

基于电商评论文本情感分析的新疆鲜杏市场口碑分析

王思文¹, 张效瑜¹, 谢辉², 李子涵¹, 冯建英¹

(1. 中国农业大学信息与电气工程学院, 北京 100083; 2. 新疆农业科学院园艺作物研究所, 乌鲁木齐 830091)

摘要: 电商渠道是新疆鲜杏拓展市场销售、提高效益的重要途径, 而基于电商评论文本分析的市场口碑评价能为促进新疆鲜杏电商产业发展提供重要的市场反馈信息。首先获取电商平台新疆鲜杏的消费者评论, 并进行预处理; 随后对评论文本进行深入分析和挖掘, 包括判定评论语料情感倾向、基于 LDA(隐含狄利克雷分布)主题模型的评论主题提取、基于依存句法分析和情感计算的消费者口碑定量评价。研究表明, 消费者对新疆鲜杏的评价以积极情感为主, 提取的 3 个核心主题包括果实品质、物流速度和产品价格, 消费者最关注的产品属性是果实口感和物流速度, 对果实感官品质满意度最高, 对物流基本满意, 对价格满意度最低。根据研究结果, 为鲜杏电商产业提出发展建议。

关键词: 电商评论; 情感分析; 市场口碑; 鲜杏; 新疆杏

中图分类号: TP391.1 **文献标志码:** A **文章编号:** 1671-1807(2025)01-0143-07

杏产业是新疆具有优势的特色林果产业。独特的自然环境和气候条件赋予新疆鲜杏卓越的口感和优良的品质, 深受广大消费者的青睐。随着国内生鲜电商产业的发展, 电商渠道成为推动新疆鲜杏出疆、进入内地消费市场的重要销售途径。然而, 由于主产区距离内地消费市场遥远, 电商平台销售的新疆鲜杏需经过较长距离物流运输才能到达消费者手中。鲜杏果实本身娇嫩且不耐储运, 很容易在物流过程中品质受损, 从而导致消费者满意度降低。因此, 基于消费者反馈信息研究新疆鲜杏的市场口碑、消费者的评价和情感倾向, 是促进新疆鲜杏电商发展、提升新疆鲜杏消费者满意度的重要途径。

在线评论是用户利用在某种网络平台针对某个或某些对象所发表的用户自生成内容^[1]。电商平台的用户在线评论是重要的客户反馈信息, 也是商家了解消费者对其商品和服务情况的重要途径。此外, 已购消费者发布的评价体验信息直接反映了商品和服务的优缺点以及消费者对于商品和服务的态度和情感, 能够帮助潜在消费者评估产品品质和商家服务^[2], 从而做出购买决策。不同的评价信息反映了产品的不同特征属性, 为了实现评论文

本中蕴涵的情感进行自动、高效的分析, 文本情感分析技术应运而生^[3]。

关于电商平台在线产品文本挖掘的研究近年来受到了许多学者的关注。林伟振等^[4]通过分析亚马逊平台上两款 Fitbit 产品的在线评论, 采用隐含狄利克雷分布(latent Dirichlet allocation, LDA)算法自动提取了 98 个主题, 随后将这些主题手动归类为 13 个影响顾客满意的核心主题。刘佳锴和李敏^[5]以京东商城的蚕丝被为例, 使用文本挖掘方法对商品评论进行情感分析和共现聚类分析, 研究表明消费者主要关注产品质量、价格和促销、包装和快递、客服服务 4 个方面, 并从行业和企业角度提出相关建议。张梦婷等^[6]通过分析 3 大汽车网站用户评论, 利用情感分析、词云图、语义网络图和 LDA 主题建模, 挖掘新能源汽车用户的关注点和产品特性, 为厂商提供优化生产和营销策略的依据, 以促进产业发展和提升用户体验。具体到水果领域, Sari 等^[7]采用描述性统计方法和朴素贝叶斯方法对消费者评论进行分析, 探究消费者对电商平台上水果的主要关注属性, 研究结果显示消费者最关注的属性是服务属性和产品属性。Wang 等^[8]通过选择实验方法和统计分析技术, 以苹果为研究对象探究

收稿日期: 2024-07-30

基金项目: 新疆杏产业技术体系专项资金项目(XJCYTX-03)

