**探究中国电影市场的观影取向**

**--基于第35届金鸡奖提名获奖电影的弹幕情感分析**

黄杰瑞1)秦子麟1) 宋倍峤2) 陈雨翔3)

王一先4) 罗轶致5) 王艺辰6) 张钰梁7)

1)(华南师范大学附属中学, 广东, 中国)

2)(岱山实验学校, 浙江, 中国)

3)(北京师范大学实验中学, 北京, 中国)

4)(苏州外国语学校, 江苏, 中国)

5)(武汉市光谷实验中学, 湖北, 中国)

6)(北京师范大学昌平附属学校, 北京, 中国)

7)(广州市第八十六中学, 广东, 中国)

**摘 要** 本项目以第35届金鸡奖提名获奖电影为对象，应用Python技术，对主流视频媒体平台的弹幕数据进行情感分析，获得观众弹幕的舆论导向和舆论形象，用于探究中国电影市场的观影取向，给国产电影未来发展方向提供部分参考。分析表明，目标电影的弹幕均以积极的舆论导向为主，钟情于英雄、军人以及真实鲜活的舆论形象。观众对历史、战争重大题材影片和贴近生活的现实主义题材影片表现积极，获奖电影基本上是高票房和主旋律電影。由此可得，中国电影市场的观影取向是积极向上的，未来的国产电影应以弘扬主流价值观、讴歌人性人生的影片为主打方向。

**关键词** 电影；弹幕；Python；情感分析

**Explore the viewing orientation of the Chinese film market**

**--** **the sentiment analysis based on the barrage texts of nominated films at the 35th Golden Rooster Awards**

Jierui Huang1) Tszlun Qin1) Beiqiao Song2) Yuxiang Chen3)

Yixian Wang4) Yizhi Luo5) Yichen Wang6) Yuliang Zhang7)

1)(The Affiliated High School of SCNU, Guangdong, China)

2)(Daishan Experimental School, Zhejiang, China)

3)(The Experimental High School Attached to BNUl, Beijing, China)

4)(Suzhou Foreign Language School, Jiangsu, China)

5)(Wuhan Optical Valley Experimental School, Hubei, China)

6)(Changping School Attached to BNU, Beijing, China)

7)(Guangzhou No. 86 Middle School, Guangdong, China)

**Abstract** This project takes the nominated award-winning films of the 35th Golden Rooster Awards as the object, and uses Python technology to conduct emotion analysis on the bullet screen data of mainstream video media platforms, so as to obtain the public opinion orientation and public opinion image of the audience, so as to explore the movie-watching orientation of the Chinese film market and provide some references for the future development direction of domestic films. The analysis shows that the bullet screen of target films is mainly oriented to positive public opinion, and they are fond of heroes, soldiers and real and vivid public opinion images. The audience showed a positive attitude towards historical and war films and realistic films that are close to life, and the winning films were mostly high box office and theme films. It can be seen that the movie-watching orientation of the Chinese film market is positive, and the future domestic films should focus on promoting mainstream values and euloguizing human life.

**Key words** films; bullet screen; Python; sentiment analysis

**1** 引言

随着网络的迅速发展和大数据时代的到来，互联网用户已发生了巨大变化，他们已从过去那种被动接受外界信息的人群，变成了主动在网上表达观点和分享心得的群体。近几年，弹幕视频与直播行业蓬勃发展，弹幕已经变成了用户之间在网络上互相交流与讨论的常用工具。弹幕作为一种新媒体时代下的短信息表达方式，以实时评论的方式表达了用户对于当前视频的即刻认知与行为倾向，相比于传统评论方式具有更强的情感色彩和时效性，这对于网络舆情情感分析研究具有独特的研究价值。目前，针对弹幕文本情感分析的主要方法有基于情感词典的方法和基于机器学习的方法[1]。通过对弹幕内容进行数据可视化、情感分析以及主题分类，有助于动态把握网络舆情态势走向，追踪网民关注热点，寻找弹幕背后所蕴含的情感倾向和舆情热点[2]。目前，针对弹幕评论数据的研究并不多，弹幕数据价值尚未被充分挖掘。

