

电商平台助农产品的消费者偏好挖掘与分析

——以京东平台新疆灰枣为例

晏发发^{1, 2}, 许旭^{1, 2*}, 张剑凯¹, 钟梅琳¹, 陈颖彤¹

1. 广州理工学院经济管理学院, 广东 广州 510540

2. 广东特色金融与高质量发展研究中心, 广东 广州 510540

DOI:10.61369/SE.2026020005

摘 要 : 本研究聚焦社交媒体大数据对农产品电商消费行为的驱动作用, 采集电商平台评论数据与社交媒体互动信息, 运用潜在狄利克雷分配 (Latent Dirichlet Allocation, LDA) 模型, 引入高频词分析、共现网络分析、情感分析等方法, 在京东平台上深度挖掘消费者对农产品口感、品质、价格、包装等维度的偏好特征。研究结果表明, 消费者在关注助农产品时, 主要聚焦于产品质量与价格合理性、商家服务水准与物流配送效率、个人对产品的综合感知价值、助农活动本身的吸引力, 以及地方官员 (如县长) 对助农事业的支持力度等方面。

关 键 词 : 社交媒体大数据; 农产品电商; 消费者偏好; 乡村振兴

Mining and Analysis of Consumers' Preferences for E-Commerce Platforms' Agricultural Products — A Case Study of Xinjiang Grey Jujubes on JD.com

Yan Fafa^{1,2}, Xu Xu^{1,2*}, Zhang Jiankai¹, Zhong Meilin¹, Chen Yingtong¹

1.School of Economics and Management, Guangzhou Institute of Science and Technology, Guangzhou, Guangdong 510540

2.Guangdong Research Center of Characteristic Finance and High-Quality Development, Guangzhou, Guangdong 510540

Abstract : This study focuses on the driving effect of social media big data on consumer behavior in agricultural product e-commerce. It collects comment data and social media interaction information from e-commerce platforms, and uses the Latent Dirichlet Allocation (LDA) model, combined with methods such as high-frequency word analysis, co-occurrence network analysis, and sentiment analysis, to deeply explore consumers' preference characteristics for agricultural products in dimensions such as taste, quality, price, and packaging on JD.com. The results show that when paying attention to poverty-alleviation agricultural products, consumers mainly focus on product quality and price rationality, merchant service level and logistics distribution efficiency, personal perceived comprehensive value of products, the attractiveness of poverty-alleviation activities themselves, and the support of local officials (such as county magistrates) for the poverty-alleviation cause. Although consumers are generally satisfied with the poverty-alleviation agricultural products sold on JD.com, there are still areas that need to be optimized.

Keywords : social media big data; agricultural product e-commerce; consumers' preferences; rural revitalization

基金项目:

2024年广州理工学校级项目“数字信贷对服务业创业的影响效应和机制——基于广东县域面板数据的经验证据”(2024KY052);
2024-2025年度广州理工学校级教学质量与教学改革工程项目“多语言程序设计+数字经济专项人才培养计划(实验班)”(2024XZLGC39);
2025年广州理工学校级项目“数字经济对中国经济不平等的影响效应与影响路径研究”(2025KYQ003);
2025年广州理工学院工作室项目铭心博士工作室; 2025年广东省教育科学规划课题(高等教育专项)“产学研赛+混合编程+数字经济跨学科人才培养探究”(2025GXJK0660);
2025年广东省教育科学规划课题(高等教育专项)“AI虚拟助教助力高校国贸专业互动式教学的实践探索”(2025GXJK0664);
2025-2026年度广东省教育评估协会一般课题“学生视角下高校课程思政实施效果评价体系构建研究”(BDPG25100);
2025年度广东省普通高校特色创新类项目哲学社会科学类“广东省低空经济产业绿色发展的路径优化:基于多维系统关联的研究”(2025GXJK0664)。

作者简介:

晏发发, 广州理工学院经济管理学院, 副教授, 研究领域: 计量经济学理论和应用、宏观模型构建和应用、农业经济、数字经济、教育经济、资产定价、机制设置、博弈论等;
张剑凯, 广州理工学院经济管理学院, 本科生在读, 研究领域: 互联网金融;
钟梅琳, 广州理工学院经济管理学院, 本科生在读, 研究领域: 互联网金融;
陈颖彤, 广州理工学院经济管理学院, 本科生在读, 研究领域: 会计。

通讯作者: 许旭 (第二作者), 广州理工学院经济管理学院, 副教授, 研究领域: 产业经济、国际贸易、跨境电商。

引言

传统的农产品产销模式更注重中间生产环节而忽视产后销售能力（徐向梅，2022）^[1]，而农村电商“去中介化”的特点直接提高了农业经济运行模式中的产后销售能力（王雅丽和杨雪云，2024）^[2]。在2020年中央一号文件提出“实施数字乡村战略”和2025年中央一号文件进一步聚焦“提升乡村产业发展水平”政策引导下，农产品电商成为乡村振兴的重要抓手，例如，缩小城乡居民收入差距（李宏兵在，2021）^[3]、促进发展性消费支出（刘明辉和卢飞，2023）^[4]、带动非农产业就业（田野和夏杰长，2024）^[5]、促进共同富裕（曹增栋，2024）^[6]、降低地区基尼系数并促进农村收入包容性增长（陈华师和谢可琴，2024）^[7]。

电商巨头京东平台开展“新农人计划”助力无数农民实现从“土特产”到“品牌商品”的转型升级，其“光伏+物流园区”为主的物流模式减轻农村物流成本、提高配送效率，实施直播带货等新型电商营销模式使消费者更加详细地了解助农产品的信息，农民基于市场需求准确地进行农业生产，最终全面提高助农产品的销售能力（林莉和杨承亮，2022）^[8]。

因此，以京东平台新疆灰枣为例，对电商平台助农产品的消费者偏好进行挖掘与分析，这一主题具有深刻的理论和实践意义，其实际贡献在于构建“主题挖掘-关键词频次分析-关联网络构建-情感倾向剖析”的综合研究框架，该框架不仅能揭示整体评论的高频词共现模式，还可通过主题细分量化消费者情感极性，为电商助农策略优化提供理论支撑与实践指导以帮助带动农民共同富裕、助力农产品电商的可持续发展。

一、文献综述

LDA模型常用于构建“主题抽取-全局情感分析”的研究框架（程谦益和姜娟，2023）^[9]进行消费者情绪分析（周婷玮，2023）^[10]、消费者偏好识别与用户画像构建（杨梦琳和卢益清，2022）^[11]、消费者满意度评估以及影响因素研究（王二朋和倪郑宇，2020）^[12]等多个核心领域。然而，LDA在呈现高频词汇共现网络时无法直观且清晰地呈现各个主题之间的相互联系（殷彤等，2022）^[13]，并且局限于整体评论的情感倾向分析，无法捕捉到不同评价主题所特有的情感色彩。因此，本文设计的“主题挖掘-关键词频次分析-关联网络构建-情感倾向剖析”的综合框架对该领域做出补充。

此外，主题聚类结果中部分特征词语义关联性有待加强，精确主题命名存在挑战，这提示基础主题建模方法可能未充分掌握消费者关注点的细微差异。因此，本文扩展研究方法，在LDA模型后结合高频词分析、共现网络分析以揭示主题间内在联系，探索LDA模型可能未捕捉的关键信息，并整合情感分析技术识别不同主题背景下的消费者情感倾向。

二、研究设计

（一）研究数据

选择京东商城“三只松鼠”官方店铺中，依据消费者评价数量较多且销售排名在前十大的新疆灰枣商品的橱窗展示评论当作数据源，运用Python软件采集到总计10432条产品评价数据，剔除冗余数据、机械压缩去词优化、筛选有效评论，最终保留了具有足够信息量的评论样本。

