

微博营销对电影票房影响的实证研究

陈莉, 李林

(上海理工大学 管理学院, 上海 200093)

摘要:随着微博逐渐成为电影宣传的主要平台,研究微博数据与电影票房之间的关系显得尤为重要。运用实证研究的方法,收集3个平台上电影上映前一周和上映四周这段期间每天每个影响因素的数据。将这些因素分别与票房进行相关性分析,发现每个因素都与票房成正相关,经过主成分回归分析建立了回归模型,将模型在统计的数据中进行实证研究。结果表明,该模型有很好的预测性,可以为电影制片方的营销策略提供更具有实用价值的参考。

关键词:微博营销;电影票房;实证研究;主成分回归模型

中图分类号:F201 **文献标志码:**A **文章编号:**1671-1807(2017)07-0050-05

微博——一个由推特演变而来的信息传播平台,虽然发布信息的上限仅有140字,但是有极强的实时性,并且方便微博用户之间的互动。因此,迅速在人们生活中活跃起来。由于微博面向的是所有在微博注册账号的用户,信息受众面广,并且信息的实时更新较快,微博也逐渐成为各个行业进行产品营销的渠道。在电影行业,微博营销作为一种全新的营销手段,受到广大电影发行公司的青睐。《2015年微博用户发展报告》显示,从2014年12月至2015年11月,共有268部电影开通官方微博,发布136 180条微博,获得6 452 435条评论、2 3120 759条转发、25 193 300条点赞。在微博迅速走红的同时,基于微博的微博营销逐渐受业界认可,电影官方微博逐年递增。因此,了解微博营销、微博口碑传播和电影票房之间关系具有重要的意义。

笔者通过分析国内外文献^[1-8]发现,对票房的影响和预测通常使用定性和定量分析。定性分析一般基于营销学、消费者行为学等学科;有些文献用定量分析的方法分析一部电影的官方微博账号发布的总的微博条数和最终票房的相关性,有的分析电影预告期内活跃度与票房的相关性,这些文献中数据大多在一个平台上进行数据收集,平台和指标过于单一,并且没有体现在电影上映期间每天的变化。这就存在一些问题:仅在一个平台对数据收集,忽略了其他平台的影响;仅仅研究一个因素的总数与总票房的关系,忽略了其他因素的影响;进行多元线性分析时,各

个因素之间可能存在多重共线性;电影的关注情况,很大一部分情况是在电影上映后,不考虑这部分的数据可能会使结果与实际存在偏差。

针对上述可能产生的问题,本文在以往学者研究的基础上,对电影票房的影响因素做了全面的补充,跨平台收集了电影上映前一周和上映四周这段期间每天每个影响因素的数据。对指标之间进行相关性分析,发现每个因素都与票房成正相关,对比使用了主成分回归分析方法,最终建立了回归模型,对模型进行实证分析。结果表明,该模型有很好的预测性。

1 基于微博的电影营销分析

微博营销^[9]是指企业或非盈利组织通过微博这种新兴社会化媒体来宣传公司,推广产品等,进而影响受众。微博营销是病毒式营销的创新方式。所谓病毒式营销,是指企业通过网络传播的方式在全球网络社群发动营销活动,利用口碑传播成为与消费者交流强有力的营销方法。微博的转发和评论具有病毒式传播的特点,而病毒式营销正是建立在病毒式传播基础上的营销模式,通过利用公众的积极性和人际网络,让营销信息像病毒一样传播和扩散,营销信息被快速复制并传向数以万计的受众,从而使信息的传播和影响呈指数式爆发。

在过去的2015年,中国电影市场再攀高峰,另外,微博完成了“认知—售票—口碑”服务闭环的构建,成为中国最大的电影营销平台和用户社区,微博成为各大电影宣传的首选平台。

收稿日期:2017-03-24

作者简介:陈莉(1993—),女,河南周口人,上海理工大学,硕士生,研究方向:管理科学与工程,数据挖掘,数据分析;李林(1966—),女,辽宁宁口人,上海理工大学工业,副教授,硕士生导师,工程研究所副所长,研究方向:工业工程,质量管理。

笔者选取了 2016 上映的三部贺岁片的数据,用 EXCEL 绘制柱状图,从图 1 中可以看出,电影票房较高时,它的微博数、评论数、转发量、赞数都相对较高。微博的信息可能与票房存在正相关的关系,为了验证这一想法,笔者收集更多的电影数据来进行分析,研究电影票房和这些因素之间的联系。