在中国经济、文化大发展，国民的精神消费日益增长的背景下，大众对影视作品的需求也在不断变化。什么样作品更能迎合大众口味，获得高票房？什么样的作品能得到专家青睐，专家评选结果和国人的观影取向有何异同？国产电影如何实现名利双收？这些问题引起了我们的兴趣，着手探究中国电影市场的观影取向，希望给国产电影未来发展方向提供参考建议。

2 研究设计

应用Python爬虫和情感分析技术，针对主流视频媒体平台的弹幕数据，分析观众弹幕的舆论导向和舆论形象。恰逢第35届中国电影金鸡奖揭晓，我组选取了最具有代表性的6部最佳故事片提名影片的观众弹幕内容进行情感分析探究活动。

3 实现流程

3.1数据源

从爱奇艺、腾讯视频、芒果TV等观影平台进行数据采集，是因为这些平台活跃用户群体大，弹幕数量丰富且具有较好的包容性。

3.2 数据获取

3.2.1 爬虫

通过对目标网页数据来源地解析，使用Python的requests库中的requests.get()向视频地址发送请求并用re库中的re.findall()解析出cid号,再根据cid号拼出xml接口地址。最后解析得到的xml页面获取原始弹幕数据[3]。获取数据原代码参见附录A。

3.2.2 清洗、存储

剔除无关数据，并将弹幕数据进行存储。通过剔除原始弹幕文本中的表情符号、数字、空白值等无效信息，可以有效提升数据质量。

3.3 情感分析

将有效数据进行可视化处理。

3.3.1 舆论导向

用SnowNLP给弹幕内容打标:积极、消极，并统计占比情况，绘制饼图。原始数据见附录B。

3.3.2 舆论形象

用jieba.lcut()分词，并统计高频词；用WordCloud库中的cloud()绘制词云图。原始数据见附录C。

3.4 文字结论

基于6部提名获奖电影的弹幕情感分析，结合专家评审结果及市场票房情况，提出分析结论。

4 分析与结论

4.1《长津湖》弹幕情感分析结论

4.1.1 舆论导向

积极——79.3%、消极——8.3% （图1a）

4.1.2 舆论形象

高频词——中国、致敬、英雄、军人、伟大、好、真、可爱（图1b）

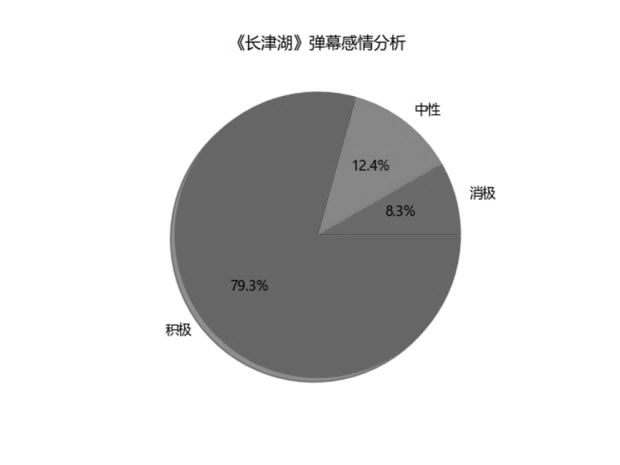
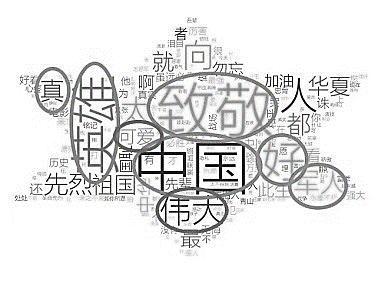
 

图1a 《长津湖》弹幕饼图 图1b 《长津湖》弹幕词云图

4.1.3 结论

《长津湖》是一部以历史、战争为题材的主旋律的电影，商业性、教育性和口碑都达到最高峰；占据目前中国电影票房最高纪录，近58亿；获得最佳电影和最佳导演这两个重要的奖项获，实至名归。

