

采用数学建模的统计学分析方法，将已有影片的各项数据作为数据库，再将影响票房的因素划分为几个变量，通过数学模型来分析一部电影可能获得的票房。关于票房预测的研究从 20 世纪 80 年代开始于美国，美国电影经济学家巴瑞·李特曼在 1989 年发表了论文《电影经济成功预测：基于八十年代人的经验》，<sup>(6)</sup> 开创了电影票房预测的先河，奠定了电影票房研究的基本模型和方法。之后国外有多项关于电影票房预测的研究，而国内也开始进行相关的探索。

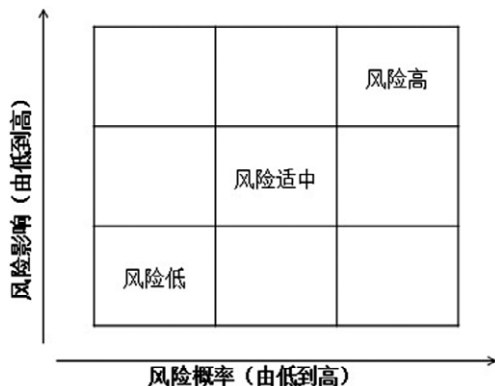


图 2. 简单风险矩阵图

此外，风险矩阵是一种风险管理工具，可用于分析项目的潜在风险。在进行风险评估时，将潜在问题的后果按严重程度（风险影响）定性分为若干级，再将潜在问题发生的可能性（风险概率）也相对地定性分为若干级，

然后以严重性为列，以可能性为行，制成图表（见图 2）。潜在问题如果处在右上角，就要想尽一切办法阻止其发生；如果潜在问题在中间区域，则应该给予一定的控制，并尽可能降低其发生时带来的影响；如果潜在问题在左下角，应该尽可能避免其发生。

以上是风险矩阵的简单应用，其根本还是定性分析，只是较为直观。风险矩阵还可以通过较为复杂的计算算出风险等级。风险矩阵中风险影响程度、风险概率可以通过咨询专业人士来确定。在电影项目中，可以邀请相关领域的专家来讨论决定，让专家们通过投票选择每项潜在问题的风险影响程度和风险概率大小，再用 Borda 排序法计算出潜在问题风险等级的排序，<sup>(7)</sup> 排序最高的潜在问题风险等级最高，相反则最低。这种较为复杂的风险矩阵，能够将定量分析与定性分析相结合，给出一个比较合理的风险评定。这种风险矩阵的方法给电影风险评估提供了可能，其研究价值和市场价值有待开发。

（司若，山东省新闻出版广电局与清华大学联合培养博士后 / 副教授，中国传媒大学艺术学部戏剧影视学院；张强，2013 级硕士研究生，中国传媒大学艺术研究院，100024）

(6) 张家林、钟一安《电影投资分析及风险管理手册》，北京：中国经济出版社 2014 年版。

(7) Borda 排序法是一种较为简单的排序投票法，每个选项借由选票上的排序来取得积分，积分最高者获胜。

## 基于大数据的电影舆情数据挖掘与风险监控研究 \*

The Study of Data Mining of Public Opinions towards Films Based on Big Data and that of Risk Control

倪万 / Text/Ni Wan

**提要：**随着网络与新媒体特别是社交媒体的发展，电影舆论的信息量具备了大数据的特征。基于大数据的电影舆情数据挖掘，是目前国内外进行电影项目决策、主创团队选择、营销策略制定的重要方式，在电影生产的风险监控中有着特殊的作用和价值。当然，大数据分析不能解决电影生产风险监控的所有问题，其价值主要体现在通过数据优化的结果作为提升风控质量的参照系，起到辅助预测与决策的作用。

**关键词：**大数据 电影舆情 数据挖掘 风险监控

从产业的角度讲，电影的生产总要符合投入与产出的对比逻辑，这就必然追求产品投入市场后能够获取最大收益；而电影作为文化艺术产品，其生产的风险及风险监控有着极强的特殊性。

即“投资的预期收益偏离实际结果的可能性”。<sup>(1)</sup>但从电影生产的全过程看，可能面临的风险会来自于四个方面：一是电影投资环境的风险：这种风险是由于

