

Predicting Commodity Sales Based on Big Data of Online Public Opinion

Jiayi Li, Xuan Ao, Shiyu Zhou, Jiaxun Dong, Zengrui Zhang, Kang Zhao

School of Mathematics and Statistics, Changsha University of Science and Technology, Changsha Hunan
Email: zhaokangmath@126.com

Received: Oct. 5th, 2019; accepted: Oct. 22nd, 2019; published: Oct. 29th, 2019

Abstract

In this paper, we firstly divide the Internet Public Opinion into five categories: adrift public opinion, customer evaluation of public opinion, economic consensus, social public opinion and cultural opinion. Then, we screen out the commodities which are directly affected by the above five kinds of public opinions. According to the micro blog index of public opinions, we determine the burst time of public opinion and get the commercial taobao index at that time. By comparing the significance of different curves which is obtained by SPSS, we can confirm the relationship between themicroblog index of public opinions and commercial taobao index with the specific formula expression. Finally, we get the relationship between the public opinion of the 88 shopping festival and the sales of four categories of common goods: women's clothing products, daily necessities, shoes and clothing accessories.

Keywords

Classification of Internet Public Opinion, SPSS, Regression Analysis, Significance Analysis

基于网络舆情大数据预测商品销量

李嘉依, 敖璇, 周诗雨, 董佳勋, 张增瑞, 赵康

长沙理工大学数学与统计学院, 湖南 长沙
Email: zhaokangmath@126.com

收稿日期: 2019年10月5日; 录用日期: 2019年10月22日; 发布日期: 2019年10月29日

摘要

本文对网络舆情分为5大类: 从众舆论、客户评论舆论、经济舆论、社会舆论以及文化舆论, 根据不同类型舆论选取其直接影响的商品。根据舆论微博指数确定舆论突发时间, 再获取相同时间的商品淘宝指

数,运用SPSS曲线回归,通过比较不同曲线的显著性确定舆论微博指数与商品淘宝指数间的关系及具体关系式。最后,我们得到88购物节舆论与女装、日用品、鞋、服饰配件四大类常见商品销量的关系。

关键词

网络舆情分类, SPSS, 回归分析, 显著性

Copyright © 2019 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着计算机技术的日新月异和互联网的快速普及,自媒体的发展可谓乘风破浪。如今微博、微信等为代表的自媒体公众平台已无孔不入,进一步推动了互联网作为网民表达利益诉求和社会关切的重要渠道。“人人都有麦克风”催化网络舆情的生成发酵,自媒体网络舆情时代悄然将至。

2016年1月22日,中国互联网络信息中心,在京发布的第37次《中国互联网络发展状况统计报告》显示,截至2017年12月,我国网民规模达6.88亿,互联网普及率为50.3%,超过半数中国人已接触互联网。同时,移动互联网正在塑造一个全新的社会生活形态,随着“互联网+”的提出,互联网对社会发展的影响空前繁荣。在我们生活中的方方面面都可看到互联网的影子,它给人类带来的进步不容置疑。与此同时,社会突发事件通过互联网传播进而产生衍生影响的情况数不胜数。

我国目前还处于对突发网络舆情研究的初级阶段,研究理论体系、网民意识素质、立法制度等方面还落后于发达国家水平。但基于我国社会阶级层次、网民知识结构、社会稳定程度以及庞大的人口基数等方面的因素,我国的网络舆情场所较之发达国家有过之而无不及,且造成的衍生影响及网络舆论的影响力同样不可小觑。

当突发事件发生时,政府和传媒公信力的缺失,网民参与意识的不断提高,再加上自媒体亲民、开放的特点及碎片化的信息传播方式,都能推动一件鸡毛蒜皮的小事演化成为全国人民茶余饭后的讨论话题。同时,如今的网上购物可谓风生水起,各种购物网站应运而生,购物形式也是丰富多彩。网购已成为人们必不可少的一种生活方式,很多突发网络舆情都会对某些商品的销售情况有影响[1]。因此我们从实际问题出发,通过分析突发网络舆情,研究网络舆情与商品销量间的关系,方便商家及时掌握消费者的需求。

2. 网络舆情分类

2.1. 从众舆情(从众、网红、明星同款)

消费者通过网络渠道,获取到他人对于产品的评价、购买意愿或购买行为的信息以及体验感分享之后,改变了自己的产品评价、购买意愿或购买行为,努力与他人保持一致或跟随大众消费产生想要尝试此类产品的心理与现象。比如:某明星代言了某种化妆品或某一网络舆情的衣服,当这种消费行为被认可,评价较高时,其他客户就会效仿。网络舆情的产品,从而引发了一种流行趋势[1]。吴亦凡代言 Burberry 品牌的销售量就有了很大的变化。在2016年上半年集团利润还低至7200万英镑,自代言起统计到2016年12月31日止,销售收入就达到7.35亿英镑,同比增长22%。

