

# Analysis of Consumers' Evaluation of Domestic Cosmetics Brand Based on Python Text Mining

Lin Lin, Jie Gu, Yang Jiao, Huqin Yan

Xiamen National Accounting Institute, Xiamen Fujian  
Email: 903310185@qq.com

Received: Apr. 1<sup>st</sup>, 2020; accepted: Apr. 15<sup>th</sup>, 2020; published: Apr. 22<sup>nd</sup>, 2020

---

## Abstract

By using the big data mining method of Python beautifulsoup, word frequency analysis method of Python Wordcloud, and the Chinese character analysis method of Python Jieba, this paper analyzes the existing problems of domestic cosmetics brands and the most important consumption hot spots for consumers based on the feedback from consumers in various platforms. We select three domestic cosmetics brands with different positioning—Judydoll, Perfect Diary and MAOGEPING as reference. We use a large number of comments for text mining, word frequency statistics, analysis of consumption hot spots, and then analyze the characteristics of each brand, target population and suggestions for further development. Finally, according to word frequency calculation of the proportion of different consumption hot spots, we find out the most important direction of consumers. The research results can be used as a reference to improve the use experience of domestic cosmetics brands and further develop local brands.

## Keywords

Text Mining, Domestic Make-Up, Focus of Consumption

---

# 基于Python文本挖掘的消费者对中国产彩妆品牌评价的分析

林琳, 顾洁, 焦阳, 阎虎勤

厦门国家会计学院, 福建 厦门  
Email: 903310185@qq.com

收稿日期: 2020年4月1日; 录用日期: 2020年4月15日; 发布日期: 2020年4月22日

## 摘要

通过应用Python BeautifulSoup大数据挖掘方法、Python Wordcloud词频分析方法、以及Python Jieba 汉字词库分析方法，本文根据各大平台中消费者的反馈，分析了国产彩妆品牌存在的问题及消费者最看重的消费热点。通过选取定位不同的3个国产彩妆品牌包括橘朵、完美日记、MAOGEPING作为参考。利用大量的评论进行文本挖掘、统计词频、分析消费热点，进而分析每个品牌的特点、目标人群及能够进一步发展的建议。最后，根据词频计算不同消费热点所占的比重，以找出消费者最看重的方向。研究结果可作为参考，以此改进国产彩妆品牌的使用体验，进一步发展本地品牌。

## 关键词

文本挖掘，国产彩妆，消费热点

Copyright © 2020 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

近几年，中国美妆市场增长势头强劲，中国成为全球第二大化妆品消费国。一些小众新锐品牌频频爆红，各种之前被嫌弃的国产化妆品频繁出现在各大“种草”平台，更存在国产品牌谈力度甚至力压国际知名品牌的态势。高颜值、高性价比、快速迭代的“国潮”文化持续推动着国产彩妆不断升级，短视频和在线直播等社交工具可视化技术和交互共享需求持续推动着中国彩妆品的热销。这些“新国产美妆”品牌诞生于线上，成立时间或走红时间均不足三年，但天猫旗舰店的单品月销量都达到几万以上的级别，国货美妆热度空前。尽管如此，国内市场对新近崛起的美妆品牌的质疑声依旧不断，从抄袭嫌疑，到纯贴牌模式、没有技术，初期“平价替代”模式已经难以满足品牌长远发展的需求，在竞争不断加剧的市场中，国产品牌的研发能力和品牌影响力上仍与国际大牌存在巨大差距。在各种媒体与社交软件中，消费者对各个国产品牌的产品质量仍存在褒贬不一的声音。消费者对国货彩妆的态度到底怎么样，消费者的评判标准和关注热点是什么，哪些因素可能对营销品牌的口碑产生影响，这些问题在市场分析中都是必须回答的。