作者简介: 王思文(2001—), 男, 山西太原人, 硕士研究生, 研究方向为农业大数据分析; 张效瑜(2001—), 女, 河北沧州人, 研究方向为文本挖掘、情感分析; 谢辉(1984—), 男, 山西运城人, 博士, 研究员, 研究方向为特色果树栽培与加工; 李子涵(2001—), 女, 山西长治人, 硕士研究生, 研究方向为农业大数据分析; 通信作者冯建英(1982—), 女, 陕西宝鸡人, 博士, 副教授, 研究方向为大数据分析、智能决策、数据挖掘。

消费者购买行为的偏好,结果发现消费者对富士苹果的偏好表现出明显的消费分层,倾向于选择便宜或昂贵的苹果,而不喜欢中等价格。Chien-Hua 等^[9]以电商平台上茂名沙糖桔的在线评论为研究对象,采用 LDA 确定评论主题的数量,TF-IDF(term frequency-inverse document frequency,词频-逆文档频率)得到评论主题的相关词汇进行共性分析,进一步完善沙糖桔电商销售带来的问题。但目前还未有学者开展鲜杏的电商评论文本挖掘与分析,无法为鲜杏生鲜电商发展提供基于消费者反馈信息的经营管理决策建议。

因此,本文采集主流电商平台的新疆鲜杏评论数据,通过 LDA 主题模型提取出消费者评论的核心主题,并采用依存句法分析和情感词典方法计算消费者对新疆鲜杏的关注度和满意度,得到消费者对新疆鲜杏的口碑评价,从而帮助商家更好地了解消费者需求和市场情况,并提供产品改进和服务优化等决策依据^[10]。

1 研究方案

1.1 数据及预处理

1.1.1 数据来源

本文采用 Python 爬虫技术获取京东平台、淘宝平台和拼多多平台有关新疆杏的评论数据,共计 2 149 条。部分获取到的评论内容如下。

(1)小红杏很新鲜,基本没有坏果,包装也好。吃起来有普通杏子的味道,还夹带了一点李子的那种酸甜口感,非常不错。

(2)确实很甜,但皮比较厚,吃起来好像桃的味道。有那么三五个坏果。

(3)这家的小红杏品质不错的,收到没有坏果完好无损,成熟度刚刚好,收到就可以直接吃,很新鲜。儿子特喜欢这种软糯香甜的口感。

(4)杏是一个个单独包装的,收到无损坏,很甜,很好吃,会回购。

(5)小红杏收到了,一打开箱子就能闻到一股很浓的杏香,很清新的感觉,拆开洗了一盘给家里人尝了一下,都说味道挺好的,很快就吃完了。很软糯的口感,特甜,汁水很多。好吃。

(6)今年第一次吃到小红杏,一直在他们家买新疆的这些水果,到货收到都很新鲜,皮薄薄的肉也很厚实,放到嘴里汁水四溢,很软糯,非常甜,顺丰快递也很 OK,收到没有坏的。

1.1.2 数据预处理

评论数据属于非结构化数据,其中存在大量冗

余词汇、网络用语、表情符号和无效评论,若直接进行分析会导致文本挖掘效果不佳,因而需要对文本进行数据去重、删除短评论、结巴(jieba)库分词、去除同义词和常用词替换等预处理^[11]。本文的替换标准见表 1。

表 1 中英文常用词替换表

原表达	替换后的表达
good、nice、奈斯、奶思、绝绝子	好
辣鸡、拉胯、达咩	差
双 11、11.11	双十一
6.18、618	六一八

去除停用词时,由于目前常用的停用词表无法适用于所有文本数据,所以本文根据实际情况自建了更适合的停用词表^[12],新增的停用词包括京东、淘宝、拼多多、新疆、鲜杏、收到、一箱、孩子。

1.2 评论文本分析方法

1.2.1 基于情感指数的情感倾向分类方法

消费者对于新疆鲜杏购买之后的感觉和评价呈现差异。因此,有必要深入细致分析评论中反映的消费者情感倾向。将情感倾向分成 3 类:积极情感倾向、中性情感倾向和消极情感倾向。对于预处理后的数据集中每一条评论,首先进行分词操作,然后依次判定该评论中每个词语的情感倾向,最终计算该条文本的情感倾向。