4.2 《狙击手》弹幕情感分析

4.2.1 舆论导向

积极——70.4%、消极——12.5% （图2a）

4.2.2 舆论形象

高频词——致敬、英雄、先烈、中国川军、伟大、好、真、可爱、漂亮（图2b）

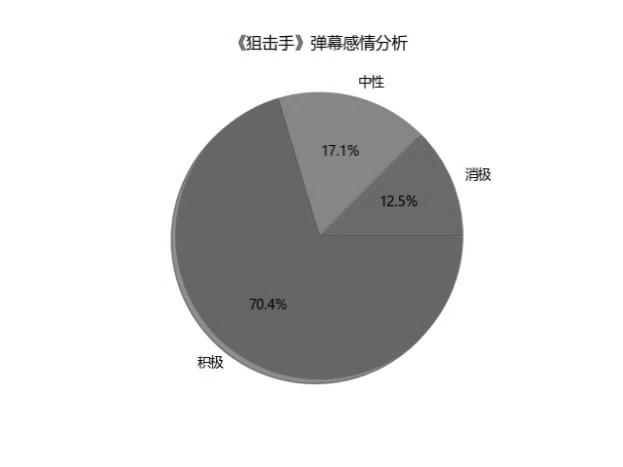
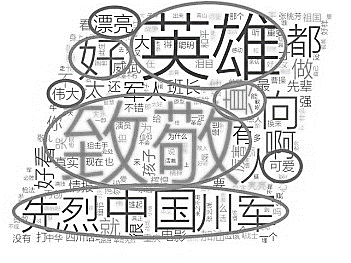
 

图2a《阻击手》弹幕饼图 图2b《阻击手》弹幕词云图

4.2.3 结论

《狙击手》是一部以历史、战争为题材的主旋律电影，闪耀着爱国主义精神和人性的光辉；6亿票房，获最佳摄影和最佳录音两项技术大奖。

4.3 《我和我的父辈》弹幕情感分析

4.3.1 舆论导向

积极——57.5%、消极——20% （图3a）

4.3.2 舆论形象

高频词——致敬、祖国先辈、繁荣昌盛、哈哈哈、好、真、哭（图3b）

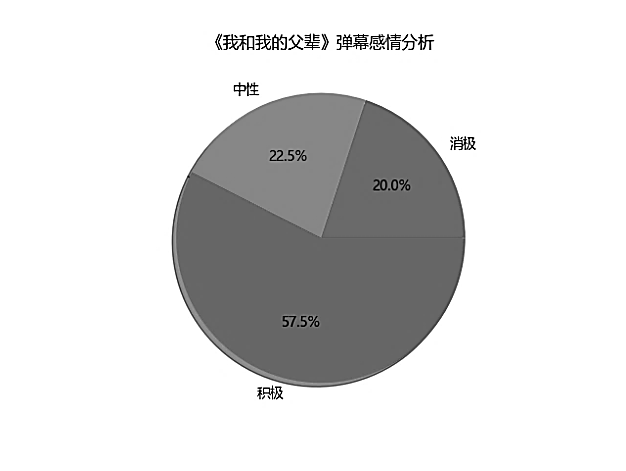
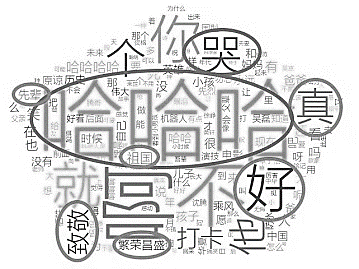
 

