

舆情数据分析与商业化应用——以《小丑》豆瓣短评为例

摘要：本案例使用《小丑》的豆瓣短评做舆情统计分析。结果显示，评论热度先升后降，演员演技和电影的社会价值观被评论提及最多，且差评最爱讨论后者，舆情的正向情感分数在未来 14 天会保持在较低水平。在商业应用层面，本案例对于同题材影片的拍摄和引进以及电影口碑的维护有着一定的指导作用。在未来，结合舆情统计数据，可以制定一个影片市场口碑的预测系统。

一、背景介绍

2013 年至今我国网民规模呈不断增加的趋势，截至 2020 年 3 月已达 9.04 亿人，比 2019 年增加了 5.7%。在网络上有越来越多的信息在流通，观点被交换。

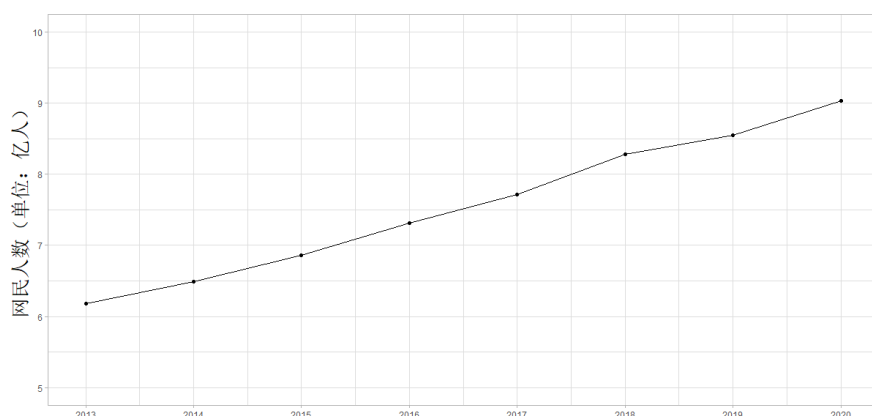


图 1 我国网民数量变化图

互联网舆情因而不断发展，并且通过商业口碑影响着品牌。在传媒界尤其如此，当一部影视作品上线之后在豆瓣，微博，知乎等信息平台上形成的海量评论会进一步影响下一个观众的选择，这种绵延不断的口碑传播很大程度上决定了一部作品的市场表现（比如票房）。因此检测舆情有助于企业把握影视作品的质量和市场的偏好，进而调整自己的出品计划以及和合作方合作时的议价策略，达到规避作品的“暴死”风险和利润增长的目的。

互联网上的舆情监测依托于大数据技术，可以被分为热度和内容两个角度。其中，热度分析旨在量化社会对一部影视作品的关注度随时间的变化，企业利用好其中的规律便能延续旗下作品的社会生命力。而内容分析包括对主题，情感倾向的研究。主题分析给出舆论关键词，帮助企业把握互联网上主要在讨论的话题，配合宣发能起到引导舆论，炒热话题，吸引流量的作用。情感倾向分析涉及到大

众对作品的观感，往往以一个正向情感分数的形式呈现，企业通过监控这一指标便能及时发现那些负面舆论并及时作出反应，减少损失，控制风险。

本文将以 DC 影业出品的电影《小丑》为对象进行舆情监测和分析。《小丑》以 DC 漫画中最著名的反派角色小丑为主角，讲述了一个精神病患者亚瑟(Arther)是如何在社会的重重压力下走向犯罪的故事。电影被评为 R 级，因此并未在国内上映，但光靠线上盗版资源的播放就引发了广泛的关注和讨论。通过舆情监控我们能分析国内观众对此类题材的观感，为拍摄，引入类似的作品和影片的口碑维护策略提供商业参考。

二、 数据来源与说明

本案例所使用的数据是 2019 年上映的美国电影《小丑》的评论信息。数据来自豆瓣短评，采集范围从 2019-9-1 至 2021-12-27，共 1625 条。数据共包含 7 个变量，其中豆瓣短评中采集三个（发布内容，评分，点赞数），评论者用户页面采集四个（发布者 ID，发布者加入时间，发布者被关注数，发布者阅片数）。详细的变量说明如表 1 所示。