2.2. 客户评论舆情(评价, 好评、差评)

发掘引领潜在客户。通过在线客户评论提供导向性信息, 以网络舆情的形式由以前的客户向当前的客户提供建议。在线客户评论是指客户在购物网站或者其他评论网站、论坛对某种产品或服务发表的正面或负面的看法, 可以根据自己的亲身体验也可以是他人的经历。例如, 在购买百乐 p500 时, 淘宝内多家店面均有售卖此款类型的碳素笔, 买家可能会根据好评率、店铺评价、月销量决定在哪一个店铺购买。

在线客户评论的质量正在影响消费者的购买意向[1]。通过分析发现, 如果评论的内容与产品密切相关、评论的内容真实可靠、评论观点比较中立, 并且评论包含了大量有用的信息, 这样的评论对于消费者的购买意向的影响越大。也就是说, 消费者更倾向于阅读那些高质量的评论, 即客观、具体、逻辑性强且能够基于产品的具体特征给出推荐原因的评论。

2.3. 经济舆情(性价比、假日折扣)

客户购买一类心仪的物品时, 搜索相关的关键词, 淘宝通过数据分析后, 向客户推荐搜索相关性高的商品。消费者根据商品价格、规格、优惠力度等条件对商铺进行筛选, 而相加比高、优惠力度大的商品往往具有更大的吸引力, 例如: 双十一、双十二、6.18 等各大类折扣节日以及办理会员积分购买商品将享受的独特优惠。

2.4. 社会舆情(气候、地域、政治、突发流感)

1) 根据不同地区的温度、湿度、天气质量情况, 在某一时间段推荐某类产品。如: 高原地区紫外线辐射强, 对于防晒等产品有需求; 多雨气候湿润的地区雨伞、内衣裤等商品的需求量更大; 北京的冬天由于雾霾会使用 3 m 口罩; 南方地区在梅雨季节对于内衣等用品量的需求增加; 在三亚或沿海地区推荐海上用品等等。

2) 由于政治因素, 一些恐怖主义极端分子存在着分裂国家的错误思想。这些极端分子制造混乱, 令市民处于水深火热之中。例如: 港独事件发生之后, 港独分子在香港大街上制造混乱, 严重影响市民的正常生活, 昔日繁华的街道现在几乎没有人逛街购物, 对商家的影响是致命的。

3) 猪流感在国内爆发时, 猪肉相关产品价格攀升, 消毒物品的销售量一度高攀淘宝榜首。

2.5. 文化舆情(文化节日、旅游出行)

庆祝文化节日, 需要购买一些相应的礼物或是特色食品。比如: 端午节期间, 对于粽子、咸蛋黄等食品购买量将增加。旅游出行时, 对于日用品、收纳包、衣物以及防晒或御寒类产品的需求会增加。当到达特色旅游景点时, 游客对于旅游纪念品的购买量会增加。

3. 舆情数据与销售情况数据获取

本文采用 Python 语言编写数据获取程序, 数据来源为第三方(idataapi.cn)微博和小红书 API (Application Programming Interface)接口。程序采用通用的第三方库 request 库连接 API 接口。Request 是一个采用 Apache2 Licensed 网络开源协议的 HTTP 库、基于 urllib 的 Python 库。

4. 问题分析

根据分析, 我们发现商品的销售量与价格、信誉、服务质量和客户评论有很大关系, 因此通常情况下, 并不会受网络舆论的很大的影响。但是, 当一个网络舆论的讨论量突然上升较多时, 则会对某些商品的销售情况产生较大影响, 例如, 最近关于电影哪吒之魔童降世的舆论话题讨论量突然上升, 相应的哪吒发箍、哪吒发夹等商品的销售情况都出现较大涨幅, 其中哪吒发箍的搜索趋势增长 373%, 搜索指数也是其他发饰商品的几倍, 哪吒发箍和哪吒发夹分别占据发饰、头饰商品类新词搜索榜第 2、4 位(见图 1)。

