

基于Apriori算法的国货彩妆产品在线评论数据关联分析

李颖

上海工程技术大学, 上海

收稿日期: 2022年7月11日; 录用日期: 2022年8月4日; 发布日期: 2022年8月12日

摘要

本文旨在分析以完美日记为代表的国货品牌的在线评论, 得出消费者重点关注的内容, 以及评价内容中可能存在的关联规则, 从而进一步推动国货彩妆品牌向国际品牌的发展。首先通过webscraper获取某网购平台上该品牌口红的相关评价数据。其次利用SPSS Modeler分析工具, 通过Apriori算法对整理好的数据进行关联分析, 得出研究结论。研究结果表明: 现阶段由于信息渠道增多, 电子商务提供的客服服务重要性下降; 产品的价格、品牌、视觉评价、外包装都会影响消费者的购买评论; 产品视觉评价受到多方面因素影响, 包括产品内外包装、品牌力、嗅觉评价、触觉评价、价格和触觉的综合评价。品牌商可以借鉴消费者在线评论中的关联内容, 通过改进相关产品属性或服务, 进一步提升消费者对产品视觉上的评价, 进而提升整体消费者满意度。

关键词

在线评论, 国货彩妆品牌, 关联分析, 网页爬虫, Apriori算法

Correlation Analysis of Online Review Data of Domestic Makeup Products Based on Apriori Algorithm

Ying Li

Shanghai University of Engineering Science, Shanghai

Received: Jul. 11th, 2022; accepted: Aug. 4th, 2022; published: Aug. 12th, 2022

Abstract

This article aims to analyze the online reviews of domestic brands represented by Perfect Diary,

draw the content that consumers pay attention to, and the possible correlation rules in the evaluation content, so as to further promote the development of domestic makeup brands to international brands. First of all, through webscraper, the relevant evaluation data of the brand lipstick on an online shopping platform was obtained. Secondly, using the SPSS Modeler analysis tool, the correlation analysis of the sorted data is carried out by the Apriori algorithm to draw the research conclusions. The results show that at this stage, due to the increase in information channels, the importance of customer service services provided by e-commerce has decreased; The price, brand, visual evaluation, and outer packaging of the product will all affect the consumer's purchase review; product visual evaluation is affected by many factors, including product internal and external packaging, brand power, olfactory evaluation, haptic evaluation, price and comprehensive evaluation of touch. Brands can learn from the relevant content in consumer online reviews to further enhance consumer visual evaluation of products by improving relevant product attributes or services, thereby improving overall consumer satisfaction.

Keywords

Online Reviews, Domestic Makeup Brands, Association Analysis, Webscraper, Apriori Algorithm

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

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



Open Access

1. 引言

进入 WTO 至今, 中国成为全球第二大美妆市场。一方面, 欧莱雅、雅诗兰黛等国际一线品牌纷纷进驻中国, 投资设厂、培育人才、开展创新, 不仅打开了巨大的中国市场, 也为中国美妆产业带来人才、技术、经验与要素资源的交流融合。随着电子商务的快速崛起, 一大批国货品牌快速发展壮大, 比如完美日记、花西子等。但是国货品牌与国际品牌在综合实力上仍然差距悬殊。国货品牌要想真正的实现品牌的壮大发展, 必须进一步接近消费者市场。电子商务背景下, 在线消费者评论成为免费的“市场调查员”。一方面, 在线评论可以帮助消费者确定最能满足其需要的产品[1], 另一方面, 也是企业进行市场调查、市场营销等管理活动的重要信息来源[2]。但是, 现阶段在线评论信息的数量急剧增加, 大大增加了数据分析的难度。数据挖掘技术的出现, 可以很好地解决信息庞大、杂乱等问题[3]。

2. 数据来源与研究方法

2.1. 数据来源

通过网页数据抓取工具 webscraper 分别获取淘宝和拼多多两个网购平台上关于完美日记销量最高的一款口红的最新 1000 条产品评价, 共 2000 条。在谷歌浏览器中安装 webscraper 插件之后, 选择选取数据的网页位置信息和文本数据类型, 自动抓取网页信息。为了减少无效数据筛选的工作量, 选择包含图片和文字的评论类型。由于选择了两个消费者常用的购物平台, 加上 webscraper 一次性只能获取 500 条数据, 因此, 共采集了四次数据。为了避免数据自动收集过程中出现重复的用户信息, 在两个平台上共选取同一款产品的四家销量最高的店铺。