通过知网 HowNet 发布的情感词语集进行情感判断,并采用导入否定词来进行判罚处理,取积极倾向的情感词语记为+1,消极的情感倾向词语记为-1。计算情感倾向词语前面的否定词个数:如果是偶数个否定词,则情感指数不发生改变;如果是奇数个否定词,则情感指数记为之前的相反数。

方法逻辑步骤如下:

Begin

Step1:for 循环遍历所有的评论数

Step2:获取第 i 条评论拆分后的所有词语

Step3:依照顺序对词语进行情感倾向的判断

Step4:进行判罚操作:

if 词语前面拥有偶数个否定词:

当前情感指数不变

if 词语前面拥有奇数个情感词:

当前情感指数取相反数

Step5:对该条评论拆分出的情感词语的情感指数进行累加计算

Step6:计算评论最终的情感指数,大于零标记为积极评论,小于零标记为消极评论,等于零标记

为中性评论。

1.2.2 LDA 主题建模方法

LDA(隐含狄利克雷分布)主题建模早在 2003 年被 Blei 等^[13]提出,是由主题词、主题和文档组成的 3 层贝叶斯概率模型。与其他模型不同的是 LDA 模型不需要事先对文本进行标注,只需要提供文档集和指定的主题数就可以有效提取不同主题的主题词以及主题词的生成概率^[14]。LDA 的主要思想是对于给定的文本集合,每一篇文档都是隐含的话题组成,每一个话题又是由一些特定的词汇组合而成。因此,LDA 主题模型可以将文本集合中的每一个文档都表示成为多个话题的混合,同时每个话题还由一些特定的词汇组合而成^[15]。具体训练过程^[16]如下。

(1)评论 m 包含的特征词数量 N_m 服从泊松分布,即 $N_m \sim \text{泊松}(\xi)$ 。

(2)对于评论 m 生成主题分布,其中 $m \in \{1, 2, \dots, M\}$,即 $\theta_m \sim \text{Dirichlet}(\alpha)$,其中 M 表示数据集评论的总数量, θ_m 表示第 m 个评论的主题概率分布, α 为每个评论下主题的多项分布的 Dirichlet 先验参数。

(3)对于主题 n 生成特征词分布,其中 $Z \in \{1, 2, \dots, K\}$, $\phi_k \sim \text{Dirichlet}(\beta)$, K 为总的主题数, β 为每个主题下的词多项分布的 Dirichlet 先验参数。

(4)评论 m 中的特征词 $\omega_{m,n}$ ($n \in \{1, 2, \dots, N_m\}$) 的生成过程, N_m 为第 m 个主题包含的特征词。
①根据主题分布 θ_m 生成评论 $\omega_{m,n}$ 的特征词主题,即 $Z_{m,n} \sim \text{Multinomial}(\theta_m)$, $Z_{m,n}$ 表示的是第 m 个评论的第 n 个词的主题。
②根据词项分布 $\phi_{Z_{m,n}}$ 生成所选词主题此项,即 $\omega_{m,n} \sim \text{Multinomial}(\phi_{Z_{m,n}})$ 。

1.2.3 依存句法分析

依存句法分析是自然语言处理领域的一种语法分析方法。其以句子作为分析对象,依存关系是一个中心词与其从属之间的二元非对称关系,一个句子的中心词通常是动词,所有其他词依赖于中心词或通过依赖路径与中心词关联^[17]。

依存句法解释了句子中词与词之间的依存关系以及搭配关系,句子中词与词的依存关系包括主谓关系(SBV)、动宾关系(VOB)、间宾关系(IOB)等,但句子中仅包含一个核心关系(HED)^[18]。依存句法关系存在 5 条公认理论^[19],总结如下。

(1)一个句子中有且只有一个句子成分是独立的。

(2)除独立的句子成分外,其他句子成分有唯

一的父节点存在,即依存于某一成分。

(3)任何一个句子成分都不能依存两个或者两个以上的其他句子成分。

(4)一个句子中,如果 A 成分依存于 B 成分,而句子中的 C 成分在 A 和 B 之间,那么 C 依存于 B ,或者依存于存在于 A 和 B 之间的某一其他成分。