发饰、头饰 上升榜			发饰、头饰 新词榜		
关键词	搜索趋势	搜索指数	关键词	搜索指数	全站商品数
11 欧美发箍	↑391%	134	1 婚纱 纱菲凡品婚纱	878	18
12 哪吒发箍	↑373%	599	2 哪吒发箍	599	49
13 淘宝 赠品	↑371%	73	3 大肠发圈 韩秀	340	23
14 双层头纱	↑364%	10	4 哪吒发夹	171	37
15 玫瑰花发夹	↑340%	51	5 iu同款发夹德鲁纳酒店	150	40
16 金属发夹 欧美	↑340%	66	6 水果糖果色简约一字夹	123	44
17 头饰发圈	↑309%	44	7 韩版新款布艺发饰发箍	123	11
18 秀禾服头饰	↑303%	145	8 scrunchies	73	39
19 大抓夹	↑300%	80	9 hair bling	66	20
20 彩虹头带	↑300%	44	10 hair pins	66	32

Figure 1. Nezha hair ornaments micro-blog index

图 1. 哪吒发饰微博指数

本文主要研究突发舆论对其对应商品的销售情况的影响。首先定义，当一个舆论当天的讨论量较前一天相比增长 120%以上时为突发舆论，且舆论突发的时间段为从增长超过 120%的日期到与讨论量最高日期相比下降超过 60%的日期，即我们所讨论的是商品在舆论突发时间段的销售情况。

我们用微博讨论微指数(以下简称微博指数)来表示舆论的讨论情况，用淘宝指数[2]来表示商品的销售情况(淘宝指数指根据淘宝集市和天猫里所有行业的成交量计算而成的一个综合素质，指数越高表示淘宝市场的采购量越多，销售情况越好)。

经过对过去一段时间商品销售情况的观察，大部分商品的销量都会在网络购物节出现大幅度上升，所以我们将网络购物节作为影响商品的一类舆论来分析其对不同类商品销售情况的影响，其他时间段，则分别考虑不同舆论对不同商品销售情况的影响。

对于网络购物节舆论，首先收集网络购物节舆论微博数据和网络上受购物节影响的商品大类的淘宝指数，进而确定舆论突发时间段。对突发时间段的微博指数与淘宝指数用 SPSS 作曲线回归，通过比较不同曲线的显著性来确定购物节舆论对不同类商品的曲线关系与具体关系式。

对于其他时间段的网络舆论，首先根据收集的数据通过折线图来确定某一个舆论的突发时间段和其直接影响的商品，收集其影响商品在突发时间段的淘宝指数。通过观察估计其之间可能的曲线关系，运用 SPSS 软件对这几种可能的关系进行曲线回归，分析不同曲线的回归结果。最后，通过比较显著性来确定此舆论微博指数与对应商品淘宝指数的关系并得到最终关系式。

5. 模型建立

首先利用微博指数和淘宝指数绘制对应事件的关系折线图，得到网络舆情突发变化时间段并分析微博指数和淘宝指数之间的相关性。然后，通过建立曲线回归分析模型，以淘宝指数做因变量，微博指数为变量，分别对其进行线性、二次以及指数曲线的回归分析[3] [4]，运用 SPSS 软件进行曲线回归，判断得到舆论关键词与对应主要影响商品之间的数据关系，分析结果中的参数估计量进行总结概括。

5.1. 曲线回归分析模型

5.1.1. 一次线性回归模型

$$y = \beta_0 + \beta_1 \cdot x + \varepsilon$$

其中, β_0, β_1 为回归系数, ε 为随机误差项。假设 $\varepsilon \sim N(0, \sigma)$, 则随机变量 $y \sim N(\beta_0 + \beta_1 x, \sigma)$ 。对 y 和 x 进行了 n 次独立观测, 得到 n 对观测值 (y_i, x_i) , 均满足 $y_i = \beta_0 + \beta_1 \cdot x_i + \varepsilon_i$ 。

5.1.2. 二次线性回归模型

$$y = \beta_0 + \beta_1 \cdot x + \beta_2 \cdot x^2 + \varepsilon$$

其中, $\beta_0, \beta_1, \beta_2$ 为回归系数, ε 为随机误差项。假设 $\varepsilon \sim N(0, \sigma)$, 则随机变量 $y \sim N(\beta_0 + \beta_1 x + \beta_2 x^2, \sigma)$ 。对 y 和 x 进行 n 次独立观测, 得到 n 对观测值 (y_i, x_i) , 均满足 $y_i = \beta_0 + \beta_1 \cdot x_i + \beta_2 \cdot x_i^2 + \varepsilon_i$ 。

5.1.3. 指数回归模型

$$y = \beta_0 \cdot e^{\beta_1 x} + \varepsilon$$

其中, β_0, β_1 为回归系数, ε 为随机误差项。假设 $\varepsilon \sim N(0, \sigma)$, 则随机变量 $y \sim N(\beta_0 e^{\beta_1 x}, \sigma)$ 。对 y 和 x 进行 n 次独立观测, 得到 n 对观测值 (y_i, x_i) , 均满足 $y = \beta_0 \cdot e^{\beta_1 x} + \varepsilon$ 。

