

# 基于多特征的柳州螺蛳粉购买评论情感倾向性分析

余 婷, 黄李韦

广西科技大学理学院, 广西 柳州

收稿日期: 2021年10月18日; 录用日期: 2021年12月29日; 发布日期: 2022年1月6日

---

## 摘 要

新冠疫情催生了“宅经济”，袋装柳州螺蛳粉强势崛起，成为全国热销的方便食品之一。目前网购平台上开设的螺蛳粉网店高达1.2万家，在竞争如此激烈的情况下，挖掘大众的喜好变得尤为重要。本文拟采用关键词提取、词云图可视化分析、LDA主题分析以及评论分词主题模型等情感分析方法对螺蛳粉购买评论数据进行分析，挖掘大众对螺蛳粉的喜好，为商家改进商品提供有利参考，提升产品的用户体验，从而提高销量。

## 关键词

文本挖掘, 情感分析, 评论分词模型, 螺蛳粉

---

# Analysis on Emotional Tendency of Liuzhou Snail Powder Purchase Comments Based on Multi Features

Ting Yu, Liwei Huang

School of Sciences, Guangxi University of Science and Technology, Liuzhou Guangxi

Received: Oct. 18<sup>th</sup>, 2021; accepted: Dec. 29<sup>th</sup>, 2021; published: Jan. 6<sup>th</sup>, 2022

---

## Abstract

COVID-19 has spawned the “home economy”. The bagged Liuzhou snail powder has risen vigorously and has become one of the most convenient foods in the country. At present, there are

**12,000 snail powder online stores opened on the online shopping platform. In the case of such fierce competition, it is particularly important to tap the public's preferences. This paper intends to use emotional analysis methods such as keyword extraction, word cloud visualization analysis, LDA topic analysis and comment word segmentation topic model to analyze the snail powder purchase comment data, mine the public's preference for snail powder, provide favorable reference for merchants to improve products, improve product user experience, and improve sales.**

## Keywords

Text Mining, Emotional Analysis, Comment Word Segmentation Model, Snail Powder

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

在国家高度重视创新,大力支持创业背景下[1],以广西柳州螺蛳粉为代表的方便食品产业迅速崛起。有关数据显示,2016年柳州螺蛳粉年产值16亿元;2017年上升至30亿元;2018年上升至45亿元;2019年全产业链产值突增至135亿元,其中预包装螺蛳粉占比约为46%;因受新冠疫情影响,2020年预包装螺蛳粉销量持续增长,截至2020年10月,柳州预包装螺蛳粉产值已超87亿元。目前网购平台开设的螺蛳粉网店高达1.2万家,在阿里巴巴米粉特产类销量排名第一。市场竞争日益激烈,如何满足大众的喜好变得尤为重要。面对众多口味及多种配料的螺蛳粉,我们利用多特征情感分析的方法从大量的螺蛳粉购买评论中挖掘出大众的喜好,同时挖掘出商品的不足,为商家改进商品提供有利参考。

近年来,文本情感分析的研究成果越来越多。例如,王亚[2]从京东商城爬取了几种手机的评论数据,对文本进行预处理后,通过建立情感分类模型对其进行情感分析,并提取出对商家改进产品有用的信息。毕春光[3]等提出基于评论特征词构建评论分类模型以及基于朴素贝叶斯的特征情感分类器,从用户的评论出发,对人参的多个特征进行情感分析。Ghose [4]等利用构造的情感分类器对预处理后的文本进行情感分类,在用户进行购买时可以起到参考作用。杨鑫[5]等通过基于领域词典的分析法对贵阳名宿评论数据进行情感分析,同时利用LDA主题挖掘,分析民宿评论中正负面情感主题并分析其原因,为民宿管理者做出相应决策提供理论性依据。周欢[6]等通过对评论文本进行词频分析和LDA主题模型分析来对商品的物流服务质量进行研究,并对商家或企业提出相关建议。张公让[7]等利用情感分析方法挖掘出影响客户服务评价和满意度的关键因素,为客户情感关键影响因素识别提供数学科学的研究范式。

