

学校代码： 10255

学 号： 2201230

中图分类法： F272



学术学位硕士学位论文

融合弹幕及字幕情感的电影视频播放量影 响因素研究

学位申请人： 雷宇

指导教师： 王扶东

一级学科： 管理科学与工程

研究方向： 文本情感分析

所在学院： 旭日工商管理学院

提交日期： 2022 年 11 月

University Code: 10255

Student ID: 2201230

CLC Index: F272



ACADEMIC MASTER DISSERTATION

RESEARCH ON THE FACTORS AFFECTING THE PLAY VOLUME OF MOVIE VIDEO COMBINING BULLET SCREEN AND SUBTITLE EMOTION

Author: LeiYu

Supervisor: Prof. Wang Fudong

First-level Discipline: Management Science and Engineering

Research Field: Text emotion analysis

College: College of Donghua

Date of Submission: Dec, 2022

硕士学位论文答辩委员会名单

姓名	职称	单位	备注
段永瑞	教授	同济大学	主席
沈滨	教授	东华大学	委员
王长军	副教授	东华大学	委员

答 辩 人：雷宇

答辩地点：线上答辩

答辩日期：2023.1.5

融合弹幕及字幕情感的电影视频播放量影响因素研究

摘要

近两年,我国电影行业迅猛发展,也在发展中产生了新的趋势和特点,尤其是电影从院线下映后在网络视频平台上播放已经成为一种趋势,并且以爱奇艺、腾讯视频、优酷等为首的各大视频网站已经积累沉淀了大量的网络视频忠实用户,视频的播放量对这些网络视频平台的发展也尤为重要。同时在互联网背景下,视频播放量的影响因素不仅与视频本身特点有关,视频平台的新颖用户互动形式的作用也不容忽视,其中以弹幕和评论为主,且字幕与弹幕的情感关联性影响也非常值得研究,因为弹幕可以直观了解用户在观看视频时的情感,而通过字幕可以大致判断剧情的发展方向,对两者进行结合研究也更能还原用户在观看过程中的情感波动。因此,本文认为有必要结合传统影响因素以及视频平台的新型互动特点,对电影视频在平台上的播放量影响因素进行研究。

基于以上背景,本研究依据 SOR 理论、感知风险理论及感知价值理论,结合电影本身特点以及网络视频平台的互动特点,对电影视频播放量的影响因素进行了研究。首先基于以上理论构建了理论模型,提出了相关假设,包括电影类型、主创阵容、续集或 IP 改编、电影评分、电影票房、弹幕、评论以及弹幕与字幕情感关联性对电影视频播放量的影响假设;其次对假设中涉及的弹幕、评论、字幕等文本因素的度量进行了研究,同时考虑到弹幕是随着视频播放的时间轴滚动加载的,所以提出了结合情绪感染机制及衰减效应的弹幕情感值修正方

法，并结合自然语言处理等文本分析技术对弹幕、字幕及评论中所包含的用户情感进行分析研究。

最后，本研究以爱奇艺平台在 2020-2021 年间的电影视频相关数据为例，对提出的电影视频播放量影响因素的假设进行验证，研究方法包括多元线性回归分析以及主体间效应检验分析。研究表明：弹幕、评论的数量以及票房对电影视频播放量有正向影响，且电影类型对播放量的影响也较为显著，其中惊悚类型的电影对播放量会产生显著负向影响。同时评论的异质性会在一定范围内对播放量产生正向影响，并且当评论数量较少时，负向评论对播放量的影响较小；用户更偏好与弹幕与字幕情感为弱正相关或负相关的电影视频，往往因为这类电影视频具有情节的反转性或讽刺意味，能满足用户的好奇心及观看欲望。

关键词：电影视频；播放量；弹幕情感；情绪感染；情绪衰减

RESEARCH ON THE FACTORS AFFECTING THE PLAY VOLUME OF MOVIE VIDEO COMBINING BULLET SCREEN AND SUBTITLE EMOTION

ABSTRACT

In the past two years, China's film industry has developed rapidly, and new trends and characteristics have also emerged in the development. In particular, it has become a trend for films to be played on online video platforms after they are released from cinemas. In addition, a large number of loyal users of online video have been accumulated on major video websites led by iQIYI, Tencent Video, Youku, etc. The amount of video played is also particularly important for the development of these online video platforms. At the same time, in the context of the Internet, the influencing factors of video playback are not only related to the characteristics of the video itself, but also the role of the new user interaction form of the video platform can not be ignored, which is dominated by bullet screens and comments, and the emotional relevance of subtitles and bullet screens is also worth studying, because bullet screens can clearly understand the audience's likes and dislikes of the video, and subtitles provide a basic judgment of the good or bad trend of the plot, The combination of the two can also restore the emotional fluctuations of users in the viewing process. Therefore, this paper believes that it is necessary to combine the traditional influencing factors and the new interactive characteristics of the video platform to study the influencing factors of movie video playback on the platform.

Based on the above background, this study, based on SOR theory,

perceived risk theory and perceived value theory, combined with the characteristics of the film itself and the interactive characteristics of the online video platform, studied the influencing factors of movie video playback. First, based on the above theories, a theoretical model is constructed and relevant assumptions are put forward, including the assumptions about the impact of movie types, main creative lineups, sequels or IP adaptations, movie ratings, movie box office, bullet screens, reviews, and the emotional correlation between bullet screens and subtitles on movie video playback; Secondly, the measurement of text factors such as bullet screen, comments, and subtitles involved in the hypothesis is studied. Considering that bullet screen is loaded in a rolling manner along the time axis of video playback, a bullet screen emotion value correction method combining emotional infection mechanism and attenuation effect is proposed, and the emotional information contained in video bullet screen and subtitles is extracted and visualized using text emotion analysis technology.

Finally, this research takes the film video related data of iQIYI platform from 2020 to 2021 as an example to verify the hypothesis of the factors influencing the movie video playback. The research methods include multiple linear regression analysis and inter subject effect test analysis. The research results show that the number of bullet screens, reviews and box office have a positive impact on the amount of movie videos played, and the movie type has a significant impact on the amount of movies played, among which the thriller type movies have a significant negative impact on the amount of movies played. At the same time, the heterogeneity of comments will have a positive impact on the playback

volume within a certain range, and when the number of comments is small, negative comments will have a small impact on the playback volume; Users prefer movie videos with weak positive or negative correlation with bullet screen and subtitle emotions, often because such movie videos have plot reversal or irony, which can satisfy users' curiosity and viewing desire.

Keywords: movie video ; Play volume ; Bullet screen emotion ; Emotional infection; Emotional decay

目 录

第一章 绪论.....	1
1.1 研究背景及意义.....	1
1.1.1 研究背景.....	1
1.1.2 研究意义.....	4
1.2 研究内容.....	5
1.3 研究方法.....	6
1.4 研究难点及创新点.....	6
1.5 论文结构安排.....	7
第二章 文献综述及相关理论概述	10
2.1 相关理论概述.....	10
2.2 国内外文献综述.....	10
2.2.1 院线电影票房影响因素研究现状.....	10
2.2.2 网络视频播放量影响因素研究现状.....	13
2.2.3 网络视频弹幕、字幕及在线评论研究现状	15
2.2.4 文献述评.....	16
第三章 理论模型构建及研究假设	17
3.1 理论适用性分析及模型构建	17
3.1.1 理论适用性分析.....	17
3.1.2 理论模型构建.....	18
3.2 研究假设.....	18
3.2.1 电影类型对电影视频播放量的影响.....	18
3.2.2 主创阵容对电影视频播放量的影响.....	19
3.2.3 续集或 IP 改编对电影视频播放量的影响	20
3.2.4 电影评分对电影视频播放量的影响.....	20
3.2.5 电影票房对电影视频播放量的影响.....	21
3.2.6 弹幕对电影视频播放量的影响.....	22
3.2.7 评论对电影视频播放量的影响.....	23
3.2.8 弹幕和字幕情感关联性对电影视频播放量的影响	24
3.3 本章小结.....	25
第四章 电影视频弹幕和字幕及评论文本情感分析	26
4.1 文本数据来源及预处理.....	26
4.1.1 文本数据来源.....	26
4.1.2 文本数据处理.....	27
4.2 弹幕及评论文本情感分析	28

4.2.1 基于 PaddleHub 的弹幕及评论情感分析.....	28
4.2.2 考虑情绪感染机制及衰减效应的弹幕情感修正	32
4.3 弹幕与字幕情感关联性分析	34
4.3.1 弹幕与字幕情感种类关联分析.....	34
4.3.2 弹幕与字幕情感趋势关联分析.....	37
4.4 本章小结.....	42
第五章 电影视频播放量影响因素实证分析	43
5.1 数据获取及变量定义.....	43
5.1.1 数据获取.....	43
5.1.2 因变量.....	43
5.1.3 自变量.....	43
5.2 播放量影响因素模型构建及分析	45
5.2.1 描述性统计分析.....	45
5.2.2 多元线性回归分析.....	45
5.2.3 逐步多元线性回归分析.....	47
5.3 变量主体间效应检验分析	49
5.4 本章小结.....	56
结论与展望.....	57
研究结论.....	57
管理启示.....	58
展望.....	59
参考文献.....	60

第一章 绪论

1.1 研究背景及意义

1.1.1 研究背景

随着社会的发展和科技的进步，人们的物质生活水平日益提高，也开始对精神生活、社交等心理需求有了更多的向往。而电影则是人们精神文化领域需求的首选，其作为文化娱乐的主要衍生产品，不仅丰富了人们的业余生活和精神世界，也促进了我国的经济发展以及各国之间的文化交流，并且随着消费水平的不断提高，以及政府对文化产业的大力扶持，影视行业的发展前景是不可估量的。

与此同时，随着互联网技术的迅猛发展，网络视频与其融合速度加快，整个行业的规模也在不断扩大，如图 1-1 所示，由中国产业研究院的数据可知，网络视频用户的规模从 2016-2021 年一直呈上涨趋势，且近 6 年来，网络视频用户的使用率一直高于 87%，截止 2021 年底更是达到了 94.5%，说明网络视频正逐步成为大多数用户日常休闲娱乐、茶余饭后消磨时光的重要手段。

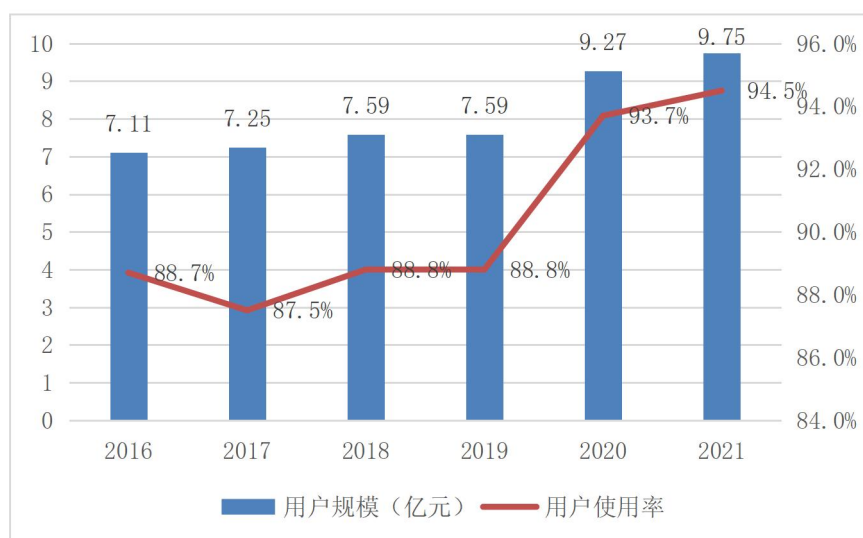


图 1-1 网络视频用户规模及使用率

Fig 1-1 Scale and utilization rate of network video users

并且，互联网的发达也导致了网络 IP（Intellectual Property）的爆火，网络电影、自制剧、用户生产内容（User-generated Content，简称 UGC）、专业生产内容

（Professionally-generated Content，简称 PGC）以及两者融合而成的专业用户生产内容（Professional-User-Generated Content，简称 PUGC）也以此为依托从而不断崛起，用户也能够通过各大视频网站随时随地观看丰富多彩的视频内容。

电影是网络视频不可或缺的种类，如今，电影从院线下映后在网络视频平台上播放已经成为一种趋势，并且以爱奇艺、腾讯视频、优酷等为首的各大视频网站已经积累沉淀了大量的网络视频忠实用户，促使其成为了互联网信息共享与传播的重要载体。在此种背景下，网络视频的播放量成为了衡量一个视频网站盈利能力的重要指标，并且还能够在一定程度上反应该视频的流行度，因此研究其播放量的影响因素有着重要意义。一般来说，网络视频可按其时长分为三类，如表 1-1 所示，电影作为典型的长视频类型，在这些网络视频平台的播放量除了会受到上映过程中的传统因素如类型、演员等的影响，还会被网络视频平台的新型互动形式如弹幕、评论等影响，且电影在上映过程中产生的票房也会成为影响因素之一。即电影视频从院线进入网络视频平台后，其播放量的影响因素与上映过程中的传统影响因素有所不同，因为电影进入网络视频平台后，其不存在排片这一影响因素，用户对于电影的细节可以进行深入的思考，经典的电影也会多次观看；特别是上映过程中高票房的电影可能导致从众心理，致使观看过的用户产生褒贬不一的评论，而未观看的用户也会受到不同的影响，所以无论用户是否在上映时观看过，他们是否选择在网络视频平台上观看也为不确定因素，是具有一定的研究价值。

表 1-1 网络视频类型分类
Table 1-1 Network video type classification

类别	短视频	中视频	长视频
时长	1 分钟以内	1-30 分钟	30 分钟以上
生产模式	UGC	PUGC	PGC
展现形式	竖屏	横屏	横屏
国内产品代表	抖音、快手	西瓜视频、哔哩哔哩	优酷、爱奇艺、腾讯视频
海外产品代表	TikTok	YouTube	Netflix、Disney+

目前，由于互联网背景下各大视频平台都引入了很多新颖的用户互动形式，这也导致除了视频本身的内容之外，播放量还会受很多其他因素影响，其中非实时的评论区互动以及实时的弹幕互动则是视频网站最主要两大互动形式，都承载着大量的用户反馈信息。但是两者仍然有着明显的区别，由于评论区的展示区域更大，其相对于弹幕在表达的内容上更具有完整性、客观性；而由于弹幕是伴随时间轴出现的用户互动内容，尽管信息的传递更加分散，但是其具有更强的时效性和互动性。此外，由于弹幕和评论的表现及互动方式不尽相同，弹幕多为碎片化的信息，交流氛围浓厚，一般是用户在观看过程中对电影内容的实时讨论，相比于评论，用户通常会随时因为某个片段、某一句话、甚至某个表情而发送弹幕，从而能够更加及时和具体地反应用户当下的情感 and 态度，有着更强的实时性、及时性与交互性；而评论一般是在视频之外，内容较为完整，有很强的阅读性，通常是对整部电影的评价，情感趋势较为缓和。所以弹幕和评论的情感强度和情感倾向也会在一定程度上存在差异，并且其数量和表达的情感对于视频播放量的影响也有所不同。同时，网络视频的字幕也不容被忽视，对字幕本身进行情感分析能够揭示视频播放过程中用户的情感变化，又因为字幕和弹幕一样关联着相同的时间轴，处在同一时间轴上的弹幕和字幕的情感会存在一定的契合或差异，对两者进行结合研究也更能还原用户在观看过程中的情感波动。弹幕、字幕及评论的特征归纳总结于表 1-2 中。

表 1-2 弹幕和字幕及评论特征
Table 1-2 Bullet screen and subtitles and commentary features

	弹幕特征	评论特征	字幕特征
时间	结合时间轴，发表于视频播放时	无时间限制，一般发表于视频结束后	结合时间轴，视频播放时同步出现
展示	通常展示与视频上方区域，同时段的弹幕同时展示	展示于视频外，排列顺序取决于发表时间	展示与视频下方区域，结合时间轴出现
内容	实时吐槽，信息交流，碎片化	内容完整性强，有阅读价值	与视频内容相关
发布者	匿名，以视频人物名发布，无意见领袖	公开发布者昵称、等级等信息，有意见领袖	官方字幕

表 1-2 弹幕和字幕及评论特征（续）
Table 1-2 Bullet screen and subtitles and commentary features

	弹幕特征	评论特征	字幕特征
氛围	互动氛围，交流氛围浓烈	互动、对话氛围较弱，有交流氛围	—
语言	以文字、表情、符号等为主，字体颜色、大小、位置等可改变，存在网络流行语、网络热梗	多以文字为主，无变换形式，语言通俗易懂	均为文字类型，可能存在中英文双语不同视频语言体系不同
长度	字数较少	字数较多	依视频而定

综上所述，院线电影下映后作为网络长视频的一种特殊的类型，在互联网背景下进行研究时不仅考虑到网络视频的特点，如新型的互动形式等，还需要与传统票房影响因素相结合，以此探究该类网络视频播放量影响因素，是非常值得思考和关注的。

1.1.2 研究意义

本研究结合传统的电影票房影响因素和视频网站新型的互动特点，探索了曾经在院线上映过的电影在网络视频平台上的播放量影响因素，该研究内容对于未来探究不同影响因素对此类网络视频播放量的作用机制提供了新的思路，具有一定的理论意义。同时，本文的研究也有助于电影的制片方以及网络视频的运营方等深入了解不同因素的影响效果，为用户提供更好的观看体验，因此也具有重要的实践意义。

（1）理论意义

首先，本研究聚焦于网络长视频的特殊类型，即曾在院线上映后在网络视频平台上播放的电影，同时基于网络视频平台的用户互动特点对弹幕、评论等因素进行研究，结合电影上映过程传统的影响因素如电影类型、是否为续集或 IP 改编等，并且还考虑到票房这一因素对播放量的影响，为该类特殊网络长视频的研究提供了一定的参考支持。

其次，本文对弹幕、字幕以及评论的文本进行分析，并且对弹幕与评论、弹幕与字幕的差异进行对比分析，基于分词、文本情感分析等理论和研究方法对不规整及琐碎的文本表达进行分析研究，这也是对相关理论和分析方法的进一步

实践和验证。

（2）实际意义

第一，基于收集到的电影网络视频数据，对数据中的不同数据特征如文本、数字等进行统计分析和文本分析，建立相关的播放量影响因素模型，有助于探究在视频传播过程中不同因素的影响效果及影响机制，研究结果对于管理者制定电影这类的长视频营销策略有重要参考意义。

第二，通过挖掘得到的网络视频播放量影响因素的作用机制，不仅能为网络视频平台的同类视频以及其他类型视频提供新的平台运营思路，甚至有助于更多不同行业打造用户与社区生态沉淀有效的方法论。

1.2 研究内容

本研究涉及了消费行为学及心理学，营销学、统计学等多领域的知识，首先对电影票房、网络视频播放量影响因素以及与弹幕、评论、字幕相关的国内外文献进行梳理研究；又结合本文研究背景进行理论的适用性分析，提出相关的研究假设；再通过对收集到的数据进行相关处理，构建电影视频播放量影响因素模型进行假设验证，具体包括以下研究内容：

（1）依据本文背景，结合 SOR 理论、感知风险理论及感知价值理论进行理论适用性分析，对电影本身的特点如类型、明星阵容、评分、导演以及网络视频的互动特点如弹幕、字幕、评论等对播放量的影响做出相应的研究假设，并构建了电影播放量影响因素的理论模型。

（2）对腾讯视频平台的中国大陆院线电影的数据进行收集，对弹幕、字幕、评论等相关文本数据进行情感分析，并对电影相关评论进行情感异质性分析，以及对字幕与弹幕进行情感种类及趋势关联性分析。

