

# 基于文本挖掘的国货美妆电商消费者满意度影响研究

杨婷婷, 田书格

上海工程技术大学管理学院, 上海

收稿日期: 2023年2月6日; 录用日期: 2023年6月5日; 发布日期: 2023年6月12日

## 摘要

国妆品牌凭借“国潮”浪潮赢得年轻消费者青睐, 为中国消费市场注入充沛活力。本文基于电商在线评论, 通过BTM主题模型、情感分析、回归分析等方法, 探究在线评论中隐含的各影响因素对消费者满意度影响大小, 为国货美妆企业电子商务提出管理建议。研究表明使用感受、产品功效和感知价值是影响消费者满意度的三大因素。

## 关键词

BTM模型, 国货美妆, 情感分析

# Research on the Influence of Consumer Satisfaction of National Beauty E-Commerce Based on Text Mining

Tingting Yang, Shuge Tian

School of Management, Shanghai University of Engineering Science, Shanghai

Received: Feb. 6<sup>th</sup>, 2023; accepted: Jun. 5<sup>th</sup>, 2023; published: Jun. 12<sup>th</sup>, 2023

## Abstract

China cosmetics brands win the favor of young consumers with the tide of “China tide” and inject vitality into the Chinese consumer market. Based on e-commerce online reviews, this paper explores the impact of various influencing factors implied in online reviews on consumer satisfaction through BTM topic model, sentiment analysis, regression analysis and other methods, and puts forward management suggestions for national beauty cosmetics enterprises in e-commerce.

**Research shows that usage experience, product efficacy and perceived value are the three major factors affecting consumer satisfaction.**

## Keywords

**BTM Model, Domestic Beauty Makeup, Sentiment Analysis**

Copyright © 2023 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

长期被国外美妆品牌挤压的国妆品牌,在“国潮”背景下迎来宝贵的发展机会,与先前一味跟随、模仿国外品牌不同,而是选择主动出击,在保证产品质量的同时,以创新理念将中国传统文化、经典元素融入到产品设计当中。凭借国潮之风崛起的品牌,为中国消费市场注入了活力,也为中国品牌国际化打下了坚实的基础。为此,为更好地理解国内化妆品品牌的消费偏好,本研究利用 BTM 主题模型、情感分析、回归分析等方法研究电商在线评论,探讨不同因素对消费者满意度的影响程度,为国货美妆企业提供相关管理建议,以提高我国美妆商品的市场竞争力。

## 2. 文献综述

借着“国潮”青年的热情和国内公司的主动变革,国内化妆品行业涌现出了完美日记、花西子等知名的国货美妆品牌,这些品牌在外国化妆品的围攻中脱颖而出,也引起了国内学术界的关注。于晴等[1]以“完美日记”为例剖析了国货美妆品牌的创新宣传战略,并通过剖析了完美日记在小红书平台采用的宣传战略,并总结了其创新宣传模式,从而为国货美妆品牌推广宣传提供建议与策略;李梦梵[2]以花西子为例,研究其如何利用社会化媒体营销,通过分析其视频号发布内容,总结其社会化营销成功经验。完美日记、花西子为成立不久的新锐国货美妆品牌,而柏莱雅、百雀羚等我国传统国货品牌也在顺应时代潮流,积极求变。刘雨欣[3]以柏莱雅为例,分析其在网红经济背景下的营销传播策略,指出柏莱雅的网红营销转型是成功且具有前瞻性的,但由于网红自身不确定性较高,此营销策略对于品牌形象和产品口碑都具有一定风险。

近几年,自然语言处理技术取得广泛发展,文本挖掘技术也被国内学者用于文档主题提取。赵桂红等[4]运用 LDA 主题模型对评论进行主题提取作为旅客满意度的影响因素,研究旅客在航班延误后的满意度影响因素;张道海等[5]基于语义分析技术,采用特征提取、情感分析、主体挖掘与事件构建方法,对消费者网购在线评论进行分析,探究电商平台物流服务质量的主要影响因素;李冠等[6]运用 LDA 模型挖掘政府数据开放平台初建期和发展期用户需求主题,分析其需求热点,选取朴素贝叶斯算法研究用户需求主题的情感倾向,计算两个时期的平台用户需求主题相似度,揭示用户增量需求的动态演进路径。

目前,学术界对影响顾客满意的因素进行了深入的探讨,并提出了相应的改进意见。张碧芸等[7]通过实证分析探究女性消费者的消费倾向、购买满意度和购买行动三者间的关系,帮助企业了解女性消费群体,制定营销策略;朱艳娇等[8]通过问卷调查法获得智能手表满意度影响因素,利用多元回归模型分析影响因素对消费者满意度影响大小,并运用模型进行回归分析;陈文玉等[9]使用 IPA 分析方法调研消费者,构建体验型酒店评价指标体系。

综上所述, 国内化妆品行业的相关理论主要是针对市场营销方式进行探讨, 缺乏消费者满意度的研究; 对于消费者满意度的调查多采用问卷、访谈等方法进行小规模样本研究, 但这些方法由于数据有限, 得出的结果也是有限的。相较于传统问卷等方法, 在线评论包含消费者对产品及服务的真实态度, 通过对其进行深入分析, 找出潜在需求特征, 帮助企业快速、全面地获取消费者反馈信息。因此本文采用文本挖掘方法, 对影响国货美妆消费者满意度的因素进行分析。