\* 本文系山东大学文科实验室培育建设项目“大数据与精确传播实验室”（项目编号：HSSLAB1504）的阶段性研究成果。

(1) 黄亮《文化产业投资基金的风险与风控机制》，《东南学术》2013 年第 2 期，第 114 页。

电影产业所处的社会环境具有不确定性而产生的,主要包括与电影相关国家政策、法律法规的变动以及市场结构的变化。二是电影项目管理的风险:这是由于电影项目在运作过程中由于自身管理方面的各种问题形成的风险,主要包括信息不对称和项目运行风险等。三是电影资金流动的风险:这是由于资产投资周期较长、流动性不足而引发的风险,一旦所投项目状况不佳或者退出渠道不畅,资金运作就有可能陷入困境。四是电影舆情引发的风险:这主要是公众意见出现了对电影产品不利的倾向而导致影片口碑不佳,进而影响票房。

对于前三种风险而言,电影产业有着较为成熟的风险监控策略,如投资环境风险需要提前做好前期调研与市场考察工作,就有可能在事前规避投资环境风险;管理风险可以通过制度安排得到有效控制;资金流动的风险可以通过完善的退出机制和资转债的方式规避。但对于电影舆情引发的风险而言,由于媒体形态的快速演变,社会舆情的整体状况发生了很大的变化,其监测难度大幅增加。

所谓电影舆情,是社会舆情在电影生产领域的延伸,特指社会公众对某部或多部电影乃至整个电影产业形成的认知、态度、情感与行为倾向等方面舆情的总体情况。在传媒大发展的时代,对社会公众或是广大的不特定电影观众而言,其对电影的认识不仅通过人际传播、(小)团体传播形成,更依赖于具有广泛传播效果的大众传播媒介,其评价电影更是需要大众传播渠道才能形成舆论场,进而产生更为广泛的影响。

这种影响,是由电影作为特殊文化产品而决定的,因为“社会公众需要观影后才能完整地知道影片的实际情况,但在观影前,一般依靠电影广告、评论、口碑等方式获取影片信息,这些信息不仅直接影响社会公众的观影选择,并且会在一定程度上影响观众对影片的价值判断。调查显示,超过三分之一的美国人积极寻求影评人的意见,而国内观众去影院观影的驱动因素中影片口碑占59.1%”。<sup>(2)</sup>

从目前的情况看,制片方普遍注重上映前影片的舆情态势,如通过大规模宣传开机仪式、关机仪式,上映时的首映式、剧照及预告片进一步拓展舆情空间。但在电影上映后,关于影片的社会舆论出现褒贬不一的时候,制片方则少有作为,影片取得商业成功但最终舆论不佳的情况也比比皆是。

当前传媒正在经历从传统媒体向网络与新媒体时代的过渡。传统媒体在电影舆论引导方面的影响力明显减弱,在电影上映的前期主要起信息宣传和议程设置作用,为舆论的最终形成提供信息源;对上映时的舆论也有一定影响,但总体来看影响力有限。而网络在电影舆论形成与传播过程中所发挥的作用日渐凸显。门户网站与电影相关的频道或栏目、专业电影网站、博客、微博、电影论坛等都成为舆论生成和传播的渠道,而且通过社交媒体进行的公众与公众间的口碑营销更

是能够取得信息传播较高的可信度与满意度。

因此,当前网络与新媒体对电影舆情的影响主要体现在的两个核心层面:

一是从舆情平台来看,虚拟社区与社交网络在电影舆论引导方面的作用日益凸显。如就国内影迷来说,时光网和豆瓣网是最为重要的电影交流阵地。大量具有较高互动频度的影迷聚集在一起,成为国产电影的重要舆论场。其影片评分为影片整体的舆论做出定位,成为社会公众观影选择的参照系。社会公众可通过此类社区来生产内容,进行信息的搜索与分享,形成网络公共舆论场。

二是从社会公众意见构成来看,网民意见与网络新意见领袖的意见相互交织。网络与新媒体的开放性使社会公众在网上更为方便地发表自己的意见,导致网络舆情的数据量急剧增长,并且随着现代社会价值观念多元化的趋向而快速变化。这时,网上的新意见领袖起着引导舆论、重塑主导意见的作用,他们一方面喜欢发表观点且观点独到,为他人能够提供很多信息;另一方面其本身有一定的社会影响力,其观点可以影响更多的人。如职业影评人,他们能够通过内部试映体现审阅影片,公映前便发布影评,进而影响了社会公众的态度和行为趋向。在这一过程中,新意见领袖们有着明显的双面性,也就是观点独到的同时又不陷于孤立。因此网民意见与网络新意见领袖的意见交织构成了舆情发展的整体态势。

