

基于文本分析的水产品预制菜消费者情感倾向研究

江展蓉, 郭菊喜*

岭南师范学院数学与统计学院, 广东 湛江

收稿日期: 2025年3月23日; 录用日期: 2025年4月14日; 发布日期: 2025年4月25日

摘要

水产品预制菜具有易于保存、方便携带和食用等特点, 是当前市场上备受欢迎的一种食品, 近年来颇受关注。本文利用八爪鱼爬取京东商城上水产品预制菜的文本评论数据, 通过R语言建立LDA模型分析消费者对水产品预制菜的偏好与期望, 以期对相关行业的生产厂家和餐饮企业提供有益的参考和建议, 以促进水产品预制菜市场的健康有序发展。

关键词

水产品预制菜, 文本分析, 情感分析, LDA模型

Research on Consumer Emotional Tendency of Prepared Aquatic-Products Dishes Based on Text Analysis

Zhanrong Jiang, Juxi Guo*

School of Mathematics and Statistics, Lingnan Normal University, Zhanjiang Guangdong

Received: Mar. 23rd, 2025; accepted: Apr. 14th, 2025; published: Apr. 25th, 2025

Abstract

Prepared aquatic-products dishes are a popular food in the current market due to the characteristics of easy preservation, easy to carry and eat, and have attracted considerable attention in recent years. In this paper, the Octopus harvester was used to extract the text review data of pre-fabricated aquatic products on Jingdong Mall, and the LDA model was established through R language to analyze consumers' preferences and expectations of prefabricated aquatic products, in

order to provide useful references and suggestions for manufacturers and catering enterprises in related industries, and promote the healthy and orderly development of prepared aquatic-products dishes.

Keywords

Prepared Aquatic-Products Dishes, Text Analysis, Emotion Analysis, LDA Model

Copyright © 2025 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

随着现代人生活节奏的不断加快,人们对于食品的要求也越来越高,食材供应多元化,国人烹饪和饮食观念在逐渐发生变化,“宅经济”、“懒人经济”渐兴。在这种情况下,预制菜应运而生,成为了人们饮食生活中不可或缺的一部分。水产品预制菜作为其中的一个重要品类,为消费者提供了更加丰富的选择,同时也降低了餐饮行业的成本和人力资源压力。人们对餐饮多元化需求变化,便捷、简洁、营养、多样成为人们对餐饮的要求,加上我国大力深化农业供给侧改革,市场需求与政府政策的推动使得预制菜成为热点。然而,其行业发展仍处于初级阶段,对于水产品预制菜的研究很少且比较分散,与其他种类的预制菜相比,水产品预制菜行业存在的问题显著,其原料产地区域差异明显,价格受季节波动大,对生产装备、生产技术要求更高。我国预制菜企业数量虽多,但大规模企业的数量较少,缺乏龙头企业,行业格局分散,市场竞争较为激烈,产品出品不稳定[1]。

我国是渔业养殖大国和水产品消费大国,具备水产品消费基础,且水产品具备丰富的营养价值,未来水产品消费占比仍可持续提升。随着居民生活工作节奏加快、线下餐饮消费场景缺失、消费者对方便快捷的水产预制菜产品需求提升、对预制菜的认知度提升等因素,促进了预制菜行业蓬勃发展。根据艾媒咨询数据显示,由水产品预制菜行业市场规模及预测图,可以看出 2022 年中国 5 水产预制菜行业规模为 1047 亿元,同比增长 22.3%,预计未来中国水产预制菜市场保持较高的增长速度,2026 年水产预制菜市场规模将达 2576 亿元[2]。水产品预制菜食材品种广泛,表现空间很大,开发价值高。我国水产品预制菜行业处于发展初期,消费市场潜力大,我国预制菜产业发展仍然处于上升阶段。

水产品预制菜作为新兴产品也存在一些问题,例如制作工艺不够科学、配料不够透明、保存时间不足等问题。这些问题可能会导致水产品预制菜的品质下降,从而影响消费者的食用体验和健康安全。因此,对水产品预制菜的调查和研究显得十分必要。本文基于水产品预制菜消费者评论数据,通过文本分析手段研究消费者对水产品预制菜情感倾向,包括偏好、态度及期望等。