以往，由于受分析工具的限制，商家对于消费者市场消费行为的分析并不容易。一是收集数据难，对于网络上散见于不同网页和论坛上的消费意见寻找、查看、分析都很不容易。二是数据统计分析难，面对海量的数据信息，难以抓住要领。三是信息识别难，特别是彩妆品牌商品，中文外文夹杂在一起，处理起来颇费周折。开源计算机语言 Python 语言正好提供了解决这三个问题的工具方法。其中，Beautifulsoup 方法，主要用于网页数据爬虫，它可以帮助我们将不同网页上的信息转化为文本信息，并且通过文本挖掘来处理。Wordcloud 词云方法，专门处理文本词汇，它可以将文本中不同的词汇按照出现的频率进行排列和统计，那些出现频次较高的词汇一般表现了消费者所关注的焦点问题。Jieba 汉字库方法专门用于汉字词汇处理，由于汉字的字符之间没有空格将其分隔开来，所以在词频分析时需要专门识别汉字词库的方法。所以，将 Python 的 BeautifulSoup + Wordcloud + Jieba 方法应用于市场分析，可以很好地帮助我们理解和分析消费者的消费行为。

本文选取了几个具有代表性的国产彩妆品牌的网络评论进行文本处理。先提取高频词，分析消费者的关注热点，对几个品牌横向比较的同时，在内部纵向比较消费者对哪个消费属性比较敏感，从而得以在大数据的支持下探讨消费者对国货彩妆品牌的总体态度。

## 2. 研究背景与文献综述

### 2.1. 研究背景

本文选取的第一个品牌是橘朵；橘朵所属上海橘宜化妆品有限公司，成立于2016年4月5日。2017年4月进驻，2018年8月在天猫开设旗舰店。在2018年有过两轮投资，分别是顺为资本和新宜资本。自橘朵在天猫上线之后的销售数据都呈上升趋势，尤其在经历2018年双十一之后，销售数量大幅上升。第二个品牌是完美日记；完美日记的成立于2016年，创始人黄锦峰是逸仙电商创始人的CEO，也是为御泥坊前CEO。其主要针对的消费人群为学生党或刚迈入职场的年轻女性，短短三年便在美妆市场占据了巨大份额，近日完美日记完成了最新一轮的融资，其融资估值高达10亿美元，稳坐国产第一彩妆品牌的宝座。我们选取的第三个国产美妆品牌是MAOGEPING，取自其创始人毛戈平的拼音。毛戈平先生是国内著名化妆师，被誉为魔术化妆师，2000年他倾力创建了美妆品牌MAOGEPING。该品牌致力于秉承光影美学理念，针对全球女性不同的面部特征、肤质与肤色，提供多种产品及色彩解决方案，并为每位女性提供由专业彩妆师所打造的量身定制妆容。虽然MAOGEPING彩妆品牌是国内本土为数不多的高端彩妆品牌，但是这个品牌之前并不是如此的广为人知。在近三年的国货风潮中，毛戈平先生令人赞叹的化妆技术加上多渠道的推广营销，MAOGEPING才真正走进大众视野。

### 2.2. 文献综述

我国学者从不同角度对目前国产彩妆的营销做了分析。例如刘慧(2015) [1]是通过挖掘分析公司外部环境中的威胁和机会，以及内部环境中的优势和劣势来对营销策略现状进行分析。张茜(2018) [2]是通过资生堂的案例来探讨我国彩妆市场该如何完善营销模式。李天淋、胡健(2019) [3]通过对新零售特征和彩妆行业的特性探索，探讨彩妆行业在新零售的推动下要如何做出改变才能借势消费升级吸引消费者。钱丽艳、陈雅婷(2019) [4]分析了国产化妆品“种草”营销本身存在的问题及国家政策，并对“种草”营销引导方面提出了相应的对策。陈红、金秀玲(2019) [5]在消费者行为视角下来进行Z世代消费行为的剖析，并结合产品策略、价格策略、渠道策略等营销方式寻找适合化妆品行业的营销策略。王灿(2019) [6]从营销渠道的基本理论入手，梳理了国产化妆品企业营销渠道混乱与资源浪费的困境，最后提出了细分化、精准化、融合化的渠道设计意见。