5.2. 最小二乘法估计

5.2.1. 一次线性回归模型

用最小二乘法估计 β_0, β_1 的值, 即取 β_0, β_1 的一组估计值 $\hat{\beta}_0, \hat{\beta}_1$, 使 y_i 与 \hat{y}_i 的误差平方和达到最小。记 $Q(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2$, 则

$$Q(\hat{\beta}_0, \hat{\beta}_1) = \min_{\beta_0, \beta_1} Q(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

$Q(\beta_0, \beta_1)$ 关于 β_0, β_1 可微, 则有多元函数存在极值的必要条件得:

$$\frac{\partial Q}{\partial \beta_0} = -2 \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) = 0, \quad \frac{\partial Q}{\partial \beta_1} = -2 \sum_{i=1}^n x_i (y_i - \beta_0 - \beta_1 x_i) = 0$$

求解后得到:

$$\begin{cases} \hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \end{cases}$$

称 $\hat{\beta}_0, \hat{\beta}_1$ 分别为 β_0, β_1 的最小二乘法估计[5], 其中, \bar{x}, \bar{y} 分别为 x_i 与 y_i 的样本均值, 即

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i.$$

5.2.2. 二次非线性回归模型

取 $\beta_0, \beta_1, \beta_2$ 的一组估计值 $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$, 使得 y_i 与 \hat{y}_i 的误差平方和达到最小。

利用换元法, 令 $z_i = x_i^2$, 二次非线性方程转化为一次线性方程 $y = \beta_0 + \beta_1 x_i + \beta_2 z_i + \varepsilon$ 。再利用一次线性回归模型进行求解, 得到的 $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$ 为 $\beta_0, \beta_1, \beta_2$ 的最小二乘法估计[5]。

5.2.3. 指数回归模型

取 β_0, β_1 的一组估计值 $\hat{\beta}_0, \hat{\beta}_1$, 使得 y_i 与 \hat{y}_i 的误差平方和达到最小。对 $y = \beta_0 \cdot e^{\beta_1 x} + \varepsilon$ 两边同时取

对数, 近似得到 $\ln y = \ln \beta_0 + \beta_1 x_i$ 。指数非线性方程转化为一次线性方程。利用一次线性回归模型进行求解, 得到的 $\hat{\beta}_0$, $\hat{\beta}_1$ 为 β_0 , β_1 的最小二乘法估计[4]。

5.3. 显著性检验

假设 $H_0: \beta_1 = 0$, $H_1: \beta_1 \neq 0$ 。若假设 H_0 成立, 则有 $y_i = \beta_0 + \varepsilon$ 。假设检验水平为 δ 。利用最小二乘法拟合得到: $\hat{\beta}_1 = 0$, $\hat{\beta}_0 = \bar{y}$, 故对任意 i 均有 $\hat{y}_i = \bar{y}$ 成立。

$$\begin{aligned} \text{令 } SSE &= \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad SST = \sum_{i=1}^n (y_i - \bar{y})^2, \\ SSR &= SST - SSE = \sum_{i=1}^n (y_i - \bar{y})^2 - \sum_{i=1}^n (y_i - \hat{y}_i)^2. \end{aligned}$$

则有 $SSE/\sigma^2 \sim \chi^2(n-2)$, $SSR/\sigma^2 \sim \chi^2(1)$, 且 SSE/σ^2 与 SST/σ^2 是独立的随机变量。构造 F 检验统计量:

$$F = \frac{SSR/1}{SSE/(n-2)} \sim F(1, n-2)。$$

若 $F \leq F_{\delta}(1, n-2)$, 则肯定假设 H_0 , 此时认为 β_1 显著为 0; 若 $F > F_{\delta}(1, n-2)$, 则否定假设 H_0 , 肯定假设 H_1 , 此时认为 β_1 显著不为 0。

6. 实际应用

6.1. 购物节舆论对商品销量的影响

关于经济舆情中购物节的情况, 虽然日用品, 服装, 鞋等物品受平常网络舆论影响不大, 主要受季节气候、生活需要和人的主观意识影响, 但是这些商品的销售情况都会在购物节日期间出现突发性的较大波动, 因此, 我们收集了 2019 年 88 购物节的微博指数和商品在该时间段的销量, 分析购物节对日用常见商品销量的影响, 当购物节的微博指数与商品销量有显著相关性时, 进一步利用 SPSS 软件进行曲线回归, 得到不同类商品与购物节舆情的函数关系。