2.2. 研究方法

本文旨在分析以完美日记为例的国货彩妆品牌的在线评论, 得出消费者重点关注的内容, 以及内容

中可能存在的关联规则。首先通过 webscraper 获取淘宝网和拼多多平台上完美日记口红的相关评价数据。其次,对文本数据进行预处理。一是,通过初步筛选,删除明显无效评论;二是预先定义产品评论的类别以便下一步数据转换;三是将文本数据转换为表格格式;四是利用 SPSS Statistics 对初步整理好的 Excel 数据的残缺值进行处理。接着,对预处理之后的数据进行描述性分析,对消费者的评价内容进行初步描述统计分析。最后,利用 SPSS Modeler 分析工具,通过 Apriori 算法对整理好的数据进行关联分析,得出研究结论。

3. 在线评价文本数据预处理

3.1. 初步处理

在线评论数据的初步处理包括 3 个步骤:首先,过滤系统默认评论。当买家未能及时做出评价时,网站会给出默认评论,如“评价方未及时做出评价,系统默认好评”、“买家未在 15 天内未做出评价”等或没有填写任何评论。这些评论不是买家给出的,因此需要对此类评论进行过滤。其次,过滤无效评论。有些评论只是标点符号或表达的内容没有实际意义,几乎无法提供有用的信息,且对评论文本的主题分析没有帮助,因此需要将其删除。最后,过滤重复评论。当买家一次购买多件同一商品时,经常会出现重复的评论。信息经过预处理后,发现 945 条无效评论,剩余 1055 条评论予以保留,以供下一步分析。

3.2. 预先定义产品评论的类别

通过对数据集进行分析,根据在线评论的文本内容,将相似特征或相关性较高的常用术语合并在一起,所有评论归入为触感、嗅觉、视觉、价格、外包装、品牌认可、客户服务、物流这八个类别。触感评价包括质地不错、产品很滋润等表达;嗅觉评价包括好闻、淡香等表达;视觉评价包括显白、颜色正、绝美等词语;价格包括划算、物美价廉、促销活动等;外包装包括产品包装高级、包装完好等;品牌认可主要包括会回购、推荐购买、送朋友、复购等表达;客户服务指客户耐心等方面;物流,比如收货快、顺丰包邮等。

3.3. 预先转换文本数据

本文先在 Excel 中录入数据,将收集的文本信息转换为 SPSS 能处理的数据类型。在建认了这些主题类别之后,对于所搜集到的每条评论,确定该特定评论是否包括与相应主题类别高度相关的单词。具体来说,可以将客户评论由句子分解为一组分句(分解),每个分句又可以被分到 8 个类别之中。通常以标点符号和连词(如“和”“与”等)作为分句的边界。如果某条评论有与某个特定类别高度相关的单词,则在该主题类别下录入“1”,代表某个消费者比较关注产品该类别的属性;如果没有,则将该评论在相应的主题类别中表示为“0”,代表某个消费者不关注该属性。

3.4. 残缺值和异常值处理

将 Excel 数据导入 SPSS Statistics 中进行描述分析,结果表明数据范围均为 0~1,不存在异常值,但有效数据小于 1055 条。在对原数据进行检查发现有几处缺失,结合初始文本资料,补全残缺值。再次进行描述分析,有效数据等于 1055 条,不存在异常值和残缺值。

4. 描述性统计分析

根据统计资料,在触觉评价方面,消费者对触觉上使用体验的关注度为 51.2%,略高于不关注度。在嗅觉评价方面,消费者不关注度为 65.6%,即消费者对于彩妆产品的气味属性的关注度较低。在视觉评价上,消费者关注度高达 93.1%,即消费者购买该产品的目的主要是为了体现自我形象在视觉上的美化。在价格评价内容方面,消费者较少提及对价格的偏好,这可由于于消费者默认国货产品性价比较高,

因此较少提及价格内容。但仍有超过四分之一的消费者提及对价格的关注。品牌相关评价内容的关注度和不关注度几乎持平,说明消费者的整体国货品牌意识较强。在频数统计中,发现客户服务的关注度仅为 6.6%。这可能于网络信息渠道扩宽,消费者对于客服的需求降低有关。物流作为独立于产品生产商的第三方企业提供的商业辅助服务,对产品整体评价的影响力较低,消费者关注度较低。