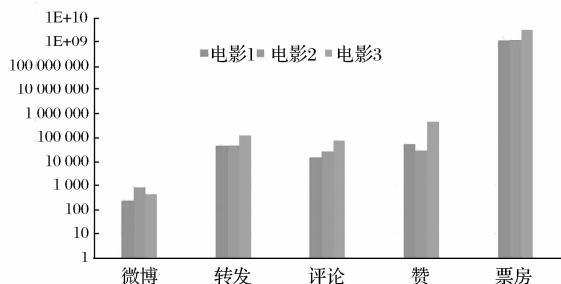


图 1 电影票房影响因素图

2 数据收集

2.1 数据来源与样本选择

微博数据的大数据量、实时性、短文本等特性给微博数据的搜集带来了困难。有论文采用数据爬虫的算法进行数据收集,该方法对收集者要求较高,需要学习编程语言。本文通过微博平台直接搜集每天微博数、评论数、转发数,相对简单,可操作性较好。热词趋势的数据来源于微博风云榜网站,“微博风云”是一家基于微博开放平台进行数据统计、分析与挖掘的网站。关于搜索指数,笔者采用百度指数中的数据,百度是一个家喻户晓的浏览器,从百度指数中统计到的数据具有代表性。关于票房的统计,本文选择相对权威的电影票房排行榜的数据。

在样本选取方面,本文采用统计学的抽样方法,收集从 2015 年 1 月至 2016 年 2 月期间上映电影中的 24 部电影,以每月两部进行收集,由于电影宣传方式及电影数据不全等问题,最终收集 17 部影片的详细信息。这些样本选择主要基于两方面的考虑:①这些电影在票房和微博数据上有差异,便于研究不同电影的票房情况。②17 部电影数据量达到 400,数据量较大,从而保证研究结论的可靠性。

2.2 指标选取

在指标选取方面,除了选择以往文献中使用的微博数这一数据,笔者还考虑微博本身涉及到的特征数据(如转发数、评论数、赞数),百度指数提供电影名称的搜索指数以及微博风云所提供的热词趋势等数据,使用多平台、选择多指标来研究票房的影响因素。最终选择微博数 X_1 ,转发数 X_2 ,评论数 X_3 ,赞数 X_4 ,搜索指数 X_5 ,

热词趋势 X_6 共 6 个指标,各指标的含义如下:

- 1) X_1 为各电影的官微每日发布的总微博数;
- 2) X_2 为各电影官微每天所发微博的总转发数;
- 3) X_3 为各电影官微每天所发微博的总评论数;
- 4) X_4 为各电影官微每天所发微博的总赞数;
- 5) X_5 为各电影名称的每日百度搜索量;
- 6) X_6 为各电影名称的每日热词量。

3 实证分析

3.1 数据标准化

虽然在以往研究中,没有文献将数据标准化,但是为了消除量纲的影响,笔者首先将 X_1 、 X_2 、 X_3 、 X_4 、 X_5 、 X_6 和 Y 进行标准化处理。

3.2 相关系数分析

本文运用 SPSS22.0 对样本数据首先进行相关性分析,相关性结果如表 1。其中皮尔森相关系数^[10]的定义如下:

设 $(x_i, y_i), i = 1, 2, \dots, n$ 是 (x, y) 的 n 组样本观测值,称 r 为 x 和 y 的皮尔森相关系数。

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

用 SPSS 可以对因素做相关性分析,相关性结果均在 0.01 的标准上显性相关,具体情况如图 2。从图 2 中可以看出,各个指标与票房的相关系数最小的,是 0.273,最大的是 0.709,且均为正,则在 0.01 的标准下都正相关。分析其他各个指标间的相关性发现,各指标之间均正相关,其中 X_2 与 X_3 的相关系数是 0.683, X_2 与 X_4 的相关系数为 0.645, X_3 与 X_4 的相关系数为 0.712,三者之间有着很强的相关性。