(5)位于中心成分左右两边的句子成分相互独立。

1.2.4 基于情感词典的情感计算方法

情感词典能够体现文本的非结构化特征,它是文本情感分析所需的重要资源^[20]。情感计算旨在创建一种能感知、识别和理解人的情感,并能针对人的情感构建智慧、灵敏、自然反应的计算机系统^[21]。

在情感词典的基础上,基于依存句法分析提取词对,对于词对中的属性词和观点词分别进行数目累加。其中,观点词为正向词语满意指数+1,负面词语+0,累积得到的满意度为满意指数和观点数目的比值。

方法步骤逻辑如下:

Begin

Step1:基于依存句法分析,进行关系提取

Step2:for 循环提取出的数组

Step3:对词语进行情感倾向的判断

Step4:进行打分操作

Step5:导入否定词库

Step6:进行判罚操作

if 该词语为积极情感倾向:

满意度+1;属性+1

if 该词语为消极情感倾向:

满意度+0;属性+1

Step7:单个属性关注度等于属性词出现次数

单个属性满意度等于属性情感得分与情感基数的百分比。

2 结果与分析

2.1 鲜杏评论数据描述分析

2.1.1 数据集词云分析

对原始评论数据集进行数据预处理操作后的词云统计图如图 1 所示。

根据词云图显示,在新疆鲜杏的消费者评论中,出现频率最高的为“新鲜”“口感”“味道”和“坏果”。表明消费者对鲜杏的“口感”和“味道”属性高度关注,“新鲜”属于对果实品质属性的积极评价;“坏果”则反映出经过电商物流后果实有损坏现象。

表 2 LDA 积极评价

类型	LDA 积极评价
第 1 类	0.051 * “物流” + 0.040 * “及时” + 0.025 * “满意” + 0.025 * “好” + 0.015 * “精致” + 0.014 * “发货” + 0.012 * “迅速” + 0.011 * “值得” + 0.011 * “店家” + 0.010 * “朋友”
第 2 类	0.079 * “价格” + 0.073 * “合适” + 0.050 * “值得” + 0.034 * “好” + 0.032 * “不错” + 0.020 * “好评” + 0.019 * “水果” + 0.018 * “真的” + 0.016 * “物美价廉” + 0.013 * “很快”
第 3 类	0.092 * “好吃” + 0.062 * “甜” + 0.053 * “很甜” + 0.038 * “味道” + 0.028 * “口感” + 0.027 * “回购” + 0.024 * “喜欢” + 0.020 * “软” + 0.017 * “果实” + 0.014 * “一点”

注:数值表示该词在相应主题下的概率权重,反映了每个词在该主题下的重要性和出现频率;*表示乘法运算,用于表示词的概率权重与其对应的词之间的关系。

2.2.2 消极评论的 LDA 主题寻优结果

负面评论 LDA 主题数寻优结果如图 6 所示,横坐标代表不同的主题数量,纵坐标表示模型的困惑度。困惑度越低,说明模型的拟合效果越好。观察发现,主题数为 3 时出现了明显的拐点,因此最佳主题数应确定为 3。

