幕交互的师生在线教学探究[J]. 汉字文化, 2023(4): 31-33.

[6] 洪庆, 王思尧, 赵钦佩, 等. 基于弹幕情感分析和聚类算法的视频用户群体分类[J]. 计算机工程与科学, 2018, 40(6): 1125-1139.

[7] 金丹丹, 于干. 基于多维情感词典的 B 站视频弹幕倾向性分析[J]. 阜阳师范大学学报(自然科学版), 2022, 39(2): 99-105.

[8] 王文韬, 陈千, 张肖, 等. 弹幕视角下的网络热搜健康视频关注度与情感分析[J]. 图书馆论坛, 2022, 42(3): 98-107.

[9] 付兵. 基于双模态的弹幕情感分析[D]: [硕士学位论文]. 南昌: 江西财经大学, 2022.

[10] 郑飏飏，徐健，肖卓. 情感分析及可视化方法在网络视频弹幕数据分析中的应用[J]. 现代图书情报技术, 2015(11): 82-90.

[11] 辛雨璇，王晓东. 基于文本挖掘的电影评论情感分析研究[J]. 牡丹江师范学院学报(自然科学版), 2021(1): 25-28.

[12] 马梦曦. 基于弹幕文本挖掘的情感极性分析研究[D]: [硕士学位论文]. 武汉：武汉理工大学, 2019.

[13] 陆霞，武善锋. 基于神经网络的在线课堂弹幕评论的情感分析与研究[J]. 无线互联科技, 2021, 18(6): 167-168.

[14] 陈志刚，岳倩，赵威. 弹幕文本情感分类模型研究——基于中文预训练模型与双向长短期记忆网络[J]. 湖北工业大学学报, 2021, 36(6): 56-61.

[15] 曾诚，温超东，孙瑜敏，等. 基于 ALBERT-CRNN 的弹幕文本情感分析[J]. 郑州大学学报(理学版)，2021，53(3): 1-8.

[16] 周卫桐. 基于深度学习的弹幕文本情感分析[D]: [硕士学位论文]. 大连：东北财经大学, 2021.