## 二

电影的舆论“通常以显舆论、潜舆论和行为舆论等形式体现,显舆论即以语言、文字形式传递的信息、表达的意见,潜舆论以情绪形式表达,行为舆论则是以购买、点击等形式体现的。因此,除了口头评价、文字影评、影片评分之外,票房信息、电影奖项、视频点击、文字信息的浏览、转发等都是电影舆论的一部分”。<sup>(3)</sup>也就是说,分析电影舆情可以利用的电影数据是非常多,其中不仅包括了舆论研究所需的社会公众主观认知、态度、情感与行为倾向部分,还包括了反映市场情况的客观数据及社会公众观影行为的统计结果;在此基础上结合社会公众的背景信息等控制变量,就可以完整地描绘电影生产舆情的总体状况。从目前的情况看,可以直接为电影舆情数据挖掘所利用的数据类型有:“搜索数据、媒体热议数据、社交网站提及数据、视频网站用户数据、在线购票数据、影院观众消费数据”<sup>(4)</sup>等。

随着网络与新媒体特别是社交媒体的迅速发展,电影舆情的数据集也在向大数据方向演进。这一趋势正在影响着电影产业的生产方式和创作思维,无论是

<sup>(2)</sup> [3] 凌燕《新媒体时代的国产电影舆论生态》《现代传播》2013年第4期,第53页。

<sup>(4)</sup> 蒋波《大数据让电影不再“闭门造车”》, <http://media.people.com.cn/n/2015/0623/c14677-27193705.html>, 2015年12月21日访问。

电影管理部门、电影企业、电影人还是影评人，在电影舆情监测、舆情研判方面都面临着大数据的挑战，特别是对于投资人来讲，舆论引发的风险更是应该受到足够的重视。

所谓“大数据”，按照维克托·迈尔·舍恩伯格和肯尼斯·库克耶编写的《大数据时代》一书中的解释，指的是不用抽样调查为基础的随机分析法，而采样所有数据形成数据集的方法，现在“一般是指互联网范畴内可采集的各种类型的海量数据，尤其是与平台、渠道和用户相关的数据集。应该说无论在剧本创作，也就是电影创意的开端，还是在整个电影制作流程的管理上，大数据都有着非常多的对接可能和使用价值”。<sup>(5)</sup>

大数据其实不是一个新概念，从操作方法而言，大数据的运用历史已经很久了。而现在之所以对“大数据”这一概念有了更为广泛的认知和接受，主要是两点原因：一是随着数字化进程的加速，能够被数字化形成计算机可以处理的数据的信息源也越来越多；二是随着计算机存储和计算能力的不断提高，数据本身也就越来越多。

这一趋势对我们的思维方式有了颠覆性的变革。从技术上讲，大数据带来的数据管理的革命性变化：一是在计算机信息管理方面，数据流动方式从最初的中央式计算方式转变为现在以“云计算”与“云和端互动运行”为特征的分布式计算方式，使得数据处理能力海量增加；二是统计学和现在大规模的数据相结合，从大样本的描述性统计与中样本的推断性统计相结合，转向超大规模样本数据对总体的直接描述分析；三是通过不同数据集之间的交互，直接建构了事物之间的相互关联，为产业投资的预测决策提供依据。

就电影产业而言，大数据带来的冲击也是多方面的，对政府决策部门而言，可以获得更加高效的工具，以进行科学的、理性的决策；对投资者而言，一切用数据来说话，运用专业而又科学的大数据工具，使投资人对电影行业的分析，特别是影片的投入与产出的预期更加清晰；对从业者而言，无论是制片方还是创作团队或者影评人，有了大数据的参考，能够建构比较明确的创作方向，使整个行业更加专业化；对电影观众而言，观众在选择观看影片时有了更加直观的参考数据，进而也直接影响到电影票房。