文本情感分析(Sentiment Analysis)是指利用自然语言处理和文本挖掘技术,对带有情感色彩的主观性文本进行分析、处理和抽取的过程。Pang 等人(2002) [3]第一次使用机器学习的算法对电影评价数据进行情感分析,通过实验证明了使用机器学习进行情感分析是可行的。Taboada (2011) [4]等人在情感倾向分类的过程中使用情感词典进行分类,考虑和总结了各类有可能会影响情感值的因素,比如考虑对情感词级性有影响的否定词,以及语气词对情感词的权重的作用等等因素。杨奎、段琼瑾(2017) [5]提出了基于情感词典的情感倾向分析方法,用来获取网络舆情里面的观点,提出了一种基于 How Net 概念词典,它是一种可以通过分析计算词汇相似度用来构建社会情感字典的方法。同时设计了情感得分的策略。根据分数来挖掘出人们对于舆论的贬损态度,从而准确分析文本的情感动向。黄奇景(2020) [6]参考了基于

情感词典的方法, 创建了情感向量指引神经网络获取到文本的序列信息, 这一提议让情感分类模型更加可靠。同时对卷积神经网络与循环神经网络进行了交融, 然后对于神经网络这一基础引进了注意力的机制。吴洁(2020) [7] 基于深度神经网络技术、循环神经网络等技术方法, 提出了有效的微博情感分析方法, 还针对微博数据稀缺、用户情感动态变化问题等问题, 提出了新方法。提出的方法名叫微博情感分类新方法。

2. 研究思路与数据采集

2.1. 研究思路

商品评论文本长度差异过大, 以往研究中对过长文本进行截断处理, 传统的情感分析在文本上提取的信息并不完整, 对消费者所关注商品的某些属性无法做出分析。根据分析的问题, 使用了绘制文本词云图、情感分析模型、LDA 主题分析模型来对评论文本进行情感分析。利用网购平台上该产品积累下来的海量真实评论数据, 采用八爪鱼采集器对数据进行挖掘, 利用 R 语言对其进行去重、机械压缩去词、短句过滤等文本预处理, 基于预处理后的数据进行情感分析, 文本分词以及制作词云图, 对词云图进行详细的结果分析, 最终提出改善产品的建议。

2.2. 数据采集

本文针对京东商城上预制菜的虾蟹类、鱼肉类的文本评论数据, 在对文本进行基本的机器预处理、中文分词、停用词过滤后, 通过对预处理后的数据进行情感分析, 并将评论文本数据按照情感倾向分为正面评论数据(好评)和负面评论数据(差评), 随后分别对正、负面评论数据进行分词, 然后制作词云图, 从对应的结果分析文本评论数据中有价值的内容。

为了让采集的数据更有代表性, 选取了在京东电商平台上预制菜的虾蟹类、鱼肉类产品销量第一的商家进行评论数据抓取, 采集完毕后统一输出为 Excel 文档得到原始评论文本。共抓取了评论数据 3973 条, 得到原始数据部分如评价内容图 1 所示。

评价内容									
国联烤鱼一直入手, 价格实惠, 品质和外面店家的质量差不了太多, 京东特价实惠采购, 次发货很快, 物流给力, 虽然不是自营的, 也很快, 日期很新鲜, 还没吃, 吃了再来追评, 看这款烤鱼味道真心不错! 特地吃了以后再来评价! 制作非常方便! 多种加热方法都可以! 青我家冰箱常备的烤鱼哦, 每种口味都好吃? 不能吃辣的可以试试蒜香的, 味道特别赞?? 放点葱物流很快, 没两天就收到了~ 这个?? 青花椒烤鱼? 味道特别好? 吃起来没有那么辣, 青花椒的物流很快, 没两天就收到了~ 这个烤鱼? 我家常备的, 特别好吃? 还很省事, 每次都放些爱吃白很好吃很好吃啊, 很满意的一次购物, 冰没有退, 商家服务好, 发货神速, 物流也很给力,									