### 3. 研究方法

本文研究框架如图 1 所示, 对爬虫获取的在线评论进行 BTM 主题挖掘, 对挖掘结果总结归纳为消费者满意度影响因素, 通过情感分析对影响因素进行量化处理, 最后对量化结果进行多元回归分析, 研究各影响因素对满意度影响因素大小, 针对分析结果为商家提供管理建议。

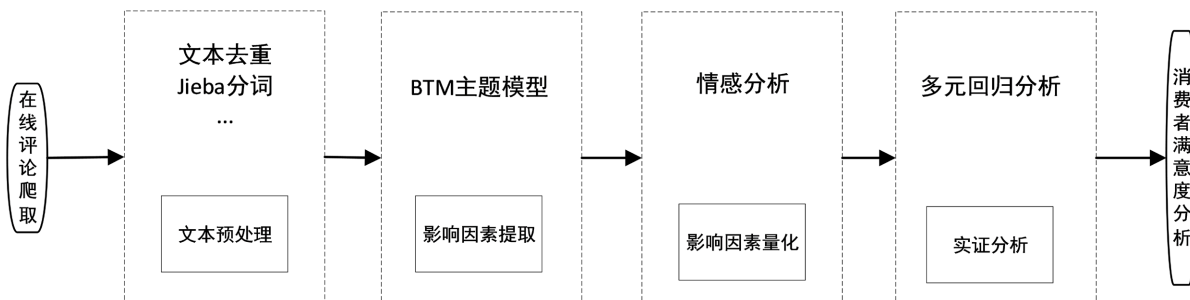


Figure 1. Research framework  
图 1. 研究框架

#### 2.1. BTM 主题模型

随着社交媒体的飞速发展, 大量用户产生的短文本数据呈指数级速率增长。使用为传统长文本而设计的 LDA 模型处理短文本数据存在着数据稀疏度等问题。Yan 等[10]在 2013 年提出了一种词对主题模型 (Biterm Topic Model, BTM), 该数据模型是在构建一个词对的时候直接针对文本数据进行建模, 实验结果可以不受文本长度的限制。BTM 模型原理如下:

词对是指文中共现的无序词组, 在含有 3 个不同单词的文件中, 词对产生程序的表达式如下(1), 则拥有  $n$  个字的短文本语料库可以生成  $C_n^2 = n(n-1)/2$  个词对。

$$(w_1, w_2, w_3) \rightarrow \{(w_1, w_2), (w_1, w_3), (w_2, w_3)\} \quad (1)$$

给定包含  $N_D$  个文档的语料库  $D = \{d_1, d_2, \dots, d_{N_D}\}$ , 其中每篇文档  $d_k$  可表示为  $d_k = \{w_1, w_2, \dots, w_m\}$ , 并引入一个隐含主题集合为  $Z = \{z_1, z_2, \dots, z_k\}$ 。BTM 生成图模型如图 2 所示。

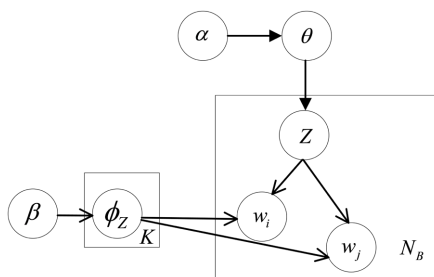


Figure 2. BTM generates a model diagram  
图 2. BTM 生成模型图

- 1) 假设整个词对集全局的主题分布  $\theta \sim Dir(\alpha)$ ;
- 2) 对于每一个主题  $z$ : 假设该主题下词分布  $\phi_z \sim Dir(\beta)$ ;
- 3) 对于词对集合  $B$  中每一个词对  $b = (w_i, w_j)$  执行下列操作:
  - a) 抽取主题  $z \sim Multi(\theta)$
  - b) 从主题  $z$  对应词分布中抽取单词  $w_i \sim Multi(\phi_z)$   
从主题  $z$  对应词分布中抽取单词  $w_j \sim Multi(\phi_z)$

$$Biterm: z \sim Multi(\theta), w_{i,1}, w_{i,2} \sim Multi(\phi_z)。$$

根据上述流程, 词对  $b = (w_i, w_j)$  的联合概率计算方法见公式(2):

$$\begin{aligned} P(b) &= \sum_z P(z) P(w_i|z) P(w_j|z) \\ &= \sum_z \theta_z \phi_{w_i|z} \phi_{w_j|z} \end{aligned} \quad (2)$$

则整个 BTM 语料库的词对集合  $B$  生成概率如式(3):

$$P(B) = \prod_{(w_i, w_j)} \sum_z \theta_z \phi_{w_i|z} \phi_{w_j|z} \quad (3)$$