在基于Python文本挖掘的应用上，我国也有许多学者进行了不同方面的探讨。周建、刘炎宝、刘佳佳(2020) [7]为了解国内外情感分析领域的研究状况，揭示该领域的知识结构、研究热点与发展动态，借助各类软件，对数据库收录的以情感分析为主题的相关文献进行计量分析与知识图谱绘制，张敏(2020) [8]以京东商城某品牌热水器的购买用户文本评价数据为基础，通过构建主题模型进行主题的提取，实现对其倾向性判断以及所隐藏信息的挖掘与分析，为生产商提供指导性建议。拓凯渊(2020) [9]运用python网络爬虫技术对美团网上陕西省六个最具代表性的5A级景区的评论进行爬取，然后通过数据进行处理，挖掘出游客们对陕西省景区评价隐含的语义，为旅游景点管理者提供决策依据。敖长林、李凤佼、许荔珊、孙宝生(2020) [10]采用两种典型的文本分析方法，分析游客对哈尔滨6个冰雪旅游景点的认知形象、情感形象和总体形象感知，丰富了旅游大数据研究方法，为哈尔滨市冰雪旅游形象塑造提供建议。

### 3. 数据描述和方法

#### 3.1. 数据选取及预处理

我们使用数据采集器从微博、淘宝和豆瓣等平台采集所需的评论词条组成数据。本文一共采集 284 条分别关于完美日记、毛戈平和橘朵三个国产彩妆品牌的数据用 jieba 软件预处理。首先将采集到的评论进行分词处理，即依据语义将一句话切分成一个个的词，例如，“我觉得很好用”分词成：“我”“觉得”“很”“好用”。虽然分成了一个词，但是句子里有一些没有意义的停词，如例子中的“觉得”，要将其过滤掉，该过程使用一个含有 1000 多个停词的词库，可以实现停词过滤。分词这一步尤为重要，分词效果的好坏直接影响了后面分析的准确性。依据 jieba 里自带的词典，文本被分为一个个的词。最后统计了三个品牌的评论词频。截取排序前 20 的词，见表 1。依据统计的各品牌的词频制作词云图，让统计结果更加直观，便于寻找主要属性特征。筛选掉无意义的高频词，制作得到词云图，如图 1~3 所示。

**Table 1.** Comment frequency of three brands

**表 1.** 三个品牌的评论词频

词序	完美日记		橘朵		MAOGEPING	
	词	频数	词	频数	词	频数
1	眼影	46	眼影	62	粉底	33
2	营销	27	粉丝	29	化妆	27
3	好看	26	消费者	26	故宫	18
4	颜色	23	小红书	22	手法	16
5	质量	21	明星	21	特别	16
6	一般	20	美妆	16	性价比	16
7	腮红	20	营销	16	感觉	15
8	飞粉	18	微博	16	效果	15
9	不好	17	小狗	16	不会	15
10	单色	17	动物	16	粉膏	14
11	好用	15	好用	15	上妆	13
12	价格	14	李佳琦	15	视频	13
13	晕染	13	彩妆	14	毛孔	12
14	睫毛膏	12	联名	14	比较	12
15	亮片	11	KOL	14	好用	12
16	粉质	10	种草	14	使用	12
17	持久	10	合作	12	产品	12
18	口红	9	线下	12	高光	12
19	显色	9	天猫	11	东方	12
20	眉膏	8	渠道	11	喜欢	11