（3）对收集到的电影相关的数据进行清洗处理，基于研究假设，对影响因素进行描述并建立相应的评价指标体系，以此为基础构建电影视频播放量影响因素模型以验证研究假设，分析电影视频中不同互动特点的弹幕和评论对播放量的影

响，并研究弹幕与字幕种类的关联及情感变化趋势的一致性，最后对研究结果进行分析并提出管理启示。

1.3 研究方法

本研究基于网络视频平台的以及院线电影的特点，结合 SOR 理论、感知风险及感知价值理论，探究了影响电影类网络视频播放量的因素，包括其作用机制及影响效果。基于此本研究采用多学科的方法进行综合性研究，主要涉及了以下研究方法：

（1）文献和理论研究的法。通过 CNKI、Science Direct、SpringerLink 等中外文数据库查询与电影下映后视频播放量影响因素研究的相关文献以及理论资料，对获取的文献资料进行整理分类、对比阅读，整合分析其中有价值的信息，并按照一定的逻辑整理成文献综述，总结后进行文献评述，依据相关理论提出假设，确定本文的基本框架以及创新点。

（2）文本情感分析法。结合本文研究的背景，利用文本情感分析的方法对电影相关的文本因素进行量化分析，作为研究的自变量之一。

（3）实证研究法。收集腾讯视频平台的电影视频相关数据，对数据进行整理清洗后构建相应的电影视频播放量影响因素模型，运用 SPSS、R 语言等方法进行统计分析，其中包括样本和变量的描述性统计、回归分析、主体间效应检验等，通过对研究假设进行验证而得出研究结论。

1.4 研究难点及创新点

依据本研究的实际情况，本研究的难点在于：

（1）数据收集。本研究要收集与电影相关的多种类型的数据，包括比较容易获得的类型、评分等，但是对于弹幕和评论的获取，目前还没有很成熟的批量爬虫技术，只能根据每部剧不同的存放 ID 进行获取，且部分电影的字幕也存在缺失的情况。

(2) 弹幕数据处理。弹幕作为新型的文本型数据，具有口语化、碎片化等特点，其中还不乏一些网络热词和流行梗，对其进行识别较为不易。

而本文研究的创新点在于：

(1) 弹幕文本分析。弹幕作为网络视频平台新型的评论交互模式，一度成为近几年国内外的研究热点，对其传播效应及影响机制的探究是研究网络视频播放量影响因素中十分重要的一个环节。

(2) 弹幕与字幕关联性分析。弹幕和字幕具有相同的时间轴这一属性，而现有的研究对于弹幕与字幕的关联性分析较少，但是在用户观看视频时其实会同时接收来自视频内容、弹幕以及字幕三方面的信息，通过研究弹幕与字幕的情感异同以及情感趋势，是网络视频播放量影响因素中的重要指标。

1.5 论文结构安排

根据本文的研究内容，将全部内容分为六大部分，依次为：绪论；文献综述及相关理论概述；理论模型构建及研究假设；基于弹幕、评论及字幕的文本分析；实证研究与假设验证；总结与展望，具体内容如下：

第一章：绪论。本章节包含了本研究的研究背景、研究的理论和实际意义、研究内容和方法、研究难点和创新点，经过一系列的分析 and 论述，引出本文研究的主题以及文本研究的必要性。

第二章：文献综述及相关理论概述。本章节首先对本研究运用的相关理论以及研究方法进行阐述，包括 SOR 理论、感知风险理论和感知价值理论、文本分析方法、多元回归分析法等。又回顾了国内外与电影票房影响因素、网络视频播放量影响因素以及与弹幕、评论、字幕有关的研究，对其进行梳理并总结评述。

第三章：理论模型构建及研究假设。本章节依据本文研究背景，对前一章节所涉及的理论进行适用性分析，并结合理论提出相关研究假设，包括与电影传统因素如类型、导演、明星、评分等对播放量影响的假设，还包括网络视频平台特点如弹幕、评论、字幕等对播放量影响的假设。

第四章：基于弹幕、评论及字幕的文本分析。本章节利用网络爬虫技术获取电影视频相关的文本信息，通过学习并实践第二章所提到的文本分析方法，对弹幕、评论以及字幕进行文本分析，主要包括弹幕文本情感分析、评论情感异质性分析以及弹幕与字幕情感关联性分析。

第五章：实证研究与假设验证。本章节介绍了实证研究所需要的电影视频相关数据的来源及处理，还包括实验设计的步骤和变量的选取及指标体系的建立，在此基础上进行播放量影响因素模型的构建并初步分析，以检验所提出的假设是否被支持；在此基础上详细介绍了本研究关于电影视频播放量影响因素的相关结论，并从不同角度给予尽量全面客观的分析。

总结与展望。本章节对全文进行总结阐述，同时依据研究结论提出了本研究的管理启示及研究贡献，最后指出本研究的不足之处，希望能对其他学者未来继续研究的方向提供参考。

本文的研究技术路线图如图 1-2 所示：

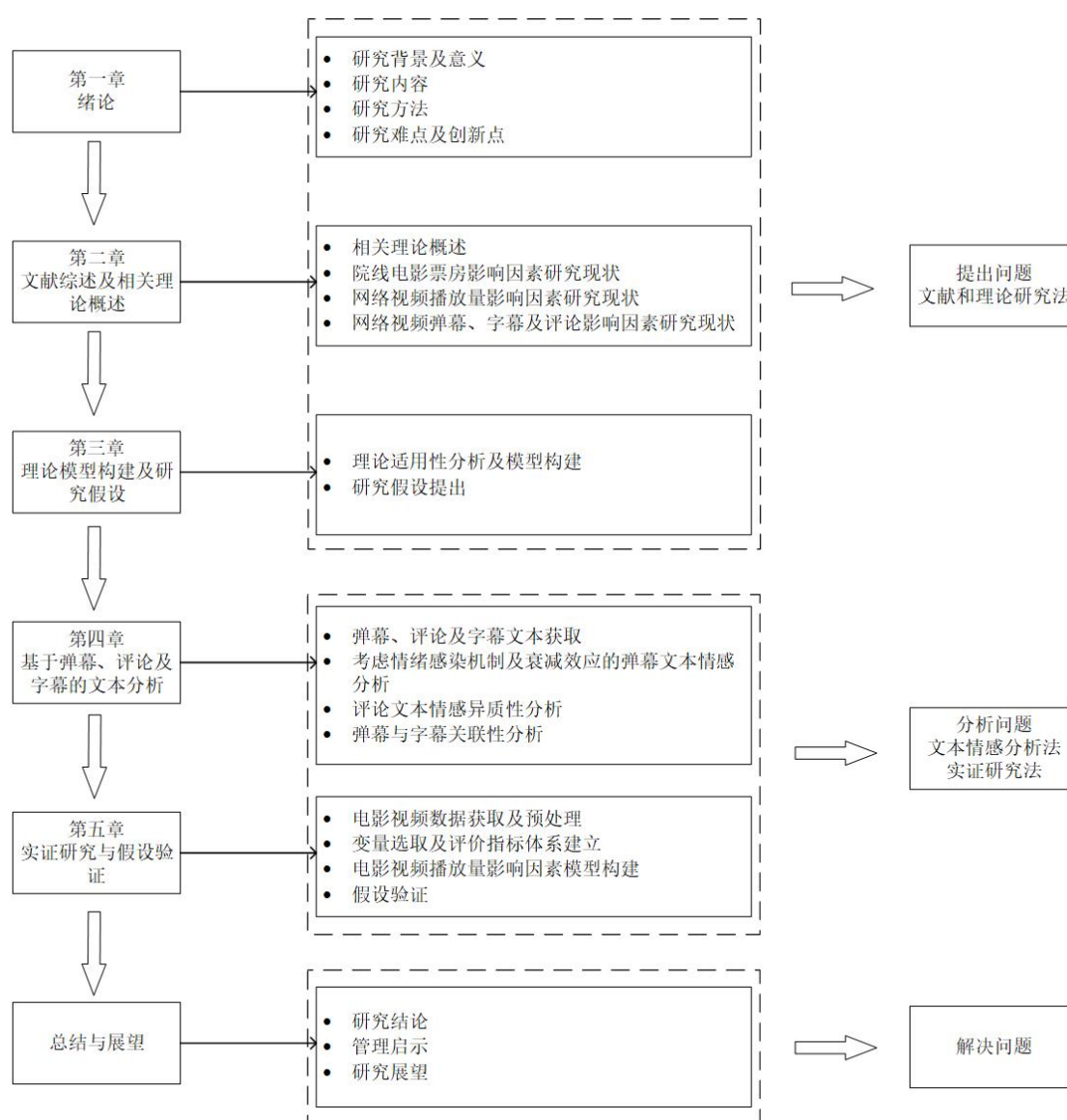


图 1-2 论文技术路线图

Fig 1-2 Paper Technology Roadmap

第二章 文献综述及相关理论概述

2.1 相关理论概述

结合本文研究背景以及所涉及的知识领域，主要的相关理论与研究方法包括：

（1）SOR 理论。SOR 理论表明了外界环境会对个体的内在造成一定的影响，使个体产生不同的态度或行为。其中 S（Stimulus）是外在刺激变量，代表了外界环境中会激发个体或影响个体机体状态的因素；O（Organism）是个体的机体变量，代表了个体内在变化的过程；R（Response）是个体的反应变量，代表了个体在受到外界刺激后所产生的态度或者行为结果。

（2）感知风险和感知价值理论。感知风险理论是从心理学延展而来的，其中包括用户所做出的决策结果的随机性以及做出错误决策后需要付出的代价，即当用户对能满足自身期待的消费行为不确定或者消费行为发生后产品和服务未能达到预期时，都会产生感知风险；同时用户在进行消费或者选择服务时，购买或选择的不止是产品和服务本身，更期望获得其中的价值，其本质也是感知，而这种“感知”会对用户的行为产生影响。

（3）文本分析。文本分析主要包括文本情感分析、文本情感种类统计等，指结合爬虫、自然语言处理、文本情感分析等技术手段，对文本中的主客观咨询、主题、情感强度及倾向等进行提取和分析。

（4）多元回归分析。多元回归分析是一种常用的统计学方法，其主要目的是通过构建模型，研究多个自变量与因变量之间的影响关系，对样本数据进行分析研究。

2.2 国内外文献综述

2.2.1 院线电影票房影响因素研究现状

传统背景下的影响因素研究主要包括电影类型、电影名称、排片量、上映日

期、电影评分、导演、参演人员等，但是随着时代的发展，由于很多电影相关数据的可得和量化，国内外学者不再局限于传统的电影票房影响因素研究，以及在大数据时代下很多新兴的研究方法逐渐得到普及，越来越多的学者也更加倾向于利用神经网络、机器学习等方法建立票房预测模型，以处理更加复杂的电影票房问题。

（1）引入网络口碑的院线电影票房研究

Thorsten 等和 Kim 等研究了网络口碑对电影票房的影响，Thorsten 等对 105 部电影在上映的周末中在 Twitter 上的 400 万条口碑消息进行分析研究，发现了电影票房在 Twitter 上的口碑效应，负面口碑过多的电影将直接影响观众对电影的选择，进而影响电影票房^[1]；Kim 等考虑了不同电影之间的竞争关系以及社交网络上的口碑效应分析其对电影票房的影响，用社交网络上正负评论的数量来衡量口碑的好坏，以遗传算法为基础对机器学习算法进行了训练，发现好的口碑对电影的总票房会产生显著影响^[2]。华锐等和黄文清等在传统的电影影响因素的基础上，引入了网络口碑、宣传力度、电影关注度以及上映时长等指标，华锐等基于 252 部电影数据，使用了 C-D 生产函数模型并进行了分位数回归，研究结果表明，当电影的票房达到一定程度时，口碑的积极作用会有所衰减，而此时的导演对于票房的影响大于主角，且电影的上映时间也会对票房有一定影响^[3]；黄文清等构建了多元线性回归模型和随机森林模型对票房进行预测，最后结果表明在电影上映的前期大众关注度、首周票房以及档期等对票房有着很大的促进作用，而对着市场的成熟，网络口碑的作用则会变得更加重要，且非线性的随机森林模型预测准确度更高^[4]。

（2）引入消费者社交网络行为的电影票房研究

Liu 等通过用户在社交媒体上的行为挖掘出了每部电影的关注度、受欢迎程度以及用户的消费意愿等，建立了线性回归模型和支持向量机模型，基于用户的发帖行为对票房进行预测，研究发现用户的社交网络行为与电影票房息息相关，且引入用户的购买意愿提高了预测模型的准确率^[5]。Ding 等认为用户在 Facebook 上的点赞行为会对其他消费者产生社会影响，通过建立横截面回归模型研究电影上

映前用户点赞行为对上映后票房的影响,研究表明在电影上映一周前的点赞次数的增加能明显促进票房的增加,具体表现在如果点赞数每增加 1%,那么首周票房能增加约 0.2%^[6]。Sehwan 等统计了从 2013 年到 2014 年的电影预告片的观看数据,并与相应的电影票房进行匹配,建立了面板联立方程模型,研究结果表明,与电影预告片的评论等其他社交数据对票房收入的影响相比,电影预告片的积极作用更大,并且此影响在电影上映的前期更为明显^[7]。Chong 等通过普通最小二乘回归模型对消费者在 Facebook 和 YouTube 上的社交参与行为进行分析,并提出了一套消费者参与行为在这些社交平台上的指标,来预测其对票房总收入的影响,发现呈正相关^[8]。Senhui 等认为用户在社交平台上的相似性会影响用户对电影票的购买意愿,基于 SOR 理论等对用户的社交行为进行了分析,并通过调查问卷的方法收集了相关的数据,利用偏最小二乘法对数据进行了分析,发现用户社交行为的内部和外部相似性都在很大程度上影响电影票的购买意愿,从而影响票房的收入^[9]。

(3) 引入社交平台影评形式及情感分析的电影票房研究

史伟等和孙春华等认为社交媒体上影评的情感极性会影响消费者对电影的观看决策,史伟等将微博作为观点输出和传播的载体,并结合微博文本中的情感因素,提出了新型的票房影响因素模型,构建了相应的模型对微博文本情感进行分析,研究发现微博影评的情感因素对票房有显著影响^[10];孙春华等收集了 2013-2015 年间不同类型电影预告片的评论,分别对首周票房建立了截面数据模型以及对后续票房建立了面板数据回归模型,进行模型的参数估计,以预测预告片的关注度以及评论的情感极性对票房的影响,研究发现评论中带有强烈情感色彩的词对票房有显著影响,且预测较为准确^[11]。石文华等认为在互联网背景下在线评论的形式越来越丰富,将其分为了较长的专业影评以及较短的非专业短评,利用豆瓣电影数据为样本进行分析,发现影评的数量对票房无显著影响,而短评对其影响呈倒 U 型变化,同时短评对票房的影响比影评更大,且持续影响的时间更长^[12]。Ya-Han 等从 IMDb 收集的电影数据作为样本,通过情感分析工具对电影评论进行量化处理,结合电影自身的影响因素以及外界的影响因素,采用数据挖掘

技术、M5 决策树、线性回归以及支持向量回归模型对首周票房进行预测，研究结果表明，对首周电影票房影响最大的是电影数量，而影评的内容和情感等只针对特定的电影有作用^[13]。

由上述文献可知，随着电影市场的发展以及各大社交媒体平台的普及，国内外学者在关注传统影响因素对票房的研究的同时，也引入了各种互联网背景下的因素进行综合研究，例如网络口碑，观众在社交平台上发表的影评的情绪、网络搜索量等，同时也会考虑消费者行为与心理等因素，也采用了更合适的科学方法例如神经网络、随机森林等构建票房预测模型，使得对电影票房的影响因素研究变得越来越丰富、全面以及准确。

2.2.2 网络视频播放量影响因素研究现状

近年来，由于各大网络视频平台的兴起，网络视频在国内外迅速崛起及流行，而且用户不再仅仅是视频的观看者，在视频平台上他们也是视频的创造者及传播者，也为各种不同类型的网络视频的传播创造了有利条件。同时各国研究学者对于视频播放量影响因素的研究也逐渐展开，主要从视频本身特征、用户参与行为等方面展开研究；除此之外，也有学者通过建立视频传播模型对视频播放量进行了预测^[14, 15]。

（1）视频本身特征及视频平台对网络视频播放量的影响

Borghol 等学者和 Yoganarasimhan 等学者探究了与视频内容无关的影响因素对视频播放量的影响，结果发现视频发布的社交平台，视频上传的时间，发布者的特征，视频中所提及的关键词以及视频发布平台的推荐系统等都会对视频播放量产生影响^[16, 17]。Cha 等通过实证研究分析了 YouTube 和 Damu 视频系统上 UGC（用户生成内容）和 VOD（视频点播）的不同，探究了 UGC 视频的播放量与视频年龄之间的关系^[18]。Welbourne 等对 YouTube 上不同频道的 390 个视频的内容进行了分析，研究发现内容越具有吸引力的视频播放量更高，同时用户自制视频相比专业视频更容易获得高播放量，并且当发布视频的用户或者视频内的角色保持不变时，也更容易得到较高的播放量^[19]。Florenca 以注意力经济为切入点，通

过研究 Bubz beauty 发布的视频,将美妆类视频分为了面向内容、面向市场、鼓励型以及关系型四大类,探索了视频播放量与内容特征和受众的注意力维持度之间的关系^[20]。Jeong-Sun 等通过发放调查问卷的方式,对在校大学生观看 YouTube 上的美妆视频的体验进行了研究,发现用户选择观看该美妆类视频主要取决于其原有的播放量及关注度^[21]。

(2) 用户参与行为对网络视频播放量的影响

Zhou 等为了探究如何利用 YouTube 的推荐机制来提高视频播放量,基于用户的浏览行为生成了视图传播模型,探究了视频之间如何影响各自的播放量^[22]。Li 等收集并分析了人人网的视频数据,建立了模拟社交网络中用户对视频的请求过程,研究了视频播放量的分布及演变,利用该仿真模型获取了高播放量视频的用户行为特征^[23]。Vandersmissen 对国外最先兴起的短视频 Vine 进行了研究,结果发现如果一个视频经常在 Twitter 内部进行分享,相比那些未被分享的视频在发布之后会赢得更多的喜爱。并且这类视频一经发布便会有很强的生命力,会迅速受到众多用户的注意。随着时间的推移,其播放量虽然会逐渐回落但是不会停止,最后趋于稳定^[24]。Wu 等人提出了随机流体模型 EvoModel,并以腾讯视频的数据为研究对象,综合考虑用户的传播行为以及首页推荐等影响因素,通过拟合真实的视频播放量数据验证了该模型^[25]。张海涛等从多个不同维度的视角切入,并结合全信息理论和超级 IP 理论,构建了短视频传播效果评价模型,对用户偏好与视频内容特征进行分析,研究结果表明官方认证与视频的播放量呈强相关^[26]。龚艳萍等结合用户参与行为,通过实证研究对短视频的播放量进行研究,发现用户的参与行为会显著受到视频是否为原创及是否具有个性的影响,进而影响短视频播放量^[27]。

综合上述研究文献,目前对于网络视频播放量的影响因素研究多集中于视频本身的特征如时长、内容等,以及观看视频的用户行为特征如点赞、转发等,虽然都取得了很多优秀的成果,但是对于当今时代视频播放过程中许多新兴的互动方式的出现,例如弹幕、评论等的情感和数量因素也应该被综合考虑。