确定最佳主题数后,针对负面评论数据集训练主题模型,提取主题,结果见表 3。

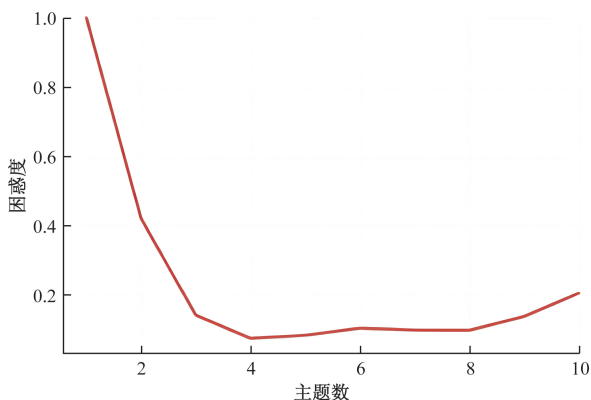
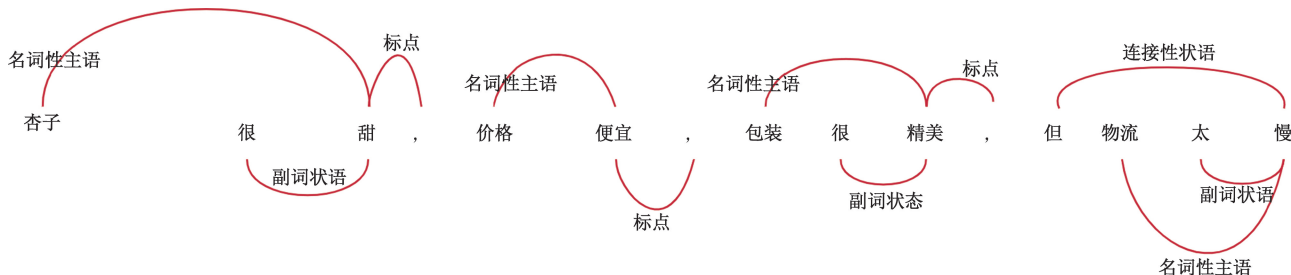


图 6 消极评论 LDA 主题数寻优



曲线表示依存关系,每个曲线的两端分别代表两个词语之间的依存关系,曲线上的标签表示依存关系的类型

图 7 依存分析结果图

表 3 LDA 消极评价

类型	LDA 消极评价
第 1 类	0.053 * “包装” + 0.015 * “不好” + 0.015 * “坏” + 0.015 * “下单” + 0.013 * “物流” + 0.012 * “果子” + 0.012 * “破损” + 0.012 * “太慢” + 0.012 * “水果” + 0.010 * “大小”
第 2 类	0.036 * “差劲” + 0.024 * “贵” + 0.017 * “价格” + 0.016 * “高” + 0.015 * “果实” + 0.015 * “糯” + 0.014 * “太慢” + 0.014 * “坏果” + 0.014 * “味道” + 0.014 * “独立”
第 3 类	0.083 * “坏果” + 0.045 * “不好” + 0.035 * “吃” + 0.031 * “涩” + 0.022 * “特别” + 0.013 * “太小” + 0.010 * “中” + 0.009 * “生生的” + 0.009 * “质量” + 0.009 * “下次”

注:数值表示该词在相应主题下的概率权重,反映了每个词在该主题下的重要性和出现频率;*表示乘法运算,用于表示词的概率权重与其对应的词之间的关系。

通过 3 个主题下的词语分析,第 1 类主要是产品的包装和物流相关主题,第 2 类主要是产品的价格相关主题,第 3 类主要是果实的质量相关主题。

综合 LDA 对积极评价和消极评价的分析,可以发现物流速度、价格、品质是消费者最关注的主题。

2.3 基于依存句法关系的鲜杏评价指标体系与情感词典的构建

2.3.1 电商鲜杏评价指标体系构建

依存句法关系抽取规则有以下两部分。

(1)根据由依存句法分析所构成的句子语意关系树,判断句子中分词词性为形容词的词语用以修饰某个分词词性为名词的词语,则名词为句子中的主语,认为形容词是句子中消费者对于名词评论属性的观点。

(2)如果存在否定修饰词且用以来修饰句子中的形容词,则将否定关系纳入考虑,并将否定意义加入观点词前。

以评论“杏子很甜,价格便宜,包装很精美,但是物流太慢为例”,分析结果如图 7 所示。

基于以上句法依存关系分析,本文提取并构建出二级电商鲜杏关键指标体系,指标见表 4。

表 4 电商评论文本属性指标词表

一级指标	二级指标	同义词
果实感官属性	口感	口味、感觉、味道、甜度、质量、品质、总体
	形状	大小、个头、长相、品种
物流属性	速度	物流速度、配送、送货、时间
	包装	外包装、纸盒、纸箱、箱子、盒子
价格属性	价格	价位、性价比、价值

2.3.2 领域情感词典构建

将知网 HowNet 情感词典作为基础情感词典,并根据电商平台鲜杏的评论语料,进行人工补充,部分新增领域词典见表 5。

表 5 领域情感词典表

基础情感词典倾向	情感词
积极	水分多、精致、严谨、足斤足两、实惠、回购、好评、点赞、爽口、紧实、足、饱满、细滑、推荐、巴适、信誉、新鲜、靠谱、鲜艳、红彤彤、细腻、清冽、惊艳、物美价廉、黄、诚信、核小、均匀
消极	硬、酸、涩、青、普通、一般、贵、烂果、坏果、小、生、损坏、慢、淡、袖珍、少