## 2. 螺蛳粉购买评论的特征分析

特征分析是通过评论文本进行分析,从大量的评论文本中分析出用户重点关注的特征,从而对这些关注度高的特征进行情感分析。本文采用TextRank算法、词云图的特征可视化、LDA主题特征分析三种特征分析方法找出用户对商品特征关注度的分布情况,对关注度高的商品特征进行情感分析研究。

### 2.1. 数据准备

利用八爪鱼软件爬取了京东商城上客户对螺蛳粉的评论,并选取了国内排名前八的螺蛳粉品牌,包括好欢螺、李子柒、螺霸王等,爬取的评论总数为20,640条。通过对原始评论数据进行去除重复评论、

机械压缩去词、删除短句评论等预处理操作, 得到 14,400 条有效数据。按照标点符号对评论进行分割, 每条评论可以分成若干个子句, 将其格式化保存。

## 2.2. 基于 TextRank 算法的文本特征提取

采用 TextRank 方法提取评论关键词, 得出顾客关于螺蛳粉评论关注度较高的前 20 个特征词如下:

{螺蛳粉、味道、牌子、感觉、物流、价格、口味、配料、酸笋、腐竹、配料、汤、品牌、速度、朋友、辣椒油、辣、份量、青菜、质量、性价比}。

其中, 除了“螺蛳粉”外, “味道”、“牌子”、“感觉”、“物流”和“价格”等词脱颖而出, 说明这几个特征词在螺蛳粉的评论文本中具有比较重要的意义。因此, 顾客非常关注螺蛳粉的味道、物流信息以及螺蛳粉的价格等问题。同时可以看到这 20 个特征词中关于配料的词就出现了 5 次, 说明顾客都比较关注螺蛳粉的配料, 比如“酸笋”“腐竹”“青菜”等配料的质量及份量。通过对 8 家品牌的螺蛳粉评论进行关键词提取分析, 可以得到影响顾客满意度的因素可以分为四类, 分别是味道、物流、配料以及价格。

## 2.3. 基于词云图的特征可视化分析

为直观地展现出购买螺蛳粉顾客关注的焦点和主题, 本文通过绘制词云图实现特征的可视化。将所有品牌的螺蛳粉汇聚在一起生成词云图, 如图 1 所示。

在图 1 中, 大部分都是正向的信息, 比如“好吃”、“不错”、“喜欢”、“回购”等。并且注意到影响顾客满意度的因素主要包括味道、配料、物流以及包装, 由于螺蛳粉会有很多配料, 顾客可以根据自己的口味进行选择, 所以相对于粉来说, 顾客会更加关注配料里面有什么, 除了螺蛳粉本身的因素外, 物流和包装也会是影响顾客满意度的重要因素。



Figure 1. Total comment cloud

图 1. 总评论词云图

由于总评论数量较大, 能展现的关键词信息不多, 对单个品牌的螺蛳粉评论进行可视化分析可以挖掘出更多的有用信息。于是通过分别绘制 8 家螺蛳粉的评论词云图, 得出 8 家螺蛳粉顾客评论词云图中出现频数前 14 的特征词, 如表 1 所示。

首先, 从整体上来看, 排名前 14 的特征词都是积极正面的, 说明在这 8 家螺蛳粉的评论中好评是占大部分的。其次, 8 家螺蛳粉排名前 3 的特征词都是“好吃”、“味道”及“螺蛳粉”, 由于螺蛳粉是食品, 顾客对于食品关注的因素无疑是它的味道, 所以它好不好吃就成了顾客首要关注的因素。从每个

特征词在各品牌下出现的次数可以发现,“包装”在每个品牌下都有出现,说明顾客除了关注螺蛳粉的味道外,最关注的是它的包装。其次有“物流”、“配料”和“价格”。通过对 8 家品牌的螺蛳粉评论进行词云图的可视化分析,可以得到影响顾客满意度的因素可以分为五类,分别是味道、包装、物流、配料以及价格。