2.2.3 网络视频弹幕、字幕及在线评论研究现状

(1) 网络视频弹幕研究现状

Shen 等认为弹幕的情感会影响用户的判断, 利用情感反应模型对相应数据进行研究分析, 探究弹幕是否会引起用户更加愉悦和兴奋的观看视频的状态, 进而影响用户的观看决策^[28]。王霞等通过研究 Bilibili 这一视频平台上与电影相关的视频弹幕, 对其进行情感分析, 研究发现弹幕情感强度及数量与视频的播放量、点赞量等具有正相关的关系^[29]。袁海霞等以消费者的决策进程模型为基础, 收集了 Bilibili 视频播放平台上的 1019 部电影, 构建了时变效应模型对其进行分析, 研究结果表明弹幕的数量对视频播放量有正向影响, 且影响效果随时间呈倒 U 型变化, 弹幕的情感对视频播放量的负向影响效果随时间呈 U 型变化^[30]。陈明红等收集了 Bilibili 视频网站上的数据, 以 ELM 理论为基础, 构建了弹幕视频播放量影响因素的模型, 通过层次回归方法进行了分析, 结果表明弹幕数、评论数、点赞数、转发数都会对视频播放量产生正向影响^[31]。

(2) 网络视频在线评论研究现状

Cha 等通过对国外视频网站 YouTube 上的视频数据进行分析, 获取了大量的视频样本, 得出了各参数的分布函数, 最后分析发现视频网站的播放量与评论有很强的关联性, 系数达到了 0.8^[32]。Mcauley 通过将潜在的系统推荐因素与潜在的评论主题相结合, 通过 LDA 主题模型等方法对评论中的文本信息进行分析, 研究结果表明其对网络视频的评分和播放量都有影响^[33]。徐晓枫等为了预测某电视剧在网上单集的平均播放量, 通过对其在微博上的搜索量、微博发布的数量、在线评论的情感、导演知名度等进行分析, 结果表明电视剧首次播出之后, 其播放量会受到微博上与电视剧相关的评论中积极或消极情感趋势的影响^[34]。周良等从用户的视角出发, 收集了来自优酷视频网站在 2014 年的数据作为研究样本, 探究视频的口碑与播放量之间的关系, 研究结果表明评论数量会显著正向影响网络视频的播放量, 而与视频的差评及顶踩数等没有显著关联性^[35]。

(3) 网络视频字幕研究现状

由于弹幕和字幕都与视频播放的时间轴相关, 处于同一时间轴的字幕和弹幕

很有可能存在较强的关联性，所以目前针对字幕的研究多与弹幕相结合。何楚佳等人通过文本情感分析的方法，获取了弹幕和字幕数据作为数据集，对其文本内容进行了关联性分析并总结为了六大类，且发现关联分析更贴近观众对视频的真实想法^[36]。王敏等以情感词典为基础，构建了相应的比较模型对弹幕与字幕情感进行分析，同时结合时间序列对弹幕和字幕文本的情感值、情感种类等进行研究，结果表明虽然在情感种类的分布上，弹幕与字幕所占比重不同，但是具有相似的分布规律，并且弹幕的情感强度普遍高于字幕^[37]。

结合上述文献可知，对于网络视频中弹幕、字幕以及评论的分析多集中于文本情感分析以及文本主题分析，探究其数量及情感强度等对视频播放量的影响；并且还有部分学者根据弹幕和字幕存在统一时间序列的特点，对其进行了关联性分析，都得出了很多优秀的结果。

2.2.4 文献述评

纵观国内外的相关文献研究，对于处于上映阶段的电影票房相关的研究已经十分成熟了，但是电影从影院下线之后在各大视频网站上则会成为网络视频的一种类型，而对于网络视频的研究国内外学者也颇有建树，从国外的 YouTube 到国内的 Bilibili 等平台，研究了与网络视频播放量的各种因素，同时在当下新型互动方式出现的基础上，对弹幕、评论以及字幕等都进行了深入的分析研究，但是涉及院线电影的网络视频播放量的研究不多，同时对视频弹幕与字幕关联性分析的研究较少。

根据文献梳理的结果，本研究从多角度切入，以网络视频播放量的影响因素研究为基础，将传统电影上映过程中的影响因素进与网络视频平台上的互动性较强的弹幕与评论相结合，又考虑到与其他同类型的网络视频有区别，因为研究对象为院线下映后在网络视频平台上播放的电影，影响因素会涉及该电影上映时所取得的票房，由此运用科学的理论和方法对电影视频的网络播放量影响因素进行综合分析研究。

第三章 理论模型构建及研究假设

3.1 理论适用性分析及模型构建

3.1.1 理论适用性分析

（1）SOR 理论适用性分析

在本研究中，用户在视频平台观看网络视频时，其中视频内容、弹幕、字幕等是用户个体以外的事物，即是一种外部条件，也就相当于 SOR 理论中的 S。对用户来说，视频及弹幕、字幕、评论等内容都是一种外部刺激，它们从视觉、听觉等方面刺激用户的认知，影响用户后续的对视频的认知过程及行为。

用户在受到外界刺激后，一般会生成对刺激的认知，进而产生相应的情绪，在本研究中，用户在观看视频时，会受到来自字幕、弹幕、以及视频本身等多种刺激，而用户自身会产生对应该刺激对自身影响的调节机制，即内在的过程会产生不同的变化，也就是 SOR 理论中的 O。

在用户受到刺激经过个体内心的反应及调节机制后，就会产生相应的态度和行为，如心理反应包括满意、厌恶等，机体行为包括是否购买商品、是否选择观看视频等，对应 SOR 理论中的 R。在本研究中，用户经过个体内在分析刺激后产生的行为又分为对电影视频的趋近或者规避，即用户在观看视频时收到的刺激使其获得满足感与幸福感，就会产生趋近行为，从而选择继续观看，甚至有可能之后会再次观看；相反，如果收到的刺激让用户产生厌恶的感觉，就会产生规避行为，导致用户会退出视频的播放。

（2）感知风险理论和感知价值理论适用性分析

在本研究中，用户期待在观看电影视频的过程中体验到某种价值，而这种价值的本质就是“感知”，其包括用户在选择观看某一视频之前时，会对视频相关信息如等类型、评分、票房等产生一种主观的感知。用户的感知心理活动是进行其他消费等活动的基础，对用户的行为和选择具有重要意义。当用户在这种感知价值的利弊之间进行比较和权衡时，就产生了对该视频的感知风险，而这种感知风

险则会影响用户接下来是否观看该视频的选择，进而对视频的播放量造成影响。

3.1.2 理论模型构建

综上所述，基于 SOR 理论、感知风险理论以及感知价值理论，并结合众多学者在相关领域的研究，依据本研究的背景，分别考虑到电影在上映过程中的特点以及网络视频独有的特点，构建了本研究的理论模型，具体如图 3-1 所示。该模型表示，用户在选择某一电影视频观看之前，会感知到来自电影相关信息如类型、评分、票房等的风险；同时在与视频进行交互时，也会受到来自视频弹幕、字幕等的外部刺激（S），而感知到的风险与外部刺激因素都会对用户内在的感知价值造成一定的影响，从而影响用户的反应（R），即用户对该电影视频的满意度，最终对其播放量造成影响。

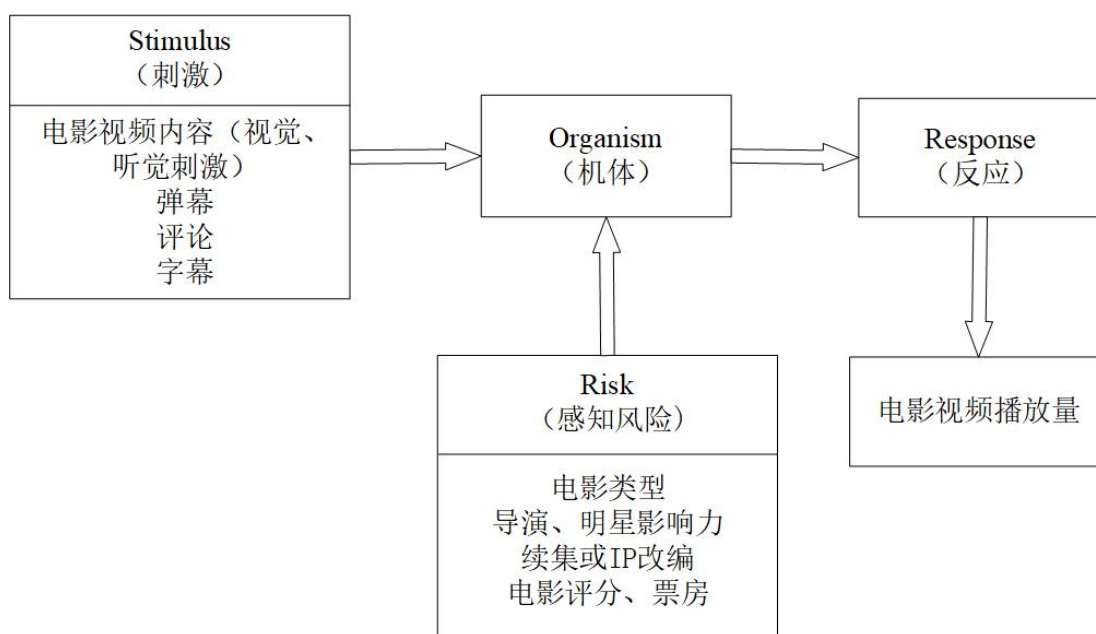


图 3-1 理论模型构建

Fig 3-1 Theoretical model construction

3.2 研究假设

3.2.1 电影类型对电影视频播放量的影响

电影的类型是由观众的喜好决定并演变而来，同一类型的电影一般都具有相

同的基本特征。在猫眼发布的《2021 中国电影市场洞察》报告中显示，爱情、动画等类型的电影受到 00 后的偏爱较多；而 95 后除了此类型的电影之外，还会选择悬疑题材的电影进行观看；90 后则更偏好动作、冒险类型的电影；80 后主要以观看亲子类型的电影为主；而剧情、历史战争类型的电影则更受 70 后的青睐。虽然我国电影类型的发展相比国外其实还不够成熟，但是很多类型片、亲子片等仍然是我国高频刚需的电影类型，并且存在一批特定的群体具有稳定的观看需求。

对于电影类型的研究最早起源于国外，但是受制于不同的地区和国家，所得出的结论不尽相同。在 Litman 的研究中曾指出，科幻类型的电影会对消费者的决策造成显著影响，而其他类型的电影则与之没有显著的关联^[38]；我国学者张立基于信息理论对影响电影消费的因素进行分析，最后的研究结果表明电影类型与电影的消费情况呈现很强的关联性^[39]。综上，结合国内外各学者的研究，表明电影类型对于其视频播放量的影响作用是毋庸置疑的，特定的电影类型往往存在一批忠实的观众群体，所以用户是否选择某一电影视频进行观看时会对电影类型进行考虑。因此，本研究提出以下假设：

H1:不同的电影类型对播放量的影响程度不同。

3.2.2 主创阵容对电影视频播放量的影响

一部电影的成功是离不开主创阵容的心血和努力的，主创阵容包括导演、演员、编剧、监制、场记等人员，涉及影片的筹备、拍摄、上线等多个阶段。但是，目前结合我国的影视产业，除了导演和演员之外，大多电影相关人才暂时还没有得到较大的重视，甚至存在其贡献和所获得的社会关注度不相符的现象。因此，针对我国电影发展的实际情况，为了对主创阵容这一指标进行衡量，本文将选取“导演”和“演员”两个代表性人物。

目前很多相关研究均表示导演和演员对电影票房存在一定程度的影响，Devendra 等通过研究 2007-2013 年间宝莱坞上映的 334 部电影，构建了多元线性回归模型，发现参演人员这一因素会显著影响票房^[40]；Duan 等利用高斯回归模型对多方面的影响因素进行研究，发现如果一部电影中获得奥斯卡的明星和导演的

数量越多,那么票房也越高^[41]。因此,根据学者的相关研究,首先对于一部电影来说,导演是电影视频的核心人物,其对于电影艺术品质的把控往往起到决定性的作用。而就演员而言,往往演员的明星影响力能够为电影带来一定的话题讨论度,并且演员也是电影的诠释者,在各种因素的加持下有可能会对视频播放量产生很大的加持。基于此,本研究提出以下假设:

H2a:电影视频播放量与明星影响力呈正相关。

H2b:电影视频播放量与导演知名程度呈正相关。

3.2.3 续集或 IP 改编对电影视频播放量的影响

续集和 IP 改编都是对某一品牌的,续集即系列影片,与其他的商品不同,电影属于一次性的体验式消费商品,很难产生品牌影响力,用户是否真正对一部电影感兴趣往往都需要在观看电影后才能做出判断;IP 改编则是知识产权电影,也就是所谓的翻拍,也是当前主要的电影创作形式,通常会结合不同的时代背景对原有的故事框架和人物形象等进行适当改编。通常来说,IP 改编的电影会自带流量,因为其在未进行影视化之前就有一定的受众群体,还未对电影视频进行观看的观众也会对其比较熟悉,例如郭敬明的《小时代》系列电影,在拍成电影之前其同名小说就具有很高的知名度以及广泛的阅读群体,也导致了这一系列电影都取得了很大的成功。胡小莉等学者通过建立相关的用户决策影响因素模型,对两年间上映的两百多部国产电影进行研究,表明翻拍和续集会对消费者是否选择观看电影的决策产生一定的正面影响^[42];刘志新学者也曾对 2018 年国内上映的 292 部电影构建多元线性回归模型进行票房影响因素的实证分析,研究结果表明续集对票房有显著的正向影响^[43]。综上,续集和 IP 改编电影都可以认为是对经典的延续或再现,本身都拥有一批忠实的受众,本研究提出以下假设:

H3:续集或 IP 改编的电影视频播放量更高。

3.2.4 电影评分对电影视频播放量的影响

电影评分通常被视作网络口碑中的一种,网络口碑指的是用户自主在网络平台上生成的内容,很多学者将其分为口碑数量和口碑效价,口碑效价包括评论和

评分, 评分属于量化性质的效价, 现在许多网络视频平台都设有给视频打分的系统, 用户能在观看视频之前通过评分的高低来衡量该视频的质量, 从而决定是否进行观看。对于电影评分对消费者决策的影响, 很多学者都进行了相应的研究, 王铮等通过在传统影响因素上引入包括“电影评分”、“票价”等在内的因素对 2007-2012 年国内上映的 500 多部国产电影进行研究, 实证研究结果表明影片得分对电影票房有比较显著的正向影响^[44]。宋子川等和王一帆等经研究发现电影评分均对票房有正向影响, 宋子川等利用实证分析研究法, 通过对 3 年间从国外引入, 而后在国内上映的几十部影片的票房影响因素进行研究, 包括评分、影片热度等, 结果表明评分对票房的影响最为显著^[45]; 王一帆对国内在 2013 年内上映的电影进行研究, 探究电影类型、电影导演、参演人员、电影评分、上映时间、出品方以及国内生产总值和居民可支配收入等因素对电影票房的影响, 利用多元线性回归模型对票房进行预测分析, 研究发现电影评分对票房的影响最显著^[46]。由此可以得出结论, 电影评分对于用户是否观看视频的决策具有一定的影响力, 因为用户能够通过评分的高低直观地初步了解某部电影视频的好坏, 于是作出以下假设:

H4: 电影视频播放量与电影评分呈正相关。

3.2.5 电影票房对电影视频播放量的影响

通过电影票房可以在电影上映的时期衡量电影的质量, 但是当电影下映后在网络视频平台上播放时, 其对播放量的影响机制可能会发生变化。对于高票房的电影, 很多未观看的用户可能会产生从众心理从而去网络视频平台上进行观看, 但是当该电影无法引起其共鸣时, 用户则不会因为高票房就认可这部影片, 反而会输出负面的评论, 从而负向影响视频的播放量, 同时也会有用户因为在影院观看过而不会选择再次在网络视频平台上观看, 也对播放量有一定影响。但是对于经典影片或是高口碑低票房的影片, 很多用户在网络视频平台上没有了约束, 对于很多细节可以细细品味, 甚至会进行多次观看, 由于目前网络视频的播放量每观看一次都会计入播放量次数, 所以很多影片就算在上映的时候并没有取得很好的票房成绩也有可能获得较高的播放量。基于此, 本研究提出以下假设:

H5:电影视频播放量与票房不相关。

3.2.6 弹幕对电影视频播放量的影响

如今几乎所有的网络视频平台都设有弹幕的功能，弹幕是一种新形式的基于真实消费情境的动态的瞬间反馈信息，能够反映用户对视频消费产品的喜好程度，而弹幕的数量和弹幕情感一般都会对视频播放量产生一定影响。

一般而言，弹幕的数量越多，网络视频的播放量很大可能就会越高。一方面，因为视频在播放时就会提示当前视频的弹幕数量，这一提示通常会被用户用来衡量该视频质量的好坏，即用户会根据弹幕的数量来判断某一视频的用户参与度，认为弹幕数量越多的视频质量越高；另一方面，弹幕数量越多也表明视频的用户讨论互动性更强，在大量的弹幕中往往会蕴含很多与视频有关的有价值的信息，对于用户来说观看视频的趣味性和有用性也会更强，因此弹幕数量多的视频更有可能获得更多的播放量，但是根据信息过载理论，弹幕数量太多的视频会提供给用户过多的信息，可能会造成一定程度上的认知过载或信息污染，从而对视频的播放量造成负向影响。^[47]此外，也有研究表明弹幕过多可能会造成剧情泄露，反而会降低视频的播放量^[48]；同时弹幕也属于网络口碑，有研究指出口碑具有衰竭效应，即当口碑的数量不断增多时，用户在决策过程中带来的边际效用呈递减趋势，即视频的播放量可能会逐步下降^[49]。

除了弹幕的数量之外，弹幕所表达的情感也会对视频的播放量造成影响。目前很多用户都倾向于通过弹幕来表达自己在观看视频时当下的情感，因为弹幕独有的特性，导致处在不同时间和空间的个体也能够进行互动，也就是说用户会受到其他个体发表的弹幕中表现出的情感强度和倾向的影响，从而影响用户观看视频的决策。有研究认为，当用户被一些情感强度强烈的弹幕所感染时，有可能会发表同样倾向的弹幕，也有可能会因为观点不和或者彰显自己的独特性而发表情感倾向相反的弹幕^[50]；同时梁栋等人研究发现，弹幕数量情感均会正向影响网络视频的热度，并且弹幕情感对其的影响程度更大^[29]。此外，即使在观看视频的过程中不发送弹幕，用户只是同时进行弹幕和视频的观看也会对用户的情感倾向造

成影响，即用户对视频的情感倾向会受到视频中与其意见相同或相左的弹幕的影响，从而影响用户的观看体验，导致影响视频的播放量^[51]。

通过上述分析，本研究认为，弹幕数量对电影视频的播放量有正向影响，同时，当弹幕数量较少时，也会对电影视频播放量产生一定的正向影响，但是由于存在边际效用递减规律，过多的弹幕信息会令用户感觉到视觉混乱或是剧透等不适，且积极弹幕信息过多会导致用户潜意识对其真实性存在怀疑，所以会削弱对播放量的正向影响。基于此，本研究提出以下假设：

