

文章编号: 2095-2163(2019)05-0338-03

中图分类号: TP311.13

文献标志码: A

基于数据挖掘对电商促销活动的销量预测研究

刘 帅, 李相如, 赵润方, 徐 旭, 张文杰, 李艳娟

(东北林业大学 信息与计算机工程学院, 哈尔滨 150040)

摘 要: 通过研究以往电商促销活动基本情况及历史数据, 运用数据挖掘与数据分析技术对某次促销活动销量进行预测和分析, 针对淘宝店是否参加这次电商促销活动的决策问题, 基于整体销量的角度进行决策建模分析。经过对销量的预测分析, 可得到促销活动对销量增长的促进作用是否可以达到商家的要求, 是否可以帮助商家盈利或者可以获得比不参加促销活动更大的利润, 从而帮助商家做出正确的销售决策。

关键词: 电商促销活动; 销售效应; 数据挖掘

Sales forecast of electric business promotion activities based on data mining

LIU Shuai, LI Xiangru, ZHAO Runfang, XU Xu, ZHANG Wenjie, LI Yanjuan

(College of Information and Computer Engineering, Northeast Forestry University, Harbin 150040, China)

[Abstract] By studying the basic situation and historical data of previous e-commerce promotion activities, this paper forecasts and analyzes the sales volume of a promotion activity using data mining and data analysis method, and conducts decision-making modeling analysis from the perspective of overall sales volume for the decision-making problem of whether taobao stores participate in this e-commerce promotion activity. After the above forecast analysis on sales, it is feasible to learn whether the activities' promotion of sales growth can meet the requirements of the merchants, whether participating in the activities could help the business to make a profit or get a bigger profit than not to participate in the activities. Therefore, the research is helpful for merchants to make the right sales decisions.

[Key words] e-commerce promotional activities; sales effect; data mining

0 引 言

随着互联网大数据的发展, 基于数据挖掘的相关研究在电子商务中的应用正日渐趋于广泛, 一方面呈现出高速增长态势, 另一方面可以预见的是其未来的研究前景也必然更加广阔。因而近年来, 已经受到了业界众多学者的高度重视与关注。时下, 电子商务网站的交易量以及越来越多的使用人数也都在体现着电子商务的强大行业引领能力。毋庸置疑, 其优越性是显而易见的。电子商务可使人们的生活购物变得便捷, 可以使人们足不出户就能买到千里之外的商品; 可使人们无需跑路就能买到物美价廉的商品。然而站在电商平台以及卖家的立场来说, 就需要考虑顾客的消费心理, 即如何才能卖出更多的商品, 这就需要精准定位消费者的购物意向。现如今国内外的电商平台都已经启动了对消费者心理的精准预测。比如国内的阿里巴巴, 国外的亚马逊平台, 就都会记录下每天在其平台上的用户消费

记录, 凭此来预测顾客的消费习惯, 同时有针对性地推出一些促销活动, 将用户潜在所需商品精准地推荐给目标客户, 用户也会省下搜寻各类中意商品的时间, 可谓是一种双赢之举。但仍有研究指出, 目前的各大商家竞争日益激烈, 从而催生出种类多样、良莠不齐的促销活动。而在有些促销活动中, 活动举办商家并不能取得预期的利润, 甚至还会导致亏损。基于此, 如果能够预先知道该类促销活动是否会盈利, 或者是否可以获得比不参加活动更大的利润, 那么商家就可以利用这一点来判断是否开展该类促销活动。本次研究就是利用数据挖掘和朴素贝叶斯分类算法来进行销量预测, 从而帮助商家尽可能获取最大化利益。对此, 拟做研究论述如下。

1 研究方法

研究发现, 吸引客户和商家注意力的各类指标中, 首推的 2 项就是商品评价和销售量。因此, 本文通过结合活动期间以及活动期外这 2 个方面的效应

作者简介: 刘 帅(1996-), 男, 本科生, 主要研究方向: 数据挖掘分析; 李相如(1996-), 男, 本科生, 主要研究方向: 数据挖掘分析; 赵润方(1996-), 男, 本科生, 主要研究方向: 数据挖掘分析; 徐 旭(1997-), 男, 本科生, 主要研究方向: 数据挖掘分析; 张文杰(1996-), 女, 本科生, 主要研究方向: 数据挖掘分析; 李艳娟(1979-), 女, 博士, 副教授, 主要研究方向: 机器学习、生物信息学。

收稿日期: 2019-04-24

来进行分析。研究选用销售量来分析销售效应,选用商品评价来提取客户对商品的好坏评价数量并加以分析。综上可得,整体效应分析模式详见表1。

表1 效应分析
Tab. 1 Effect analysis

	活动前后期	活动期间
销售	正/负效应	正/负效应
评价	正/负效应	正/负效应

这里,主要采用假设性分析方法,提出如下假设:

- H1: 活动销量能够达到日常销量的 n 倍
 - H2: 活动销量不能够达到日常销量的 n 倍
- 其中, n 为商家设定的一个确定值。

2 数据准备

本文主要选取某淘宝店历史数据进行分析。利用网络爬虫来取得所需数据,考虑到活动预期通常情况下均为2天,且所能爬取到的数据只有月销量,所以本文将活动对应月份作为活动期进行标记处理,其它月份的数据标记为非活动区间,即日常期间。针对同一种类商品(衣服)的同一季度的不同品牌,绘出销量对比图,即如图1所示。同时,研究中还以评价和折扣作为影响销量的因素,因而将商品评价和商品折扣爬取下来,绘制后结果具体如图2、图3所示(图1~3中的数据均为所选取的示例数据)。

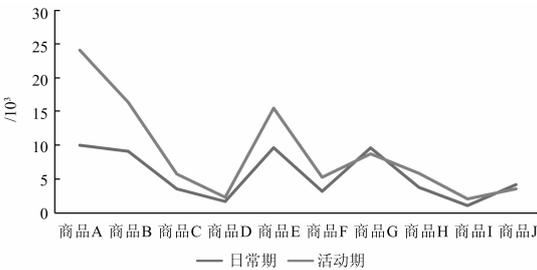


图1 销量对比
Fig. 1 Sales comparison

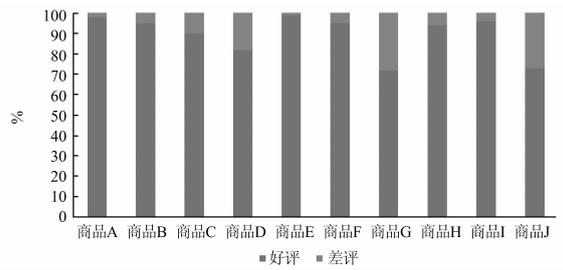


图2 评价对比
Fig. 2 Comparison of evaluation

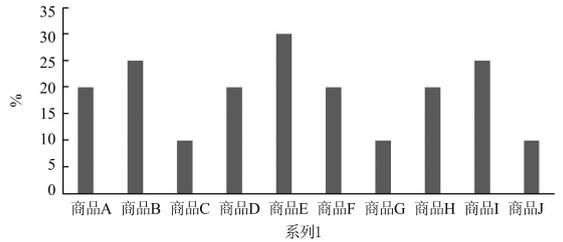


图3 折扣率
Fig. 3 Discount rates

3 数据分析

本次研究中,采用朴素贝叶斯算法作为分类算法。将评价和折扣率作为条件,活动期间销量和日常销量的比例作为结果,即结果为活动销量能够达到日常销量的 n 倍和日常销量的比例作为结果,即结果为活动销量能够达到日常销量的 n 倍和日常销量的比例。结合示例数据,条件为好评占比90%以上和折扣率10%以上;预测结果为活动销量能够达到日常销量的1.5倍和日常销量的1.5倍。

利用前文爬取到的数据作为训练样本,将示例的评价和折扣率数据作为输入参数,预测得出最终结果,再将其与实际情况进行对比。深入分析后可知,这是典型的二分类问题,依据朴素贝叶斯进行求解,其数学公式可表示为:

$$p(\text{类别} | \text{特征}) = \frac{p(\text{特征} | \text{类别})p(\text{类别})}{p(\text{特征})}, \quad (1)$$

在此基础上,结合特征个数,可知共有4种结果,这里只举一种为例。假设好评占比大于等于90%,折扣率大于10%,则转换为以下结果:

$$p(\text{活动销量能够达到日常的1.5倍} | \text{好评占比大于90\% 折扣率10\%以上}) = \frac{p(\text{好评占比大于90\% 折扣率10\%以上} | \text{活动销量能够达到日常的1.5倍})p(\text{活动销量能够达到日常的1.5倍})}{p(\text{好评占比大于90\% 折扣率10\%以上})}$$

(2)

示例数据经朴素贝叶斯运算求出的分类结果详见表2。

表2 预测结果与实际情况

Tab. 2 Predicted results and actual situation

商品	好评比例大于等于90%	折扣率大于10%	达到日常销量1.5倍	实际结果/倍
A	Yes	Yes	Yes	2.4
B	Yes	Yes	Yes	1.7
C	Yes	No	No	1.6
D	No	Yes	Yes	1.3
E	Yes	Yes	Yes	1.6
F	Yes	Yes	Yes	1.6
G	No	No	No	0.9
H	Yes	Yes	Yes	1.5
I	Yes	Yes	Yes	1.9
J	No	No	No	0.8