使用 Gibbs 采样法估算“词对集合 - 主题”分布  $\theta$  和“主题 - 词语”分布  $\phi_k$ 。用公式(4)来计算当前词对的主题分布:

$$p(z|z_{-b}, B) \propto \frac{n_{-b,z} + \alpha + (n_{-b,w_i|z} + \beta)(n_{-b,w_j|z} + \beta)}{N_B + K\alpha} \frac{1}{\left( \sum_w (n_{-b,w|z} + \beta) \right)^2} \quad (4)$$

式中,  $K$  为文本文档包含的主题数目,  $B$  为文本文档的词对集合,  $\alpha$ ,  $\beta$  分别为狄利克雷先验分布的超参数。

### 3.2. 情感分析

目前情感研究主要包括情感词典和机器学习二类研究方式。情感词典研究方法主要是构建研究特定情感词典, 根据词典赋值计算情感得分, 判定其情感倾向; 机器学习的基本方法就是利用基本的算法对被检测的属性进行筛选, 然后将其引入到分类器中进行训练, 从而得出情绪趋势的分析。

考虑到本文进行情感分析目的是通过情感值的计算将挖掘出的影响因素进行量化处理, 为后文多元回归模型计算各影响因素对消费者满意度影响大小奠定基础。为此, 本文运用情感词典研究方法, 以知网情感词典为基础, 新增国货美妆专有名词, 建立本研究专有情感词典。

## 4. 研究过程与结果分析

### 4.1. 数据采集及预处理

国货美妆产品作为面向大众的消费类产品, 其在多个电商平台均有销售, 在综合考虑平台规模, 产品销售数量及评价质量等诸多因素后, 本文决定选取京东美妆馆的国货美妆产品作为研究对象。鉴于美妆产品具备的品类繁多, 品牌众多的特点, 无法对平台内所有国货美妆产品的评论数据进行收集分析, 因此本文参考《万榜 2021 中国国产美妆行业 TOP10 企业榜》, 综合考虑公司规模、品牌影响力、主流电商销售数据等因素后, 选取京东美妆馆中 5 个国产化妆品公司旗下的 20 种美妆产品的在线评论文

本数据。

采用 Python 语言编写爬虫程序, 能够快速获取研究所需要的海量格式化数据, 并根据研究目的定制爬虫, 分别爬取好评、差评内容, 这是目前市面爬虫软件难以实现的个性化需求。该程序基于京东美妆商品页面源码设置请求头参数信息, 模拟浏览器向京东服务器发送 get 请求, 获得其 response 应答, 将解析后的 json 数据格式转化为 Python 语言中的 dictionary 格式, 便于进一步数据存储。在爬取期间, 采用使用多个虚拟 User-Agent 及 time.sleep 方法设置间隔时间, 以逃避网站反爬虫机制。本文共爬取好评 20,000 条、差评 11,186 条。具体评论数据如下图 3 所示。