表 1 数据变量说明表

所属平台	细分来源	变量名称	变量类型	变量描述
豆瓣	豆瓣短评	发布内容	文本数据	例如“正式与 Rorschach，V 成为 DC 反压迫三人组。”
		评分	离散变量	分为“力荐”，“推荐”，“还行”，“较差”，“很差”
		点赞数	连续变量	取值为正整数
	用户页面	发布者 ID	文本数据	例如“王写写”
		发布者加入时间	时间变量	时间段位于 2005-4-20 到 2020-4-12
		发布者被关注数	连续变量	取值为正整数

		发布者阅 片数	连续变量	取值为正整数
注：数据来源于豆瓣公开短评，爬取总数为 1625，时间从 2019-9-1 至 2021-12-27				

三、 热度分析

(一) 密度图分析

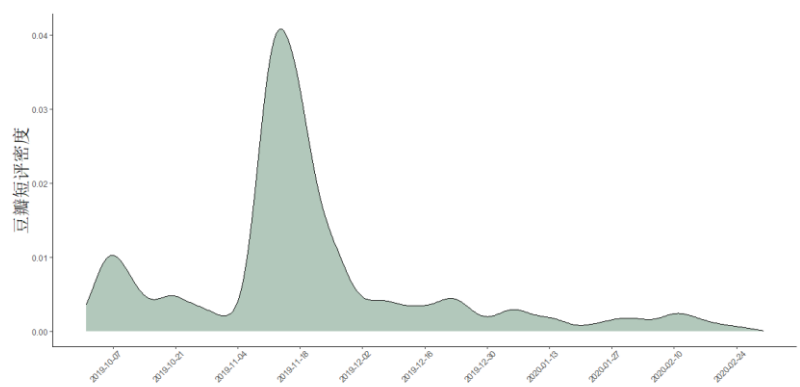


图 2 《小丑》豆瓣短评密度图

从上面的密度图中可以看出：在 2019 年 11 月 4 日到 2019 年 12 月 2 日之间豆瓣短评的数量达到了高峰，然后迅速回落，并保持在较低水平。对比密度图与百度的搜索指数可以发现，两者在同一时间段内达到峰值。作为一部未在国内上映的 R 级小众电影，小丑在 10 月上映前并没有宣发活动，而在上映后的短短 1 个月内靠口碑实现了传播。从百度指数上看，在该影片出圈的过程中网络媒体扮演了重要的助推作用。



图 3 同期百度搜索指数图

¹ 为了美观，案例在绘图时删去了短评密度较低的时间段，并将 X 轴的时间分割设置为 2 周

(二) 词云图分析

案例选取 11 月 4 日到 12 月 2 日短评中的文本数据, 经过分词后绘制词云图。从图 4 中可以看出: 高峰期中出现的短评高频词包括“小丑”, “社会”, “表演”, “故事”, “人物”, “演技”等关键词。说明《小丑》的吸引力在于主角“小丑”的人物塑造, 跌宕起伏的剧情, 扮演者的演技以及影片的社会隐喻。事实上《小丑》的主演获得了当年奥斯卡最佳男主角奖, 而电影在上映后引起了关于阶级矛盾, 贫富差距等社会问题的广泛讨论。



图 4 《小丑》短评词云图

单独对高峰期内的好评, 中评与差评绘制词云图。从图 5, 6 中可以看出好评与中评的高频关键词与所有评论的差别不大。而从图 7 中可以看出: 高峰期内的差评主要包含“小丑”和“社会”, 说明负面评价认为电影对现实的影射不够恰当。结合影片内容, 差评的观点可能包括“影片对未来社会的发展过于悲观”, “厌恶小丑过激的反社会行为”以及“反对将暴力, 民粹合理化”等。



图 5 好评词云图



图 6 中评词云图



图 7 差评词云图

总而言之,《小丑》背后所反映的社会问题和价值观念最受评论关注。较为正面的评价赞扬了人物的塑造,演员的演技以及故事情节,而差评对此提及较少,更倾向于表达对影片社会价值观的不认可。