2.4 基于情感计算的消费者关注度和满意度评价结果

根据 1.2.4 节的情感计算方法,获得消费者对于电商平台新疆鲜杏的关注度和满意度评价结果,如图 8 和表 6 所示。

图 8 显示口感是消费者评论中出现频次最高的商品属性,说明消费者对该属性的关注度最高,其次是物流速度,之后依次是果实形状、价格和包装。

表 6 显示消费者对于果实感官品质指标下的果实形状指标满意度最高,对口感也高度满意;对于物流指标下的包装和物流速度基本满意;而价格指标的满意度最低。可以得出,新疆鲜杏的品质口碑很好,目前生鲜电商的包装盒物流速度也达到了大多数消费者的需求,但远距离运输成本推高了终端商品的售价,导致消费者的价格满意度不高。

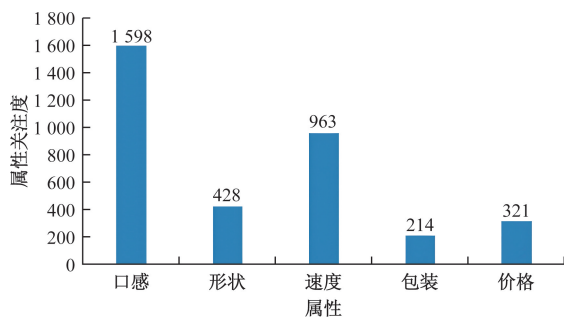


图 8 属性关注度

表 6 属性满意度

一级指标	二级指标	得分(满分 10 分)
果实感官品质	口感	9.4
	形状	10
物流属性	速度	8.8
	包装	9
价格属性	价格	7.2

3 结论与建议

本文基于文本分析技术挖掘消费者对电商平台新疆鲜杏的评价和口碑,以期为新疆鲜杏电商发展提供经营决策建议。基于 LDA 主题模型挖掘出消费者评论语料中的 3 个重要主题,涉及果实品质、物流速度和商品价格;随后采用依存句分析进行了消费者关注度和满意度计算。结果显示:消费者对于果实口感和物流速度关注度最高;从满意度来看,消费者对于果实感官品质(包括口感和果实形状)的满意度最高,对于物流包装和速度基本满意,对商品价格满意度低。

对于鲜杏电商从业者来说,应在保持商品品质和物流效率的前提下,进一步完善仓储配送体系,拓展运输渠道和资源等,优化运输方式与运输环境,降低物流成本。电商商家应充分重视消费者关于产品品质、客户服务、品牌认同、价格接受度等方面的评价和反馈,优化产品供给,提高消费者满意度,促进新疆鲜杏的外销,提高整个产业的效益。

参考文献

- [1] 李欣颖,徐恺英,张响,等. 基于情感视角的国内外在线评论研究动态及发展趋势研究[J]. 情报科学, 2020, 38(11): 148-154.
- [2] 冯建英,吴丹丹,王博,等. 中文在线评论文本分析对生鲜农产品电商影响研究综述[J]. 农业机械学报, 2021, 52(S1): 504-512.
- [3] 张小艳,白瑜. 基于加权融合字词向量的中文在线评论情感分析[J]. 计算机应用研究, 2022, 39(1): 31-36.
- [4] 林伟振,刘洪伟,陈燕君,等. 基于在线评论的顾客满意度研究:以健康监测穿戴产品为例[J]. 数据分析与知识发现, 2023, 7(5): 145-154.
- [5] 刘佳锴,李敏. 基于文本挖掘的蚕丝被在线评论分析:以京东商城为例[J]. 丝绸, 2023, 60(8): 11-20.
- [6] 张梦婷,叶楠,范黄健. 基于情感分析的新能源汽车用户评论语义网络分析与主题建模[J]. 科技和产业, 2022, 22(12): 364-369.
- [7] SARI A K, ANGGASARI H, SETIYANI R. Analysis of fruit consumer satisfaction in an online purchasing system using an E-commerce platform: a naive Bayes Approach[J]. West Science Interdisciplinary Studies, 2024,