1	content	nickname	creationTime	productColor	score
2	同事推荐说韩束效果很好, 抱着试试看的想法下单了, 现在是用后特意过来评价的, 效果果然很好, 用后皮肤依... 吸		2022/4/10 15:18	【紧致淡纹】聚时光臻颜7件套	5
3	吸收效果: 很好。保湿程度: 很润的, 我购买了一个套盒, 里面水乳霜都有, 眼霜洗面奶, 韩素这个套盒价格:***z		2022/4/21 7:10	【紧致淡纹】聚时光臻颜7件套	5
4	这已经是第二次购买韩束的这款护肤品了, 价格不会, 但是很好用, 我用其他有的产品很多都过敏, 用这个倒... 年		2022/4/6 12:04	【紧致淡纹】聚时光臻颜7件套	5
5	韩束礼盒包装很有档次。上次购买完后继续购买。价位挺实惠的, 而且是正品。比商场专柜便宜很多。很划... k***o		2020/3/8 0:00	【改善细纹】聚时光臻颜7件套	5
6	宝贝收到了, 没想到有那么多东西! 洗面奶用着很温和, 洗过后皮肤不干不紧绷, 爽肤水很好用, 吸收好, 不... d***3		2022/3/8 20:05	【紧致淡纹】聚时光臻颜7件套	5
7	产品包装: 包装很精美 吸收效果: 很滋润, 以前就用过, 冬天用非常滋润, 精华用完后脸蛋干爽滑滑的 保湿程... 瓜***		2021/11/9 20:36	【紧致淡纹】凝时五件套	5
8	给妈妈买的。礼盒高档, 上档次。妈妈说好用。滋润皮肤。看上去也很大气。还送了一个正装精华水。很划算... Y***活		2021/10/30 16:34	【紧致淡纹】凝时五件套	5
9	产品包装: 包装结实大气 吸收效果: 非常不错一直在用 保湿程度: 很润的感觉 适合肤质: 任何皮肤都可以用 ...**6		2020/11/3 16:25	【紧致淡纹】凝时五件套	5
10	快递真的很快晚上下单第二天中午就到了, 迫不及待的打开快递, 盒子非常有质感, 很高大上, 整整一套护肤... 小***曼		2019/9/26 21:00	【紧致淡纹】聚时光臻颜礼盒	5
11	包装挺严实的, 没有破损, 内包装很高档, 看起来高端大气上档次, 用了一段时间, 觉得还不错, 比较水... S***		2022/3/4 12:48	【紧致淡纹】聚时光臻颜7件套	5
12	产品包装: 很高大上, 送礼也挺合适的, 包装完好 吸收效果: 洗完澡用洗面奶按顺序用能感受到皮肤很滋润, 对... 霜		2022/4/17 14:57	【紧致淡纹】聚时光臻颜7件套	5
13	给妈妈买的, 包装看起来不错 瓶子比较有质感 两瓶水一瓶乳搭配刚刚好 妈妈试了下面霜 味道挺好闻的 质感不... 米***6		2021/6/23 11:08	【紧致淡纹】凝时五件套	5
14	用了韩束很多次了, 以前年轻的时候用补水的效果很好, 现在年纪大了, 应该用点营养更丰富、成分更安全的... 7***7		2021/7/9 11:20	【紧致淡纹】凝时五件套	5
15	第一次买这个牌子, 产品包装高大上, 性价比, 用了一片送的面膜, 皮肤很光滑, 补水效果非常好, 下次就... F***2		2021/1/18 6:40	【紧致淡纹】凝时五件套	5
16	一直在用这个牌子, 真的很好, 物有所值, 活动买的, 给了很多赠品真的好用超级划算 ...**6		2021/6/16 19:43	【紧致淡纹】凝时五件套	5
17	韩束套装, 一直用韩束化妆品, 用后很舒服, 不紧绷, 味道也很清新, 不是那种味道很重的化妆品, 趁着618活... h***h		2021/6/19 12:06	【紧致淡纹】凝时五件套	5
18	给我心爱的女人, 买了一套化妆品, 打开包装的那一刻, 真好, 绝对的正品, 打开一瓶也是香气扑鼻, 很细腻... E***v		2021/8/1 9:57	【紧致淡纹】凝时五件套	5
19	韩束的化妆品我一直在用, 大家都说我的脸最近都很润, 有光泽。吸收效果: 很不错 保湿程度: 一天脸部都很... **o		2021/5/2 18:50	【紧致淡纹】凝时五件套	5
20	婆婆有一套, 妈妈用了用了一两次, 说超级好用, 果断我也给妈妈买了一套! 物美价廉, 性价比超高! 主要的... S***6		2022/4/15 12:32	【紧致淡纹】聚时光臻颜7件套	5
21	整套产品主打保湿紧致肌肤, 保湿效果还不错, 这是本人购入的第三套, 而且价格很实惠, 用后也么什么不适... H***+		2021/4/22 14:01	【紧致淡纹】凝时五件套	5
22	疫情期间能这么快收到货已经很满意了, 包装箱完整没破损, 化妆品看着很好, 很大气, 送人没有问题的, ...**9		2022/4/5 9:48	【紧致淡纹】聚时光臻颜5件套	5
23	产品包装: 包装完整, 很大的一个盒子, 保护得很好, 没有破损 吸收效果: 水吸收的很快, 不会很油腻, 用着... 子		2022/3/16 18:47	【紧致淡纹】聚时光臻颜7件套	5

Figure 3. Comment data of the positive part

图 3. 好评部分评论数据

未经处理的原始数据存在字母、数字、字符以及系统的自动评价等噪声, 因此为了研究后续更好地对文本进行分析, 提高结果精准性, 需要对爬取的原始文本数据进行文本预处理。

本文从两方面进行文本清洗, 首先使用 Python 中的 drop\_duplicates 方法, 删除掉由于买家未及时做出评论而系统自动评论的重复文本内容。其次对于文本数据中的数字、字母、符号等无效、无意义并干扰结果的语料。数据清洗后得到好评 19,935 条, 差评 10,744 条。

中文与英文中使用空格将单词间分隔开的词语特性不同, 所以对中文语料进行分词对于中文文本分析是十分必要的。本文利用当前中文文本挖掘领域广泛使用的“jieba”分词工具对清洗好的文本数据进行文本切分。在对“jieba”分词结果进行分析时发现, 原始文本中关于美妆领域的词语由于该 Python 导入库内置词库缺乏该领域的分词, 同时由于彩妆产品名称原创性较高, 极少被收录进常规词库, 从而得到一些错误的分词结果。基于以上原因, 本研究通过使用 Python 调用用户自定义词库方法 load\_userdict, 结合搜狗细胞词库中包含美妆品牌及美妆产品常用成分等的美妆领域词库及本文原始数据中彩妆产品名称, 进一步丰富本文分词词库, 使分词结果更为精准。

#### 4.2. 基于 BTM 模型的满意度影响因素提取

目前学者确定最优主体数量主要采用主题间余弦相似度和困惑度两种方法, 其中主题间余弦相似度计算时间较长, 不适用大量数据, 因此本文将 BTM 模型设置为[5, 10], 分别运行模型计算困惑度。评论困惑度曲线如图 4 所示。

由图可知, 困惑度随着主题数增加而下降, 但是在主题数达到 9 后, 下降趋势趋于平缓, 同时, 虽然主题数增加会降低困惑度, 但是设立过多的主题数却偏离的研究的目的, 因此, 本文将 BTM 模型主题参数设置为 9, 各个主题下词语及权重如下表 1 所示。