四、 主题模型的建立

(一) 文本主题展示:

案例通过 LDA 建模,将《小丑》1625 条豆瓣短评的主题归纳为“演员表现”,“电影情节”以及“社会内涵”三个大类。“演员表现”里的关键词聚焦于男主角菲尼克斯的演技,将其与同类型电影里的经典角色相提并论;“电影情节”主题描述影片细节,包含了一些暴力,血腥的关键词;“社会内涵”里的关键词主要涉及《小丑》的现实隐喻。表 2 展示了前 6 个文本中包含的主题比率,可以发现:序号为 1, 2 的文本主题是社会内涵,序号 3, 4 的文本主题是演员表现,6 是电影情节,而 5 号短评侧重于社会内涵的同时也兼顾探讨了演员的表现。

表 2 部分文本主题概率表

文本序号	演员表现	电影情节	社会内涵
1	0.0032	0.0032	0.9936
2	0.0189	0.0189	0.9623
3	0.6773	0.0957	0.2270
4	0.9683	0.0159	0.0159
5	0.4038	0.0021	0.5941
6	0.1068	0.8835	0.0097

(二) 计算文本主题的平均概率

计算三个主题在所有文档中的平均概率。从玫瑰图中可以发现,文本平均出

现概率最大的主题依次为：社会内涵，演员表现以及电影情节。这说明人们最关注影片背后的价值观念和现实意义，结合短评内容，观众主要讨论的主题有“社会进一步撕裂的可能”以及“对暴力，民粹问题的态度”。此外比起剧情细节，《小丑》的卡司阵容受关注度更高，演员（尤其是荣获奥斯卡奖的男主角）的演技受认可度高。

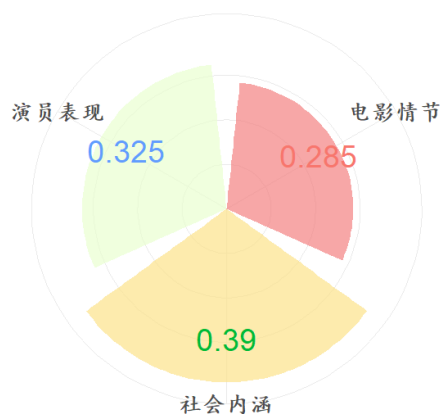


图1 文本主题平均概率玫瑰图

（三）星级与文本主题的交叉分析

豆瓣短评的星级可被分为“很差”，“较差”，“还行”，“推荐”，“力荐”五档。从图2中可以看出：“演员表现”主题评价越接近“还行”出现概率越高。结合短评内容可以发现：观众往往在表达“影片本身一般”的同时，会补充“表扬”男主的演技，因此这一现象不意味着观众对男主角持褒贬不一的态度。评价越低，“社会内涵”主题出现的概率就越大，这说明《小丑》的争议性主要在其背后的价值观念。主角的暴力行为最后非但没有遭到惩罚还被民众所膜拜，这在崇尚和谐稳定，社会矛盾没有到达不可调和地步的中国是不受到认可的。所以对文艺作品的评价也会被文化因素所影响。“电影情节”主题的占比总体上比较稳定。结合短评内容可以发现：评论若提及影片的某一场景，最终还是会落到其他两个主题上。