- 2(3): 528-540.
- [8] WANG E, LIU Z, GAO Z, et al. Consumer preferences for agricultural product brands in an E - commerce environment[J]. *Agribusiness*, 2022, 38(2): 312-327.
- [9] CHIEN-HUA W, CHIA-HSUAN Y, CHIN-TZONG P. Using text mining to evaluation for online shopping rural fruits and vegetables[C]//*Proceedings of the 8th International Conference on Communication and Information Processing*. Beijing: ACM, 2022: 56-60.
- [10] 王安宁, 张强, 彭张林, 等. 在线评论的行为影响与价值应用研究综述[J]. *中国管理科学*, 2021, 29(12): 191-202.
- [11] 毕达天, 楚启环, 曹冉. 基于文本挖掘的消费者差评意愿的影响因素研究[J]. *情报理论与实践*, 2020, 43(10): 137-143.
- [12] 张公让, 鲍超, 王晓玉, 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.
- [13] BLEI D M, NG A Y, JORDAN M I. Latent dirichlet allocation [J]. *Journal of machine Learning Research*, 2003, 3: 993-1022.
- [14] 冯智依. 基于 LDA 模型的小米 12 手机在线评论研究[D]. 上海: 上海师范大学, 2023.
- [15] 皮婧文, 王海燕, 吴保国. 基于 LDA 模型的大兴区都市农业发展方向研究[J]. *农业科研经济管理*, 2020(4): 43-47.
- [16] 王珠美, 胡彦蓉, 刘洪久. 基于 LDA 主题模型和直觉模糊 TOPSIS 的农产品在线评论情感分析[J]. *数据采集与处理*, 2020, 35(5): 965-977.
- [17] 楼雯, 马昕钰, 苏子龙. 融合依存句法网络和 PageRank 的检索词推荐方法研究[J]. *情报学报*, 2023, 42(11): 1358-1368.
- [18] 范涛, 王昊, 吴鹏. 基于图卷积神经网络和依存句法分析的网民负面情感分析研究[J]. *数据分析与知识发现*, 2021, 5(9): 97-106.
- [19] 王昊, 刘丹, 刘硕. 基于句法分析及主题分布的关键词抽取模型[J]. *计算机应用研究*, 2022, 39(9): 2603-2607.
- [20] 杨书新, 张楠. 融合情感词典与上下文语言模型的文本情感分析[J]. *计算机应用*, 2021, 41(10): 2829-2834.
- [21] 情感计算发展与应用研究[J]. *软件和集成电路*, 2023(8): 72-84.

Market Word-of-mouth for Xinjiang Fresh Apricots Based on Sentiment Analysis of E-commerce Reviews

WANG Siwen¹, ZHANG Xiaoyu¹, XIE Hui², LI Zihan¹, FENG Jianying¹

(1. College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China;

2. Institute of Horticultural Crops Research, Xinjiang Academy of Agricultural Sciences, Urumqi 830091, China)

Abstract: E-commerce channels are a crucial means for Xinjiang fresh apricots to expand market sales and improve efficiency. Analyzing market reputation based on E-commerce review texts can provide valuable market feedback for promoting the development of the Xinjiang fresh apricot E-commerce industry. Consumer reviews of Xinjiang fresh apricots from e-commerce platforms were collected and preprocessed. Subsequently, an in-depth analysis and mining of the review texts were conducted, including determining the sentiment orientation of the review corpus, extracting review themes based on the LDA (latent Dirichlet allocation) topic model, and quantitatively evaluating consumer reputation based on dependency syntax analysis and sentiment calculation. The study shows that consumer reviews of Xinjiang fresh apricots are predominantly positive. The three core themes extracted are fruit quality, logistics speed, and product price. Consumers are most concerned with the attributes of fruit taste and logistics speed, with the highest satisfaction in fruit sensory quality, general satisfaction with logistics, and the lowest satisfaction with price. Based on the research results, development suggestions were proposed for the fresh apricot e-commerce industry.

Keywords: E-commerce reviews; sentiment analysis; market reputation; fresh apricots; Xinjiang apricots