6.1.1. 购物节对女装、日用品、鞋、服饰配件四大类常见商品销量的影响

将收集的购物节微博指数缩小 10 倍后与商品销量放入同一个折线图中, 得到图 2~图 5, 分别表示该时间段四大类商品销量和 88 购物节的微博指数。根据购物节舆情的特殊性, 可知一般情况下, 节日前一天商品销量会达到这段时间的最小值。由图 2~图 5 可知四大类商品在节日前一天(8 月 7 日)微博指数达到最大值, 8 月 8 日商品销量达到峰值, 代表销量有延后舆论一天的情况, 因此用 SPSS 进行曲线回归时, 将销量数据与前一天微博指数结合分析, 分析结果如下:

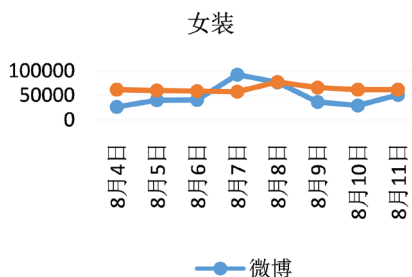


Figure 2. Micro-blogt index and Women's clothing sales during shopping

图 2. 购物节期间的微博指数和女装销量

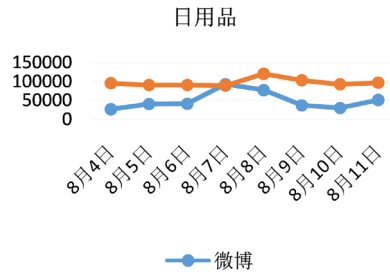


Figure 3. Micro-blog index and commodity sales during shopping
图 3. 购物节期间的微博指数和日用品销量

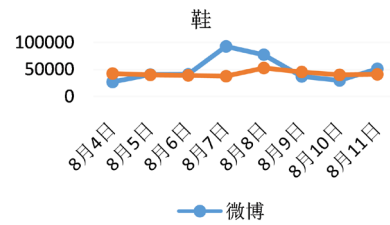


Figure 4. Micro-blog index and shoes sales during shopping
图 4. 购物节期间的微博指数和鞋的销量

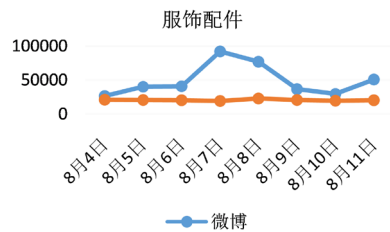


Figure 5. Micro-blog index and clothing accessories sales during shopping
图 5. 购物节期间的微博指数和服饰配件销量

1) 利用 SPSS 对女装数据进行曲线回归, 得到女装模型摘要和参数估计值, 如表 1 所示。根据表 1 显著性可知, 线性函数更适合用来描述女装和购物节微博指数的关系, 关系式为

$$y = 29.453x - 1312048.657$$

其中 y 表示微博指数, x 表示女装销售量。

2) 利用 SPSS 对日用品数据进行曲线回归, 得到日用品模型摘要和参数估计值如表 2 所示。根据表 2 显著性可知, 线性函数更适合用来描述日用品与微博指数的关系, 关系式为

$$y = 18.416x - 1239557.802$$

其中 y 为微博指数, x 为日用品销量。

Table 1. Model abstracts and parameter estimates of women's clothing
表 1. 女装模型摘要和参数估计值

方程	模型摘要					参数估计值		
	R 方	F	自由度 1	自由度 2	显著性	常量	b1	b2
线性	0.829	9.697	1	2	0.089	-1312048.657	29.453	
二次	0.843	2.685	2	1	0.396	-4462243.651	123.688	-0.001
指数	0.769	6.657	1	2	0.123	25415.742	4.752E-5	

Table 2. Model abstracts and parameter estimates of commodity
表 2. 日用品模型摘要和参数估计值

方程	模型摘要					参数估计值		
	R 方	F	自由度 1	自由度 2	显著性	常量	b1	b2
线性	0.892	16.501	1	2	0.056	-1239557.802	18.416	
二次	0.942	8.037	2	1	0.242	-7089715.103	131.882	-0.001
指数	0.840	10.505	1	2	0.083	27918.968	2.994E-5	

3) 利用 SPSS 对微博指数和鞋的销量进行曲线回归, 得到鞋的模型摘要和参数估计值如表 3 所示。根据表 3 显著性可知, 线性函数更适合用来描述鞋与微博指数的关系, 关系式为

$$y = 38.873x - 1076184.578$$

其中 y 为微博指数, x 为鞋的销量。

Table 3. Model abstracts and parameter estimates of shoes
表 3. 鞋的模型摘要和参数估计值

方程	模型摘要					参数估计值		
	R 方	F	自由度 1	自由度 2	显著性	常量	b1	b2
线性	0.902	18.484	1	2	0.050	-1076184.578	38.873	
二次	0.919	5.678	2	1	0.284	-3684560.517	156.812	-0.001
指数	0.854	11.661	1	2	0.076	36197.484	6.334E-5	