H6a： 弹幕数量越多的电影视频电影播放量更高

H6b： 弹幕数量较少时，弹幕的正向情感越强烈，电影视频的播放量越高；当弹幕数量过多，正向影响会减弱。

3.2.7 评论对电影视频播放量的影响

与弹幕不同，评论是用户在观看视频后的评价，与视频的时间轴没有必然的关联，且以留言板的形式呈现，一般包含对视频的整体解读，在线评论的数量、情感以及异质性等都会对视频播放量造成一定的影响。首先，在线评论的数量越多，表明观看电影视频的活跃用户数越多，高度活跃的用户数会加深其他用户对该电影视频的印象并提高其知晓程度，增加了更多用户对该电影视频进行讨论的可能。王晓耘等对美国市场和国家市场的 169 部影片进行了分析，实证研究结果表明消费者的评论数量是影响票房收入的重要因素^[52]，即评论的数量可以通过消费者对视频的知晓程度来影响其播放量评论数量越多，表明该视频受到的关注度越高，继而才会引发更多的消费者关注并且选择观看，影响视频的播放量。

而评论情感对视频播放量的作用机制则与评论的数量不同，评论的情感是通过影响用户对视频的态度进而影响视频播放量，积极的评论有助于提高用户对于视频的期望值，提升用户对视频的好感度以正向影响视频的播放量，否则会造成负向影响。KIM 等通过实验探究评论的情感倾向对消费者的影响，研究发现当评论数量较少时，负面评论对于用户的影响会减弱，因为用户会因为评论数量过少而将其归于非理性评论；而当评论数量较多时，正向和负向的评论则会相应地影

响用户的选择^[53]。Chen 等通过研究在线评论的情感倾向对用户决策的影响发现, 正面和负面评论都会影响用户的选择, 并且正面评论的影响更大^[54]。同时, 评论的异质性也会对播放量产生影响, 但是学者对于其影响的看法略有不同。Quaschnig 等认为异质性低的评论比异质性高的评论更容易得到用户的认可, 此时消费者购买产品的意愿更强烈^[55]; Lee 等认为如果用户在电影评论中呈现出多种不同的情绪, 对于没有观看的用户来说会更可行, 因为用户会认为有一定异质性的评论更加真实^[56], 李亮等学者也认为对于体验性产品的在线评论来说, 极端性越高用户能感知到的价值越大, 越能够促进用户的消费^[57]。

综上所述, 评论数量越多越能引起消费者的关注, 从而提高电影视频的播放量, 而评论中的正向情感倾向有助于降低消费者的感知风险, 促使消费者观看该视频, 但是负向评论也并非一定对播放量造成负向影响, 一定评论的异质性有助于得到消费者的信任, 也会正向影响视频的播放量。因此, 本研究提出以下假设:

H7a: 评论数量越多的电影视频电影播放量更高。

H7b: 积极评论会正向影响电影视频播放量, 当评论数量较少时, 消极评论对电影视频的负面影响较小。

H7c: 评论的异质性在一定范围内会正向影响电影视频播放量。

3.2.8 弹幕和字幕情感关联性对电影视频播放量的影响

弹幕和字幕有明显的特征关联性, 即都与视频的时间轴相呼应, 但是弹幕还存在一个本身的发表时间, 同时弹幕与字幕的字数长度都有所限制, 并且弹幕还会带有强烈的情感表达, 而字幕往往与视频剧情的走向相关。弹幕的情感反映了观看视频的用户对于剧情所产生的情感, 字幕的情感则是随着视频的情节而改变的, 结合弹幕和字幕的情感分析可以归纳出用户对于视频情节的喜恶, 同时可以分析哪种情节会引起更多的用户进行讨论^[36]。除此之外, 视频中用户发表的弹幕在与字幕情感极性相同且演变趋势形同的情况下, 或者弹幕与字幕的情感种类大致形同, 可以说明该视频的内容与预期效果较为相符; 相反, 如果弹幕和字幕的情感极性和演变趋势都相反, 则表示可能有部分剧情的走向和大部分观众的想法

不相符，也有可能是视频的制作人员特意为之或者后期具有反转剧情^[37]。综合上述分析，结合现在电影视频的现状，很多电影都不是单纯的叙事风格，会在很多地方埋下伏笔，最后出现剧情反转，以达到出人意料的效果，获得更高的讨论量和播放量，因此本研究提出以下假设：

H8a:弹幕和字幕情感种类大致相同的电影视频播放量更高。

H8b:弹幕和字幕情感倾向及演变趋势有反转的电影视频播放量更高。

3.3 本章小结

本章结合本研究内容，对 SOR 理论、感知风险理论和感知价值理论进行适用性分析，并在国内外学者的研究基础上进一步梳理了电影视频播放量与各个变量之间的关系，同时提出了相关假设，为后续的研究打下基础。

第四章 电影视频弹幕和字幕及评论文本情感分析

本章节通过弹幕情感分析，能够了解观众对于剧情走向、画面呈现等方面的真实想法，对于评论的情感分析也能了解观众看完影片的整体喜恶程度，有助于从用户的角度分析电影视频引起的情感变化及效果。由于在电影视频中字幕与人物的动作、表情、台词等同步呈现，其内容与剧情发展密切相关，能够传达人物的内心世界及思想感情，所以本章节对字幕进行情感分析能梳理电影视频情节发展过程中情感的变化脉络，同时将弹幕与字幕的情感进行种类及趋势变化的关联性分析，有助于分析两者情感的契合及差异程度，为该类视频的制作、宣发、改进等提供新的方向和评估工具。

4.1 文本数据来源及预处理

4.1.1 文本数据来源

(1) 弹幕、评论文本数据爬取

本文依托于 Python 的爬虫框架对爱奇艺的电影视频相关弹幕及评论文本进行爬取，大致的爬取流程如图 4-1 所示。

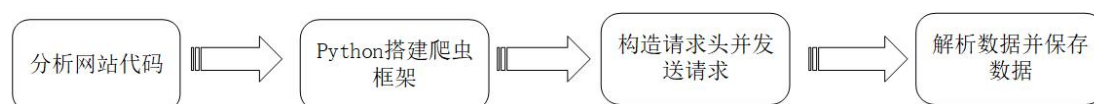


图 4-1 文本数据爬取流程
Fig 4-1 Text data crawling process

①分析网站代码。通过对要爬取的电影类视频的网页进行分析，发现页面请求头的方式为 get。由于弹幕是浮于视频表面的，且是随着视频的播放滚动加载的，所以能得知弹幕为 JS 异步加载。在开发者模式下进行抓包，发现了特殊的请求“danmu”，通过审查发现其中存放着需要的弹幕数据，于是根据 URL 提取出与每一部电影对应的弹幕数据的 target_id、vid 和 timestamp，timestamp 是用来控制页数的，首条 URL 为 15，根据数据包的更新时长，得出时间戳每 30 会更新一次弹幕信息，单位为秒，；而评论数据的获取也需要进入开发者模式抓包，其存放在

“comment”请求的 URL 中，跳转至 URL 即可获得该电影对应的 comment_id。

②搭建爬虫框架。导入爬虫需要使用的 python 相关库，并构建用户代理池，利用 random.choice 函数从用户代理池中随机选择一个用户代理，设置好用户对应的 headers 信息，以伪装成浏览器。

②发送请求。导入 requests 请求模块，并通过该模块里的 get 请求方法，对每一部电影视频对应的 url 发送请求，同时携带上 headers 请求头。

③解析数据。网页数据格式为标准的 json 格式，因此可以用 json.loads 直接解析数据，并使用 strict 参数解决部分 json 内容的格式解析报错问题。

④保存数据。对于弹幕文本，需要遍历要提取的信息，包括弹幕内容、发送时间、会员等级、点赞数、弹幕 id 等，并写入 csv 文件；对于评论文本只提取评论的内容并写入 txt 文件。

（2）字幕文本数据获取

本研究的字幕数据来源于一个分享交流下载字幕的网站（subhd.tv），下载的字幕文件类型主要为 srt 和 ass。

4.1.2 文本数据处理

（1）缺失值、重复值处理。对于文本内容来说，缺失和重复的数据均被认为是必须删除的无效数据。利用 drop_duplicates 函数对重复数据进行删除，并 dropna 函数删除存在缺失值的行，从而完成初步的数据清洗。

（2）机械压缩去重。机械压缩去重即对每条文本中的重复数据进行删除，弹幕内容存在例如“哈哈哈哈哈”的文本数据，而对文本进行情感分析时，只需要一个“哈”即可。针对这种情况需要定义一个机械压缩函数，并对弹幕和评论文本进行应用。

（3）特殊字符过滤。部分文本内容包含表情包、特殊符号等，这些脏数据也会对情感分析产生一定影响，这种情况需要通过正则表达式过滤，本文运用 extract(r"([\u4e00-\u9fa5]+)")函数匹配出文本的中文内容并保留。

（4）数据类型转化。爬取的弹幕相关数据中，发送时间字段是以秒为单位，

需要改成时间，本文自定义了一个 `time_change` 函数进行处理，将 `time_change` 函数应用于发送时间字段，将时间格式设置为时分秒；并将弹幕文本数据类型由 `object` 更改为 `str`。

（5）其他文本处理。针对弹幕和评论文本的现状，考虑到发送的内容果断可能不具有实质的分析效果，所以剔除了在 4 个字以下的弹幕文本内容；对于评论文本而言，有可能存在粉丝控评（网络流行词，即操控评论，通常指粉丝群体有组织地对社交平台上正面的评论点赞回复，使其保持前列，多用来描述粉丝控制明星相关评论的走向）的现象，所以将评论文本的长度限制在 4-200 个字以内。对于从字幕网站上下载的字幕来说，虽然已经经过前期处理整合，但是为了便于分析，通常使用 `ass` 类型的字幕文件，需要利用字幕转换助手将部分 `srt` 类型的字幕文件转换为 `ass` 类型。

最终处理完成的部分文本数据示例如图 4-2 所示，以弹幕文本为例。

	content	showtime
0	我最期待的一步电影终于上映了	2022-11-01 00:00:01
1	这跑步的细节演的好	2022-11-01 00:00:01
2	终于上映了	2022-11-01 00:00:01
3	倪妮演的很好	2022-11-01 00:00:01
4	我去眼神完全变了	2022-11-01 00:00:01
...
44798	请大家评分	2022-11-01 01:59:59
44799	片尾曲叫什么	2022-11-01 02:00:00
44802	比肩战狼	2022-11-01 02:00:00
44803	说黑化到底的真是恶魔	2022-11-01 02:00:00
44805	满分拿走	2022-11-01 02:00:00

37027 rows × 2 columns

图 4-2 处理后文本数据示例
Fig 4-2 Example of processed text data

4.2 弹幕及评论文本情感分析

4.2.1 基于 PaddleHub 的弹幕及评论情感分析

PaddleHub 是依托于 PaddlePaddle 框架的预训练模型应用工具，可以结合 Fine-tune API, 便捷地使用高质量的预训练模型快速完成模型迁移到部署的全流程工作。PaddleHub 提供的预训练模型涵盖了包括文本情感分析在内的很多主流模型。PaddleHub 的基本框架如图 4-3 所示，其中 Parameter 和 Arguments 表示神经网络中所有的数据和参数；Parameter 表示神经网络中层与层之间的连接参数，而 Argument 表示每一层的输入和输出。如图 4-4 所示，Parameter 表示图中的连接的实线，而 Argument 表示下图中的输入(Input)和输出(Output)，并且 Parameter 与 Argument 并不只保存了参数值，还同时保存了梯度，动量等信息。

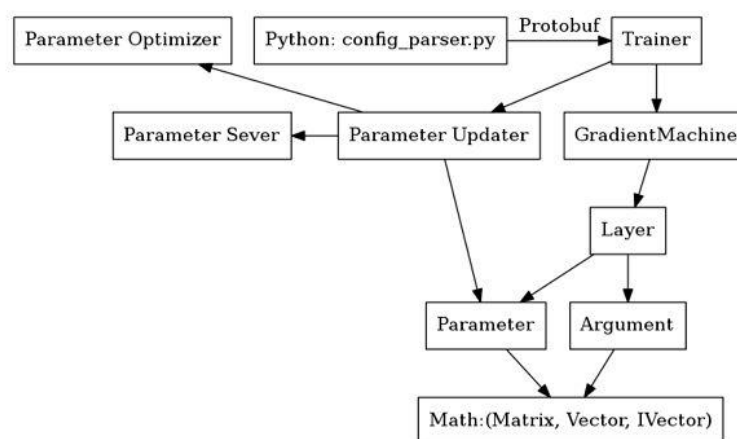


图 4-3 PaddleHub 模型基本框架（来源于 CSDN）
Fig 4-3 PaddleHub model basic framework (Origin from CSDN)

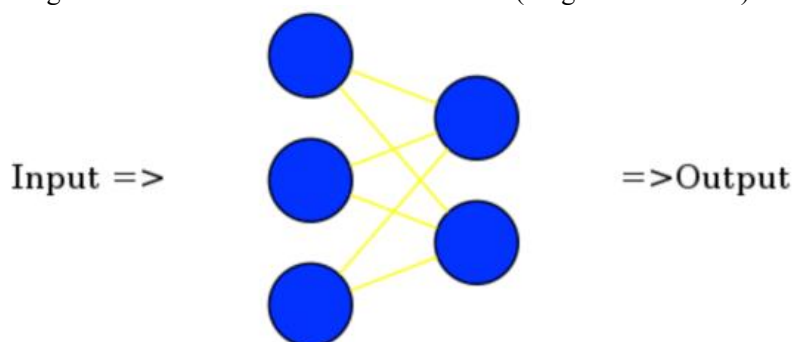


图 4-4 神经网络处理框架（来源于 CSDN）
Fig 4-4 Neural Network Processing Framework (Origin from CSDN)

本研究利用 PaddleHub 进行文本情感分析的主要包括以下几个步骤：

- （1）搭建模型运行的环境。首先需要进行 paddle-GPU 环境配置，需要安装

CUDA (Compute Unified Device Architecture) 和 cuDNN。CUDA 是一个通用的并行计算架构,该架构使 GPU 能够解决复杂的计算问题,其安装的版本依赖于电脑显卡驱动的版本;cuDNN 则是用于深度神经网络的 GPU 加速库,其安装依赖于 CUDA 的安装版本。安装完成后即可配置环境变量并在 Python 中安装 paddle-GPU,便可在 Python 中导入 PaddleHub 包进行分析。

(2) 加载预训练模型。Senta 开源了一系列预训练模型,包括 BOW (Bag Of Words)、CNN (Convolutional Neural Networks)、GRU (Gated Recurrent Unit)、LSTM (Long Short Term Memory)、Bi-LSTM (Bidirectional Long Short Term Memory)。由于在实际场景中,对于文本情感的预测需要考虑到词语在句子中前后顺序,即结合上下文进行分析,才能更好地得到语义之间的依赖关系。因此本研究采用 Bi-LSTM (Bi-directional Long Short-Term Memory) 模型,为序列模型,因为虽然 LSTM 通过训练过程可以学习到记忆和遗忘的信息,但是无法编码从后到前的序列信息,即 LSTM 模型只能考虑上文信息,所以本研究采用双向 LSTM 结构,该结构在 LSTM 的基础上结合了一个逆向的序列,组成双向长短时记忆网络,能更好地捕获句子中的语义特征。

(3) 文本情感值预测。运行 Bi-LSTM 模型后,针对输入的文件中的每条文本都会进行情感倾向的分析,会得到 sentiment_label (情感倾向值,1 或 0)、sentiment_key (情感倾向,积极或消极)、positive_probs (积极情感的概率)、negative_probs (消极情感的概率)。其中 sentiment_label 为 1 表示积极,为 0 表示消极,本研究以积极情感预测的准确率代表每一条文本的情感分值,并写入原始数据中以进行分析。

本研究总共收集了 160 部影片的弹幕和评论文本,并依据影片的名称将其存储为独立的工作表格,将待分析的文本数据单独存储为一行,每一行为一条独立的弹幕或评论文本数据。

利用 PaddleHub 模型进行弹幕和评论文本情感分析的具体过程如图 4-5 所示。

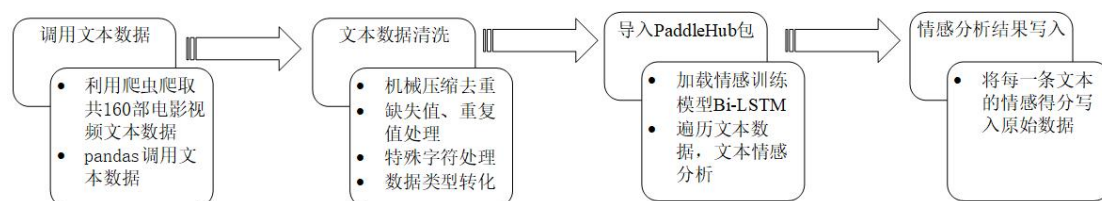


图 4-5 文本情感分析流程
Fig 4-5 Text emotion analysis process

首先利用 python 中的 pandas 函数读取文件，经过数据清洗后，调用情感分析需要的 PaddleHub 包以及情感训练模型 Bi-LSTM，利用 其对弹幕和评论文本数据进行遍历，经过模型分析训练，将会得到每一条弹幕及评论对应的情感倾向分类以及他们属于某一分类的概率。本研究以积极情感的概率作为每条文本数据的情感分值并写入原始数据中，因为概率的取值范围为 0-1，该值越接近 1 则表示该条文本积极情感更强烈，越接近 0 则表示消极情感更强烈，最后部分结果示例如图 4-6 所示。

	content	showtime	情感分值
showtime			
2022-11-11 00:00:01	我最期待的一步电影终于上映了	2022-11-11 00:00:01	0.9955
2022-11-11 00:00:01	这跑步的细节演的好	2022-11-11 00:00:01	0.8740
2022-11-11 00:00:01	终于上映了	2022-11-11 00:00:01	0.9344
2022-11-11 00:00:01	倪妮演的很好	2022-11-11 00:00:01	0.9874
2022-11-11 00:00:01	我去眼神完全变了	2022-11-11 00:00:01	0.5959
...
2022-11-11 01:59:59	请大家评分	2022-11-11 01:59:59	0.4002
2022-11-11 02:00:00	片尾曲叫什么	2022-11-11 02:00:00	0.1383
2022-11-11 02:00:00	比肩战狼	2022-11-11 02:00:00	0.8801
2022-11-11 02:00:00	说黑化到底的真是恶魔	2022-11-11 02:00:00	0.0598
2022-11-11 02:00:00	满分拿走	2022-11-11 02:00:00	0.9674

37027 rows × 3 columns

图 4-6 文本情感分值计算结果示例
Fig 4-6 Example of text emotion score calculation result

每条文本的情感分值计算完成后，需要计算每一部视频弹幕和评论文本整体的情感得分，本研究使用每条文本情感分值的均值代表整体情感得分，对于评论文本而言，以所有文本情感分值的方差表示其离散程度，即可代表评论之间的异

质性。

4.2.2 考虑情绪感染机制及衰减效应的弹幕情感修正

由于用户所处的环境会在一定程度上对其行动或选择造成影响，而弹幕具有随视频播放滚动加载的特点，且包括点赞、加一（其他用户对赞同的弹幕点击加一后会直接发送相同的弹幕）等功能，使用户能在短时间内接收到大量的不同情感倾向的信息。情绪感染理论认为他人的情绪存在传播性，并且为了匹配他人的情绪还会产生自动化传递的过程，结合该理论与视频弹幕的特点，本研究认为弹幕的情绪感染机制为：初始的传播者会根据电影视频内容产生不同情感倾向的信息，这些信息则会在视频播放的过程中以弹幕的形式进一步传播给其他用户，而后续观看电影视频的用户不仅会收到来自电影内容的刺激，同时也会被其他用户积极或消极的情绪感染，从而会增强初始的情感强度；但是用户同时也会对弹幕信息的内容以及接收次数进行判断，如果同质化的情绪信息被过多传播，其他用户也会产生免疫机制，又会对情感强度产生抑制。