Figure 1. Word cloud diagram of Judydoll  
图 1. 橘朵词云图



Figure 2. Word cloud diagram of Perfect Diary  
图 2. 完美日记词云图



Figure 3. Word cloud diagram of MAOGEPING  
图 3. MAOGEPING 词云图

### 3.1.1. 橘朵热点词分析

从图 1 中可以看出，消费者对于橘朵评论的关键词基本集中在眼影、质量、价格、营销等字眼上，可以看出橘朵的产品定位于年轻女性，特性就是走性价比、平价路线。

眼影可以说是橘朵的当家产品。在之前各大品牌的眼影盘都集中在 9 色或者 12 色，而橘朵却另辟蹊径，避开了这大热品类，把焦点放在了单色眼影和 4 眼影盘上。相比 9 色或者 12 色的眼影盘合集，单色眼影的成本造价势必会降低很多，而且橘朵也因为单色眼影的成功推出避开了彩妆眼影盘合集的这片红海，通过低单价吸引到了第一批学生党、小白领、化妆小白等群体。

消费者同样十分关心橘朵的营销。橘朵在网络上的营销是非常成功的，其原因是其巧妙的利用了小红书和微博上的 KOL 进行营销。我们知道网络 KOL 上的账号级别是金字塔型的，分为四级，从上到下依次是明星、大 V、中小 V 和草根。橘朵的投放基本都集中在腰部及尾部的 KOL，而这种小成本的不断投放试错能最快的根据消费者反馈来进行内容投放上的调整输出，而小 KOL 对于内容的把握和沟通上来说



品牌更有可控性。除此之外，橘朵在双 11 和 618 这两大淘宝、天猫的重大节点开始前的两个多月前已经在开始慢慢预热。对比一些品牌在大节点前几天才开始集中投放做声量，成长于电商的橘朵已经对于线上的套路和打法十分得心应手。

从词频分析中可以看出，对于橘朵这类新兴国产彩妆来说，价格定位十分重要。而橘朵是如何实现产品的低单价呢？其独特之处就是发现了单色眼影的这个趋势，然后匹配到了目标人群，决策上的超低单价，让其选择度更高，所以在目标人群确定之后，再开始讲故事他们是乐于接受的。再者就是高效率的媒体投放和分发，不断试错找到一批转化率不错的 KOL 让他们把内容分发出去，这里面所有的环节串联起来需要足够快。

### 3.1.2. 完美日记热点词分析

如图 2，完美日记的飞速走红依靠的是出色的全网内容营销，完美日记以 KOL 带动全民种草，通过直播、短视频、图文等多种形式的灵活运用，在以小红书、抖音、微博为主的社交媒体平台，与消费者进行持续且深入的沟通，从而促成高效转化。不管是眼影还是口红，完美日记都邀请了明星与头部 KOL 如李佳琦、小猪姐姐等演绎核心创意视频及海报，众多美妆博主发挥创意产出妆教内容，进行广泛种草，打造产品口碑。其次，完美日记持续推动跨界 IP 联名，联合大英博物馆、探索频道，以及中国国家地理等品牌推出联名款眼影盘，聚焦于品牌与 IP 内在的文化价值，以这种文化价值内涵来赋能产品，更加能够引起消费者情感共鸣的艺术表达，也由此积累品牌价值资产。

品牌通过营销迅速发展的同时，也带来了一些负面效果。越来越多的消费者对各大平台频繁出现的推广感到疲倦和厌烦，同时国产低价的优势逐渐失去其吸引力，消费者对产品质量仍存在很大争议，回购率和好评率较国际品牌仍有很大差距。当单纯依靠网红“种草”越来越困难时，如何提升品牌影响力和培养忠实顾客便成为完美日记日后需要面对的难题。

### 3.1.3. MAOGEPING 热点词分析

从图 3 中可以看出，针对 MAOGEPING 品牌的评论两极分化比较严重。一部分用户表示该品牌的产品非常符合自己的期望，会进行重复购买以及延伸购买其他产品和推荐。另外一部分用户则表示产品和自己的预期相差甚远，不能满足他们的需求。有用户表示 MAOGEPING 性价比太低，同等价格自己更愿意购买其他品牌或者增加预算购买欧美大牌彩妆。也有用户表示希望 MAOGEPING 同步提升价格和质量，栖身一流彩妆品牌。