图3a《我和我的父辈》弹幕饼图 图3b《我和我的父辈》弹幕词云图

4.4.3 结论

《我和我的父辈》是一部以历史事件为题材的主旋律电影,讲述各时期的代表性群体们不忘初心、砥砺前行，传承民族精神的奋斗历程；15亿票房，荣获评委会特别奖。

4.4 《人生大事》弹幕情感分析

4.4.1 舆论导向

积极——61.4%、消极——20% （图4a）

4.4.2 舆论形象

高频词——三哥（朱一龙）、真、好、可爱、笑、演技、武汉（图4b）

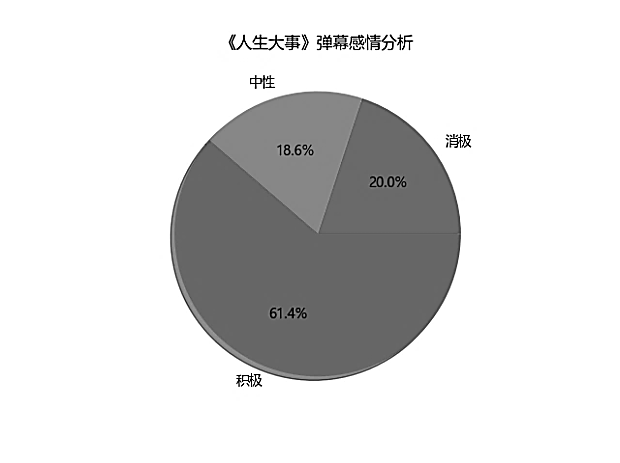
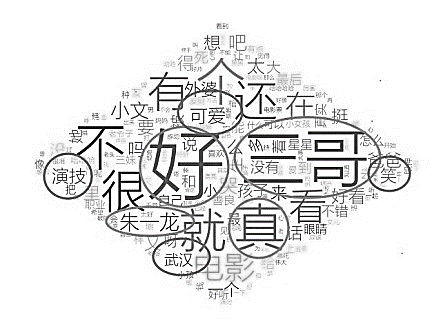
 

图4a《人生大事》弹幕饼图 图4b《人生大事》弹幕词云图

4.4.3 结论

《人生大事》是一部家庭题材的主旋律电影，贴近生活，带有浓郁的（武汉）市井气息，同时也蕴含着一定的哲理意味，让观众在哭过笑过之后，深思人生的价值意义；票房黑马（17亿），获得最佳导演处女作奖及最佳男主角奖。

4.5 《奇迹**.**笨小孩》弹幕情感分析

4.5.1 舆论导向

积极——66.2%、消极——15.8% （图5a）

4.5.2 舆论形象

高频词——（易烊）千玺（景浩）、真、好、可爱、努力、成功（图5b）

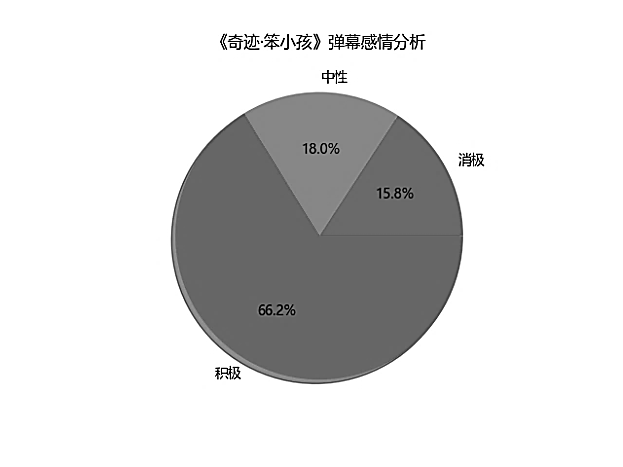
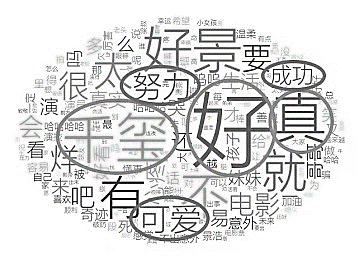
 