目前业界广泛认可大数据具有4V特征，即：大量(Volume)、多样(Variety)、高速(Velocity)、价值(Value)。而这四个特征的前三者为客观性特征，是需要通过各类计算机工具进行获取，并可以进行客观性描述。第四个特征“价值”是最重要的，没有价值的大数据没有实际意义的，而价值需要通过数据挖掘加以人工与机器的研判才能得以体现。

于是通过对大数据的意义与特征的分析，给与我们通过电影舆情的数据挖掘进行风险监控有了一些基本的认识。

一是大数据是当前电影舆情数据挖掘的基础。

大数据要求电影信息的来源要广泛，不仅来自电影管理部门与电影制片方、投资方的舆论场，还包含社会公众的舆论场，既有正面褒扬还有负面批评，从而形成纷繁杂乱甚至真伪难辨的海量信息。特别是虚拟空间中贴吧、论坛、新闻跟帖、微博、微信等形成的互动数据，是研判舆情发展趋势的基础。

二是要做数据挖掘才能体现大数据的价值。

美国学者纳特·西尔弗所写的《信号与噪音》一书中论证，在海量信息中，信号才是有用的信息，而还有很多没用的信号是干扰决策的噪音。“我们总觉得信息量越大，能够得到的真理就越多。其实不是的，信息增长的速度要远远快于真理增长的速度。”<sup>(6)</sup>所以大数据越大，有用信息越多的同时，噪音也越多。如何发挥其数据的作用，关键要有洞察之心进行数据挖掘才能获取大数据中蕴含的价值。

三是电影产业应用大数据舆情分析的目标是为电影生产提供前瞻与预测，进而进行风险监控。就大数据舆情而言，包含了社会公众客观性信息的同时，更重要的是对电影的关注程度、意见情况、情绪变化、行为倾向等。在杂乱的电影舆情信息中，发现趋势、预测走势成为关键。

### 三

从实际情况看，基于大数据的电影舆情数据挖掘对电影风险监控能够起到什么作用？效果如何呢？我们从数据来源开始分析。当前电影的舆情数据整合，主要整合了前述三类数据：一是社会公众的人口特征等背景信息，二是反映市场情况的客观数据及社会公众观影行为的统计结果，三是舆论研究所需的社会公众主观认知、态度、情感与行为倾向数据。这样才形成了电影舆情数据挖掘所需的大数据。这些数据的获得方式是多方面的，如淘宝用户分析报告分析了消费市场的人群特征和用户行为，以此为根据定向推送信息、选择不同的营销策略等；谷歌等科技企业主要基于网络舆情的抓取和分析，“要么基于来自搜索引擎等工具的公共数据，要么使用网站自己注册用户的行为数据”。<sup>(7)</sup>这对电影生产的风险监控有着特殊的意义，主要体现在以下几个方面：

第一，大数据为电影上映前提供了舆论分析的条件，对潜在风险有着预判的作用。

从理论上讲，电影项目主管可以根据用户数据对整个电影创意流程进行有针对性的策划和调整，这一点对电影产品的目标与社会公众的对接非常重要。在

(5)李迅、王义之《大数据对电影创意和内容管理的意义》，《当代电影》2014年第8期，第4页。

(6)《谈谈大数据思维》，[http://www.360doc.com/content/14/0922/16/2088748\\_411510444.shtml](http://www.360doc.com/content/14/0922/16/2088748_411510444.shtml)，2015年12月25日访问。

(7)罗政、张玉洁《百度大数据预测“首秀”缘何“失算”》，《经济参考报》2014年10月21日。



以往的电影商业策略中,只有抽象的“观众”概念,但在具体的电影生产实践上,无法精准地知道“观众”想的是什么。而大数据的意义首先在于可以通过舆情数据挖掘将那些“抽象的观众”活灵活现地呈现,“引导着以用户—观众—消费者为主导的整个电影创作和管理观念的转变”。<sup>(8)</sup>可以说,网络与新媒体为电影创作者提供了更真实全面的渠道,去了解观众、分析观众。这样可以分析什么样的题材可能市场上比较受欢迎。通过大数据分析能捕捉消费者的观影爱好、习惯,甚至付费方式。如《小时代》大概有2400万的小说读者,具备进行电影创作的基础;该片选择演员的时候,曾对每个演员的粉丝群体特征做过数据分析,发现粉丝数超过一亿,具备了社会公众基础。再如韩寒在微博上拥有4000万粉丝,其电影《后会无期》的影片主创人员的微博粉丝数累加达5000万,这部以微博为营销核心阵地的电影赢得了6.5亿元的票房。