针对用户们的评价以及所生成的词云图，可以看到有相当大的群体对 MAOGEPING 的满意度很高，这一部分是品牌的忠实顾客。还有一部分用户会选择性的使用某些产品，但是对另外的产品则不会尝试。而对品牌的负面评价主要集中在使用感不好，性价比不高以及需要一定的化妆技巧和手法才能画出想要的效果。这样的评价是可以解释的。通过 MAOGEPING 彩妆的发展历史不难看出，之前它主要服务于专业化妆师，后来才更多的面向大众。即便如此，它面向的很大一部分群体依然是那些有一定化妆基础的、能够在化妆上投入一定的时间的人。同时，MAOGEPING 品牌的理念是创造东方新时尚。而现在很多客户受欧美和日韩妆的影响，不是很喜欢古典的东方妆感。这也在一定程度上解释了为什么会有一部分用户觉得无法达到预期的效果。

针对 MAOGEPING 彩妆的发展，通过词云图的提取分析，该品牌可以从以下两个方面着手：一方面针对专业化妆师的产品，可以进一步提升质量和使用效果。另一方面 MAOGEPING 如果想做成国际一流彩妆，提升用户的使用感是非常重要的。MAOGEPING 可以开拓一系列无门槛产品，使其不拘泥于专业手法和精力的限制，普通的涂抹同样可以达到预期效果。

#### 4. 热词情感分析

利用 python 抓取近期有关三大品牌评论的高频词，并合并成词频表。这些高频词可以清晰地展示每个品牌自身的情况和消费者对其的感受。为了进一步了解在美妆消费市场中，消费者更看重品牌哪一方面的品质，我们将这些高频词进行分类。关于性价比、价格相关的描述词分类为“价格定位”。“小红书”、“合作”、“视频”等词汇分类为“营销推广”。具体的美妆产品词汇和使用感描述词汇归类为“产品质量”。利用其他 IP 的影响力开发的联名产品或品牌理念等归类为“设计理念”，最终的分类结果如表 2。

**Table 2.** Classification and statistics of high frequency words

**表 2.** 高频词分类统计

属性	橘朵		完美日记		MAOGEPING		总计	
	词频	比重	词频	比重	词频	比重	词频	比重
产品质量	225	0.65	145	0.39	193	0.63	563	0.55
设计理念	43	0.12	58	0.12	72	0.24	1733	0.17
价格定位	51	0.15	11	0.03	28	0.09	90	0.09
营销推广	27	0.08	158	0.42	14	0.04	198	0.19

虽然现代消费群体在化妆品消费方面大体上是很理性的，但也不能够否认从众消费心理的存在。消费者通过自媒体者的 KOL 营销后，会判断该产品是否符合自己的需求并决定是否购人，但是由于自媒体者在直播测评中并未真实反产品性能，且评论区有大量枪手，导致消费者并没有很好的分辨出产品的适用与否，直接跟随着大流购买该产品，或是仅仅因为喜欢该博主而进行盲目消费。这就导致了许多产品无法达到消费者预期的情况。

且随着消费者日趋理智以及对产品的深入了解，我们可以看到很多消费者对本土品牌有了消极的评论。如脱妆、暗沉、飞粉，眼影难上色，晕妆等都是消费者反馈的重灾区，产品品质良莠不齐。因此不论营销环节怎样精心布置，营销的终极目的还是要塑造品牌形象，提升产品销量，那么产品作为直接接触消费者的“终端”是否能得到消费者的认可，就成了企业舆论正向传播与否的关键影响因素，企业在品牌传播方面的收益与产品研发的投入高低息息相关。如同常常被诟病的综艺抄袭现象一样，从他国抄袭而来的产品永远也成不了行业的领头羊，最多只能跟在大牌后面作为廉价替代品。

想要让国货化妆品受到国内外消费者的认可，需要本土企业在保证质量的同时积极研发新产品，同时设立一个良好的企业形象，才能增加消费者分享、回购的几率，促进国内彩妆市场的繁荣。