**Table 1.** The top 14 characteristic words of snail powder frequency of each brand

**表 1.** 各品牌螺蛳粉频数前 14 的特征词

品牌	好欢螺	家柳	李子柒	柳江人家	柳全	螺霸王	螺状元	嘻螺会
1	好吃	好吃	好吃	好吃	好吃	好吃	好吃	好吃
2	味道	味道	味道	味道	味道	味道	味道	味道
3	喜欢	非常	非常	不错	不错	喜欢	非常	不错
4	非常	不错	李子柒	非常	京东	非常	不错	非常
5	好欢螺	喜欢	喜欢	喜欢	非常	螺霸王	喜欢	喜欢
6	特别	京东	不错	特别	喜欢	不错	京东	可以
7	不错	包装	特别	可以	特别	京东	可以	价格
8	牌子	特别	包装	京东	牌子	特别	活动	购买
9	可以	物流	可以	柳江人家	购买	可以	包装	回购
10	京东	很快	京东	牌子	可以	牌子	回购	包装
11	回购	可以	配料	回购	物流	配料	配料	物流
12	推荐	方便	没有	包装	包装	没有	物流	推荐
13	很多	购买	购买	比较	柳全	鹌鹑蛋	螺状元	比较
14	包装	速度	回购	正宗	很多	购买	快递	实惠

#### 2.4. 基于 LDA 主题模型的特征分析

通过特征提取和词云图分析后,可以找出一些影响客户情感倾向的因素但没有进一步的挖掘评论内容的语义,而主题模型是语义挖掘的重要利器。其中,LDA 主题模型,是一种文档主题生成模型,包含文档、主题和词汇三层结构,故也被称为三层贝叶斯概率模型,它是众多主题模型中最方便有效的模型之一。通过 LDA 主题模型的特征分析,可以发现文本中潜在的主题信息。本文选择生成 5 个主题并提取 16 个特征词。LDA 主题分析生成的主题和特征词如表 2 所示。

通过对表 2 中特征词分析,第一个主题反映了顾客对螺蛳粉的包装和质量很满意并且推荐下次购买;第二个主题反映了顾客在螺霸王和李子柒之间更推荐李子柒螺蛳粉;第三个主题反映顾客对京东快递送货速度很满意并且有时候搞活动很优惠;第四个主题反映配料份量很足;第五个主题反映大部分顾客会选择回购并且认为螺蛳粉很便利。综合以上五个主题和关键词来看,顾客比较关注螺蛳粉的包装、送货速度、价格、配料的份量以及螺蛳粉带给他们的便利等问题。

**Table 2.** LDA subject analysis keywords  
**表 2.** LDA 主题分析关键词

主题一	主题二	主题三	主题四	主题五
非常	螺蛳粉	京东	配料	非常
不错	牌子	物流	里面	螺蛳粉
包装	喜欢	快递	腐竹	回购
值得	还是	很快	辣椒油	喜欢
满意	没有	活动	非常	购买
收到	特别	不错	不错	已经
喜欢	不错	非常	特别	真的
好评	但是	速度	酸笋	多次
发货	螺霸王	下单	一个	特别
购买	感觉	价格	鹌鹑蛋	这个
质量	之前	包装	还有	方便
特别	品牌	第二天	一包	正宗
终于	第一次	送货	很多	推荐
宝贝	李子柒	小哥	份量	不错
推荐	比较	优惠	比较	继续
下次	推荐	收到	米粉	每次

综合 TextRank 算法特征提取、词云图可视化分析以及 LDA 主题分析的结果, 可以得到影响顾客满意度的因素主要有味道、物流、包装、价格以及配料。

### 3. 基于多项特征的情感分析

#### 3.1. 评论分类模型

评论分类模型是利用子特征主题词库, 从所有评论中筛选出属于各个主特征的相关评论, 将所有子评论按照主特征进行分类。评论分类模型流程如图 2 所示。

#### 3.2. 子特征主题词库构建

因为表示某一特征子评论中可能不会出现那个特征词, 比如“发货快”表示的是“物流”这一特征, 但是并没有出现“物流”这个词。所以在对所有分割后的螺蛳粉子评论进行分类提取的过程中需要根据五大主特征(味道、物流、包装、价格、配料)通过计算机辅助, 构建出与之对应的子特征主题词库, 如表 3 所示。