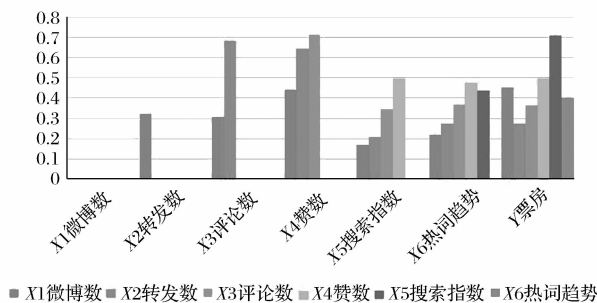


图 2 各因素相关性分析图

3.3 回归分析

对于电影票房的预测,以往研究中一般采用回归分析。因此,笔者首先对数据进行回归分析。由上节中分析图 2 时发现各指标之间存在很强的相关性,因

此笔者进行了多重共线性诊断。从表 1 中可以看出相关关系矩阵中有三个值超过 0.5,其中最大的达到 0.96,特征值中也有两个小于 0.5,这就说明可能存

在多重共线性。因此,为了消除多重共线性,笔者又进一步采用主成分回归分析法进行分析。

表 1 共线性诊断表

共线性数比例										
模型	维度	特征值	条件指数	常数	X_1	X_2	X_3	X_4	X_5	X_6
1	1	3.31	1	0	0.02	0.03	0.03	0.02	0.03	0.03
	2	1.138	1.706	0.59	0.05	0.02	0	0	0.07	0.02
	3	0.759	2.088	0.01	0.49	0.09	0.11	0.01	0.04	0.11
	4	0.663	2.234	0.28	0.38	0.02	0	0	0.03	0.42
	5	0.527	2.505	0.12	0.02	0.02	0.01	0	0.67	0.41
	6	0.37	2.991	0.01	0.02	0.57	0.68	0	0.01	0
	7	0.233	3.772	0	0.01	0.26	0.18	0.96	0.15	0.01

a. 应变数\：Y。

3.4 主成分回归分析

1)主成分分析。由图 2 可以看出,6 个指标都呈正相关关系,且有 3 个变量之间有较强的相关性,说明可以从原始指标中提取出反映某些原始指标的共

同特性的少数公因子。表 2 为研究所得到的各主成分特征值及贡献率。从表 2 中可以看出 3 个特征值

对应的方差解释度为 81.484%,满足主成分分析的一般要求。表 3 是 KMO 和 Bartlett 检验表,从表 3 中可以看出 K-M-O 值为 0.791,Bartlett 球形检验的 Sig. 取值为 0.000,表明拒绝该假设,各个变量之间不是相互独立的,具有较强的相关性,因此适合做主成分分析,同时表明研究结论可信度较高。

表 2 特征根与方差贡献率

元件	总计	起始特征值 变异的%	累加%	总计%	提取平方和载 入变异的%	累加%	总计	循环平方和载 入变异的%	累加%
1	3.119	51.984	51.984	3.119	3.119	3.119	2.181	36.351	36.351
2	0.99	16.505	68.489	0.99	16.505	68.489	1.641	27.348	63.699
3	0.779	12.991	81.48	0.779	12.991	81.48	1.067	17.781	81.480
4	0.566	9.434	90.914						
5	0.303	5.043	95.957						
6	0.243	4.043	100						

表 3 KMO 和 Bartlett 检验表

Kaiser-Meyer-Olkin 测量取样适当性		0.791
Bartlett 的球形检验	大约 卡方	854.531
	df	15
	显著性	0.000

若要对所提取的主成分进行解释,需要得到 3 个主成分的旋转矩阵,如表 4 所示。首先从表中可以看出 X_2 、 X_3 、 X_4 的载荷大于 0.5,三个指标应归为一个主成分,这也从上节中三者有较强的相关关系可以看出。分析表 4 可得到三个主成分的构成情况如下:主成分 Z_1 中, X_2 、 X_3 、 X_4 的载荷大于 0.5,该主成分主要是以上 3 种指标的反映。由指标含义可知,转发数、评论数、赞数越多,说明能够实际看到该微博的用户越多,所以将该主成分定义为覆盖度成分。 Z_2 中, X_5 、 X_6 的的载荷大于 0.5,该主成分是这两个指

标的反映。由指标含义可知,搜索指数、热词趋势都是电影本身的影响力; Z_3 中,只有 X_1 的载荷量大于 0.5,所以该主成分是 X_1 的反映。由指标含义可知该主成分是电影更新微博的程度,进行目的营销的频次,因此可以定义为活跃度成分。

表 4 旋转后的因子载荷矩阵

	元件		
	1	2	3
X_1	0.199	0.104	0.97
X_2	0.904	0.052	0.137
X_3	0.864	0.257	0.094
X_4	0.719	0.466	0.287
X_5	0.149	0.853	0.026
X_6	0.193	0.786	0.122