Figure 1. Consumer evaluation of some pre made seafood dishes

图 1. 部分水产品预制菜消费者评价

3. 文本预处理

文本评论数据里存在大量价值很低甚至没有价值的评论, 如果对这些评论数据进行分词、词频统计、提取主题乃至情感分析, 必然造成很大的干扰, 评论数据分析结果的质量也会受到很大影响。因此, 在利用这些评论文本进行数据分析之前就必须对文本进行预处理, 去除低价值、无价值的评论。

3.1. 文本去重

本文中文本去重就是去除文本评论数据中重复的部分, 部分效果如表 1。

Table 1. De duplication of some comment content
表 1. 部分评论内容去重

国联烤鱼一直入手，价格实惠，品质和外面店家的质量差不了太多，特价实惠采购，次日达物流绝对点赞！！！！可以分享！！
发货很快，物流给力，虽然不是自营的，也很快，日期很新鲜，还没，了再来追评，看评价都说不错，味道很好呢，以前不怎么买这种半成品的菜肴，现在也学着买了，比自己做的味道好，也方便
这款烤鱼味道真心不错！特地了以后再来评价！制作非常方便！多种加热方法都可以！青花椒味道纯正味好！价格也是非常实惠！物流快捷！包装完好！
我家冰箱常备的烤鱼哦，每种口味都好不能辣的可以试试蒜香的，味道特别赞放点喜欢的配菜，一家三口不完，赶上活动囤了盒种口味，不错哦
物流很快，没两天就收到了~这个青花椒烤鱼味道特别好起来没有那么辣，青花椒的味道很好，放了些爱的配菜，真心不错，赶上活动囤了盒，每种口味我都爱？
物流很快，没两天就收到了~这个烤鱼？我家常备的，特别好还很省事，每次都放些爱的菜，跟外面的一样，一次买了盒，每种口味都很赞
很好很好啊，很满意的一次购物，冰没有退，商家服务好，发货神速，物流也很给力，宝贝收到后非常喜欢，买得放心，得安心。
小霸龙品牌的系统产品风味烤鱼，口感好、味道好。这个是国联嫩滑罗非鱼麻辣青花椒口味，属于方便预制菜品，只需加热即食，青花椒烤鱼盒，包装很好，收到时冰块还没有化冻，鱼的肉质和现做的一样。
物流速度真的是没话说，就是一个字：快。商品从外包装来看非常高档，真是一分钱一分货，好不贵，快递快，值得购买，赶快下手吧

3.2. 去除停用词

停用词是一些完全没有用或者没有意义的词，停用词大致可分为如下两类。

(1) 使用十分广泛，甚至是过于频繁的一些单词。例如，英文的 i、is、what，中文的“我”、“就”之类词几乎在每个文档上均会出现。

(2) 文本中出现频率很高，但实际意义又不大的词。这一类主要包括语气助词、副词、介词、连词等，通常自身并无明确意义，只有将其放入一个完整的句子中才有一定作用的词语。例如，常见的“的”、“在”、“和”、“接着”等。运行结果如表 2：

Table 2. Remove the list of lingering words
表 2. 去除停留词表

id	word	nature	index_word
3530	产品质量	n	1
3530	值得	v	2
3530	购买	v	3
3529	价格便宜	n	1
3529	物流	n	2
3529	很快	d	3

3.3. 评论文本分词

分词就是将连续的字序列按照一定的规范重新组合成词序列的过程。分词是分析文本评论的关键步

骤, 只有分词准确, 才能得到正确的词频, 也才能通过(TF-IDF)提取到正确的关键词。本文利用 jiabe 分词对评论文本进行中文分词, 分词效果见图 2:

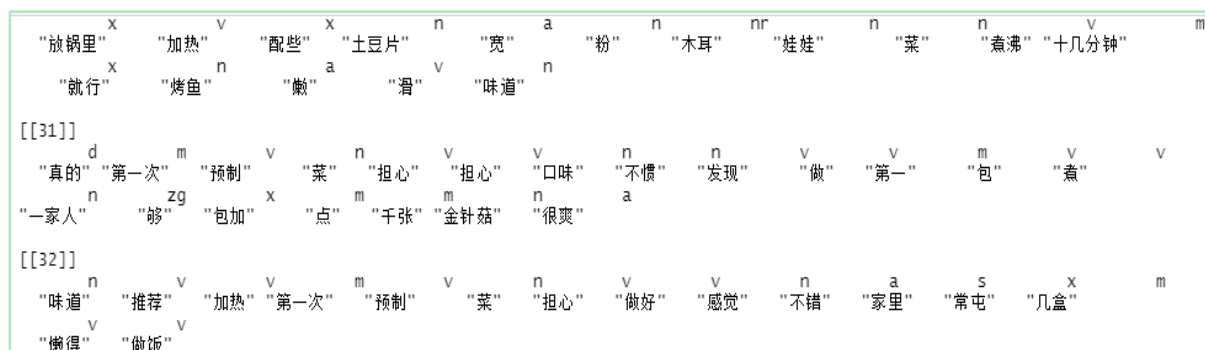


Figure 2. Text segmentation effect

图 2. 文本分词效果

4. 水产品预制菜消费者情感分析

本节将结合词云图和 LDA 主题模型进一步分析消费者对于水产品预制菜的总体情感倾向。

4.1. 基于词云图的可视化分析

词云图对高频词汇能做到突出化处理, 在对词汇的表达采取“抓大放小”的处理方式后, 浓缩了文本数据的内容, 通过文字、色彩、图形的搭配, 让高频词占据 C 位, 产生了有冲击力的视觉效果。TF-IDF 算法用于计算词语在文档中出现的次数, 即在文档中词语出现的次数越多, 词语表达文档的能力就越强。如果仅仅依据评论词汇的词频是不够足以代表有意义的评论热词, 并且由于原始数据存在缺少点赞数和关注度等相关数据的限制, 所以本文以基于 TF-IDF 算法所提取的词汇为评论关键词的热度的评判标准。根据获得的关键字词频图制作词云图模型, 以下为筛选词以及上色之后的最终词云图 (图 3 和图 4)。



Figure 3. Cloud chart of positive comment words

图 3. 好评评论词云图



Figure 4. Cloud map of negative comment words
图 4. 差评评论词云图

由图 3 可以看到“味道不错”“新鲜购买”“物流很快”等积极的词语占大多数，接着映入眼帘的还有“回购”“质量满意”“口感特别”“值得购买”等正面情感词出现的频数较高，并且没有掺杂负面情感词语，可以看出情感分析能较好地将正面情感评论抽取出来。由图 4 差评评论词云可知，最突出的点是“味道不好”“菜辣”“加热不方便”“有腥味”，这里的存在问题除了预制菜的本身，并没有具体给出其他的不足，有可能是因为顾客对产品的使用感并不满意，但由于情绪的问题没有给出过多的评价，当然，在差评词云中依然可以看到“味道”“肉辣”等部分偏褒性的词语，不过，并不能排除可能是顾客以“正话反说”的形式来表达对产品的不满，因此，对于里面蕴含的情感内容，还需要进一步分析。

4.2. LDA 情感分析模型

目前文本处理多数集中于句子级分析,本文也采用句子级分析方式,其处理过程主要分为语句提取、特定词抽取、句法分析以及计算评分。本文中模型实现算法如下:

- (1) 读取评论文本数据, 并将评论文本进行分句, 使其以句子为单位进行处理;
- (2) 从分句中抽取连词和否定词, 并标记相应连词与否定词位置;
- (3) 访问情感词汇本体, 确定词汇极性及其强度;
- (4) 针对连词(若有), 通过连词与连词位置, 确定前句与后句所占比重, 针对否定词(若有), 根据否定词位置判断双重否定, 以及临近词汇的极性反转; 若不包含连词或者否定词, 则略过该步骤;
- (5) 累加本句情感计算评分;
- (6) 循环访问步骤(2)至步骤(5), 计算该文本的评分, 若为正则为正面, 若为负则为负面, 否则为中性。算法实现结果汇总见好评评论表(表 3)和差评评论表(表 4)。

Table 3. Positive review form
表 3. 好评评论表

id	word	nature	index_word
882	给力	x	136

续表

882	价格	n	135
882	态度	n	134
882	小哥	n	133
882	快递	v	132
882	方便快捷	z	131
882	选购	v	130
882	买买	v	129
882	还会	x	128
882	粉丝	n	127
882	忠实	a	126
882	他家	r	125
882	开心	v	124
882	太	d	123