第二,大数据分析使电影的创作有了精准的电影要素分析,能够更精准有效地进行电影投资立项,降低投资风险。

电影通过大数据分析,量化了电影立项、电影投融资、电影营销、电影发行与放映以及衍生产品推广等电影元素可能产生的价值。一方面,通过大数据的舆情数据挖掘,制片人能够更精准有效地开发电影创作项目:通过大数据分析,可以更好地为导演、演员寻找符合自身的定位,如《纸牌屋》的风行是美国Netflix公司通过多年积累数据大数据分析技术的应用,得出观众最喜欢的剧情类型、演员阵容后,该公司才做出了拍摄《纸牌屋》的决策。华谊兄弟总裁王中磊谈到,互联网介入之后,“电影产业受到的最大的影响,就是让电影都变成了‘事先张扬’的故事。我能感觉这几年变化比较大的是,互联网让传统电影透明化。以往我们在策划甚至发行的时候,跟受众都没有那么多的链接。但是现在,往往一部电影还没有开拍,就已经通过互联网跟观众们有了密切的互动”。<sup>(9)</sup>另一方面,从内容源头上,大数据将成为电影项目筹划的重要技术手段。当前的中国电影界也在采用客观的数据去分析决策,爱奇艺高级副总裁杨向华指出,“现在我们做一部电影决定投资,会从两个维度判断:首先,通过关于电影预测的大数据系统,跑一下这个片子的票房、预期是多少;第二个结合宣传、发行或者做项目的团队,以专业人的角度看这个片子的票房预计大概是多少,从两个方向来决定这个片子最后要不要做”。<sup>(10)</sup>

第三,电影上映后电影舆情数据挖掘对电影发行、营销环节的决策有重要的指导意义。

除了制作领域,观影人群属性、人群分布等大数据对电影发行、营销环节有重要的指导意义。如影片《赏金猎人》,两位男主角李敏镐和钟汉良的粉丝超过5000万,因此电影营销从线下向线上发展,将社交网站作为电影信息推广以及口碑争夺的重要阵地。在这

个方面上有一个非常明显的现象,那就是“以预售为代表的在线售票直接可以影响院线排片,指导发行人力、物料的匹配”。<sup>(11)</sup>从2014年后的数据看,预售在线票务贡献约40%的电影票房。如动作电影《速度与激情7》上映前,环球公司对中国大陆票房的预估是8亿元人民币,而预售票系统的数据分析表明其已创下在线购票人数的新纪录,结合电影的口碑分析,于是环球公司在电影上映期间追加了推广预算,最终票房达到24亿元。再如,《汽车人总动员》的总票房累计只有475万元人民币,这与网络上的“抄袭”争议的舆情影响了影片的票房,豆瓣、时光网、格瓦拉、猫眼等数个较具代表性的电影网站《汽车人总动员》的评分垫底,导致各大院线不再排片。

第四,电影舆情分析可以观察电影口碑,甚至可以决定是否持续投资。

电影档期过后,随着票房等一系列数据的公布,各类媒体的持续报道,关于影片新一轮舆情将持续发酵,随着时间的推移,各种产业报告、年度评奖、研究资讯等陆续推出,关于影片的舆情还将在一定时间和范围内持续。通过对这一过程中形成的舆情数据挖掘,可以决定该类型、题材的电影是否值得持续投资。导演徐峥认为,大数据对创作者选择电影类型、题材等有重要作用,也影响着制片人或者投资方对整部电影的预期值,进而影响电影的规模,如其根据《人再囧途之泰囧》的情况,分析出《港囧》的票房,消费群体的性别、地域,甚至是年龄,决定了持续该题材的拍摄。

## 四

美国是商业电影模式最为成熟的国家,其电影的投资决策、观众定位、剧本故事及演员选择、传播策略等生产环节,均在一定程度上基于电影舆情的数据挖掘。一批商业公司通过社交媒体言论收集,分析社会公众对影片和演员的评价,以确定商业策略。