基于此，本研究定义用户在单位时间内积极情绪感染率的计算公式为：

$$\lambda_p = \frac{p_sum}{flu_max} \quad (4-1)$$

而用户在单位时间内消极情绪感染率的计算公式为：

$$\lambda_n = \frac{n_sum}{flu_max} \quad (4-2)$$

最后，用户在单位时间内的情绪免疫率的计算公式为：

$$\rho = \frac{p_sum}{flu_max} \left(1 - \frac{|n_sum|}{flu_max}\right) \quad (4-3)$$

其中用户所接收到的积极信息的影响力 $P_SUM = \sum_1^n A_{ij} * 1$ ，接收到的消极信息的影响力 $N_SUM = \sum_1^n A_{ij} * (-1)$ 。 A_{ij} 表示单位时间内系统中个体之间的亲密度，取值范围为 1-10，由于发送弹幕的用户彼此无法得知对方的身份，所以本研究认为所有系统内的用户亲密度最弱，所以该值可视为常量 1。 $flu_max = 10(N - 1)$ ，该式表明在一个系统内如果各用户之间具有最高的亲密度，且所有用户之间传播的情感信息都为积极的，那么根据从众心理，此时的情感影响力最大，N 为单位时间内系统中的用户数。

虽然发送弹幕的用户能在一定程度上对其他用户的情绪造成正向或负向的感染，但是由于弹幕随时间滚动展现，如果在一段时间内对用户没有出现新的刺激，随着时间的流逝以及环境适应性，都会产生情绪衰减效应；魏静等人认为用户之间是存在信息传播的，并且认为如果一个用户多次接收同类信息的传播会存在叠加和衰减效应，即多次相同的弹幕信息会削弱其中携带的情绪对其他用户的影响程度^[58]。基于此，本研究在弹幕情感的分析中，不仅考虑到了情绪的感染机制，同时也会对弹幕的情绪衰减效应进行分析。

本研究在殷雁君^[59]等人研究的基础上，结合弹幕情感传播的特点，定义本文情绪衰减公式为：

$$I(t) = \eta * \exp\left(\frac{-\theta * \text{sumSent}}{P_E * \text{time}}\right) \quad (4-4)$$

其中 $I(t)$ 表示 t 时刻的情绪衰减强度， η 为场合调节因子， $\eta \in [0,1]$ ，可以调节不同场合下的情绪衰减速度； θ 为情绪调节因子， $\theta \in [0,1]$ ，以对不同程度的情绪衰减进行调节，情感越强烈，越易宣泄，则衰减程度越大，即 θ 的值与初始的情感强烈程度呈比例对应； sumSent 为单位时间内用户在观看电影视频时发送的弹幕数； P_E 为用户的性格外向程度， $P_E \in [0,1]$ ，由于本文的用户均为发送弹幕的用户，为便于研究，定义该类用户的 P_E 均0.5。

因此，本研究在考虑情绪感染机制以及衰减效应的作用下，所得到的弹幕情感的整体分析方法流程如图4-7所示：

（1）爬取弹幕文本数据并进行清洗以及数据预处理后，利用 PaddleHub 中 Senta 开源的情感分析模型 Bi-LSTM 对弹幕文本进行分析，得到单句的初始情感值，记为 s 。

（2）考虑情绪感染机制对弹幕情感值的影响，与上一节中计算弹幕情感时的重采样时间保持一致，在一个单位时间的弹幕情感系统中，以0.5为正负情感的分界值，分别计算系统内的正负情感的信息影响力，从而计算积极和消极情绪的感染率以及免疫率。假设单句情感值为 s 的弹幕在叠加情绪感染机制后的情感值为 S ，计算公式如（4.5）所示：

$$S = s * (\lambda_p - \lambda_n - \rho + 1) \quad (4-5)$$

(3) 结合情绪衰减效应再次对弹幕情感值进行修正, 可以得到在单位时间 t 时最终的弹幕情感值的计算公式, 如 (4.5) 所示:

$$S(t) = [s * (\lambda_p - \lambda_n - \rho + 1)] * \left(1 \pm \eta * \exp\left(\frac{-\theta * \text{sumSent}}{P_E * \text{time}}\right)\right) \quad (4-6)$$

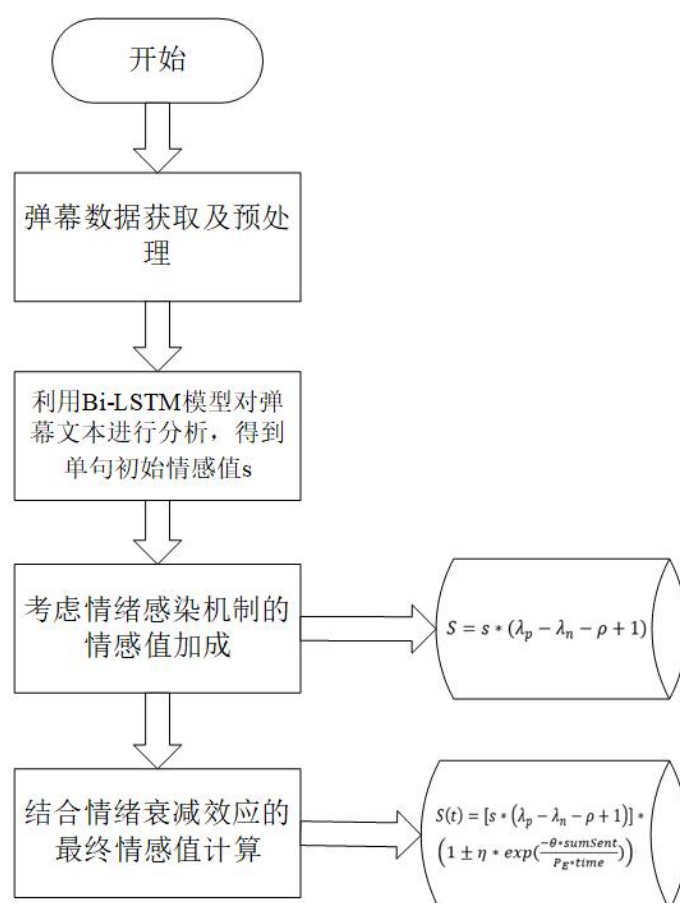


图 4-7 考虑情绪感染机制以及衰减效应的弹幕情感值分析

Fig 4-7 Analysis of bullet screen emotional value considering emotional infection mechanism and attenuation effect

4.3 弹幕与字幕情感关联性分析

4.3.1 弹幕与字幕情感种类关联分析

通过比较弹幕与字幕情感种类的关联, 能够得知电影视频所传达的情感与用户在观看的过程中产生的情感的分布情况, 可以在一定程度上对电影视频质量的

高低进行分辨,并分析其种类分布的一致程度对电影播放量的影响。

在本研究中,弹幕与字幕的情感种类是根据情感词汇本体库中的七个情感类别来划分的,目前中国研究较为成熟的词典有大连理工大学情感词汇本体库、知网的 HowNet 情感词典及 TW 大学中文情感极性词典等。本文选择的基础词典是大连理工大学情感词汇本体库,此词典将情感分为“乐”、“好”、“怒”、“哀”、“惧”、“恶”、“惊”7 个大类和 21 个小类,利用该词典可以分别统计出每种类别下的弹幕和字幕的数量。

大连理工大学情感词汇本体库是依据国外 Ekman 的原始情感分类体系构建的,也是该大学实验室整理研究、标注的中文本体资源。该资源以 Ekman 为基础,增加了情感类别“好”,将积极情感进行了更详细的分类,并从多个不同的角度描述中文词汇或词组,包含词性类别、情感类型、情感强度及极性等信息。最终词汇本体中的情感共分为 7 大类 21 小类,如图 4-6 所示。需要注意的是,在情感词典中,一个情感词可以对应多种情绪,情感分类是描述一种情绪类型的主要情感分类,而辅助情感则是这种情感词在其主要的情绪类别上所包含的其他情绪类型。

本研究引入大连理工大学情感词汇本体库对弹幕和字幕的情感种类分析的步骤包括:

(1) 调用 pandas 读取需要的弹幕或字幕数据。

(2) 导入大连理工大学中文情感词典,通过 iterrows 函数对词典的每一行进行遍历,对词典内不同情感类型的词语做标记,分别为 Happy、Good、Surprise、Sad、Fear、Disgust、Anger; 并定义积极情感的词语为“乐”、“好”、“惊”的组合,消极情感的词语为剩下种类的组合。

(3) 导入停用词典,并定义利用停用词典对文本进行分割的函数。

(4) 定义计算情感种类的函数,对分词后的文本中的七大类情感种类进行计算,写入新的工作表中,得到的结果如图 4-8 所示。

Emotion	Word	Num	
good	师父	1	
good	师父	1	
fear	可怕	1	
disgust	迷失	1	
good	好吃	1	
good	喜欢	1	
sad	痛苦	1	
disgust	古怪	1	
disgust	古怪	1	
disgust	黑心	1	
disgust	无法无天	1	
good	闻名	1	
disgust	荒废	1	
fear	不敢当	2	
good	辟邪	1	
fear	不敢当	2	
good	安静	1	
good	感谢	1	
disgust	随便	1	
good	希望	1	
good	肯定	1	
happy	放心	1	
sad	不行	1	
anger	惩罚	1	
fear	小心	1	
fear	小心	1	
fear	小心	1	
fear	小心	1	

图 4-8 文本情感种类统计表

Fig 4-8 Statistical Table of Text Emotion Types

(5) 对不同情绪种类以及积极和消极种类的频数进行计算, 得到某一文本不同类型的情绪词数量。

对本研究中涉及的弹幕和字幕文本按照以上步骤进行分析后, 得到了不同电影视频的弹幕和字幕的情绪词频分布情况, 并绘制雷达图进行分析, 以“二哥来了怎么办”、“超越”为例, 如图 4-9 所示。

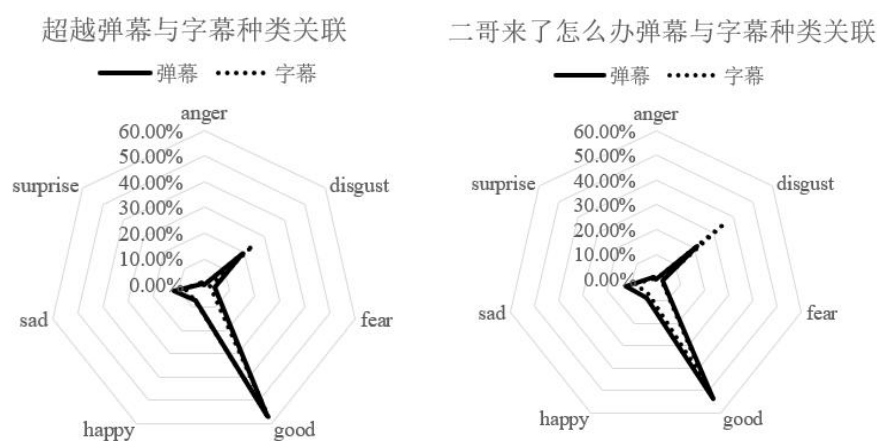


图 4-9 弹幕及字幕情绪词分布雷达图

Fig 4-9 Radar map of bullet screen and subtitle emotional word distribution

由于同一部电影视频的弹幕与字幕文本数量差异较大，所以按照不同情感所占的比例绘制相应的雷达图。从图 4-7 分析可知，尽管同一部电影视频的情感种类分布整体一致，但是各类情感的占比仍然存在一定的差异性。通常情况下，“好”和“恶”两类情感词在弹幕和字幕文本中所占的比例比较高，除此之外就是“悲”，且这几种情感种类的比例弹幕都略高于字幕，而其他种类的情感词的比例情况则相反。

利用 SPSS 软件进一步对弹幕及字幕情感种类进行关联分析，具体结果如表 4-1、表 4-2 所示：

表 4-1 二哥来了怎么办情感种类关联分析

Table 4-1 What to do when the second brother comes Association analysis of emotion types								
文本类型	anger	disgust	fear	good	happy	sad	surprise	皮尔逊卡方值
弹幕	0	1101	138	2753	439	642	63	29.561
字幕	0	90	6	114	17	22	2	

表 4-2 超越情感种类关联分析

Table 4-2 Relevance Analysis of Transcendental Emotions								
文本类型	anger	disgust	fear	good	happy	sad	surprise	皮尔逊卡方值
弹幕	0	1407	326	4183	518	875	42	11.680
字幕	0	64	6	150	19	20	3	

利用 SPSS 软件对字幕及弹幕的情感种类进行卡方检验，可以从卡方值看出不同电影视频弹幕及字幕情感种类的差异程度，电影二哥来了怎么办及超越的卡方值均分别为 29.561 和 11.680，电影视频二哥来了怎么办的皮尔逊卡方值更大，即表明其弹幕与字幕在情感种类上的差异程度也更大。

4.3.2 弹幕与字幕情感趋势关联分析

结合弹幕、字幕都与电影视频播放的时间轴相关这一特点，通过对弹幕与字幕的情感随视频播放的趋势进行分析，有助于总结弹幕与字幕的情感趋势变化规律，并得到不同电影情感趋势的一致性，从而分析一致性的高低对播放量的影响。

由于弹幕和字幕都与电影视频的时间轴高度相关，大部分电影视频的播放时长为一个半小时到两个半小时之间，又因为电影视频前 5 分钟一般为片头时间，

此时的弹幕大多与电影情节无关，因此在利用 PaddleHub 模型对每一条文本进行分析后，设定每 5 分钟进行一次情感分值的重采样，以确定随着电影视频的播放，弹幕及字幕文本情感值的变化，重采样后的情感分值数据如图 4-10 所示。

	发布时间	会员等级	弹幕点赞	情感分值
0	2022-05-28 00:00:00	1.11093	0.132107	0.673537
1	2022-05-28 00:05:00	1.40743	0.100659	0.586782
2	2022-05-28 00:10:00	1.62141	0.0596026	0.553693
3	2022-05-28 00:15:00	1.58077	0.216155	0.549155
4	2022-05-28 00:20:00	1.71643	0.280992	0.403559
5	2022-05-28 00:25:00	1.75	0.115664	0.396345
6	2022-05-28 00:30:00	1.7697	0.0960854	0.363448
7	2022-05-28 00:35:00	1.72959	0.0610772	0.492303
8	2022-05-28 00:40:00	1.81877	0.0649351	0.561664
9	2022-05-28 00:45:00	1.66405	0.0767077	0.473653
10	2022-05-28 00:50:00	1.72819	0.0641531	0.46944
11	2022-05-28 00:55:00	1.83324	0.0658524	0.439302
12	2022-05-28 01:00:00	1.79036	0.0775681	0.442192
13	2022-05-28 01:05:00	1.77369	0.059892	0.432704
14	2022-05-28 01:10:00	1.75644	0.054768	0.450265
15	2022-05-28 01:15:00	1.74428	0.0547533	0.452624
16	2022-05-28 01:20:00	1.78125	0.0852754	0.36992
17	2022-05-28 01:25:00	1.66417	0.0700535	0.473855
18	2022-05-28 01:30:00	1.58451	0.169022	0.510627
19	2022-05-28 01:35:00	1.71139	0.154984	0.392558
20	2022-05-28 01:40:00	1.62493	0.169995	0.445666
21	2022-05-28 01:45:00	1.59536	0.0508568	0.468403
22	2022-05-28 01:50:00	1.63806	0.0875212	0.493558
23	2022-05-28 01:55:00	1.56894	0.0401606	0.542596
24	2022-05-28 02:00:00	1.28926	0.0165289	0.521964

图 4-10 情感分值重采样数据
Fig 4-10 Emotional score resampling data

为了更加清晰的观察每一部电影视频弹幕与字幕文本情感随之时间的变化趋势，绘制了弹幕及字幕情感变化趋势图，以“悬崖之上”、“刺杀小说家”、“五个扑水的少年”、“拆弹专家 2”为例，如图 4-11、4-12、4-13、4-14 所示。

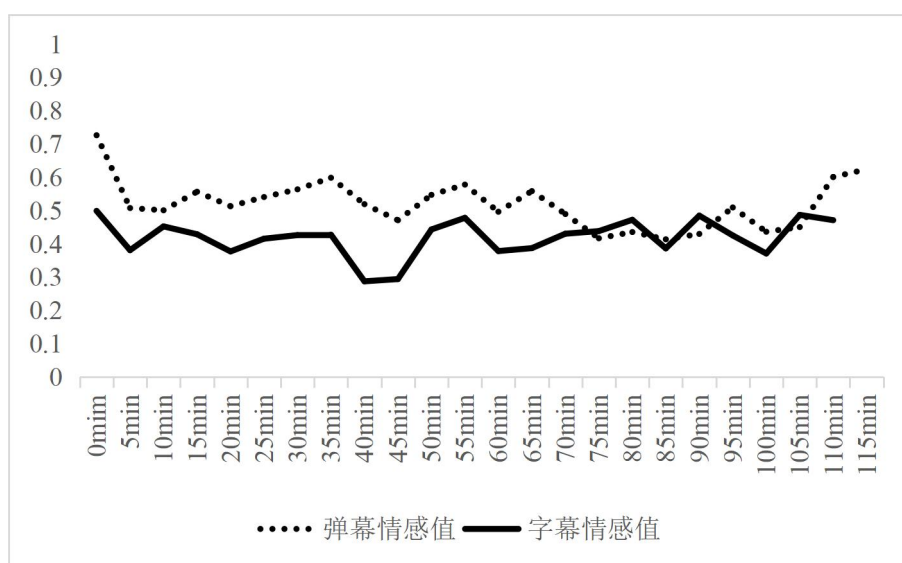


图 4-11 悬崖之上弹幕及字幕情感变化趋势

Fig 4-11 Emotional change trend of bullet screen and caption on cliff

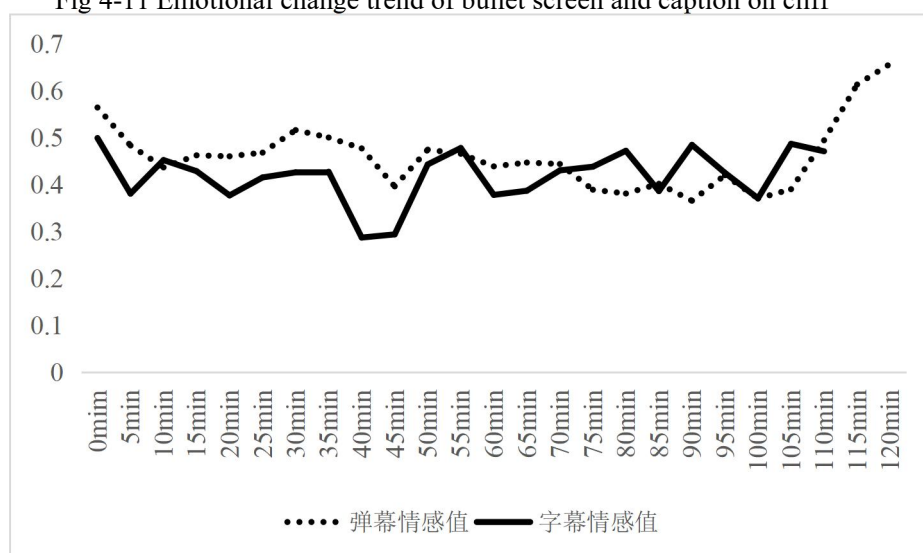


图 4-12 刺杀小说家弹幕及字幕情感变化趋势

Fig 4-12 Assassination of a novelist bullet screen and the changing trend of subtitles