#### 5. 总结

通过对网络文本数据进行挖掘和统计分析，得出游客对古镇旅游态度及相关因素的结论如下：

消费者选择美妆品牌，关注度最高的是产品质量，其次是产品设计理念。尽管树立美妆品牌的知名度需要全面的战略和营销手段，但如果无法保证产品质量则很难提升回购率并保持客户忠诚度。尤其在这个趋于饱和的市场中，美妆产品的可替代性非常高，且并没有很高的市场进入门槛，尽管解决了产品价格和包装问题，质量堪忧的产品和品牌很快就可以被替代和淘汰。与长期占据美妆市场的国际大牌相比，国产新品牌在研发能力上仍存在很大差距。要想做得比国际大牌更好用，不能仅在营销层面踩一捧一，更不能只局限于“平价替代”的层面，而是要从质量和使用体验都得到提升。

其次，盲目追求性价比对品牌发展未必是好事。如果品牌对自己的产品有充足的自信，它完全不需

要只用价格战的方式来销售，更不会用平价、便宜等标签来做自己质量一般的借口。例如，毛戈平作为国产品牌凭借过硬的产品质量和专柜服务体验，尽管有些产品的价格比很多一线国际大牌还要高，仍然保持着很好的销量和口碑。而橘朵一直走平价替代的高性价比路线，在产品质量上却始终难以超越其模仿对象，给消费者留下的抄袭、廉价等刻板印象难以迅速扭转，以至于如今也难以获得更高的品牌溢价，消费者对其逐渐涨价的行为并不买账。因此，产品质量才是品牌站稳脚跟的基础。

依靠互联网起家的国产美妆品牌，如完美日记在线上的粉丝量基本已经触顶，如今需要通过线下的渠道拓展新客户，与此同时，线下门店也是新品牌提升互动和体验的重要渠道。线下的供应链和营销方式也和线上存在巨大的差异，传统品牌已经在线下形成的渠道、市场、技术等壁垒也很难逾越。新锐化妆品牌一直是轻资产的运行模式，主要集中在品牌建设和资源整合，在线下销售并没有太大的优势。如何既保持住国产品牌的优势--高频率的上新速度，同时提高供应链能力、库存管理水平，将成为全新的挑战。

## 基金项目

本论文得到了厦门国家会计学院 2019 年“云顶课题：Python 财务数据分析”项目的支持。

## 参考文献

- [1] 刘慧. J 化妆品代理公司营销策略研究[D]: [硕士学位论文]. 哈尔滨: 哈尔滨工业大学, 2015.
- [2] 张茜. 创新中的资生堂要如何重塑中国市场[J]. 成功营销, 2018(Z3): 12-13.
- [3] 李天淋, 胡健. 关于彩妆业与新零售的分析与思考[J]. 现代商业, 2019(17): 17-19.
- [4] 钱丽艳, 陈雅婷. 国产化妆品“种草”营销的营销策略分析[J]. 现代营销(信息版), 2019(11): 226-227.
- [5] 陈红, 金秀玲. 基于消费者行为下的 Z 世代化妆品营销策略研究[J]. 现代营销(信息版), 2019(11): 227-228.
- [6] 王灿. 融合、精准、细化——对国产化妆品企业营销渠道设计的思考[J]. 艺术科技, 2019, 32(4): 227+274.
- [7] 周建, 刘炎宝, 刘佳佳. 情感分析研究的知识结构及热点前沿探析[J]. 情报学报, 2020, 39(1): 111-124.
- [8] 张敏. 基于文本挖掘的电商评论情感分析[J]. 产业与科技论坛, 2020, 19(2): 63-64.
- [9] 拓凯渊. 基于文本挖掘的线上旅游评论分析[J]. 西部皮革, 2020, 42(2): 121-123.
- [10] 敖长林, 李凤佼, 许荔珊, 孙宝生. 基于网络文本挖掘的冰雪旅游形象感知研究——以哈尔滨市为例[J]. 数学的实践与认识, 2020, 50(1): 44-54.