4) 利用 SPSS 对服饰配件进行曲线回归, 得到服饰配件模型摘要和参数估计值如表 4 所示。根据表 4 显著性可知, 线性函数更适合用来描述服饰配件与微博指数的关系, 关系式为

$$y = 155.826x - 2599801.576$$

其中 y 为微博指数, x 为服饰配件销量

Table 4. Model abstracts and parameter estimates of clothing accessories
表 4. 服饰配件模型摘要和参数估计值

方程	模型摘要					参数估计值		
	R 方	F	自由度 1	自由度 2	显著性	常量	b1	b2
线性	0.867	12.997	1	2	0.069	-2599801.576	155.826	
二次	0.940	7.860	2	1	0.245	-25060579.304	2289.889	-0.050
指数	0.811	8.563	1	2	0.100	3113.072	0.000	

6.1.2. 购物节对电脑销量的影响

将购物节微博指数缩小 10 倍后与电脑销量放入同一个折线图中, 如图 6 所示。从图中易看出数码电脑类商品与之前四种商品不同, 当购物节微博指数出现较大波动时, 电脑销量呈缓慢下降趋势, 可见购物节舆情对电脑销量无显著影响。

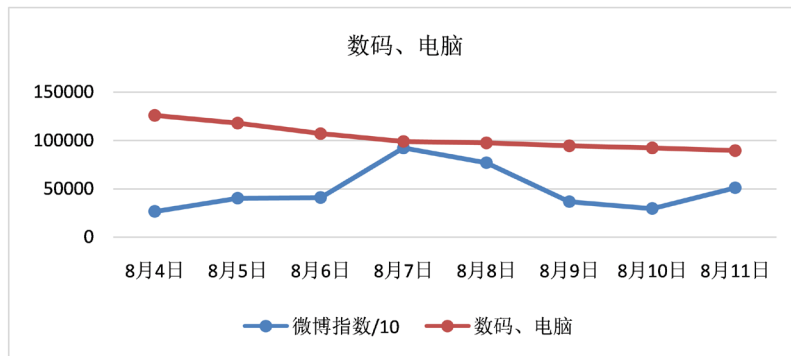


Figure 6. Micro-blog index and digital computers sales during shopping
图 6. 购物节期间的微博指数和数码电脑销量

6.2. 其他舆论对其对应商品销量的影响——以香港事件与炫雅风为例

6.2.1. 香港事件对国旗商品销量的影响

根据关键词“香港”的微博指数折线图(如图 7 所示),可以明显看出从 8 月 3 日起,微博指数才出现较大波动,所以我们选取 8 月 3 日~8 月 7 日和 8 月 11 日~8 月 13 日两个舆论突发时间段进行研究。将这两个时间段的微博指数缩小 1000 倍后与国旗销量放入同一个折线图中,得到图 8 和图 9,由图可以看出微博指数与国旗淘宝指数在时间段内呈现相同趋势,但是淘宝指数图的趋势比微博指数图趋势延后一天,说明销量有延后舆论一天的情况,因此用 SPSS 进行曲线回归时,将销量数据与前一天微博指数结合分析,分析结果如下:

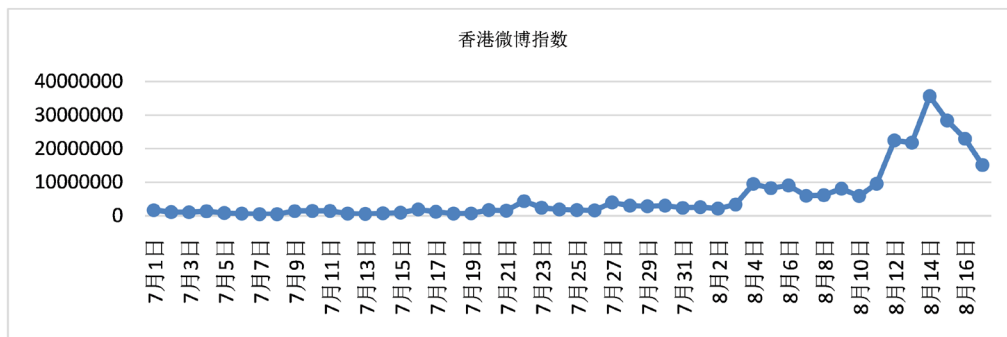


Figure 7. Hong Kong Micro-blog index
图 7. 香港微博指数

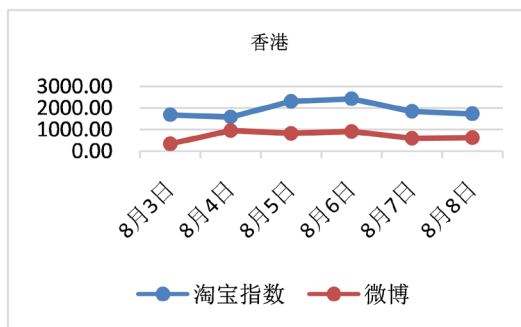


Figure 8. Micro-blog index and national flag sales from 8.3 to 8.8
图 8. 8.3~8.8 的微博指数和国旗销量