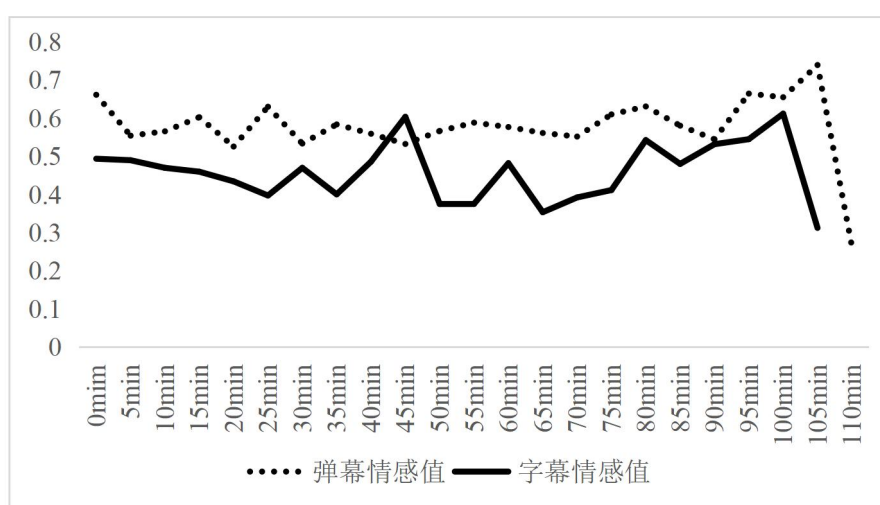


图 4-13 五个扑水的少年弹幕及字幕情感变化趋势

Fig 4-13 Emotional change trend of five juvenile bullet screens and subtitles

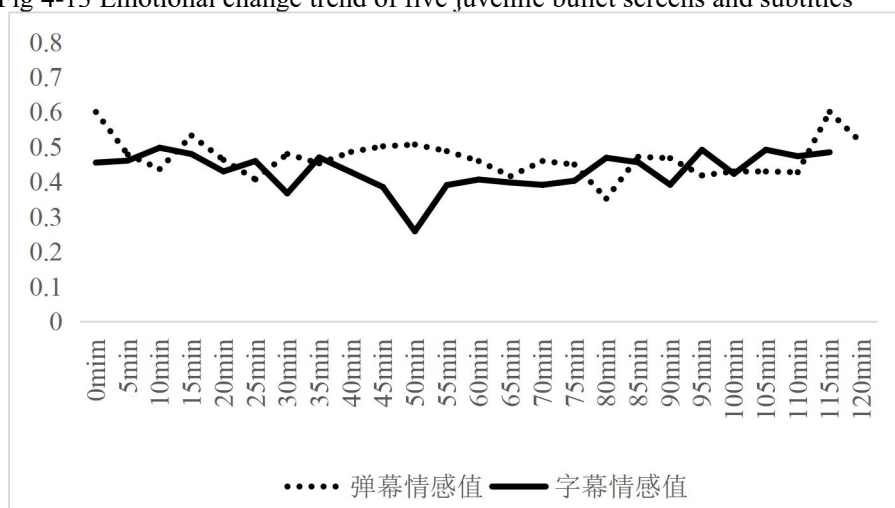


图 4-14 拆弹专家 2 弹幕及字幕情感变化趋势

Fig 4-14 Bullet demolition expert 2 bullet screen and subtitle emotional change trend

由于弹幕与字幕的情感值都与时间轴相关，所以可以其情感趋势变化的相关性可利用相关函数进行分析，以上四部影片趋势关联分析结果如表 4-3 所示：

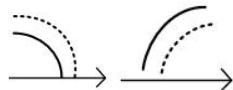
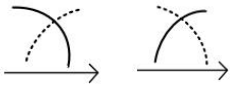
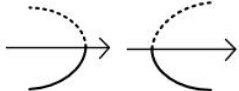
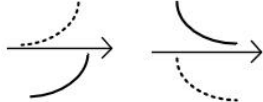
表 4-3 弹幕与字幕情感趋势关联分析

Table 4-3 Correlative analysis of emotional tendency between bullet screen and caption		
电影视频	弹幕与字幕情感极性	弹幕与字幕情感变化趋势
悬崖之上	相反	0.245
刺杀小说家	相同	0.338
五个扑水的少年	相反	-0.096
拆弹专家 2	相同	-0.127

依据上表可知，以上四部电影视频弹幕与字幕的情感趋势整体有的呈负相关性，有的呈正相关，通过比较绝对值的大小，电影八佰的相关系数最大，表明该电影视频的弹幕与字幕情感趋势的正相关程度更大。

同时，通过观察部分情感趋势图可以发现，在一般情况下，视频刚开始播放时的弹幕情感强度较高，通常是因为在视频的开头大部分观众都会签到打卡，表达自己激动的心情，而后续随着正片的开始，情感值会逐渐回落，最终随着电影的剧情进行波动。对情感趋势进行总结可以将弹幕与字幕的情感趋势变化分为四种类型，如表 4-4 所示。

表 4-4 弹幕与字幕的情感趋势变化
Table 4-4 Emotional trend change of bullet screen and subtitle

情感极性	情感变化趋势	弹幕与字幕情感变化曲线图	相关性
相同	趋势相同		正相关
	趋势相反		弱负相关
相反	趋势相反		负相关
	趋势相同		弱正相关

备注：实线代表字幕，虚线代表弹幕

以 0.5 作为积极和消极情感的分界线，如果在某时刻弹幕和字幕情感分值均位于 0.5 同侧则视为情感极性相同，否则视为相反。当弹幕与字幕的情感分值位于分界线 0.5 同侧时，且其情感变化趋势相同，表明弹幕与字幕的情感存在高度正相关性，也就是说用户在观看电影视频的过程中，与某时刻的字幕所呈现的视频内容和情感产生了高度共鸣，或者该影片的情节走向在预料之中；而如果弹幕

和字幕的情感变化趋势相反，但是情感的极性没有发生改变，则表明其情感趋势呈弱负相关性，可能是由于用户对于视频的内容有一定的不认同，但是没有产生非常强烈的抵触心理。

当弹幕与字幕的情感分值位于分界线 0.5 异侧时，如果其情感变化趋势相反，表明弹幕与字幕的情感存在高度负相关性，究其原因主要可能有以下两点：第一种是用户在观看电影视频的过程中，部分情节违背预期想法，感受到了与正常价值观或伦理违和的内容，产生了强烈的负面情绪，第二种是视频的表现手法，即是通过反向拍摄从而达到讽刺的目的；而如果弹幕和字幕的情感变化趋势相同，则表明其情感呈弱正相关性，可能是由于在某一情节的初期，观众对于视频的内容存在一定的抵触心理，而随着情节的完善，观众逐渐能够和情节内容产生共鸣，情感极性也会从最开始的相反而逐渐趋于相同。

4.4 本章小结

本章基于飞桨开发的 PaddleHub 包，调用内置的情感预训练模型 Bi-LSTM，对弹幕、评论以及字幕文本进行情感分析，不仅包括每一部电影视频整体的弹幕及评论文本情感分值计算，还涉及到评论文本情感的异质性计算，还在考虑情绪感染机制及衰减效应的基础上对弹幕情感值进行修正。同时，由于考虑到弹幕与字幕文本都与电影视频的时间轴相关，所以在利用 PaddleHub 模型进行情感分析的同时，结合电影视频的性质，设定了 5 分钟的情感重采样时间，以观察弹幕与字幕的情感变化趋势，并对变化规律进行总结；除此之外，还引入了大连理工大学情感词汇本体库对弹幕和字幕的情感种类进行分析，以雷达图及卡方检验的方式进行总结分析。

第五章 电影视频播放量影响因素实证分析

本章在前文研究假设以及弹幕、评论及字幕的情感分析的基础上，构建了电影视频播放量影响因素的模型，并利用 SPSS 进行回归分析及主体间效应检验分析，以验证前文所提出的假设。

5.1 数据获取及变量定义

5.1.1 数据获取

由于不同国家的电影产业发展速度不一，行业所处的程度区别较大，各个国家的历史及行业政策也各不相同，且国内电影近几年发展趋势尚好，所以针对国产电影播放量的影响因素，有必要进行单独研究。因此本文的电影数来源于 2020-2021 年间在我国国内公开上映的电影，利用后羿采集器在爱奇艺平台爬取了每部电影的基本信息，包括上映的年份、视频播放量、电影类型、导演及演员等；并建立电影视频播放量影响因素模型，对国产电影视频播放量影响因素的自变量与播放量之间的关系进行深入研究。

5.1.2 因变量

本文的因变量为截止到数据爬取日期，电影在腾讯视频平台上的播放量，由于该平台隐藏了播放量在前台的显示，所以用视频热度值代替播放量。同时结合大部分视频热度值的变化趋势，发现热度值的变化符合二八法则，即一般来说，当电影视频进入平台的两个月左右，其播放量占整体播放量的大部分，所以本研究将电影视频刚进入平台两个月期间的热度值视作视频整体的热度。

5.1.3 自变量

(1) 电影类型：本文将电影类型分为以下几大类型：爱情片、喜剧片、动作片、奇幻片、悬疑片、剧情片、惊悚片、犯罪片，根据类型设定虚拟变量，每部电影对应的变量赋值为 1，否则为 0。鉴于大部分电影存在多种电影类型，根据影视平台中电影类型的确定，本文最多选择前两种类型。

(2) 主创阵容：本文从百度指数官网上获取两大主演及导演的百度指数整体日均值，并以其平均值作为衡量明星影响力、导演知名程度的指标，最后根据百度指数将明星影响力分为 6 个梯队，分别赋值为 1-6；将导演知名程度分为 5 个等级，分别赋值为 1-5。

(3) 续集或 IP 改编：续集或 IP 改编都可被认为是品牌延伸性，本文认为如果该电影为续集或 IP 改编，则变量值设置为 1，否则为 0。

(4) 电影评分：电影评分能在一定程度上反映电影的网络口碑，本文电影评分来源于电影视频平台的官方数据。

(5) 电影票房：电影票房表明了电影在上映过程中的火爆程度，本文依据猫眼电影专业版数据获取了每部电影在上映时期的票房，以作为研究变量。

(6) 弹幕、评论及字幕：本文通过第四章对弹幕、评论及字幕相关文本的分析得出的量化数据，以之作为弹幕及评论情感强度、评论异质性、弹幕及字幕情感种类关联的度量；并根据弹幕、字幕的情感极性及其变化趋势确定相关性，自变量的定义于描述见表 5-1。

表 5-1 自变量描述与定义

Table 5-1 Description and definition of independent variables

自变量	自变量描述
电影类型	符合某种电影类型赋值为 1，否则为 0
主创阵容	两名主要导演及演员的百度指数平均值
续集	是则赋值为 1，否则为 0
IP 改编	是则赋值为 1，否则为 0
电影评分	以影视平台官方数据为准
电影票房	以上映期间官方票房数据为准
弹幕数量	爬取并清洗后的弹幕数量
弹幕情感	文本分析得出的弹幕情感值
评论数量	爬取并清洗后的评论数量
评论情感	文本分析得出的评论情感值
评论异质性	文本分析得出的评论异质性
弹幕与字幕情感种类相关性	以皮尔逊卡方值为准
弹幕与字幕情感趋势相关性	正相关赋值为 2，弱正相关为 1，负相关为 -2，弱负相关为 -1

本研究为了消除部分数据之间的异方差性，如弹幕数量、评论数量等量级相

差较大的数据,本研究借鉴统计学常用的数据处理方式,对其取以 10 为底的对数,以对数处理后的数据进行分析。

5.2 播放量影响因素模型构建及分析

5.2.1 描述性统计分析

本文研究的样本数据为 2020-2021 年间在中国国内上映的电影,共 160 部电影,其中还包含动画类型的电影,但是由于动画类型的电影为卡通电影人物,无法度量明星影响力,因此去掉此种类型的电影,最终保留 127 部电影。所有非虚拟变量的描述性统计结果如表 5-2 所示。

表 5-2 变量描述性统计
Table 5-2 Descriptive statistics of variables

变量	Max.	Min.	Avg.	Sd.
播放量(万)	39.55	3.49	20.08	8.72
评分	9.50	6.00	7.83	0.77
票房(万)	5772424.39	1.1	28116.40	73527.28
弹幕数量	56242.00	57.00	17927.30	14506.03
弹幕情感值	0.79	0.37	0.51	0.07
评论数量	26893.00	10.00	3292.35	4223.15
评论情感值	0.85	0.26	0.57	0.13
评论异质性	0.18	0.05	0.13	0.02

5.2.2 多元线性回归分析

本文依据自变量和因变量的关系构建如下电影视频播放量影响因素模型:

电影视频播放量= $\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\beta_4x_4+\beta_5x_5+\beta_6x_6+\beta_7x_7+\beta_8x_8+\beta_9x_9+\beta_{10}x_{10}+\beta_{11}x_{11}+\beta_{12}x_{12}+\beta_{13}x_{13}+\beta_{14}x_{14}+\beta_{15}x_{15}+\beta_{16}x_{16}+\beta_{17}x_{17}+\beta_{18}x_{18}+\beta_{19}x_{19}+\beta_{20}x_{20}+\beta_{21}x_{21}+\beta_{22}x_{22}+\mu$ 。其中 β_0 为常量, $\beta_1-\beta_{22}$ 为回归系数, x_1-x_{22} 分别是爱情电影、喜剧电影、动作电影、奇幻电影、悬疑电影、战争电影、剧情电影、犯罪电影、惊悚电影、明星影响力、导演知名程度、续集、IP 改编、电影评分、电影票房、弹幕数量、弹幕情感值、评论数量、评论情感值、评论异质性、弹幕与字幕情感种类相关性、弹幕与字幕情感趋势相关性, μ 是随机误差项。

本研究利用 SPSS 对数据进行线性回归分析,模型的分析汇总结果如表 5-3、5-4 所示,由该模型的分析结果可知,模型的整体拟合程度为 84.3%,自变量整体也通过了 F 检验,其 P 值小于 0.05。但是很多变量没有通过线性 T 检验及共线性检验,导致无法进一步展开研究,所以本研究考虑到整体数据的特性,在自变量较多的情况下准备进行逐步回归分析,筛选出模型中更重要的变量以观测分析结果。

表 5-3 模型摘要
Table 5-3 Model Summary

模型	R	R ²	调整后 R ²	标准误差	F	Sig
	0.933	0.871	0.844	.0961605837	31.869	0.000

表 5-4 模型回归系数
Table 5-4 Regression coefficient of model

自变量	未标准化系数		标准化系数		共线性统计	
	B	标准错误	Beta	t	容差	VIF
评分	.147	.365	.026	.403	.290	3.453
续集	.022	.035	.025	.629	.803	1.246
IP 改编	.015	.021	.030	.718	.704	1.420
明星影响力	.016	.007	.108	2.237**	.530	1.885
导演知名程度	-.003	.008	-.015	-.345	.674	1.483
爱情	.015	.038	.031	.403	.216	4.638
喜剧	.046	.039	.087	1.188	.229	4.368
动作	.031	.046	.051	.664	.211	4.743
奇幻	-.017	.049	-.022	-.360	.337	2.972
悬疑	.026	.043	.044	.588	.218	4.579
战争	.080	.052	.070	1.519	.589	1.698
剧情	.013	.045	.026	.282	.146	6.865
犯罪	.030	.050	.035	.607	.370	2.701
惊悚	-.074	.055	-.082	-1.339	.330	3.032
票房	.014	.013	.066	1.101	.351	2.851
弹幕数量	.150	.041	.457	3.626***	.078	12.794
弹幕情感值	-.010	.025	-.022	-.391	.380	2.634
弹幕与字幕情感趋势相关性	.015	.007	.076	1.993**	.843	1.186
评论数量	.098	.041	.310	2.390**	.074	13.577
评论情感值	.003	.024	.006	.111	.393	2.546
评论异质性	.569	.507	.051	1.123	.611	1.636

弹幕与字幕情感种类关联	.026	.025	.054	1.030	.453	2.205
-------------	------	------	------	-------	------	-------

注：0 ‘***’ 0.001 ‘**’ 0.05 ‘*’ 0.1 ‘ ’ 1

5.2.3 逐步多元线性回归分析

为了进一步研究，采用逐步回归分析法中的后退法，即将模型中影响较小的变量逐步剔除，得到最终的分析结果。进行逐步多元线性回归分析后，如表 5-5 所示，第 14 个模型达到最优，其调整后的 R^2 为 0.855，F (Sig=0.000)，代表对应模型的预测变量对电影视频播放量的解释程度达到 85.5%，解释程度较高，且该最优模型的 R^2 在合理范围内。

表 5-5 电影视频播放量影响因素模型汇总
Table 5-5 Summary of influencing factor models of movie video playback

模型	R	R^2	调整后 R^2	标准误差	F	Sig
1	.933	.871	.844	.09616059365	31.889	.000
2	.933	.871	.845	.09570729473	33.726	.000
3	.933	.871	.846	.09528940048	35.720	.000
4	.933	.871	.848	.09488110756	37.915	.000
5	.933	.870	.849	.09449996335	40.331	.000
6	.933	.870	.850	.09411508343	43.029	.000
7	.933	.870	.851	.09377575401	46.037	.000
8	.933	.870	.852	.09353032653	49.336	.000
9	.932	.869	.852	.09338303858	52.981	.000
10	.932	.868	.853	.09311644667	57.356	.000
11	.931	.868	.854	.09297261036	62.274	.000
12	.931	.867	.854	.09278044260	68.169	.000
13	.931	.866	.855	.09271984745	74.999	.000
14	.930	.865	.855	.09268480032	83.293	.000
15	.928	.862	.853	.09334073058	92.063	.000
16	.927	.860	.851	.09372191929	104.081	.000

注：模型 14 预测变量：(常量)，喜剧电影，评论数量，评论异质性，惊悚类型，弹幕与字幕情感趋势相关性，弹幕数量，战争电影，票房，明星影响力

由表 5-6 的分析结果可知，VIF 均小于 10，表明自变量间不存在共线性，对于整体数据而言，表中的自变量对于电影视频播放量的影响均较为显著。其中明

星影响力对电影视频播放量有显著的正向影响,假设 H2a 得到了验证;评论数量及弹幕数量对播放量也有显著的正向影响,假设 H6a 和 H7a 得到了验证;同时,喜剧、战争类型的电影均对播放量会产生显著正向影响,而惊悚类型的电影会产生负向影响,假设 H1 得到了验证,即惊悚类型的电影在中国市场发展得可能不太成熟,对于这些类型的电影很多影视平台的用户观看意愿不高,而更愿意观看喜剧、战争类型的电影。

但是票房对播放量有一定的正向影响,假设 H5 没有得到验证,表明视频平台用户在选择电影观看时会在一定程度上注重该电影在上映时的票房表现力;并且弹幕与字幕情感趋势相关性对播放量有显著的正向影响,假设 H8b 得到了验证,即用户更加偏好弹幕与字幕情感呈正相关或弱正相关的电影视频,弱正相关表明弹幕与字幕的情感极性相反但是情感变化趋势相同,这类电影视频往往在前期因为设有伏笔而导致弹幕与字幕情感不一致,而随着情节的完善,用户逐渐理解情节,变化趋势则逐渐转为相同,因此往往能引起用户的激烈讨论,所以很多用户更想一探究竟,同时也会加入讨论,从而导致这类电影视频的播放量更高。同时研究发现评论异质性对播放量有显著正向影响,但是弹幕与字幕情感种类相关性这一自变量被模型所提出,所以假设 H8a 没有得到验证,即用户并情感种类是否关联并不敏感。