如2009年成立Fizziology公司,是一家社交媒体研究机构,其自行开发了相关软件和插件嵌入主流的社交媒体,如Twitter、Facebook、Instagram等追踪电影舆情;并采用机器分析与人工分析结合的方法,为电影制作公司有偿提供在线舆情报告。其官方介绍有超过20亿个关于媒体与娱乐的社交媒体信息源,准确度高达95%。其具体操作流程是通过监控电影舆情,将即将上映的电影与曾经上映的相似电影进行比较,以预测电影票房的情况,而且这一工作可以提前

(8)同(6)。

(9)李超《互联网+的春风吹绿了谁?内容与渠道仍在激辩》,http://www.1905.com/news/20150618/902716.shtml?fr=wwwfilm\_news\_xgxw\_1\_20141017,2015年12月28日访问。

(10)同(4)。

(11)这里的舆情分析仅指通过在线售票数据的分析,电影行业中买票房与返点的操作亦会影响观众选择和票房价值,但从另一侧面也反映出电影舆情对票房的意义。

一年开始,以分析该影片最适合的演员是谁,电影剧本中最大的亮点是什么;电影上映后,也可以对影片舆情持续追踪,并评价观众口碑。对电影制作公司而言,该舆情报告对演员选择和市场定位有着战略性的意义。

2014年3月,联合经纪公司(United Talent Agency)与电影监控公司 Rentrak 联合推出了 PreAct 系统,其官方宣称有 70% 左右的主流发行商都在使用 PreAct。这一系统主要整合各种复杂数据,可以让电影制作公司根据社交媒体言论形成的舆情数据,监测影片口碑以及营销活动的效果。其采用的方法是与全球的 2.5 万家电影院进行互动的问卷调查,通过对每部影片的百分制评分,来确定观众的规模、对影片的好感度等指标。联合精英经纪公司的研究主管大卫·赫林(David Herrin)对这一方法的解释是,“社交言论中的‘金三角’包括规模,即大范围的讨论;好感度,即人们给出的正面评价;以及那些自发性的讨论,主要由消费者自行推动,而非制片方或媒体驱动的结果”。<sup>(12)</sup> PreAct 系统也可以在电影上映的一年以前就开始监测数据,并提供情况变化的评分分析。

其他公司也开发了相关的软件或平台从事电影舆情分析业务,如分析社交媒体数据并负责数字营销活动的 Moviepilot、软件开发商 Salesforce.com 的舆情跟踪应用模块 Radian6、社交媒体分析机构 Crimson Hexagon 提供的分析服务、谷歌公布研发的基于大数据的电影票房预测模型等。其中典型的成功案例是基于美国视频网站 NetFlix 大数据分析而投资拍摄的《纸牌屋》。

从国内情况来看,“大数据”逐渐被电影行业认识与应用。2015 年被称为电影行业的“大数据”年,这主要有两个标志性的事件:“一个是 10 月份广电总局的实时票房数据平台上线,官方权威性的数据统计终于可以做到每日更新。第二个现象就是完全被‘大数据’监控的网络大电影市场的爆发,让影视行业有了全行为分析的可能”。<sup>(13)</sup> 从中国电影行业来看,有两个趋势值得我们进一步关注:

一是大数据分析平台日臻完善和成熟。百度糯米、淘宝电影、微票 BAT 三大与电影舆情直接相关的网络公司,已经具备了大数据整合和分析的能力。从其策略来看,一方面是利用线上票务拉动票房,并获取基础性数据;另一方面是通过数据挖掘和影片宣传策略相结合,提升营销效率。其核心是大数据人群属性的分析,精确地向合适的电影观众传播宣传内容,以实现口碑的快速扩散。“以《寻龙诀》为例,百度糯米利用自身大数据优势,在影片大数据平台精准勾画出《寻龙诀》用户画像,并以口碑舆情、营销指数等为《寻龙诀》影片宣发方提供参考,有效将 20% 的中立用户转化为支持用户。”<sup>(14)</sup> 再如,艺恩网、猫眼电影等专业电影大数据网站在 2015 年也建立了基于院线电影的“大数据”分析平台。

二是大数据分析向免费方向发展,降低了应用门

槛,更有利于大数据分析的普及。如艾漫数据将其电影决策系统 IFilm 向行业免费开放。该系统主要服务于电影决策,其分析了全球七亿多个社交账号,是一个典型的基于电影舆情的数据挖掘系统,其通过“消费者跟踪、影片信息热度追踪、投放物料的监测分析、影片口碑和观影期待的挖掘,实现电影全程评估”。<sup>(15)</sup> 艾漫科技表示通过零门槛的电影大数据使用,培养中国电影行业的大数据使用习惯。