旋转方法:Kaiser 正规化。

表 5 主成分得分系数矩阵

	元件		
	1	2	3
X_1	-0.177	-0.084	1.044
X_2	0.547	-0.244	-0.108
X_3	0.471	-0.063	-0.169
X_4	0.257	0.126	0.072
X_5	-0.164	0.635	-0.107
X_6	-0.142	0.557	-0.002

旋转方法:Kaiser 正规化。

由表 5 得到主成分的得分方程:

$$Z_1 = -0.177X_1 + 0.547X_2 + 0.471X_3 + 0.257X_4 - 0.164X_5 - 0.141X_6 + \tag{2}$$

$$Z_2 = -0.084X_1 - 0.244X_2 - 0.063X_3 + 0.125X_4 + 0.635X_5 + 0.557X_6 + \tag{3}$$

$$Z_3 = 1.044X_1 - 0.108X_2 - 0.169X_3 + 0.072X_4 - 0.107X_5 - 0.002X_6 + \tag{4}$$

$$Z = \sum_{i=1}^3 \frac{\lambda_i}{\lambda_1 + \lambda_2 + \lambda_3} Z_i = 3.119 \times Z_1 / (3.119 + 0.990 + 0.779) + 0.990 \times Z_2 / (3.119 + 0.990 + 0.779) + 0.779 \times Z_3 / (3.119 + 0.990 + 0.779) = 0.638Z_1 + 0.203Z_2 + 0.159Z_3 \tag{5}$$

从(5)可以看出主成分 Z_1 的权重系数为 0.638, Z_2 的权重系数为 0.203, Z_3 的权重系数为 0.159。说明主成分 Z_1 对票房的影响是最大的,主成分 Z_1 中,包含了 X_2 、 X_3 、 X_4 ,这说明微博转发数、评论数和赞数对票房的影响是很大的,也证明了微博营销对电影宣传的影响是显著的。

2)回归方程。首先,将票房与主成分进行多元回归,采用逐步回归的方法,回归过程没有剔除任何主成分。从共线性诊断中发现,相关性系数最大 0.44,最小为 0.05,说明运用主成分回归消除了多重共线性。图 3 是标准化残差的直方图,从图上可直观的看出,标准化残差服从正态分布。

回归系数表中, Z_1 、 Z_2 、 Z_3 的显著性均为 0.000,常数项的显著性为 0.029,小于 0.1,各系数的显著性概率接近于 0,都在 0.01 的显著性水平下显著。根据 t 检验原则,在显著性水平为 0.1 的条件下,估计方程模型显著,拒绝原假设 H_0 ,各项系数都通过 t 检验。据此写出回归方程:

$$Y = -0.37 + 0.583Z_1 + 0.305Z_2 + 0.111Z_3 \tag{6}$$

用 SPSS 计算出票房的标准预测值,将票房的标准预测值和实际票房做散点图(如图 4)。从图 4 中

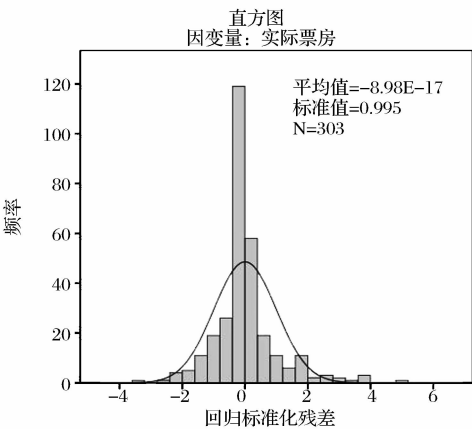


图 3 标准化残差的直方图

可以看出,票房的标准预测值与实际票房成一条直线,进一步进行两者的相关性分析,结果显示在 0.01 的显著性水平下,两者的相关系数为 0.7,说明模型的拟合度比较好。在一定程度上证明本文中选取的影响指标对电影票房的影响显著,口碑营销对电影票房有显著影响。