5. 利用 Apriori 算法进行关联分析

根据淘宝网和拼多多平台用户评论的最新记录,整理最新的 1055 条有效评论的过程中,发现有效的评论中消费者的基本态度都为相对满意。这表示所收集的数据中消费者的评价内容都是较为满意的。分析这些评价内容之间可能存在的关联规则,并结合上一章的描述性统计分析结果,可以为企业提供一些改进建议,使得消费者进一步提升产品满意程度。

首先在 SPSS Modeler 构建区构建所需的关联分析模型(如图 1)。建模的第一步是导入数据。第二步将类型中每个变量的测量属性设置为标记,角色全部设置为任意。第三步通过过滤器将不需要的字段过滤。第四步用网络图展示不同测量项目之间存在的共现关系大小,连线深度标识共现的程度(如图 2)。第四步建模,选择 Apriori 算法进行关联分析,默认支持度 10%和置信度 80%,按照提升度排序,关联结果如表 1。

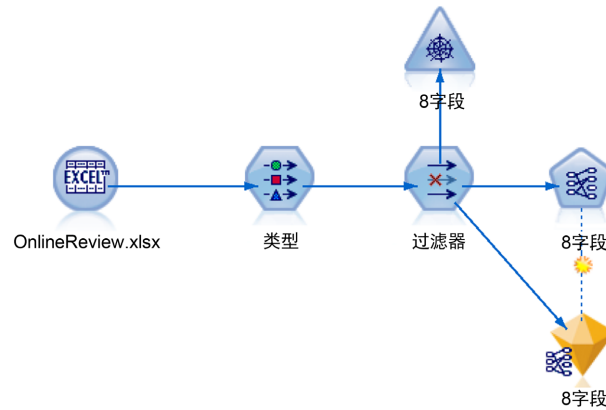


Figure 1. An associative relationship model for online comment content
图 1. 在线评论内容的关联关系模型

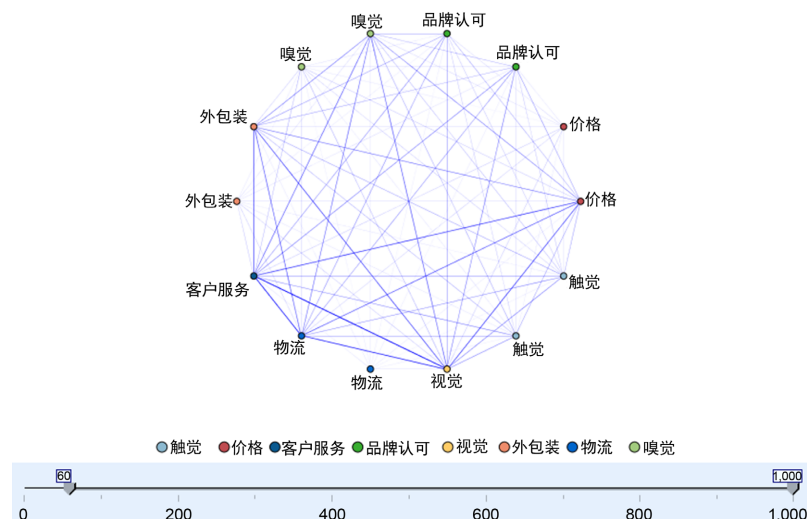


Figure 2. A co-occurrence network diagram of online review
图 2. 在线评论内容的共现网络图

Table 1. The results of the correlation analysis of the content of the online review
表 1. 在线评论内容的关联分析结果

后项	前项	实例	支持度(%)	置信度(%)	提升度
视觉	嗅觉	363	34.408	95.317	1.024
视觉	外包装、触觉	138	13.081	94.928	1.02
视觉	外包装、嗅觉	116	10.995	93.966	1.01
视觉	嗅觉、触觉	229	21.706	93.886	1.009
视觉	外包装	281	26.635	93.594	1.006
视觉	触觉	540	51.185	93.519	1.005
视觉	价格、触觉	150	14.218	93.333	1.003
视觉	嗅觉、品牌认可	147	13.934	93.197	1.001
视觉	价格	272	25.782	92.647	0.995
视觉	品牌认可、触觉	265	25.118	92.453	0.993
视觉	外包装、品牌认可	111	10.521	90.991	0.978
视觉	品牌认可	515	48.815	90.68	0.974
视觉	价格、品牌认可	155	14.692	89.032	0.957
视觉	物流	163	15.45	88.344	0.949