## 五

基于大数据的电影舆情数据挖掘,在电影项目决策、主创团队选择、营销策略制定等方面起到了风险监控的作用,但并不是说大数据能够解决电影风控的问题。

(一)从当前产业经济的风控模式来看,基于大数据的电影舆情数据挖掘在风险监控方面只能起辅助作用。

在全球经济领域里,目前主要有两种风险控制模式,“一种是信贷化模式,即通过制度安排,利用中介机构的规模优势和期限错配,让风险在经济运行的长周期中被化解。另一种是证券化模式,其本质是让核心风险在市场中寻求自我定价和平衡,最终让风险在不同的市场主体之间转移”。<sup>(16)</sup> 两者的核心区别在于信贷化属于政府能够起到强大控制力的间接性融资,在防止系统性风险方面有相当程度的积极作用,但其他风险却有不确定性;而证券化则不直接承担风险,让风险通过市场的合理配置进行分解,但控制不力则会出现极大的系统性风险。但无论怎样,解决风控问题依然是产业界的难点。对电影生产而言,引入大数据进行电影舆情的数据挖掘,可以优化风险监测的管理,在一定程度上提高风险控制的效率以及降低风控成本,但在最终的风控质量方面还需要全面的风控策略来完成。

(二)从大数据分析的数学原理来看,当前通过电影舆情数据挖掘进行风险控制其实是在以偏概全,这类分析目前能够做的是数据优化。

第一,大数据分析的核心假设条件是,网络世界能够提供全数据。维克托·迈尔·舍恩伯格教授也曾说不再需要抽样的整体数据才可以定义为大数据的集合。而据统计,“全球有能力进行所谓大数据应用的公司不会超过 50 家”,<sup>(17)</sup> 因此,对当前网络而言,能

(12) 福布斯中文网《电影舆情:网上骂烂片?电影公司全听得见!》,薄锦译、徐笑音校, <http://chuansong.me/n/1419196>, 2015 年 12 月 12 日访问。

(13)《不懂大数据的电影人都将被市场淘汰》, <http://www.aiweibang.com/yuedu/71164944.html>, 2015 年 12 月 27 日访问。

(14)《BAT 互联网大数据联合发行 撑起 寻龙诀 九成在线票务》, <http://news.sina.com.cn/o/2015-12-22/doc-ifxmszek7601348.shtml>, 2015 年 12 月 23 日访问。

(15)《艾漫数据:基础影视大数据就应该免费》, <http://www.weste.net/2015/10-30/106627.html>, 2015 年 11 月 25 日访问。

(16)《谈风控本身是很可耻的事情》,黔西南新闻网 2014 年 10 月 27 日。

(17) 中国电子商务研究中心《大数据能否带来风控革命?》, <http://www.100ec.cn/detail-6201663.html>, 2014 年 10 月 8 日。

够获得这一数据集只能说存在巨大的可能性和预期，但不是普遍的现实。因此，当前所谓的“大数据”只是“小数据”。如在电影生产领域，往往是仅仅通过问卷调查或者电话采访得来的信息，却以大数据的名义发布研究结果，这样的结论只能以偏概全。另外，对于电影舆情数据挖掘而言，也并不是数据量越大越好。对数据量规模的盲目追求，更容易使得大量非相关数据纳入分析，反而会造成由于数据收集和整理所耗费的人力、物力、财力的增高，而带来更大的成本付出，同时又由于非相关数据的增加使分析过程复杂化和分析结果呈现更高的不确定性。

第二，从大数据分析的实际内容来看，目前主要是在做数据优化。

大数据普遍认为可以有效推演和设定未来的走向，极为精准地推断某些事件出现的概率，进而可以很好地控制风险。但在大数据分析过程中，一般都将概率性事件设立了变量的幅度范围，也即进行数据挖掘事前要有先决逻辑作为分析基础。因此，当前所做的大数据分析可以称为基于一定随机数据量的数据优化，在相当多的问题上还需要设定一定概率度和抽样误差的推断性统计进行解决。