表 5-6 最优模型回归系数
Table 5-6 Regression coefficient of optimal model

自变量	未标准化系数		标准化系数		共线性统计	
	B	标准错误	Beta	t	容差	VIF
(常量)	4.098	.071		57.510		
明星影响力	.016	.006	.109**	2.474	.594	1.682
喜剧	.031	.019	.058*	1.636	.911	1.098
战争	.069	.041	.061*	1.690	.890	1.124
惊悚	-.076	.033	-.084**	-2.318	.870	1.149
票房	.018	.011	.084*	1.682	.462	2.167
弹幕数量	.136	.035	.413***	3.841	.103	9.687
弹幕与字幕情感趋势相关性	.014	.007	.074**	2.094	.925	1.081
评论数量	.123	.034	.389***	3.651	.102	9.847

评论异质性	.803	.397	.071**	2.025	.927	1.079
-------	------	------	--------	-------	------	-------

注：0 ‘***’ 0.001 ‘**’ 0.05 ‘*’ 0.1 ‘ ’ 1

5.3 变量主体间效应检验分析

由于本研究自变量较多，为了避免忽略影响因子系统中的自变量，本研究进一步对其与播放量的关系进行主体间效应检验，具体结果如下文。

（1）导演知名程度及明星影响力

根据上文模型结果可知，由于导演知名程度入选了该模型，但是对播放量的影响因素不显著，所以利用主体间效应检验进一步分析其与播放量之间的关系，如表 5-7 所示，同线性回归结果一致，明星影响力对播放量的作用较为显著，而导演知名程度均对播放量的影响不显著，其交互效应的影响也不显著。

表 5-7 明星及导演数据主体间效应检验
Table 5-7 Test of inter subject effect of star and director data

源	III 类平方和	自由度	均方	F	显著性
修正模型	3.913	25	.157	4.476	.000
截距	1575.309	1	1575.309	45050.633	.000
明星影响力	1.074	5	.215	6.142	.000
导演知名程度	.265	4	.066	1.898	.117
明星影响力*导演知名程度	.338	16	.021	.605	.873
误差	3.532	101	.035		
总计	3503.921	127			
修正后误差	7.445	126			

注：R²=0.526（调整后 R²=0.408）

同时如图 5-1、5-2 所示，可以看到明星影响力对播放量的影响整体呈正相关，而导演知名程度对播放量的影响处于波动状态，即整体关联性较弱，假设 H2a 得到了验证，H2b 未得到验证。

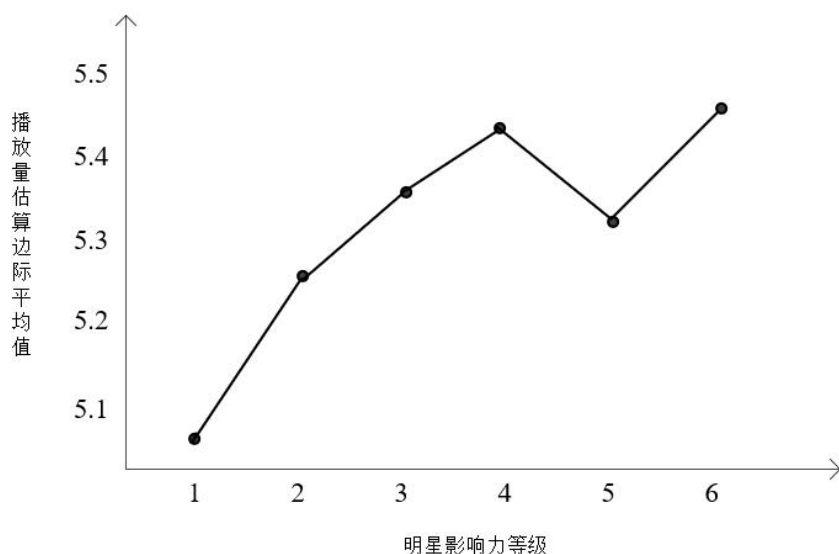


图 5-1 明星影响力对播放量估算边际平均值的影响

Fig 5-1 The influence of star influence on the marginal average of broadcast volume estimation

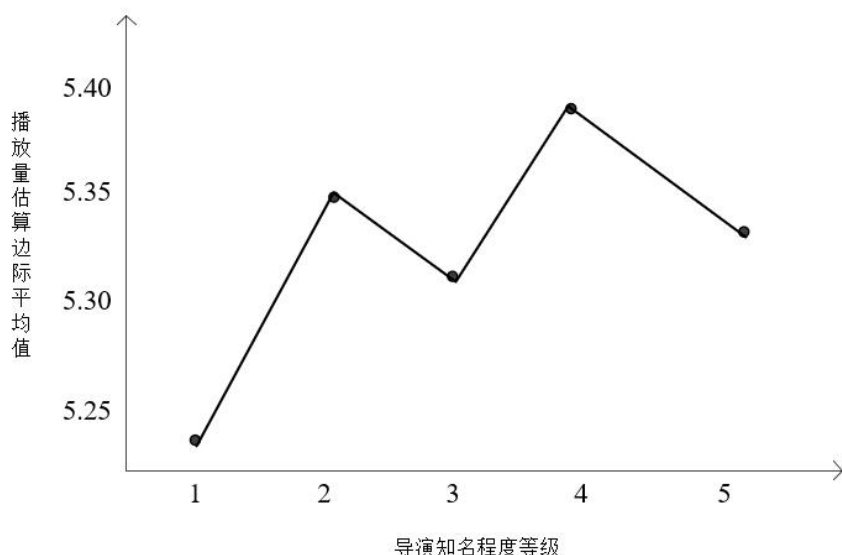


图 5-2 导演知名程度对播放量估算边际平均值的影响

Fig 5-2 The influence of director's fame on the marginal average of broadcast volume estimation

(2) 续集或 IP 改编

根据逐步多元线性回归模型可知, IP 改编及续集两个自变量均未入选模型, 同样对其进行主体间效应检验, 结果如表 5-8 及图 5-3 所示, 续集及 IP 改编对播放量的正向影响均较不显著, P 值均大于 0.05, 可能由于在上文模型中存在其他自变量的交互效应, 所以削弱了影响的显著性, 但是图 5-3 表明续集或 IP 改编仍

对播放量有一定的正向影响。

表 5-8 续集或 IP 改编数据主体间效应检验
Table 5-8 Test of intersubjectivity effect of sequel or IP adaptation data

源	III 类平方和	自由度	均方	F	显著性
修正模型	.587	3	.196	3.511	.017
截距	996.788	1	996.788	17878.558	.000
续集	.174	1	.174	3.122	.080
IP 改编	.006	1	.006	.101	.751
续集*IP 改编	.066	1	.066	1.181	.279
误差	6.858	123	.056		
总计	3503.921	127			
修正后误差	7.445	126			

注：R²=0.061（调整后 R²=-0.041）

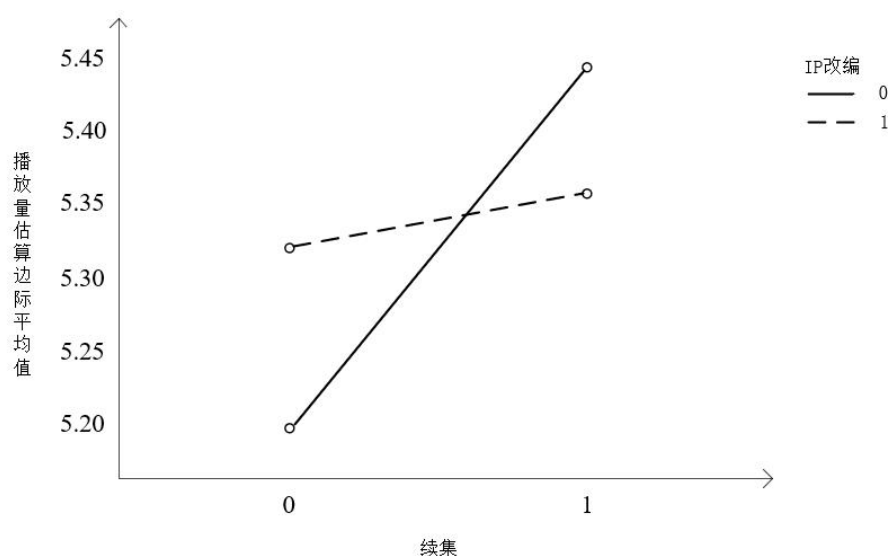


图 5-3 续集或 IP 改编对播放量估算边际平均值的影响

Fig 5-3 Influence of sequel or IP adaptation on marginal average of playback estimation

（3）弹幕相关数据

由于在本研究中，只有弹幕数量对播放量有显著影响，所以通过多因素方差分析，得出主体间效应检验结果如表 5-9 所示，分析结果表明弹幕数量等级对播放量有显著的正向影响，同时结合图 5-4 可知，弹幕数量等级为 1 时，即当弹幕

数量少于 1000 条时，弹幕情感极性从负向转为正向时，对播放量的影响最显著，而当此弹幕数量继续增多到 4 级时，此时弹幕数据范围为 20000-30000 条，积极情感对播放量的边际影响力则开始下降，可能由于弹幕信息太多对用户干扰过度，从而减弱了积极情感对播放量的正向影响，或者是用户倾向认为太多正向的弹幕是粉丝在弹幕刷屏，从而减弱了观看的意愿，因此假设 H6b 得到了验证。

表 5-9 弹幕数据主体间效应检验
Table 5-9 Test of intersubjectivity effect of bullet screen data

源	III 类平方和	自由度	均方	F	显著性
修正模型	88.956	14	6.354	4.348	.000
截距	89.418	1	89.418	61.190	.000
弹幕数量等级	68.549	7	9.793	6.701	.000
弹幕情感极性	2.205	1	2.205	1.509	.222
弹幕数量等级*弹幕情感极性	10.146	6	1.691	1.157	.333
误差	189.970	130	1.461		
总计	522.937	145			
修正后误差	278.926	144			

注： $R^2=0.319$ （调整后 $R^2=0.246$ ）

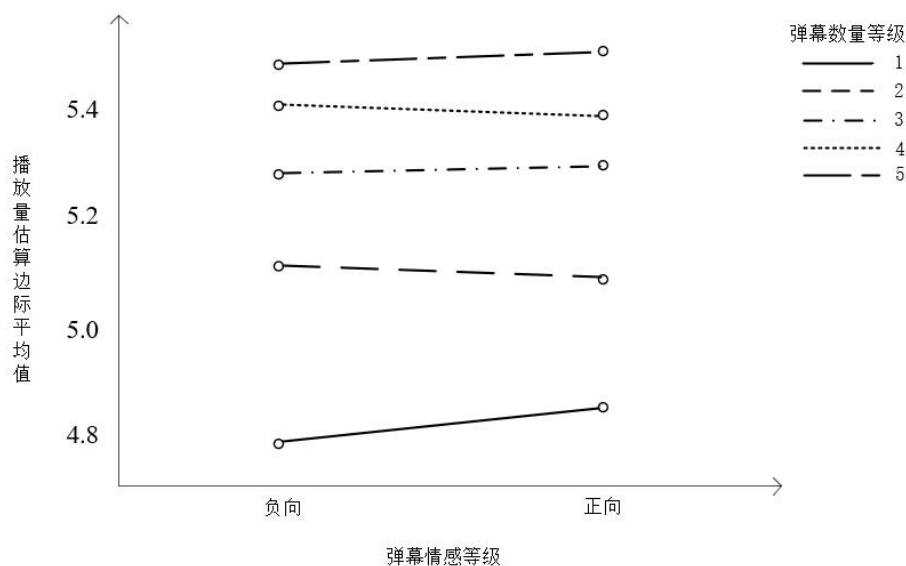


图 5-4 弹幕对播放量估算边际平均值的影响

Fig 5-4 The influence of bullet screen on the marginal average of broadcast volume estimation

同时根据弹幕与字幕情感趋势变化相关性对播放量的主体间效应检验数据可知,如图 5-5 所示,其中-2, -1,1,2 分别表示负相关、弱负相关、弱正相关和正相关,当弹幕与字幕情感趋势呈弱正相关时对播放量的影响最为显著,其次是负相关,影响程度排名为弱正相关>负相关>弱负相关>正相关。而弱正相关表明弹幕和字幕的情感极性开始相反,随情节逐渐完善趋同;负相关表明弹幕和字幕极性和变化趋势均相反,以上两类相关性均表明电影情节具有反转或讽刺意味,因此对播放量的正向影响更强烈。

表 5-10 弹幕与字幕相关性数据主体间效应检验

Table 5-10 Test of inter subject effect of bullet screen and subtitle correlation data

源	III 类平方和	自由度	均方	F	显著性
修正模型	.787	3	.262	4.847	.003
截距	2128.101	1	2128.101	39316.010	.000
弹幕与字幕情感相关性	.787	3	.262	4.847	.003
误差	6.658	123	.054		
总计	3503.921	127			
修正后总计	7.445	126			

注: $R^2=0.106$ (调整后 $R^2=0.084$)

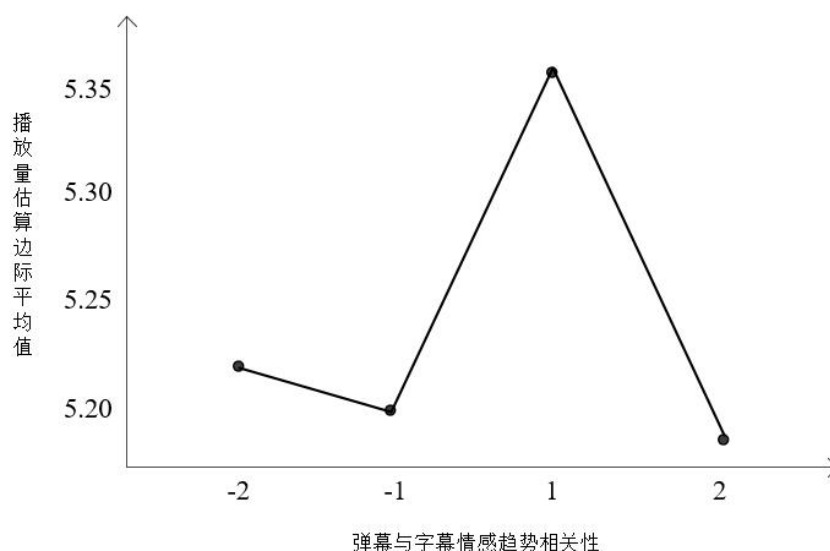


图 5-5 弹幕与字幕相关性对播放量估算边际平均值的影响

Fig 5-5 The influence of the correlation between bullet screen and caption on the marginal average of the broadcast volume estimation

(4) 评论相关数据

根据上文逐步回归分析的结果,评论数量及评论异质性整体数据对播放量有显著的正向影响,对评论相关数据如评论数量、评论异质性进行等级划分,并对评论情感区分正负向评论,以进行主体间效应检验。结果如表 5-11 所示,可知评论数量以及评论异质性对播放量的影响都较为显著,与回归分析结果一致。

表 5-11 评论数据主体间效应检验
Table 5-11 Test of inter subject effect of star and director data

源	III 类平方和	自由度	均方	F	显著性
修正模型	5.952	29	.205	13.340	.000
截距	1260.714	1	1260.714	81937.110	.000
评论数量等级	2.122	2	1.061	68.965	.000
评论情感极性	.008	1	.008	.509	.477
评论异质性等级	.324	5	.065	4.217	.002
评论数量等级*评论情感极性	.101	2	.050	3.272	.042
评论数量等级*评论异质性等级	.126	10	.013	.819	.611
评论情感极性*评论异质性等级	.179	4	.045	2.902	.026
误差	.495	5	.099	6.430	.000
总计	1.492	97	.015		
修正后误差	3503.921	127			

注: $R^2=0.800$ (调整后 $R^2=0.740$)

同时,如图 5-6、5-7 所示,根据不同评论量下及整体数据下评论异质性对播放量的影响可以分析得出,不同评论数量下,当评论异质性从较小刚开始增大时,对播放量有负向影响,而当评论异质性增大到一定范围后,对播放量会产生正向影响,对整体数据而言,也表明异质性等级增大到 4 级时,即当评论异质性处 0.13-0.14 范围内时,会对播放量有正向影响,假设 H7c 得到了验证。

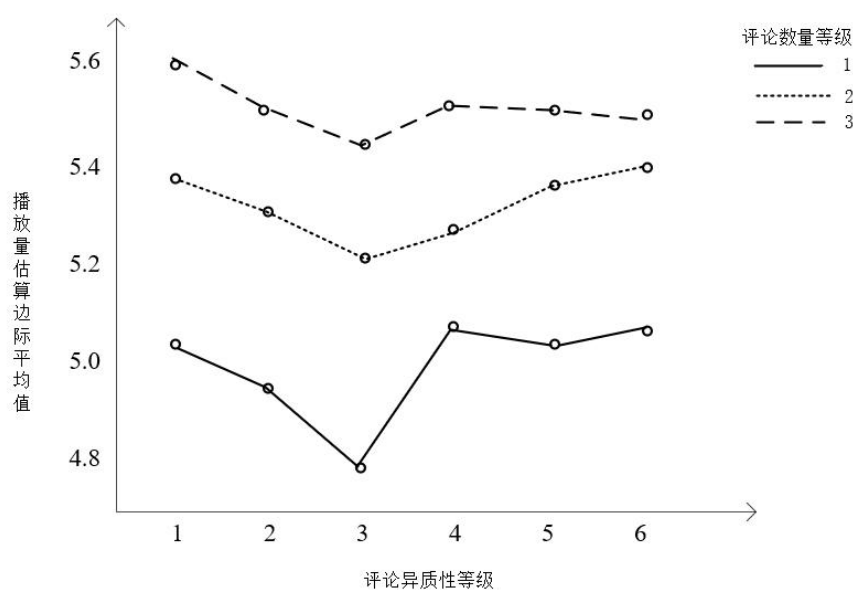


图 5-6 不同数量等级下评论异质性对播放量估算边际平均值的影响

Fig 5-6 The influence of comment heterogeneity on the marginal average of playback estimation in different quantity levels

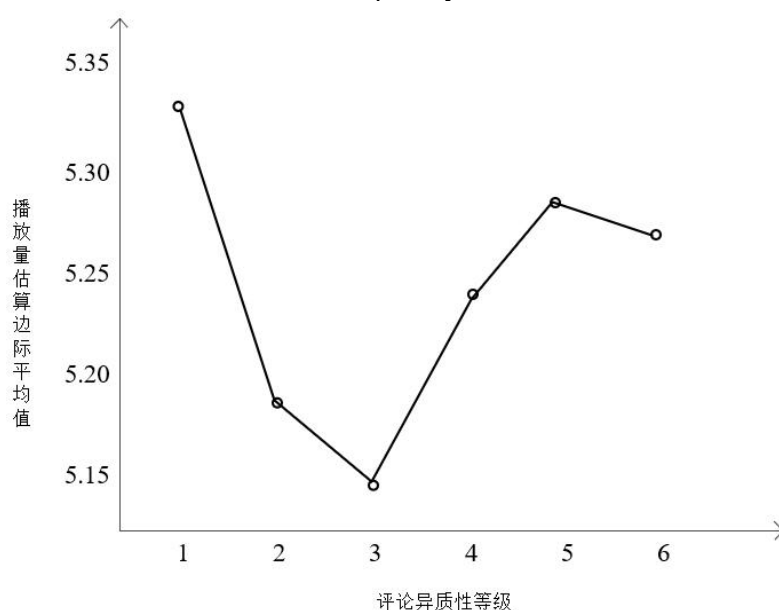


图 5-7 评论异质性对播放量估算边际平均值的影响

Fig 5-7 Comment on the influence of heterogeneity on the marginal average of broadcast volume estimation