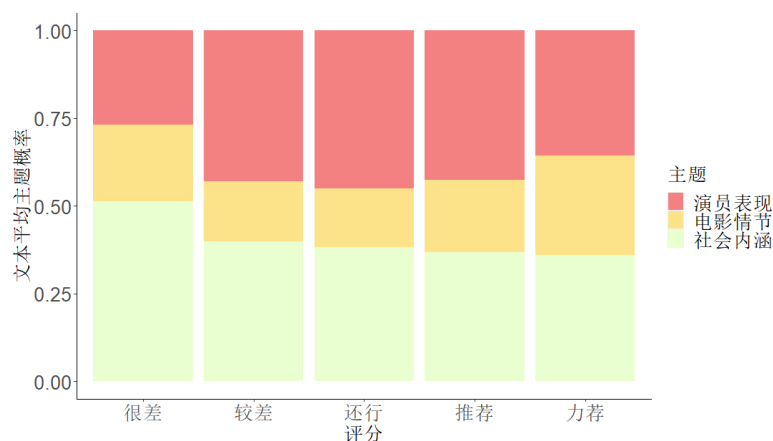


图2 各评分主题概率柱状图

五、情感倾向模型的建立

(一) 模型建立与分析

考虑到电影引起的争议，建立情感倾向的分析可以从宏观把握观众的态度。调用百度 API 完成了所有短评的情感打分后，案例以五天为一个窗口的大小，计算每个窗口内的平均情感分数。考虑到 2020 年 3 月 13 日后评论密度不高，因此只展示 2019 年 8 月 31 日到此之前的数值。从图 1 中可以发现：总体上《小丑》短评的正面性高开低走，9-10 月分数在 0.5 以上，10 月份之后分数不断下滑，最低时只有 0.1 不到。这种现象背后的原因和电影本身的特点有关联：作为没有在大陆公映，宣发的电影，出现网络资源后会第一时间观看的往往是 DC 漫画（小丑形象所属公司）的忠实粉丝，他们的影评往往是正面的。而随着《小丑》在网络上引起讨论，出圈，电影观众也不再局限于粉丝的范围。由于电影本身的基调灰暗，暴力镜头多，社会价值观的争议性会比较强，因此反映在短评的情感上就是分数不断走低。此外，考虑到行业的竞争性，DC 在漫画电影行业的主要竞争对手漫威影业也有动力引导对其作品的负面评论，但现有证据不足以支持这一观点。

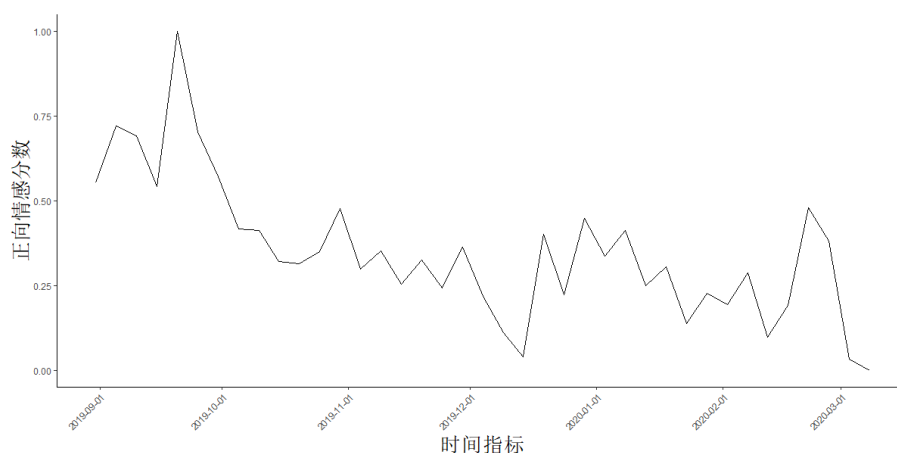


图 8 正向情感分数走势图

(二) 模型预测

对数据建立 Holt winter 模型，结果显示 $\alpha=0.91$ ， $\beta=0.02$ ， $\gamma=1$ 。这说明数据有水平因素和季节因素，而趋势因素不明显。