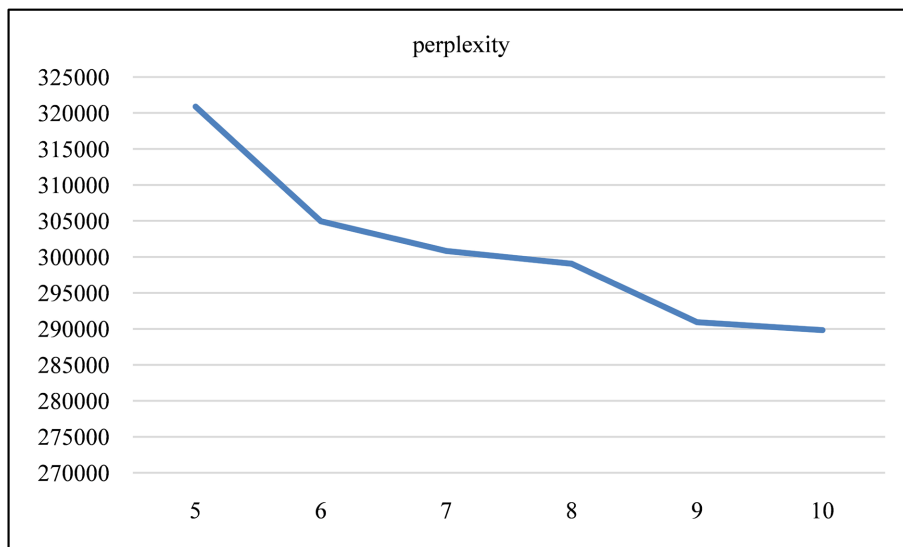


Figure 4. Comment topic perplexity curve

图 4. 评论主题困惑度曲线

Table 1. The BTM model comments on topic mining results

表 1. BTM 模型评论主题挖掘结果

主题	关键词及权重
主题 1	颜色: 0.035 保湿: 0.022 补水: 0.018 上妆效果: 0.008 修复: 0.007
主题 2	客服: 0.111 售后: 0.029 问题: 0.024 态度: 0.019 服务: 0.013
主题 3	皮肤: 0.072 吸收: 0.038 适合肤质: 0.025 油腻: 0.024 刺激: 0.009
主题 4	粉质: 0.028 细腻程度: 0.029 质感: 0.011 成分: 0.009 味道: 0.018
主题 5	妈妈: 0.033 信赖: 0.017 送给: 0.016 老婆: 0.015 礼盒: 0.013
主题 6	包装: 0.034 设计: 0.016 瓶子: 0.015 质感: 0.010 产品外观: 0.009
主题 7	快递: 0.036 物流: 0.035 速度: 0.020 发货: 0.015 小哥: 0.011
主题 8	赠品: 0.131 活动: 0.054 双十一: 0.018 价格: 0.016 划算: 0.008
主题 9	看图: 0.026 图片: 0.008 宣传: 0.007 广告: 0.006 营销: 0.006

笔者将挖掘出主题概括如下: 主题 1 为产品功效, 主题 2 售后服务质量, 主题 3 为使用感受, 主题 4 为产品性状, 主题 5 为赠礼价值, 主题 6 为外观包装, 主题 7 为物流服务质量, 主题 8 为感知价值, 主题 9 为宣传信息可靠性。

### 4.3. 基于情感分析的满意度影响因素量化

由于 BTM 主题模型是一种非监督的文本聚类方法, 其结果可能会存在不同主题下的特征词对应不准确的问题。为解决上述问题, 本文对主题词下的特征词进行人为的调节, 使得其更加清晰贴合主题, 结果如下表 2 所示。

**Table 2.** Feature word list corresponding to influencing factors  
**表 2.** 影响因素对应特征词表

影响因素	特征词
产品功效	颜色、保湿、补水、上妆效果、修复、补水效果……
售后服务质量	客服、售后、问题、态度、服务、退货……
使用感受	吸收、适合肤质、油腻、清爽、刺激、敏感肌……
产品性状	粉质、细腻程度、质感、成分、味道、刷子……
赠礼价值	妈妈、信赖、送给、老婆、礼盒、国货……
外观包装	包装、设计、瓶子、质感、产品外观、小巧……
物流服务质量	快递、物流、速度、发货、小哥……
感知价值	赠品、活动、双十一、价格、划算、实惠、便宜……
宣传信息可靠性	看图、图片、宣传、广告、营销、回购……

假设评论某条评论为  $d_j$ ，其由  $n$  个分句组成，各表示为  $s_1, s_2, \dots, s_n$ ，分句中的情感词、程度副词和否定词共同决定该分句的情感值，计算公式如下(5)所示

$$\text{score}_{s_i} = \text{score}_{\text{negative\_word}} * \text{score}_{\text{emotion\_word}} * \sum \text{score}_{\text{degree\_word}} \quad (5)$$