根据图 5-8 的分析结果可知，当评论数量等级为 1 时，即评论数量较少时，在本文中为当评论数量少于 1000 条时，负向评论对播放量的影响较小；而当评论数量较多时，随着负向评论减少，正向评论的增多，会正向影响电影视频的播放

量，假设 H7b 也得到了验证。

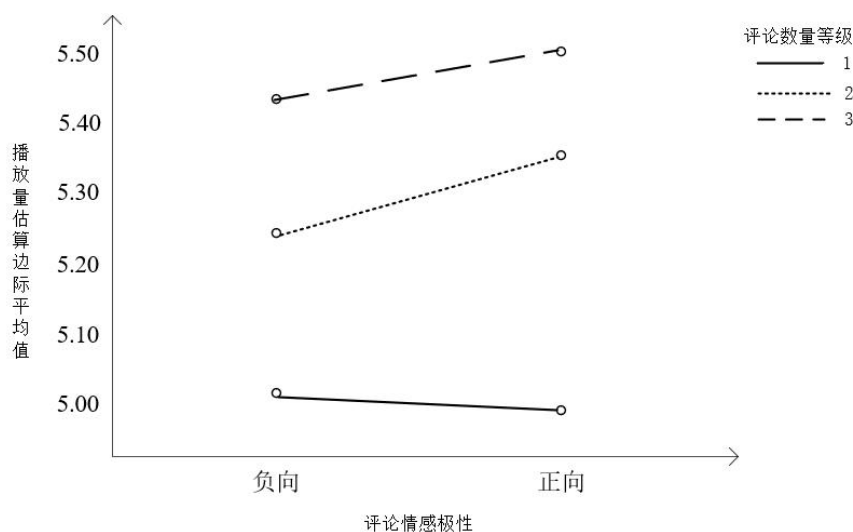


图 5-8 不同数量等级下评论情感对播放量估算边际平均值的影响

Fig 5-8 The influence of comment emotion on the marginal average of broadcast volume estimation in different quantity levels

5.4 本章小结

本章节在国内外学者研究的基础上，对电影视频播放量影响因素模型进行了构建，并利用逐步多元线性回归分析对各自变量对播放量的影响体系进行分析，同时利用主体间效应检验分析进行了进一步的研究，避免了部分自变量因影响不太显著而被剔除的影响，保证了研究的全面性和完整性。最后研究发现票房、弹幕数量、评论数量及评论异质性对播放量均有显著的正向影响，且电影视频的播放量与电影的类型也息息相关；并且当弹幕数量较少时，积极的弹幕对播放量有正向影响，而当弹幕增多到一定数量，会削弱该对播放量的影响。同时用户更偏好带有反转或讽刺意味等能引起激烈讨论的电影，而电影视频评论较少时，其情感极性对播放量影响不大，但是评论的异质性在一定范围内能正向影响电影视频的播放量。

结论与展望

研究结论

本研究对 2020-2021 年间曾在国内上映过，后在爱奇艺平台上播出的电影的播放量影响因素进行研究，不仅涉及传统电影上映过程中的影响因素，还包括视频平台上的新兴特点，如弹幕、评论等相关因素，主要得出了以下结论：

(1) 对视频平台上电影视频存在的弹幕这一新兴特点进行了深入研究，首先依托于 PaddlePaddle 框架下的 PaddleHub 包对每部电影的弹幕情感值进行初步分析，并在考虑情绪感染机制及衰退效应的基础上对其进行修正，得到最终的弹幕情感值。同时，根据大连理工情感词汇表分析每部电影视频中弹幕及字幕不同情感种类的个数，以皮尔逊卡放值进行度量；并且为了研究弹幕与字幕情感趋势的关联关系，以 5 分钟为基准对每部电影的弹幕及字幕情感值进行重采样，以相关系数对变化趋势进行衡量。

(2) 构建视频平台上电影播放量的影响因素模型，研究结果表明电弹幕数量、评论数量及电影在上映过程中的票房表现对电影视频播放量有显著的正向影响，并且当评论数量较少时，负向情感对播放量的影响较小，评论异质性也会在一定范围内对播放量造成正向影响；同时电影的类型也与电影视频播放量有显著关联，其中惊悚类型的电影均对播放量会产生显著负向影响，表明在国内市场这些电影类型的发展还不太成熟。研究结果还发现弹幕及字幕情感极性与变化趋势为弱正相关或负相关的电影视频播放量更高，即用户偏爱带有反转或讽刺效果的电影，但是弹幕与字幕的情感种类关联程度强弱对播放量的影响可以忽略。通过主体间效应检验还发现当电影在视频平台上播出时，明星影响力的影响较为显著，而导演知名程度对播放量的影响不显著，且用户相对于弹幕情感强烈程度，更在意正负弹幕的数量，太多正向的弹幕会让用户认为是粉丝刷屏，从而对播放量的边际影响减弱；同时虽然续集、IP 改编未入选模型，但是方差检验表明其对播放量存在一定的正向影响。

(3) 不同于电影在上映过程中票房的影响因素,有部分传统因素对播放量的影响并不显著。例如导演知名程度,可能是电影在视频平台播出后用户有更多判断电影是否值得观看的信息来源,同时上映过程中的票房也成为了自变量之一,用户也会以其作为重要参考因素。

管理启示

近几年来,网络视频行业的规模不断扩大,几乎所有的视频网站平台都引入了弹幕功能,其也自然而然成为了视频播放量的重要影响因素之一,特别是对于电影这一类型的视频播放量,在考虑传统因素影响的同时,结合与视频弹幕、字幕等相关的新兴影响因素,对其播放量影响因素的全面研究有重要实践意义,同时对视频平台的管理者及电影视频的制作方都具有一定的管理启示。

(1) 本研究发现用户更偏好于电影的弹幕与字幕情感极性及变化趋势呈弱正相关或负相关的电影,即带有反转或讽刺等效果的电影视频播放量更高,对于视频平台的管理者来说,可以有针对性地提高该类型电影在同一界面的排序,同时将弹幕、评论、评分、票房等加入视频推荐体系;而对于电影视频的制作者来说,则需要满足用户的需求,做好前期市场调研,精准定位,对于观众接受度不高的电影类型多加打磨,更可能创作更多具有反转、讽刺意味而又能引发观众思考和讨论的电影。

(2) 由于弹幕、评论等是视频平台的新兴特点,同时本研究也发现弹幕、评论的数量及情感均在一定程度上对电影视频播放量有影响,所以对于视频平台的运营方来说,需要注重对弹幕、评论等用户互动方式的运营,提高审核效率,适当对用户进行正向积极的引导;对于电影制作方来说,在进入平台宣传期时,可以通过传统媒体和新媒体整合,并结合弹幕、评论等适当营造正向的舆论态势。

(3) 本研究发现弹幕与字幕情感变化趋势的相关性对播放量存在显著影响,且用户更偏好弹幕与字幕情感变化趋势呈弱正相关及负相关的电影,因此平台运营方而言,要重视弹幕与字幕情感的契合或者差异,并结合平台调性,进一步为

影视作品的引入、宣传等采取针对性手段；对于电影制作方来说，也应该着重于完善电影产业生态环境，即综合考虑影响电影视频播放量的因素，结合弹幕与字幕情感信息分析影视内容与用户情感表达之间的关系，从而为电影视频生产制作、整改指明方向。

展望

本研究着眼于电影视频播放量的影响因素，结合传统的电影票房影响因素和视频网站新型的互动特点展开了研究，得到了相关的结论，但同时研究也存在以下不足之处：

（1）弹幕、评论情感粒度。本研究只是粗略地将弹幕及评论的情感分为了正向和负向，而没有进行细粒度的情感划分，在未来的研究模型中，可以进一步将情感划分为快乐、惊讶、愤怒等更细的粒度，以研究不同类型的情感对播放量的影响程度，对研究结果进行完善。

（2）电影视频来源平台。本研究电影视频数据均来源于爱奇艺平台，未能分析其他平台的电影视频数据，而不同平台的用户使用习惯以及播放量的计算方式等会有所差异，因此为了保证结果的可推广性，未来可以搜集更多不同平台的电影视频数据进行研究，进行更全面的分析。

（3）回归分析。在回归分析的过程中，对电影视频类型对播放量的影响的研究是笼统的，后续可以对不同的电影视频类型对播放量的影响单独分析。

参 考 文 献

- [1] Hennig-thurau T, Wiertz C, Feldhaus F. Does Twitter matter? The impact of microblogging word of mouth on consumers' adoption of new movies [J]. Journal of the Academy of Marketing Science, 2015, 43(03): 375-394.
- [2] Kim T, Hong J, Kang P. Box Office Forecasting considering Competitive Environment and Word-of-Mouth in Social Networks: A Case Study of Korean Film Market [J/OL]. Comput Intell Neurosci, 2017:4315419.
- [3] 华锐, 王森林, 许泱. 中国电影票房的影响因素研究 [J]. 统计与决策, 2019, 35(04): 97-100.
- [4] 黄文清, 黎嘉嘉, 宁琼敏等. 电影票房影响因素分析及预测 [J]. 信息技术与信息化, 2019, (12): 173-178.
- [5] Liu T, Ding X, Chen Y, et al. Predicting movie Box-office revenues by exploiting large-scale social media content [J]. Multimedia Tools and Applications, 2016, 75(03): 1509-1528.
- [6] Ding C, Cheng H K, Duan Y, et al. The power of the “like” button: The impact of social media on box office [J]. Decision Support Systems, 2017, 94(09): 77-84.
- [7] Oh S, Baek H, Ahn J. Predictive value of video-sharing behavior: sharing of movie trailers and box-office revenue [J]. Internet Research, 2017, 27(03): 691-708.
- [8] Oh C, Roumani Y, Nwankpa J K, et al. Beyond likes and tweets: Consumer engagement behavior and movie box office in social media [J]. Information & Management, 2017, 54(01): 25-37.
- [9] Eliashberg J, Jonker J-J, Sawhney M S, et al. MOVIEMOD: An Implementable Decision-Support System for Prerelease Market Evaluation of Motion Pictures [J]. Marketing Science, 2000, 19(03): 226-43.
- [10] 史伟, 王洪伟, 何绍义. 基于微博情感分析的电影票房预测研究 [J]. 华中师范大学学报(自然科学版), 2015, 49(01): 66-72.
- [11] 孙春华, 刘业政. 电影预告片在线投放对票房的影响——基于文本情感

分析方法 [J]. 中国管理科学, 2017, 25(10): 151-61.

[12] 石文华, 钟碧园, 张绮. 在线影评和在线短评对票房收入影响的比较研究 [J]. 中国管理科学, 2017, 25(10): 162-70.

[13] Hu Y-H, Shiau W-M, Shih S-P, et al. Considering online consumer reviews to predict movie box-office performance between the years 2009 and 2014 in the US [J]. The Electronic Library, 2018, 36(06): 1010-26.

[14] 胡颖, 胡长军, 傅树深等. 流行度演化分析与预测综述 [J]. 电子与信息学报, 2017, 39(04): 805-16.

[15] Ghosh S, Shekar S K. Video Popularity Distribution and Propagation in Social Networks [J]. International Journal of Emerging Trends & Technology in Computer Science, 2017, 6(01): 1-5.

[16] Yoganarasimhan H. Impact of social network structure on content propagation: A study using YouTube data [J]. Quantitative Marketing and Economics, 2011, 10(01): 111-150.

[17] Borghol Y, Ardon S, Carlsson N, Eager D, Mahanti A. The untold story of the clones: content-agnostic factors that impact youtube video popularity [C]// Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. August 12-16, 2012. Beijing, China: ACM, 2012: 1186- 1194.

[18] Meeyoung C, Haewoon K, Rodriguez P, et al. Analyzing the Video Popularity Characteristics of Large-Scale User Generated Content Systems [J]. IEEE/ACM Transactions on Networking, 2009, 17(05): 1357-1370.

[19] Welbourne D J, Grant W J. Science communication on YouTube: Factors that affect channel and video popularity [J]. Public Underst Sci, 2016, 25(06): 706-18.

[20] Rapp F G. The digital media phenomenon of YouTube beauty gurus: the case of Bubzbeauty [J]. International Journal of Web Based Communities, 2016, 12(04): 360-375.

[21] Jeong-Sun, Hee-Ja. Comparison of Determinants of Cosmetic Purchasing According to Execution Experience after Viewing Beauty YouTube [J]. Asia-pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology, 2019, 9(12): 417-425.

- [22] Zhou R, Khemmarat S, Gao L, et al. Boosting video popularity through recommendation systems[C]// Proceedings of the First ACM SIGMOD Workshop on Databases and Social Networks, June 12, 2011. Athens, Greece: ACM, 2011: 13-18.
- [23] Li H, Wang H, Liu J, et al. Video sharing in online social networks: Measurement and analysis[C]// Proceedings of the 22nd international workshop on Network and Operating System Support for Digital Audio and Video, June 07, 2012. New York, USA: ACM, 2012: 83-88.
- [24] Vandersmissen B, Godin F E, Tomar A, et al. The Rise of Mobile and Social Short-Form Video: An In-depth Measurement Study of Vine [J]. The Workshop on Social Multimedia & Storytelling, 2014, 2(01): 1-10.
- [25] Wu J, Zhou Y, Chiu D M, et al. Modeling Dynamics of Online Video Popularity [J]. IEEE Transactions on Multimedia, 2016, 18(09): 1882-95.
- [26] 张海涛, 张鑫蕊, 周红磊等. 融合用户偏好与内容特征的短视频传播效果评价研究 [J]. 图书情报工作, 2020, 64(16): 81-91.
- [27] 龚艳萍, 曹玉, 李见. 短视频应用的特性对用户参与行为的影响: 心理参与的中介作用 [J]. 情报科学, 2020, 38(07): 77-84.
- [28] Shen Y, Chan H C, Hung I W. Let the comments fly: The effects of flying commentary presentation on consumer judgment [J]. International Journal of Systems Science, 2014, 12(02): 1469-1475.
- [29] 王霞, 梁栋. 弹幕数量和弹幕情感强度对视频流行度的影响 [J]. 营销科学学报, 2019, 15(02): 132-156.
- [30] 袁海霞, 方青青, 白琳. 弹幕对在线消费行为过程影响的时变效应研究 [J]. 管理学报, 2020, 17(07): 1059-1066.
- [31] 陈明红, 黄嘉乐, 方世深等. 弹幕视频播放量影响因素与组态效应 [J]. 图书馆论坛, 2022, 42(06): 150-161.
- [32] Cha M, Kwak H, Rodriguez P, et al. I Tube, You Tube, Everybody Tubes: Analyzing The World's Largest User Generated Content Video System[C]. Proceedings of the 7th ACM SIGCOMM Conference On Internet Measurement. San Diego, California, USA: ACM, 2007: 1-14.

- [33] Mcauley J, Leskovec J. Hidden factors and hidden topics: understanding rating dimensions with review text [C]//Proceedings of the 7th ACM conference on Recommender systems. ACM,2013:165-172.
- [34] 徐晓枫, 贺樑, 杨静. 融合社交与搜索数据的电视剧点播排名预测研究 [J]. 计算机工程, 2015, 41(08): 6-12+7.
- [35] 周良, 王璇. 视频网站运营中用户评论的作用及实证分析 [J]. 中国流通经济, 2015, 29(02): 90-97.
- [36] 何楚佳, 姚威羽. 弹幕与字幕关联性分析对于视频发展的参考 [J]. 情报探索, 2019, (09): 45-49.
- [37] 王敏, 徐健. 视频弹幕与字幕的情感分析与比较研究 [J]. 图书情报知识, 2019, (05): 109-119.
- [38] Litman B R. Predicting Success of Theatrical Movies: the 80's experiences [J]. The Journal of Popular Culture, 1989, 23(02): 35-50.
- [39] 张立, 王湘平. 基于信息理论的电影消费影响因素研究 [J]. 消费经济, 2018, 34(01): 31-35.
- [40] Tayal D, Jain A, Gupta D, et al. A new method to predict boxoffice collections for bollywood movies using online opinion mining [J]. International Journal of Engineering Intelligent Systems for Electrical Engineering and Co, 2015, 23(2): 49-60.
- [41] Duan J, Ding X, Liu T. A Gaussian copula regression model for movie box-office revenues prediction [J]. Science China Information Sciences, 2017, 60(09): 1-14.
- [42] 胡小莉, 李波, 吴正鹏. 电影票房的影响因素分析 [J]. 中国传媒大学学报(自然科学版), 2013, 20(01): 62-7+39.
- [43] 刘志新. 中国电影票房影响因素 [J]. 合作经济与科技, 2019, (17): 114-116.
- [44] 王铮, 许敏. 电影票房的影响因素分析——基于 Logit 模型的研究 [J]. 经济问题探索, 2013, (11): 96-102.
- [45] 宋子川. 文化产业发展中电影票房收入的影响因素研究——基于 50 部进

口影片的实证分析 [J]. 浙江金融, 2013, (12): 66-68.

[46] 王一帆. 我国电影票房影响因素分析 [J]. 市场研究, 2015, (05): 30-32.

[47] Chen Y, Gao Q, Rau P-L P. Watching a Movie Alone yet Together: Understanding Reasons for Watching Danmaku Videos [J]. International Journal of Human-Computer Interaction, 2017, 33(09): 731-743.

[48] Benton A, Hill S. The Spoiler Effect? Designing Social TV Content That Promotes Ongoing WOM [J/OL]. INFORMS conference on information systems and technologyPhoenix:IEEE Morocco Section and the IEEE Morocco Computer & Communication, 2012.

[49] 邓卫华, 易明. 考虑口碑衰竭效应时追加评论对网络口碑传播的影响研究 [J]. 管理学报, 2019, 16(04): 595-602.

[50] De Vries L, Gensler S, Leeftang P S H. Popularity of Brand Posts on Brand Fan Pages: An Investigation of the Effects of Social Media Marketing [J]. Journal of Interactive Marketing, 2012, 26(02): 83-91.

[51] Seiler S, Yao S, WANG W. Does Online Word of Mouth Increase Demand? (And How?) Evidence from a Natural Experiment [J]. Marketing Science, 2017, 36(06): 838-861.

[52] 王晓耘, 袁媛, 史玲玲. 基于微博的电影首映周票房预测建模 [J]. 现代图书情报技术, 2016, (04): 31-39.

[53] Kim J, Gupta P. Emotional expressions in online user reviews: How they influence consumers' product evaluations [J]. Journal of Business Research, 2012, 65(07): 985-992.

[54] Chen Z, Lurie N H. Temporal Contiguity and Negativity Bias in the Impact of Online Word of Mouth [J]. Journal of Marketing Research, 2013, 50(4): 463-476.

[55] Quaschnig S, Pandelaere M, Vermeir I. When Consistency Matters: The Effect of Valence Consistency on Review Helpfulness [J]. Journal of Computer-Mediated Communication, 2015, 20(02): 136-152.

[56] Lee J H, Jung S H, Park J. The role of entropy of review text sentiments on online WOM and movie box office sales [J]. Electronic Commerce Research and

Applications, 2017, 22(03): 42-52.

[57] 李亮, 江晓东. 电子商务环境下的品牌资产研究——基于信息不对称的视角 [J]. 经济问题, 2017, (03): 85-90.

[58] 魏静, 丁乐蓉, 朱恒民等. 全媒体时代基于情感和耦合网络的舆情传播模型研究 [J]. 情报杂志, 2020, 39(08): 110-116.

[59] 殷雁君, 唐卫清, 李蔚清. 基于情绪感染的虚拟个体情绪模型 [J]. 计算机仿真, 2013, 30(08): 216-220.