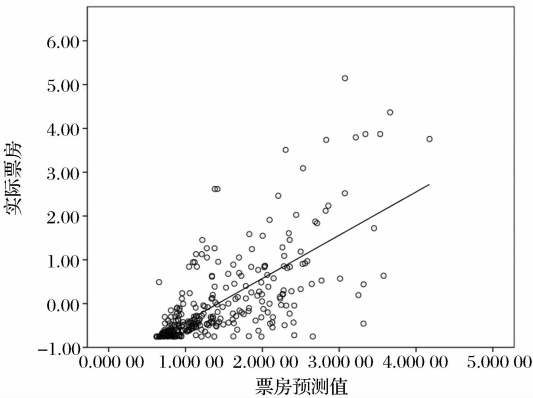


图 4 票房预测值与实际票房散点图

4 结语

本文在前者研究的基础上,突破了单一指标、单一平台,用更加方便的数据收集方法,考虑了数据变化的特性,收集电影上映前后每天的数据,用一段时间的数据提取出影响票房的主成分,并用回归方法建立了票房的回归模型。经实证分析,结果显示预测票房和实际票房变化一致,拟合程度较高,这说明本文所取的指标对票房的影响大且十分显著。在一定程度上证明微博营销有很高的商业价值。利用微博的特征数据分析影响电影票房,具有现实意义,这也与前人的研究相一致,并且本文研究的指标可以用到其他行业的营销效果的预测上。

参考文献

[1] CHEN YUBO,FAY S,WANG QI. Marketing implications of online consumer product reviews[J]. Information Technology and Management,2003,9(3):201—214.

[2] MISHNE G, GLANCE N. Predicting movie sales from blogger sentiment[C]//AAAI 2006 Spring Symposium on Computational Approaches to Analysing Weblogs. Menlo Park, CA: AAAI Press, 2006: 155—158

[3] SITARAM A, HUBERMAN B A. Predicting the future with social media[C]//2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT). IEEE Computer Society, 2010: 492—499.

[4] 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]. 华东师范大学学报:自然科学版,2015,2(49):66—72.

[5] 郑亚琴,樊鹏. 社会化媒体商业价值研究——基于票房数据的分析[J]. 吉林工商学院学报,2012,28(3):60—63.

[6] 王文文,周澍民. 社会化媒体对电影票房的预测价值研究[J]. 新闻传播,2013(12):254—255.

[7] 张玉松,张鑫. 票房的影响因素分析[J]. 经济论坛,2009(8):130—132.

[8] 周明升,韩冬梅. 基于社交媒体用户评论和关注度的电影票房预测模型[J]. 技术与方法,2014(33):73—75.

[9] 胡卫夕,宋逸. 微博营销[M]. 北京:机械工业出版社,2011,7—13.

[10] 汪冬华. 多元统计分析与 SPSS 应用[M]. 上海:华东理工大学出版社,2010:45—46.

The Empirical Research on the Influence of Micro-blog Marketing on the Movie Box Office

CHEN Li, LI Lin

(University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: With micro-blog has gradually become the main platform for film promotion, the study of the relationship between micro-blog's data and box office is particularly important. By using the method of empirical research, based on the 3 platforms to collect a week before the release of the film and the release of the data during the period of each day. Making a correlation analysis of these factors and box office, and finding that each factor has a positive correlation with the box office. The regression was establish by Principal component regression analysis, and using data collected to validate it. The result shows that the model has good predictive performance and it can provide more practical value for the marketing strategy of the film producer.

Key words: micro-blog marketing; box office; empirical research; principal component regression model

(上接第 19 页)

[4] 李孟学. 合肥市产业结构演变的城市化响应分析[D]. 蚌埠:安徽财经大学,2014.

[5] 杨波. 产业结构变迁与城市化进程研究[D]. 上海:华东师范大学,2005.

[6] 李世杰,姚天祥,杨文新. 试论产业结构演变与城市化的关系——以兰州市为例[J]. 地域研究与开发,2004(3):32—36.

Study on the Relationship between Industrial Structure Evolution and Urbanization in Shangluo City

WANG GAO-jian, WANG Ting

(Shangluo University, Shangluo Shanxi 726000, China)

Abstract: Based on the data of urbanization and industrial structure development in Shangluo from 1979 to 2015, this paper studies the development and interaction between industrial structure evolution and urbanization process by using the SPSS software analysis method, and obtains the correlation between them. Impact mode. The results show that the industrial structure of Shangluo is irrational, the development speed of urbanization is slow, and the development of industrial structure and urbanization is not in harmony. On this basis, the author puts forward some countermeasures and suggestions on the problems existing in the industrial structure and urbanization.

Key words: industrial structure; urbanization; coordinated development; Shangluo