从结果可知,这八类主题经过四次迭代,可以构成 29 个频繁项集,分别是{嗅觉}、{外包装,触觉}、{外包装,嗅觉}、{嗅觉,触觉}、{外包装}、{触觉}、{价格、触觉}、{嗅觉,品牌认可}、{价格}、{品牌认可,触觉}、{外包装,品牌认可}、{品牌认可}、{价格,品牌认可}、{物流}、{视觉}以及视觉和上述其他 14 种项集组成的 14 种集合。在此基础上,进一步筛选出提升度大于 1 的强关联规则。结果包括八个强关联规则,分别是“{嗅觉}→{视觉}”、“{外包装,触觉}→{视觉}”、“{外包装,嗅觉}→{视觉}”、“{嗅觉,触觉}→{视觉}”、“{外包装}→{视觉}”、“{触觉}→{视觉}”、“{价格,触觉}→{视觉}”以及“{嗅觉,品牌认可}→{视觉}”。

6. 结果讨论和分析

6.1. 客户服务重要性下降

Apriori 算法通过四次迭代结束,没有出现包含客户服务的频繁项集,由此可知,客户服务并不是多数消费者进行购买行为前的活动。主要原因在于消费者通过其他渠道,比如小红书、直播间等[4],已经对该爆款口红有了充足的认知,无需进一步咨询产品信息。

6.2. 价格、品牌、视觉、外包装是消费者重点关注的内容

消费者选择国货口红的评价依据仍然主要围绕其价格、品牌、视觉和外包装等方面。由于产品本身是以国货为卖点,所以产品在品牌方面专注于做国货品牌,从而进一步吸引消费者购买。而国货品牌由于刚刚起步,和国际品牌或者其他成熟品牌无法直接竞争,多采取低成本战略,售价相对优惠[5]。由于产品的独特性,视觉体验代表了产品几乎全部的特征,所以消费者对产品的评价几乎全部集中在对颜色的评价上。另外,产品的外包装作为产品的“门面”,给消费者的第一印象将直接影响消费者对产品价值的评价[6]。

6.3. 视觉评价受多方面因素影响

通过分析,发现 Apriori 算法下{视觉}全部作为后项。主要因为本研究以某国货口红产品为例,该产品的突出特征就是颜色。所以{视觉}这个项集,对于处于强关联关系中其他频繁项集来说,都可以作为后项。也可以理解为,消费者对购买国货口红的评价过程中,对其质量、包装、品牌力等等方面的评价,会进一步影响其对产品视觉体验的判断和评价。而作为唇彩产品最重要的特征之一,视觉上评价直接影响消费者的重复购买行为和潜在购买行为。

6.3.1. 包装正向影响产品视觉评价

规则{外包装} \rightarrow {视觉}的置信度是 93.598,提升度 1.006,即消费者关注产品包装的同时,也会同时关注到产品的视觉体验,并且,包装精美的产品会增强消费者心理上对于产品颜色的喜爱程度[7]。除了外包装之外,产品本身的包装也会影响消费者的视觉评价。“{视觉} \rightarrow {外包装,触觉}”这一强关联关系中触觉主要指的产品本身的包装,也称为内包装。也即是,外包装和内包装的组合能正向影响消费者对产品的视觉评价。

6.3.2. 品牌力正向影响产品视觉评价

{品牌认可,嗅觉} \rightarrow {视觉}的规则置信度较高,达到 93.197%,这也意味着对于消费者来说,知名品牌的口红往往和其特有的产品气味紧密相关,且两者的组合会正向影响消费者的视觉评价。这是因为消费者主要通过嗅觉上的感受判断是否是正品。一经确认是正品之后,消费者的关注点将往往全放在对颜色的选择和评价上,而不会考虑质量、价格等其他问题。比如,一些国际品牌的口红,消费者往往会热衷于重复某经典颜色,并且乐于尝试新品。所以说,国货彩妆要想获得更多的消费者支持,首先要提升其品牌影响力,并为其品牌产品赋予一些难以模仿的香味等特征,这样能更好地赢得消费者对其品牌的认可[8]。