图5a《奇迹·笨小孩》弹幕饼图 图5b《奇迹**·**笨小孩》弹幕词云图

4.5.3 结论

《奇迹·笨小孩》是一部奋斗题材的主旋律电影，反映新时代年轻人在深圳创业逐梦人生的精神风貌；票房黑马（超13亿），获最佳女配角奖。

4.6 《爱情神话》弹幕情感分析

4.6.1 舆论导向

积极——59.5%、消极——19.9% （图6a）

4.6.2 舆论形象

高频词——上海、爱情、真、好、可爱、哈哈哈（图6b）

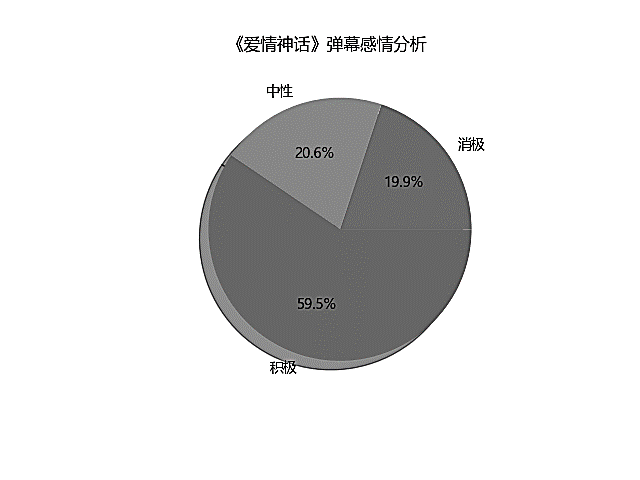
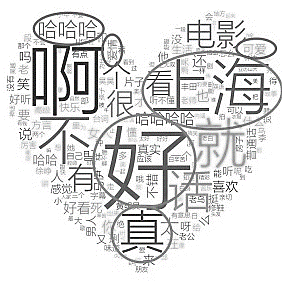
 

图6a《爱情神话》弹幕饼图 图6b《爱情神话》弹幕词云图

4.6.3 结论

《爱情神话》是一部爱情题材的文艺电影，将视角深入到上海弄堂阁楼中，近距离观察都市真实的鲜活生活，讲述了笑中带泪的爱情、友情、亲情故事，成为爱情片的新标杆；2.6亿票房，获得最佳编剧和最佳剪辑两项大奖。

5 总结

本项目在各位成员的共同努力下顺利开展，通过不断摸索改进、研究分析，汇总形成以下结论：

1、获奖电影的弹幕均以积极的舆论导向为主，说明观众对影片的认可，钟情于英雄、军人、真实鲜活的老百姓等舆论形象。

2、获奖电影基本上是高票房和主旋律电影，观众对历史、战争重大题材影片和贴近生活的现实主义题材影片表现积极，影评专家的意见与国人的观影取向趋于一致。

3、中国电影市场的观影取向是积极向上的，未来的国产电影应以弘扬主流价值观、讴歌人性人生为主打方向。

#### 参考文献

1. 朱思淼，魏世伟，魏思恒，余敦辉.基于弹幕情感分析和主题模型的视频推荐算法[J]. 计算机应用,2021,41(10):2813-2819.
2. 刘娟.大数据时代的计算机网络安全及防范措施[J].无线互联科技,2021,18(14):31-32.
3. 白健，洪小娟. 基于弹幕的网络舆情文本挖掘与情感分析[J]. 软件工程，2022，25 （11）：44-48.

#### 附录 **A** 弹幕文本的爬取证明

[获取弹幕文本的Python 代码](file:///D:\桌面\AI\2022AI探学营\自然语言处理\弹幕情感分析项目\成果\提交文件\附录文件\附录A)

#### 附录B 弹幕文本的情感分析和饼图证明

[SnowNLP打标弹幕文本和绘制饼图代码](D:\\桌面\\AI\\2022AI探学营\\自然语言处理\\弹幕情感分析项目\\成果\\提交文件\\附录文件\\附录B)

#### 附录C弹幕文本词云图分析证明

[WordCloud绘制词云图代码](file:///D:\桌面\AI\2022AI探学营\自然语言处理\弹幕情感分析项目\成果\提交文件\附录文件\附录C)