所有分句情感值之和为该评论的情感值  $\text{score}_{d_j}$ ，如公式(6)所示：

$$\text{score}_{d_i} = \sum_{i=1}^n \text{score}_{s_i} \quad (6)$$

利用 Python 分别计算了每个影响因素特征值的情感值以及每条评论的情感值，共得到了 30,679 条数据，部分计算结果如下表 3 所示。

**Table 3.** Eigenvalues of each influencing factor and sentiment value of comments  
**表 3.** 各影响因素特征值及评论的情感值

序号	产品功效	售后服务质量	使用感受	产品性状	赠礼价值	外观包装	物流服务质量	感知价值	宣传信息可靠性	评论情感得分
1	0	0	2.5	2	2.5	0	0	0	0	6
2	0	-1	0	-1	0	0	0	0	0	-5
3	0	0	1.5	1	0	1	0	1.5	0	6
4	-1	0	-1.5	0	-1.5	0	1	1	0	-1.5
5	-3.5	0	0	-2.5	-1.5	0	0	0	0	-6

#### 4.4. 在线评论对消费者满意度影响分析

##### 4.4.1. 模型建立

###### 1) 变量命名

为方便后续研究，首先进行变量命名，如下表 4 所示。

**Table 4.** Descriptive statistics of each variable  
**表 4.** 各变量描述性统计

变量类型	变量符号	变量定义	变量描述	
因变量	Satisfaction	消费者满意度	情感词典计算值, [-6, 7], 实数	
	benefits	产品功效	情感词典计算值, [-15, 24], 实数	
	feel	使用感受	情感词典计算值, [-8, 15], 实数	
	properties	产品性状	情感词典计算值, [-9, 22], 实数	
	gift	赠礼价值	情感词典计算值, [-5, 12], 实数	
	自变量	look	外观包装	情感词典计算值, [-8, 17], 实数
		value	感知价值	情感词典计算值, [-7, 14], 实数
		express	物流服务质量	情感词典计算值, [-10, 25], 实数
		service	售后服务质量	情感词典计算值, [-10, 24], 实数
promote		宣传信息可靠性	情感词典计算值, [-6, 9], 实数	

## 2) 模型构建

消费者满意度影响因素分析多元回归模型, 如式(7)所示

$$\begin{aligned} \text{Satisfaction} = & \beta_0 + \beta_1 \text{benefits} + \beta_2 \text{feel} + \beta_3 \text{properties} + \beta_4 \text{gift} + \beta_5 \text{look} \\ & + \beta_6 \text{value} + \beta_7 \text{express} + \beta_8 \text{service} + \beta_9 \text{promote} \end{aligned} \quad (7)$$

### 4.4.2. 回归过程及结果分析

针对不同的变量, 其数值区间存在很大的差别, 为提高其可解释度, 进行“最小 - 最大标准化”的归一化处理, 即原始数据减去最小值再比上最大最小值之差, 将所有数据都变成了[0, 1]之间的数据, 以便进行比较和分析。

#### 1) 描述性统计

首先进行描述性统计分析, 了解各个变量的基本数据特征, 描述性统计结果如下表 5 所示。

**Table 5.** Descriptive statistics for each variable  
**表 5.** 各变量描述性统计

变量	N	平均值	标准差	最小值	最大值
Satisfaction	30,679	0.7435	0.3213	0	1
benefits	30,679	0.3681	0.0649	0	1
feel	30,679	0.3323	0.0512	0	1
properties	30,679	0.4036	0.0422	0	1
gift	30,679	0.3722	0.0462	0	1
look	30,679	0.3672	0.0584	0	1
value	30,679	0.3279	0.0668	0	1
express	30,679	0.3456	0.0486	0	1
service	30,679	0.3779	0.0381	0	1
promote	30,679	0.4149	0.0347	0	1



从表 5 中可知, 规范化后的各变量取值变为了[0, 1], 并且每个变量的标准差都很小, 说明数据分布较集中。消费者满意度的平均值大于 0.5, 说明消费者对国货美妆产品的消费倾向偏向正面。

## 2) 回归分析

运用多元线性回归方法对消费者满意度进行实证研究, 利用 spss26.0 软件对模型进行统计分析, 结果如下表 6 所示:

**Table 6.** Regression results of factors influencing consumer satisfaction  
**表 6.** 消费者满意度影响因素的回归结果

变量	标准化系数	t 值	P 值	VIF 值
benefits	0.232	37.426	0.000	1.085
feel	0.254	41.855	0.000	1.031
properties	0.132	21.033	0.000	1.077
gift	0.113	18.826	0.000	1.028
look	0.114	18.713	0.000	1.061
value	0.231	34.411	0.000	1.273
express	0.105	17.451	0.000	1.026
service	0.092	13.712	0.000	1.234
promote	0.056	9.395	0.000	1.023
R2	0.414		F	1285.817
AdjR2	0.413		Sig	0.000

由表 6 中可知, 回归模型统计量 F 值为 1285.817,  $p < 0.001$ , 表明消费者满意度回归模型整体显著, VIF 值均低于 5, 故各变量间不存在多重共线性关系, 具体来看:

1) 国货美妆产品的功效( $\beta = 0.232, p < 0.001$ )、使用感受( $\beta = 0.254, p < 0.001$ )、产品性状( $\beta = 0.132, p < 0.001$ )、赠礼价值( $\beta = 0.113, p < 0.001$ )、外观包装( $\beta = 0.114, p < 0.001$ )都通过了显著性检验, 并且均正向影响消费者满意度。国货美妆产品在我国仍属于新兴工业消费品, 尤其出现违规添加汞等对消费者健康有害添加剂的社会新闻事件, 使得消费者在购买美妆类产品时会首先关注产品质量问题, 因此若国货美妆产品质量能满足甚至超出消费者预期需求, 则会极大提高消费者满意度, 因此产品功效、使用感受、产品性状、赠礼价值、外观包装对消费者满意度具有显著的正向影响。在上述因素中, 产品功效和使用感受对消费者满意度的影响更加显著, 说明尽管消费者会被精美的礼盒、花哨的包装吸引眼球, 但更关注的仍然是产品自身的使用感受及功效。

2) 感知价值( $\beta = 0.231, p < 0.001$ )对消费者满意度具有显著正向影响, 本文将消费者对价格的认知程度用感知价值来度量, 它反映消费者实际支付与愿意支付的价格的差异。当消费者以更低的价格获得高品质的商品时, 其感知价值就会更高, 认为购买到的商品“性价比高”、“物美价廉”, 感到满足; 当消费者花费较多的钱却买到质量低劣的商品时, 消费者就会感到花了冤枉钱, 做出“不划算”等评价; 当消费者低价购买到劣质商品时, 会形成“贪便宜”、“便宜没好货”的负面评价, 进而导致消费者产生消极的心理, 从而形成不良的口碑, 因此感知价值正向影响消费者满意度。

3) 物流服务质量( $\beta = 0.105, p < 0.001$ )正向影响消费者满意度, 本文中物流服务被分为物流配送速度和物流包装保护性。首先, 网购平台越快将商品送到消费者手中, 则消费者满意度越高; 其次, 消费者

在收到商品时看重物流包装是否完好, 尤其彩妆类产品, 运输过程中的颠簸磕碰可能会损坏产品外观甚至损坏产品性状, 例如眼影等粉质类产品, 运输过程中过度颠簸可能造成眼影块碎裂, 进而影响到消费者到手产品质量, 因此物流包装保护性越好, 消费者越满意。

4) 售后服务质量( $\beta = 0.092, p < 0.001$ )对消费者满意度具有正向影响, 对客户的良好服务, 对信息的快速答复, 会让顾客感觉到满足; 宣传信息可靠性( $\beta = 0.056, p < 0.001$ )同样对消费者满意度具有正向影响, 宣传信息包括产品营销所宣传的内容和商家在网页中展示的与产品有关的图片、产品信息和促销信息等, 由于美妆产品存在大量营销行为, 包括流媒体广告、KOL (关键意见领袖)推广等, 使得消费者对产品产生极大期待, 因此消费者接收到的商品宣传信息越符合真实商品信息, 越容易买到心仪产品, 则越感到满意。

综合来看, 根据系数将各个影响因素对满意度的影响由大到小排序分别为使用感受 > 产品功效 > 感知价值 > 产品性状 > 外观包装 > 赠礼价值 > 物流服务质量 > 售后服务 > 宣传信息可靠性, 即消费者最为关注的三个要素为使用感受、产品功效和感知价值。以上分析结果显示, 在购买国货美妆产品时, 消费者最关心的是使用感受。这是由于目前化妆品行业的状况所限, 美妆产品功效方面同质化趋于严重, 因此在功效相同时, 消费者想要更好的使用体验。对于保湿类产品, 则希望其吸收效果更好, 对于精华类产品, 则希望其干爽不油腻, 对于彩妆类产品, 则追求其易卸妆、新手小白友好等等方面。同时, 消费者选购美妆产品时, 其使用后功效越好, 消费者越满意, 越乐于在评论中分享诸如保湿效果、美白效果、显色度等等产品功效方面内容。一方面, 相较于国际美妆品牌, 国货美妆的一大优势即为价格亲民, 以柏莱雅双抗精华为代表的众多国货美妆产品以其物美价廉的竞争优势荣获“国货之光”殊荣, 因此在上述使用感受及产品功效达到消费者预期效果后, 国货产品平易近人的价格进一步引发消费者满意, 消费者倾向于在评论文本中使用“优惠”、“性价比”等词语表达自身满意; 另一方面, 美妆类产品一直是各大电商平台购物节优惠的主要品类之一, 与此同时, 各大美妆品牌不满足于单一形式的营销, 倾向于多渠道以优惠价格回馈或拓展忠实消费者, 消费者倾向于在评论文本中使用“拼单”、“双十一”等词语表达自己通过电商平台购物节或其他营销渠道以低于日常价格购买到同质量产品的喜悦及满意的心情。因此使用感受、产品功效和感知价值是影响消费者满意度的三大因素。