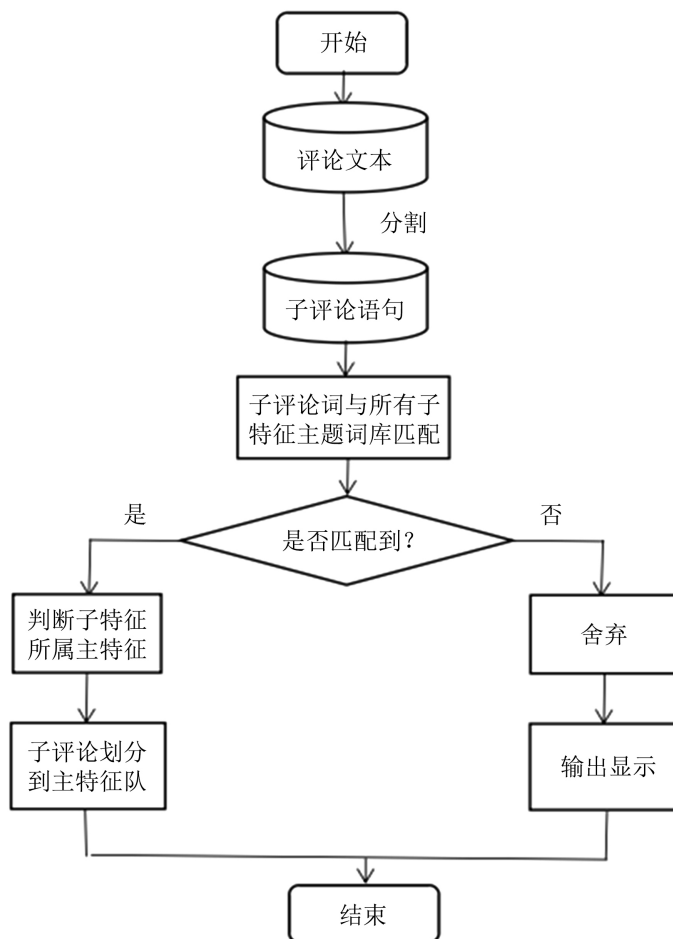


Figure 2. Comment classification flow chart  
图 2. 评论分类流程图

Table 3. Main feature—sub feature subject words

表 3. 主特征——子特征主题词

主特征	子特征主题词
味道	味道、口味、辣、辣味、口感、臭、臭臭、臭味、臭气、原味、味儿、调味、气味、香、香味、香气、劲道、酸、酸味、有嚼劲、弹性
物流	物流、小哥、快递、物流业、物流配送、配送、收到、送到、速度、太快、快速、迅捷、运货、运输、配送、发货、到货、太慢、仓库、京东快递、顺丰、圆通、申通、韵达、中通、汇通
价格	价格、价钱、钱、价位、定价、消费、实惠、经济、划算、价实、售价、降价、昂贵、较贵、便宜、价廉、市价、标价、低廉、低价、要价、性价比、超值、起价、优惠
配料	配料、辅料、料、香料、佐料、小菜、酸笋、臭笋、豆皮、榨菜腐竹、香菇、香菜、葱、汤、香肠、牛肉、鱼丸、螺蛳、螺丝、海带、青菜、白菜、蔬菜、菜、生菜、菠菜、萝卜、空心菜、酸菜、辣椒油、鸡蛋、卤蛋、豆腐泡、臭豆腐、豆腐皮、汤包、醋、酱、辣油、汤料、肉、花生、螺肉、份量、量足、面条、粉条、火腿、金针菇
包装	包装、简陋、粗陋、严密、漏气、丑、好看、精美、充气、袋装、礼盒装、塑料包装、外包装、袋子、箱子、塑料袋