4 结束语

本文以双十一活动为例,基于某淘宝店历史数据对其促销活动的效应进行分析。先利用网络爬虫爬取相关数据,再使用朴素贝叶斯算法构建贝叶斯分类器。使用之前爬取的数据对分类器进行样本训练后,再用示例数据来进行验证。在实验过程中发现商品评价和折扣率对促销活动期间的商品销量有着重要影响,所以将这2个因素作为特征。经过示例数据检验和统计分析得出,经过样本训练后的朴素贝叶斯分类器,可以大概预测出活动销量是否可达到日常销量的 n 倍,这样就能够帮助商家拟定合理商业行为决策。此外,鉴于实际条件限制,本文主

要是针对某淘宝商家进行的小范围研究。后续在条件允许的情况下则可以立足于平台视角进行长期、且大范围的活动效应分析。

参考文献

- [1] 王玉珍. Web 数据挖掘的分析与探索[J]. 电脑开发与应用, 2003,16(4):73-74.
- [2] 朱建平. 数据挖掘的统计方法及实践[M]. 北京:中国统计出版社,2005.
- [3] 周绪倩. 基于电子商务的 Web 数据挖掘系统架构的研究[D]. 邯郸:河北工程大学,2010.
- [4] 毕方明,张永平. 数据挖掘技术研究[J]. 计算机工程与设计, 2004,25(12):2242-2244.
- [5] 范治军. 基于数据挖掘的入侵检测研究[D]. 大连:大连理工大学,2012.
- [6] 陈京民. 数据挖掘与电子商务的客户满意度[C]//2004 中国(福州)国际电子商务学术交流大会暨 2004 中国国际电子商务高级学术论坛、中国国际电子商务高等教育论坛、中国国际电子商务立法研究学术论坛、中国国际电子商务应用技术研究高级论坛. 福州:中国科学技术协会学会学术部,福建省科协, 2004:36-37.
- [7] 国刚,杨青. 基于数据挖掘的客户忠诚度分析[J]. 价值工程, 2013(6):140-143.
- [8] 谭磊. New Internet:大数据挖掘[M]. 北京:电子工业出版社, 2013.
- [9] 刘军,吕俊. 大数据时代及数据挖掘的应用[N]. 国家电网报, 2012-05-15(10).
- [10] LINDEN G, SMITH B, YORK J. Amazon recommendations: Item-to-item collaborative filtering[J]. IEEE Internet Computing, 2009,7(1):76-80.
- [11] 张华忠. 贝叶斯算法研究[J]. 数字技术与应用, 2013(11):102.
- [12] 时雷,虎晓红,席磊. 朴素贝叶斯分类算法及其应用研究[J]. 光盘技术, 2018(11):54.
- [13] 赵文涛,孟令军,赵好好,等. 朴素贝叶斯算法的改进与应用[J]. 测控技术, 2016(2):143-147.

(上接第 337 页)

- [2] BENDALE J, RATANARAJ K J. Review of different IP Geolocation methods and concepts[J]. International Journal of Computer Science & Information Technology, 2014,5(1):436.
- [3] IETF Geopriv Workgroup. Geopriv status pages[EB/OL]. [2012-10-24]. <http://tools.ietf.org/wg/geopriv/>.
- [4] PADMANABHAN V N, SUBRAMANIAN L. An investigation of geographic mapping techniques for internet hosts[C]// Proceedings of the ACM SIGCOMM Computer Communication Review. New York, USA: ACM, 2001:173-185.
- [5] HUFFAKER B, FOMENKOV M, CLAFFY K. DRoP: DNS-based router positioning[J]. ACM SIGCOMM CCR, 2014, 44(3):5-13.
- [6] MOORE D, PERIAKARUPPAN R, DONOHOE J, et al. Where in the world is netgeo. caida. org? [C]// Internet Society Conference. Yokohama, Japan: CAIDA, 2000.
- [7] CAIDA, Macroscopic Internet Topology Data Kit (ITDK)[EB/OL]. [2019-05-03]. <http://www.caida.org/data/internet-topology-data-kit/>.
- [8] GIOTSAS V, SMARAGDAKIS G, HUFFAKER B, et al. Mapping peering interconnections to a facility[C]// ACM SIGCOMM Conference on Emerging Networking Experiments and Technologies (Conext). Heidelberg, Germany: ACM, 2015:1-13.
- [9] MARDER A, SMITH J M. MAP-IT: Multipass accurate passive inferences from traceroute[C]// ACM on Internet Measurement Conference. Santa Monica, CA, USA: ACM, 2016:397-411.
- [10] LUCKIE M, DHAMDHARE A, HUFFAKER B, et al. bdrmap: Inference of borders between IP networks[C]// Proceedings of the 2016 Internet Measurement Conference. Santa Monica, California, USA: ACM, 2016:381-396.
- [11] 百度. 百度翻译开放平台[EB/OL]. [2019]. <http://api.fanyi.baidu.com/api/trans/product/index>.