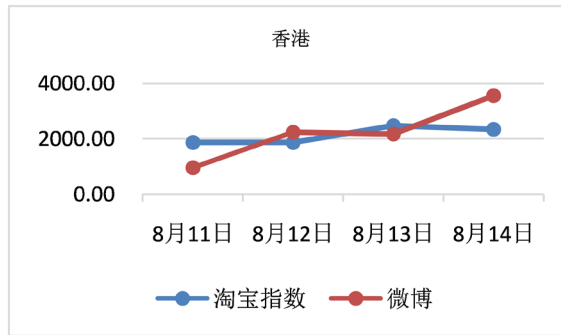


Figure 9. Micro-blog index and national flag sales from 8.11 to 8.14
图 9. 8.11~8.14 的微博指数和国旗销量

1) 利用 SPSS 对香港 8 月 3 日~8 月 7 日的微博数据和 8 月 4 日~8 月 8 日的淘宝数据进行曲线回归, 得到对应的模型摘要和参数估计值, 如表 5 所示。根据表 5 显著性可知, 指数函数更适合用来描述国旗和香港微博指数的关系, 关系式为(y 表示微博指数, x 表示国旗销售量):

$$y = 1269.483e^{5.779E-8x}$$

Table 5. Model abstracts and parameter estimates1 of national flag
表 5. 国旗模型摘要和参数估计值 1

方程	模型摘要					参数估计值		
	R 方	F	自由度 1	自由度 2	显著性	常量	b1	b2
线性	0.693	6.785	1	3	0.80	1142.030	0.000	
二次	0.758	3.129	2	3	0.242	2030.888	0.000	2.423E-11
指数	0.733	8.250	1	3	0.064	1269.483	5.779E-8	

2) 利用 SPSS 对香港 8 月 11 日~8 月 13 日的微博数据和 8 月 12 日~8 月 14 日的淘宝数据进行曲线回归, 得到对应的模型摘要和参数估计值, 如表 6 所示。根据表 6 显著性可知, 二次函数更适合用来描述国旗和香港微博指数的关系, 关系式为(y 表示微博指数, x 表示国旗销售量):

$$y = 1.239e - 11x^2 + 4095.114$$

Table 6. Model abstracts and parameter estimates2 of national flag
表 6. 国旗模型摘要和参数估计值 2

方程	模型摘要					参数估计值		
	R 方	F	自由度 1	自由度 2	显著性	常量	b1	b2
线性	0.973	35.524	1	2	0.106	1458.828	4.324E-5	
二次	1.00	0.000	2	0		4095.114	0.000	1.239E-11
指数	0.980	48.572	1	0	0.091	1544.063	2.021E-8	

6.2.2. 炫雅舆论对炫雅风发饰销售情况的影响

据关键词“炫雅”的微博指数数据(图 10 所示)可以看出, 除去网络购物节的时间, 在 7 月 29 日, 微博指数出现明显波动, 所以选取 7 月 29 日~7 月 23 日为炫雅舆论突发时间段进行研究, 将这个时间段的微博指数缩小 100 倍后与发饰销量放入同一个折线图中, 得到图 11, 由图可以看出微博指数与国旗淘