6.3.3. 价格和使用感受综合影响产品视觉评价

{价格,触觉} \rightarrow {视觉}属于强关联规则。一般来说,价格实惠的产品,会给消费者一种物美价廉的心理暗示,从而一定程度上降低对产品的高期望值[9],对于产品颜色的要求相对会较低,心理上也更容易满足。但是,消费者已经不再以价格为唯一标准来判断“物美”,只有真正在触觉,即使用体验上感到满意,才会产生真正的“物美价廉”的认知。因此,商家应该在保证产品使用体验较好的前提下,提供价格折扣,更容易提高消费者对产品的满意度。

6.3.4. 触觉和嗅觉体验正向影响产品视觉评价

从提升度上可看到,“{嗅觉} \rightarrow {视觉}”、“{触觉} \rightarrow {视觉}”和“{嗅觉,触觉} \rightarrow {视觉}”三个强关联规则的提升度都大于 1。代表商家同时注重产品在视觉、嗅觉和触觉上的消费者体验感受比单纯的产品视觉设计,会提高消费者对产品视觉上的综合评价。一般而言,颜色属于视觉上的认知,而香味是嗅觉上的认识。对于涂抹在唇部的产品,消费者也会更加关注质量安全问题。所以,对于那些在香味上较为好闻,并且质量有保障的产品,消费者对于产品整体评价会较高[10],对颜色也会有较高的接纳度。

7. 总结

本研究主要以国货品牌完美日记的口红产品作为研究对象,收集并分析其在线评论,试图挖掘其在线评价内容里可能存在的关联规则,以获取更大的市场份额,壮大国货彩妆品牌的市场竞争力。研究结果表明:现阶段电子商务提供的客服服务重要性下降,可能是信息渠道多的原因;产品的价格、品牌、视觉评价、外包装都会影响消费者的购买评论;产品视觉评价受到多方面因素影响,包括产品内外包装、

品牌力、嗅觉评价、触觉评价、价格和触觉的综合评价。这些组合与视觉评价的关联规则，都体现为提升度大于 1 的强关联规则，即消费者对这些组合的评价会进一步提升产品视觉评价。品牌商可以借鉴消费者在线评论中的关联内容，通过改进相关产品属性或服务，进一步提升消费者对产品视觉上的评价，进而提升整体消费者满意度。

致 谢

首先衷心感谢我的导师曹红苹老师对我学术上的指导和帮助。在论文的写作过程中，曹老师为我提供了研究的思路和写作意见，引导我逐步完成论文，使我的论文条理化、规范化。其次，非常感谢《应用数学进展》的期刊老师对我的论文的细心评阅，指出论文的不足之处，为论文的完整性和合理性提供了十分宝贵的建议。最后，在写作过程中，我对关联分析等数据分析方法和软件应用有了进一步的学习和掌握，为今后的学术写作奠定了一定的数据分析能力。

参考文献

- [1] 莫赞, 赵琦智, 罗敏瑶. 在线评论对消费者购买决策的影响研究[J]. 南宁师范大学学报(哲学社会科学版), 2021, 42(3): 36-44.
- [2] 史达, 王乐乐, 衣博文. 在线评论有用性的深度数据挖掘——基于 TripAdvisor 的酒店评论数据[J]. 南开管理评论, 2020, 23(5): 64-75.
- [3] 李贺, 谷莹, 刘嘉宇. 数据驱动下基于语义相似性的产品需求识别研究[J]. 情报理论与实践, 2022, 45(5): 99-106.
- [4] 宋常修, 吴雪莲, 王熙, 张善莹. “直播带货”对消费者的行为影响方式分析及其综合评价[J]. 中国集体经济, 2021(36): 64-65.
- [5] 余耀东, 刘琳. 品牌价值感知对冲动性购买行为的影响[J]. 时代经贸, 2021, 18(11): 10-17.
- [6] 曾曦. 油画元素在化妆品包装设计中的应用[J]. 日用化学工业, 2022, 52(5): 577-578.
- [7] 石胜贵, 梁鑫. 顾客网络购物满意度分析[J]. 物流科技, 2021, 44(2): 53-55.
- [8] 杨金丹. 品牌营销中情感营销的对策与价值探寻[J]. 中国集体经济, 2021(35): 59-60.
- [9] 郑海英. 新零售背景下全渠道价格管理策略研究[J]. 中国商论, 2021(18): 12-15.
- [10] 张巧可, 陈洪转. 基于质量意识的复杂产品质量控制激励策略[J]. 中国管理科学, 2021, 29(6): 105-114.