第三，从大数据为基础的统计学运用来看，当前数据无论多大都是建立在社会特征与公众特征不发生巨大变化的逻辑之上的，而在当前一个快速变革的社会里，社会特征与公众必然不断改变。这种改变，使过去任一瞬时获取数据所构建的相关性消失，因此新的数据要持续积累形成历史数据，才能不断反映当前社会的变化和推测发展趋势。大数据分析的当前数据与历史数据要有合理的结合分析，也即横向分析与纵向分析要结合进行。

第四，大数据分析电影舆情进行风险控制存在行业本身的不确定性。

电影产品不同于其他产品的最大特点，就是它是基于非理性特征的体验性消费。这很难完全用数据去分析、前瞻，它的价值可以体现对未来方向的指示，但仅是一个参考系。如施洛斯的低估值投资方法，是提高投资成功率方法，但这一方法对变化缓慢的稳定行业而言，投资预测时更容易有确定性。而像电影行业则不可预测，行业数据也非概率统计的正态分布。“如果电影票房按正态分布，那大多数电影赚到的钱都该落在某个平均值附近，有三分之二的电影票房，将落在这个平均值周围一个标准差的范围内。但在电影业中，20%的电影带来了80%的票房收入。这类由热点驱动的产业，尽管完全无法预测，却遵循了一个完全不同的分布。”<sup>(18)</sup>因此非正态分布的电影产业本身很难用统计推断的方法解决，只有基于低估值的大数法则比较成功，如狮门影业只专注于恐怖片等小成本类型，每次电影制作成本都很低，票房不佳损失很小，但成功者却可以一本万利。

(三)从电影数据的积累和运用来看，需要人工研

判和机器分析相结合。

当前电影市场希望能从简单的互联网数据中直接研判舆情，最后就变成了对数据过度依赖的局面。如布拉德·皮特(Brad Pitt)主演并担任制片人的《点球成金》获得了成功，但这部电影一开始曾向Relativity Media的老板瑞恩·卡瓦诺寻求投资，而卡瓦诺执著于用纯数学方法研究电影投资，经过分析后认为体育电影不赚钱，于是放弃了投资《点球成金》的机会。从国内情况看，电影《黄金时代》明星众多，却在国庆档票房失利。虽然《黄金时代》在上映之前百度大数据电影票房预测认为其总票房将至少达到2亿元，但却与预期相去甚远。所以即便大数据将人类对电影的所有认知都通过数据存储下来，但仍要解决数据的逻辑性梳理和人的主体性认知问题，因此人工研判过程不可替代。

(四)从舆情信息发布者的结构来看，需要提高影评人的整体素质，发挥舆论领袖的积极作用，引导舆情良性发展。

在当前社交媒体大发展的时代，各界名人都常常通过发表影评来提高人气，而能够主导电影舆论的权威“影评人”和权威艺术媒体却始终缺失，因此在电影舆情中主流话语比较弱势，出现舆论过于分散和混乱的状况，甚至出现各种势力操纵舆论的不良状况。这种舆情状况不仅不利于创作者获得对自己作品的反馈，也无益于以后的创作，更不利于广大社会公众获取客观的舆情信息。因此，在电影舆情领域也有着掌握话语权和舆论引导权的要求，改变目前非专业评论主导舆情的局面。通过社会公众与网络意见领袖的双向互动形成的舆论场来引导电影舆情的良性发展。

## 结语

2015年6月17日，上海国际电影节组委会发布全国首个《互联网+电影趋势研究报告》。报告称“借助互联网，制片公司可以深耕观众的戏份需求，找准市场空白点，从而更好地把握中国观众的心理、情感需求，提升自身的竞争力”。<sup>(19)</sup>随着技术不断进步，大数据的运用将会更加成熟，数据收集的维度将会更加多元，算法将会更加优化。基于大数据电影舆情的数据挖掘进行分析，是目前国内外进行电影项目决策、主创团队选择、营销策略制定的重要方式。当然，大数据分析不能解决电影生产风险监控的所有问题，其价值主要体现在通过数据优化的结果作为提升风控质量的参照系，起到辅助预测与决策的作用。

(倪万，副教授，山东大学文学与新闻传播学院，250100)

(18)《大数据应用：大数据就是非相关数据的相关性》，大数据观察 <http://www.shuju.net/article/MDAwMDA5M4TI0.html>，2015年4月26日。

(19)同(9)。