Table 4. Negative review table**表 4.** 差评评论表

id	word	nature	index_word
3522	教训	vn	59
3522	花钱买	ns	58
3522	很重	a	54
3522	腥味	n	53
3522	黏糊糊	n	52
3522	黏	zg	51
2122	退款	v	50
3522	一家	m	50
2122	没黄	x	49
3522	买过	x	49
2122	一家	m	48
3502	费用	n	48
3522	不好	d	48
2122	买过	x	47

评论文本的正面评价和负面评价混淆在一起, 避免直接进行 LDA 主题分析可能会在一个主题下生成一些令人迷惑的词语, 本文使用 R 语言开源的第三方 Gensim 库完成 LDA 主题分析, 应分别对正面评价和负面评价两类文本进行 LDA 主题分析。下面。通过分析及运行结果可知: 好评最佳主题为 2 个(如图 5), 差评最佳主题为 5 个(如图 6)。

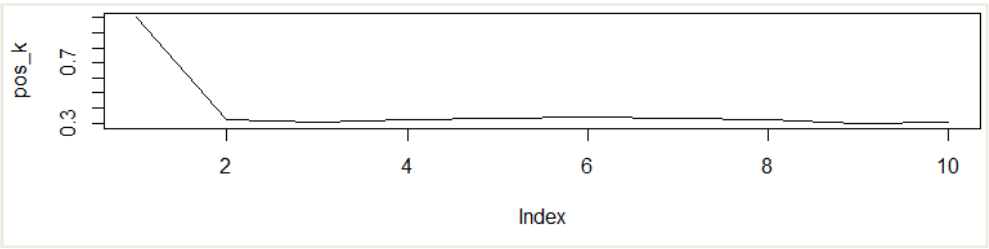


Figure 5. Number of positive review topics
图 5. 好评评论主题个数

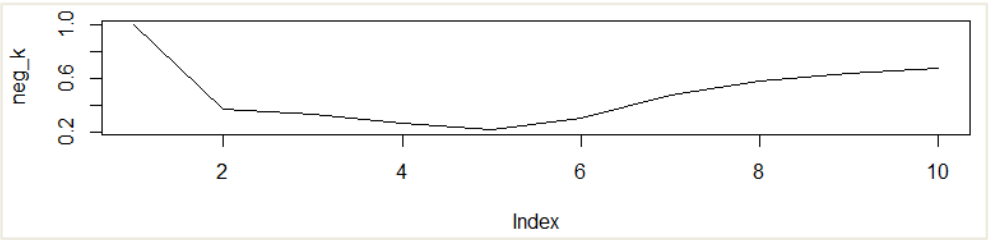


Figure 6. Number of negative review topics
图 6. 差评评论主题个数

基于 LDA 模型的模拟训练, 得到“主题 - 词组”的概率分布, 每个主题中的主题词组按其概率大小排序, 得到如下好评评论 2 个主题、差评评论 5 个主题, 且每个主题下显示 10 个最有可能出现的词语以及相应的概率。核心主题 - 词组如表 5、表 6 所示。

Table 5. Distribution of “Theme-Phrase” in positive reviews
表 5. 好评评论 “主题 - 词组” 分布表

Topic 1	Topic 2
包装	不错
购买	味道
喜欢	物流
收到	新鲜
快递	满意
特别	价格
回购	个头
质量	口感
烤鱼	值得
很快	做

Table 6. Distribution of “Topic-Phrase” in negative review comments
表 6. 差评评论 “主题 - 词组” 分布表

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
肉	辣	味道	收到	味道
不错	第一次	一只	加热	做

续表

辣	差	不好	坏	加
酸	口味	嫩	包装	很大
煮	鱼	菜	购买	蟹
蟹黄	价格	回购	罗非鱼	划算
特别	口感	感觉	卖家	个头
不划算	炒	三盒	冷冻	物流
两个	东西	麻辣	面包	真的
影响	家里	性价比	第一次	鱼肉

主题分析结果分析:

(1) 根据对预制菜的虾蟹类、鱼肉类好评的 2 个潜在主题的特征词提取, 获得以下好评分析:

主题 1 中的高频特征词有“包装”“喜欢”“烤鱼”“回购”“质量”等, 主要反映电商平台上的预制菜的虾蟹类、鱼肉类的包装、质量好, 烤鱼的预制菜味道不错等;

主题 2 中的高频特征词有“价格”“口感”“味道”“新鲜”“值得”等, 主要反映电商平台上的预制菜的虾蟹类、鱼肉类的价格便宜、口感好、新鲜等。

综合两个主题的高频词说明在商品预制菜的虾蟹类、鱼肉类中的价格便宜、口感好、新鲜、包装、质量好, 其中的预制菜烤鱼味道不错。

(2) 根据对预制菜的虾蟹类、鱼肉类差评的 5 个潜在主题的特征词提取, 获得以下差评分析:

主题 1 中的高频特征词有“肉”“辣”“酸”“特别不划算”“蟹黄”“影响”等, 主要反映预制菜的虾蟹类、鱼肉类存在一些毛病, 尤其肉不新鲜, 肉的辣度、酸度等不好, 购买后觉得特别不划算等问题;

主题 2 中的高频特征词“鱼”“价格”“口感”“口味”“第一次”等, 主要反映预制菜的虾蟹类、鱼肉类存在消费者第一次购买带来了不好的体验等问题;

主题 3 中的高频特征词“味道”“性价比”“不好”“一只”“感觉菜嫩”等, 主要反映预制菜的性价比不好, 量少, 菜较嫩等问题;

主题 4 中的高频特征词“加热”“坏”“包装”“冷冻”“购买”等, 主要反映预制菜的虾蟹类、鱼肉类存在加热存在问题, 购买后的包装损坏等问题;

主题 5 中的高频特征词“味道”“物流”等, 主要反映预制菜的虾蟹类、鱼肉类存在物流不够快等问题。

综合五个主题的高频词说明在商品预制菜的虾蟹类、鱼肉类中的线上消费体验不佳, 性价比不好, 令消费者不是很满意。

5. 结论

本文针对京东商城中水产品预制菜的评论数据, 首先通过文本预处理方法筛选有效信息, 接着利用 TF-IDF 算法提取评论关键词并分别绘制好评评论与差评评论词云图, 最后建立了 LDA 主题分析模型。模型结果表明, 消费者对于水产品预制菜最关注的是其味道、价格、包装、安全等方面。根据主题分析结果, 对于水产品预制菜的发展提出建议是: 依托优质农产品, 培育预制菜优势产区、核心企业和优质供应基地, 打造预制菜领域的龙头品牌。引导广大食品企业加强行业自律, 加强行业监管, 从食材源头

到加工流通的全过程, 建立健全可查询可追溯制度, 打造让广大消费者信得过的品牌, 抓住预制菜产业的重要发展期。

基金项目

岭南师范学院自然科学青年项目(QL1408); 岭南师范学院 2022 年度校级自然科学项目(LY2201)。

参考文献

- [1] 黄燕燕, 梁艳彤, 陆云慧, 等. 水产品预制菜行业发展现状[J]. 现代食品科技, 2023, 39(2): 81-87.
- [2] 艾媒咨询. 2022 年中国预制菜产业发展白皮书[EB/OL]. <https://www.iimedia.cn/c400/92015.html>, 2023-03-03.
- [3] Pang, B., Lee, L. and Vaithyanathan, S. (2002) Thumbs up? Sentiment Classification Using Machine Learning Techniques. In: *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, 79-86. <https://doi.org/10.3115/1118693.1118704>
- [4] Taboada, M., Brooke, J., Tofiloski, M., Voll, K. and Stede, M. (2011) Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics*, **37**, 267-307. https://doi.org/10.1162/coli_a_00049
- [5] 杨奎, 段琼瑾. 基于情感词典方法的情感倾向性分析[J]. 计算机时代, 2017(3): 10-13.
- [6] 黄奇景. Web 评论的文本情感倾向性分析[D]: [硕士学位论文]. 哈尔滨: 哈尔滨师范大学, 2020.
- [7] 吴洁. 基于用户情感倾向理解的微博情感分析方法研究[D]: [硕士学位论文]. 重庆: 重庆理工大学, 2020.