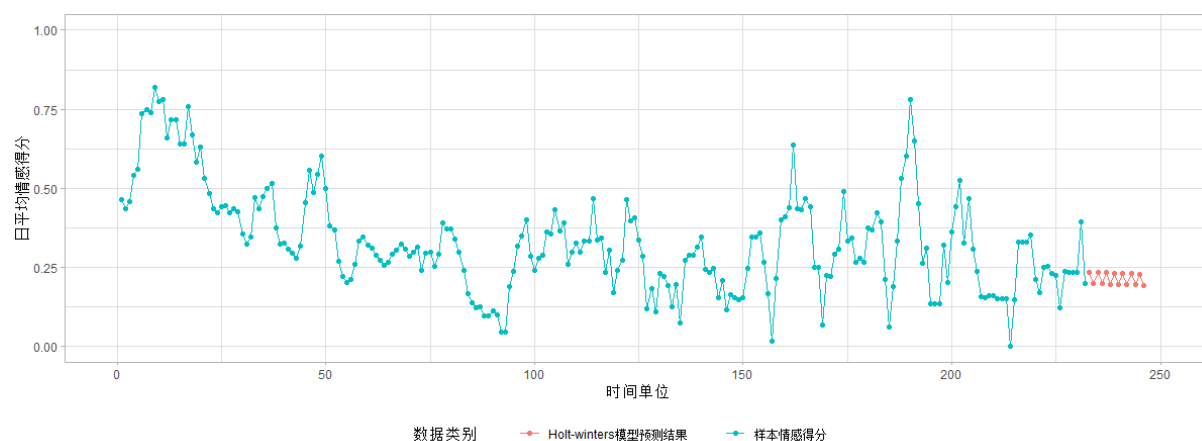


图 9 模型预测效果展示

从图 9 中可以看出：《小丑》短评在未来 14 天的评分分数将保持在 0.25 以下，这是一个较低的水准，说明观众未来对影片的正面评价不高。结合电影内容，暴力猎奇情节给人带来的新鲜感会随着时间的推移消退，但有争议的价值观可能会得到越来越客观的评价。

六、 商业应用

通过热度分析，主题分析和情感倾向分析可以得到一些很有应用价值的信息。

(一) 同题材电影的市场潜力

我国的观众强调社会和谐，且有“正义”必胜的朴素观念，因此对猎奇的犯罪剧情以及以反派为主角，以反派获胜为结局的电影普遍感到不适。在《小丑》

的评论中,对影片的讨论最终聚焦在对其反社会价值观,鼓吹暴力合法,造反有理的批判上,倘若正式登陆院线,影片很可能会遭到舆论的抵制。另一方面,同题材的电影在拍摄时很难过审,而引进时大量的剪辑又会破坏剧情,影响整体的质量。因此综合观众和审查两个角度来看,不管是拍摄还是引进,《小丑》以及同类电影都不是个好的商业选择。国内的出品方更应该选择以正面角色(如警察)作为主角,以正能量作为结局,让观众在一个相对更“安全”的心理环境下观看此类影片。

(二) 电影口碑维护

《小丑》在一开始在粉丝的助推下热度不断上升,且评价相对正面,而后来的“破圈”吸引了大量非粉丝观众,他们对影片普遍持否定态度,反映在指标上就是正向情感分数不断下降。从口碑维护的角度看,国内的舆情公关部门应该抓住两个“重点”:第一是在影片刚上线时利用好粉丝先行的优势,尽可能推高作品的评价和热度。第二是在作品“破圈”开始广泛传播时及时关注舆论的主题并加以引导。在《小丑》中,若团队能将大众对暴力和混乱的谴责引导至“如何避免社会两极分化”和“对社会边缘群体的关注”上,便能有效地控制住评论的负面情感,将讨论引到理性和人文关怀的轨道上。

七、 结论与展望

通过以上分析可以得到的主要结论:

1. 互联网对《小丑》的传播起到重要作用,影片的热度总体上先升后降,最后保持在一个较低水平上。
2. 影片的社会价值观讨论度最高,好评与中评比差评更多提及男主角的演技。
3. 评论主题可被归纳为“演员表现”,“电影情节”以及“社会内涵”。文本平均出现概率显示“社会内涵”出现概率最高,“电影情节”最低。评价越差,提及“社会内涵”的频率越高。
4. 评论的正向情感分数先升高再降低,未来 14 天还会保持在较低水准。

本案例的样本数量和变量数量较小,因此还存在一些不足可以改善:

1. 《小丑》在国内并未正式上映,且题材比较冷门,因此总体上评论样本较少。以后可以选择更有代表性的电影做舆情分析。

2. 可以将情感倾向分数分别拆成好评分数，中评分数，差评分数分别做预测，进而获得更详细的舆论信息。