宝指数在时间段内呈现相同趋势且不存在销量延后舆论一天的情况。因此用 SPSS 进行曲线回归时，将销量数据与同期微博指数结合分析，分析结果如下：

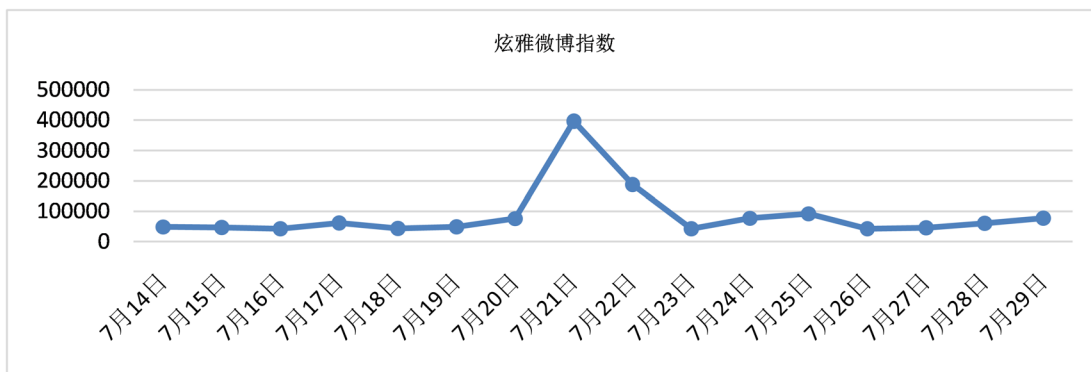


Figure 10. Micro-blog index of XuanYa
图 10. 炫雅微博指数

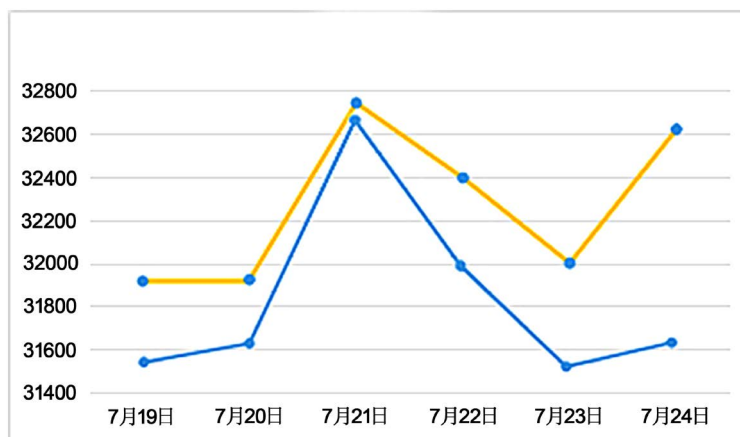


Figure 11. Micro-blog index and headwear sales from 7.19 to 7.24
图 11. 7.19~7.24 微博指数与头饰的销量

利用 SPSS 对 7 月 29 日~7 月 23 日炫雅的微博数据和发饰的淘宝数据进行曲线回归，得到对应的模型摘要和参数估计值，如表 7 所示。根据表 7 显著性可知，线性函数更适合用来描述发饰和炫雅微博指数的关系，关系式为(y 表示微博指数， x 表示发饰销售量)

$$y = 0.002x + 31863.156$$

Table 7. Model abstracts and parameter estimates2 of XuanYa headwear
表 7. 炫雅风发饰模型摘要和参数估计值

方程	模型摘要					参数估计值		
	R 方	F	自由度 1	自由度 2	显著性	常量	b1	b2
线性	0.940	31.515	1	2	0.030	31863.156	0.002	
二次	0.954	10.455	2	1	0.214	31771.552	0.004	-3.1555E-9
指数	0.939	30.773	1	2	0.031	31865.058	7.104E-8	

7. 结论

本文主要研究微博舆论指数与商品淘宝指数间的关系,并通过此关系对商品在突发舆情下的销量进行预测。我们将舆论分为从众舆论、客户评论舆论、经济舆论、社会舆论以及文化舆论。通过建立两组指数的回归模型得到两组数据间的关系,得到购物节微博舆论指数与女装、日用品、鞋、服饰配件四大类常见商品销量淘宝指数呈线性关系,对数码、电脑销量无显著影响;香港事件微博指数与国旗销量的淘宝指数呈指数和二次函数关系,炫雅舆论对炫雅风商品销量的淘宝指数呈线性关系。因此,商家可以利用此关系,在舆论突发时,对自家商品的销量进行预测,适当地进行进货补货措施,防止供不应求和货物堆积的现象发生。

基金项目

湖南省大学生创新创业训练计划项目(No. S201910536032)。

参考文献

- [1] 时文静. 影响网络购物商品销量的因素研究——基于 SEM 的淘宝网数据分析[C]//中国统计教育学会. 2015 年(第四届)全国大学生统计建模大赛论文. 北京:中国统计教育学会, 2015: 19.
- [2] 阿里指数查询网站. <http://index.1688.com/alizs/home.htm?spm=a262ha.8884014.0.0.PH9Mml>
- [3] 姜启源, 谢金星, 叶俊. 数学模型(第三版)[M]. 北京: 高等教育出版社, 2005: 1-202.
- [4] 谷恒明, 胡良平. 简单曲线回归分析及其应用[J]. 四川精神卫生, 2017, 30(6): 498-502.
- [5] 王娟. 基于最小二乘法的气温曲线回归模型分析[J]. 信息技术, 2016(2): 80-82.