售后服务质量和宣传信息可靠性相较于其他影响因素对消费者满意度影响较小。首先就行业背景而言, 美妆是以用户为主导的产品, 无论是海外品牌还是本土企业都是立足于消费者体验来发展的, 各品牌构建化妆品售后机构, 追踪与处理消费者在商品使用过程中出现的各类问题, 以满足个性化需求。其次, 京东作为国内成熟电商平台, 本身也发展了一整套专门针对售后服务纠纷、产品赔付问题的解决程序。因此消费者感受到的品牌间售后服务是无较大差别的, 其对消费者满意度影响较小。宣传信息可靠性对消费者满意度影响最小, 这是因为消费者常常买到宣传与实物不符产品, 尤其美妆类产品存在因人而异的特点, 因此对产品实际使用不及宣传效果情况有一定的容忍度; 其次, 在自媒体时代, 当新产品问世时, 品牌会联系美妆 KOL 做产品宣传推广, 即种草过程, 与此同时, 会存在另一部分美妆 KOL 做产品测评, 即拔草过程, 消费者通过两方进行信息交叉验证, 并根据商品评论, 获取到比商家宣传更全面的商品信息, 从而做出购买决策。因此对商家宣传信息可靠与否并不十分在意, 其对消费者满意度影响因素较小。

#### 4.5. 管理启示

针对上述结论, 下文为国货美妆产品电商发展提出以下建议:

1) 企业应重视在线评论内容并密切关注其变化。企业应经常关注用户的网上评价, 并运用大数据与人工智能技术对用户的评价进行分析, 从而掌握用户的需求特点, 从而适时地制定相应的市场战略, 提

高产品和服务的品质。

2) 企业应重视国货美妆产品质量, 将提升产品质量摆在首要位置, 加大科研投资。产品品质是美妆企业长期发展的基础, 而其核心发明专利是安身立命之本, 增大科研中心建设投入, 增加校企合作力度, 培养领域高技术人才是美妆企业不断发展的动力。

3) 企业应建立品牌观念, 重视口碑的累积。国货美妆企业应该从自身的产品和市场上的同类商品进行细致的区分, 针对性进行广告宣传, 塑造明星商品, 加深品牌与该细分领域的联系, 与消费者建立信任, 通过社群营销培育粉丝群体, 打造私域流量, 形成口碑传播。

4) 企业应重视服务质量, 提升客服服务能力以满足消费者售后需求。首先要引进智能客户服务助手来进行日常问题的判别和分流。针对一般的问题, 提前制定一系列的解决办法, 以达到客户的自助解决售后服务问题; 对于个性化问题则交由人工客服, 商家需要培训客服人员专业技术和服务意识, 以提供专业解答, 提升自身专业感及消费者信任感, 实现人工客服到美妆顾问的身份转变。

## 5. 结语与展望

本文采用文本挖掘技术分析国货美妆产品消费者满意度, 致力于帮助国货企业针对性提升自身产品及服务质量, 助力国货崛起。在线评论文本挖掘相较于传统问卷调查方法在数据量及时间人力成本方面具备显著优势, 是一种帮助企业更加方便快捷了解并改善消费者满意度的方法。同时本文仍存在一定局限性, BTM 模型由于其非监督文本聚类特性, 在语义理解上存在理解精度较低问题, 后续研究可选取更高精度的聚类方法。

## 参考文献

- [1] 于晴, 彭思涵. 新国货美妆品牌的创意传播策略分析——以“完美日记”为例[J]. 卫星电视与宽带多媒体, 2020(8): 239-241.
- [2] 李梦梵. 社交媒体背景下“她经济”国货的突围途径——以花西子的视频号为例[J]. 新闻研究导刊, 2021, 12(18): 254-256.
- [3] 刘雨欣. 网红经济下老牌国货品牌的营销传播新策略分析——以经典美妆国货品牌珀莱雅为例[J]. 现代营销(学苑版), 2021(10): 38-40.
- [4] 赵桂红, 宁慧慧. 基于 LDA 主题模型的航班延误服务补救意见识别研究[EB/OL]. 经营与管理: 1-13. <https://kns.cnki.net/kcms/detail/12.1034.F.20230130.1202.002.html>, 2023-01-31.
- [5] 张道海, 杨晨. 基于语义的电商平台物流服务质量主题分析[J]. 物流科技, 2023, 46(1): 36-40.
- [6] 李冠, 赵毅. 基于在线评论的政府数据开放平台用户增量需求研究[J]. 数字图书馆论坛, 2022(12): 37-46.
- [7] 张碧芸, 杨泽响, 郑涵予. 女性体育用品消费者的消费倾向、购买满意度与购买行动的关系研究[J]. 中国商论, 2022(24): 79-82.
- [8] 朱艳娇, 张水胜. 智能手表消费者满意度影响因素及产量预测[J]. 高师理科学刊, 2022, 42(12): 34-38.
- [9] 陈文玉. 基于 IPA 分析的苏州“江南文化”体验型酒店满意度研究[J]. 中国集体经济, 2022(33): 71-73.
- [10] Yan, X.H., Guo, J.F., Lan, Y.Y., et al. (2013) A Biterm Topic Model for Short Texts. *Proceedings of the 22nd International World Wide Web Conference*, Rio de Janeiro, 13-17 May 2013, 1445-1456.