根据构建的子特征主题词词库, 按照五个主要特征对子评论进行分类, 得到的分类结果如下表 4 所示。

**Table 4.** Number of sub comments of the main feature

**表 4.** 主特征的子评论数量

主特征	子评论数量
味道	15,354
物流	3731
价格	2375
配料	17,255
包装	1869

### 3.3. 顾客评论的情感倾向分析

本文进行情感倾向分析时是通过调用百度 AI 自然语言处理板块的情感分析定制 API 来分析评论的情感极性。该 API 调用百度基于深度学习训练的情感倾向性分析模型, 可以自动学习深层次的语义及句法特征, 具备较高泛化能力。在实际应用中, 我们能在通用模型的基础上, 导入一批预先处理好的的正/负向语料标注数据, 对通用模型加以优化, 提高模型在袋装螺蛳粉领域的情感分析效果。

通过 python 随机从经过剔除规则筛选后的文本数据中选取了 2040 条数据进行人工标注正负向情感。对一条评论, 如果评论者表达的内容情感为积极正向的标注为 1, 消极负面的标注为 0, 标注结果如表 5 所示。由于原数据中正负评论数量有较大的差距, 因此本文采用过采样的方法进行处理, 使得两种极性的语料数量达到平衡[8]。将过采样后的数据导入模型进行训练, 训练集和测试集的比为 7:3, 最终得到的模型精度在 92.88%, 基本符合情感倾向性分析的精度要求, 可以加以应用。

**Table 5.** Emotional distribution of comment text

**表 5.** 评论文本情感分布

情感词典	类别	数量	总计
原数据	正向	1956	2040
	负向	84	
过采样	正向	1956	3912
	负向	1956	

利用训练好的模型对各个主特征下的子评论进行情感分类, 分类结果如表 6 所示。其中, 从评论数的角度看, “配料”的正面评论数最多, “包装”的正面评论数最少; 从好评率的角度看, “配料”占比 97.12%, “味道”占比 95.81%, “价格”占比 92%, “包装”占比 83.38%, 各特征的好评率均达到了较高水平。可能是因为选取的购物平台是京东的原因, “物流”的好评数达到了 100%。

Table 6. Main feature comment emotion distribution

表 6. 主特征评论情感分布

主特征	正向情感	负向情感	好评率
包装	1435	286	83.38%
价格	2083	181	92%
配料	16,335	483	97.12%
味道	14,366	627	95.81%
物流	3504	0	100%

### 3.4. 各项特征下的词云图分析

为了了解各个主要特征的具体情况,采用词云图可视化分析方法对他们的正负向评论进行分析。

图 3 是包装的正负向评论的词云图,从包装的正面词云图可以很明显地看到“独立”、“好看”、“精美”、“完整”、“礼盒装”等描述包装的词汇,可以看出顾客对于螺蛳粉精美独立的包装很满意。然而在包装的负面评论词云图中看到“很丑”、“闻到”、“臭臭”等字眼,反映出每个顾客的审美是不一样的,厂家可以设计不同风格的包装以供顾客选择,同时加强包装的封密性,以免用户闻到气味,降低产品档次。



Figure 3. Cloud chart of positive and negative comments on packaging

图 3. 包装的正负评论词云图

图 4 是价格正负向评论的词云图,从价格的正面评论的词云图可以看到“便宜”、“划算”、“性价比”、“优惠”、“活动”等特征词,反映了顾客认为螺蛳粉具有较高的性价比,活动比较划算。从价格负面词云图中可以看到“便宜”、“优惠”、“没有”、“偏高”等特征词,通过在负面评论中查看有关“便宜”的评论中,发现前面都有否定词,说明顾客想反映的是不便宜,并且可以了解到有些顾客还是认为相对于泡面的价格来说,螺蛳粉的价格还是偏高,并且价格浮动太大,这种现象可能会对螺蛳粉的产销产生负面效应,商家应该加以关注。

图 5 是味道的正负向评论的词云图,从味道的正面词云图中可以看到“好吃”、“酸笋”、“口味”、“辣椒油”、“喜欢”等特征词,表达了一部分用户可以接受螺蛳粉酸辣的口味,然而在味道的负面词云图中可以看到“清淡”、“辣椒油”、“臭味”等特征词,说明有些顾客无法接受酸笋的臭味,也有很多顾客认为辣椒油不够辣,味道清淡。





Figure 4. Cloud chart of positive and negative comments on price  
图 4. 价格的正负评论词云图



Figure 5. Cloud picture of positive and negative comments on taste  
图 5. 味道的正负评论词云图

图 6 是配料的正负向评论词云图, 从正向词云图中可以看到“好吃”、“酸笋”、“腐竹”、“花生”、“青菜”、“喜欢”等特征词, 说明顾客对酸笋、腐竹、花生、青菜比较喜爱, 同时对配料的份量也很满意。负面评论中较多的还是对酸笋的异议, 一部分顾客无法接受酸笋的臭味。



Figure 6. Cloud chart of positive and negative comments on ingredients  
图 6. 配料的正负评论词云图

图 7 是快递正向的词云图, 从该词云图中可以看到“很快”、“速度”、“到货”、“第二天”、“超快”等特征词, 反映了顾客对京东快递物流速度快的一致好评。



Figure 7. Cloud chart of positive and negative comments of express  
图 7. 快递的正负评论词云图

#### 4. 结论及建议

首先, 我们通过 TextRank 算法对螺蛳粉评论进行关键词提取、词云图对总评论及 8 家螺蛳粉评论进行可视化分析、LDA 主题分析得出影响顾客满意度的五个主要因素分别为味道、物流、包装、价格以及配料。接着通过构建评论分类模型, 将各个评论分成五个主特征下的子评论, 并调用百度 AI 自然语言处理板块的情感分析定制 API 对各主特征下的子评论进行情感分类。最后对各主特征的正负评论进行词云图可视化分析, 得出顾客在这五个方面具体的评价信息。

本文通过对螺蛳粉评论的具体分析, 我们为商家提升顾客满意度提供以下几点建议: 第一, 商家可以设计不同款式的包装以供不同要求的顾客选择; 第二, 加强包装的密封性, 以免散发出味道和降低产品档次; 第三, 控制价格的波动, 保证给顾客带来优惠的同时稳定价格; 第四, 提供更多的调料包个数选择以满足不同顾客口味的需要。

#### 基金项目

广西自然科学基金(项目号: 2018GXNSFBA281185)资助。

#### 参考文献

- [1] 于溟. 浅谈预包装螺蛳粉行业发展与企业创业切入点[J]. 商业文化, 2021(1): 97-98.
- [2] 王亚. 手机在线评论的情感分析[D]: [硕士学位论文]. 贵州: 贵州财经大学, 2020.
- [3] 毕春光, 袁帅, 等. 基于人参购买评论的多项特征情感倾向性分析研究[J]. 东北农业科学, 2020, 45(3): 92-96.
- [4] Ghose, A. and Iperiotis, P.G. (2011) Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics. *IEEE Transactions on Knowledge and Data Engineering*, **23**, 1498-1512. <https://doi.org/10.1109/TKDE.2010.188>
- [5] 杨鑫, 杨云帆, 等. 基于领域词典的民宿评论情感分析[J]. 科学技术与工程, 2020, 20(7): 2794-2800.
- [6] 周欢, 秦天琦. 基于在线评论情感分析与 LDA 的物流服务质量影响因素研究[J]. 重庆工商大学学报, 2021, 38(6): 27-38.
- [7] 张公让, 鲍超, 等. 基于评论数据的文本语义挖掘与情感分析[J]. 情报科学, 2021, 39(5): 53-61.
- [8] 曾子明, 万品玉. 基于双层注意力和 Bi-LSTM 的公共安全事件微博情感分析[J]. 情报科学, 2019, 37(6): 23-29.