（二）模型构建及说明

1. LDA模型

LDA模型是一个有创新性的三层贝叶斯概率生成模型，其

中包含了文档、主题以及词语这样的层次架构，这个模型所做出的假设是，文档可由一种或者多种主题复合而成，每一个主题都有着其自身独特的词汇分布作为特征（更多信息见池毛毛等（2021））^[14]。由于主题数目的选取会对聚类质量产生关键影响，因此在LDA模型构建之前需对最佳主题数量进行估算（Bleidm等2003）^[15]。LDA模型的性能评估采用困惑度（Perplexity）来确定最佳主题数量，困惑度的数学表达如下：

$$Perplexity(D) = \exp \left\{ - \frac{\sum_{d=1}^M \log p(W_d)}{\sum_{d=1}^M N_d} \right\} \quad (1)$$

其中，D代表语料库；M表示文档的总数量；N_d指每篇文档所包含的单词数目；W_d表示文档中的具体词汇；p(W_d)则是文档中词汇W_d出现的概率。

2. 高频词分析

高频词分析的流程是先进行数据预处理，再统计词频，最后可视化与解读，借助可视化工具等快速提取文本核心信息，对文本进行精确挖掘，优势在于高效性和可解释性。

3. 在线评论情感分析

在线评论情感分析是由计算机技术自动识别和提取文本中的主观情感倾向，由语言学基础、机器学习与深度学习、情感本体论和心理学基础组成理论基础，妥善使用预定义的情感词典，分析文本中正反馈与负反馈词汇，综合判断情感倾向，再用标注数据训练分类模型，借助神经网络自动学习文本的语义表示，正确获得上下文信息，最终提高分析的准确性和可解释性。

4. 共现网络分析

共现网络分析在共现关系、网络科学理论、统计与概率模型和文本挖掘与自然语言处理等理论上，进行数据预处理进行文本清洗和关键词提取，随后定义共现条件，设定共现最低次数，筛选低频关联，再搭建共现网络，统计正确的元素的共现频率，组成矩阵，再将矩阵转换为图形，借助节点、边和权重分别

表示元素、共现关系和共现强度，最后在前面步骤的奠基下进行网络特征分析和可视觉解释。

三、研究结果

（一）主题聚类分析

LDA主题模型的困惑度见图1。易见，随着预设主题数增加，模型困惑度逐渐下降，这在一定程度上反映了模型在类别划分方面的效能。

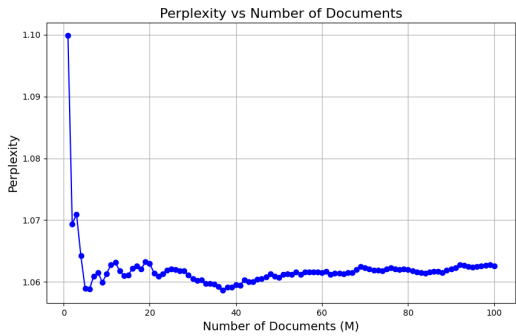


图1：主题数量 - 困惑度图

随后，对主题聚类结果进行直观展示（见图2）。当设定主题数量为5时，各主题涵盖的内容领域边界清晰、差异显著，主题间区分度达到较优状态。因此，确定主题数量为5。

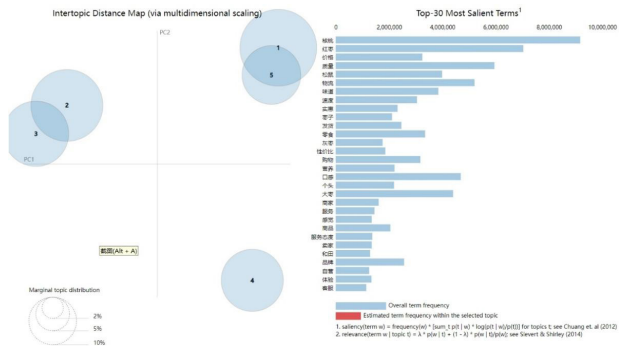


图2：主题聚类可视化结果图

接着，构建LDA模型对评论文本进行主题聚类分析，如表1所示。

表1：LDA主题聚类结果

编号	主题名称	主题强度 (%)	特征词
1	品质与口感	21.22	味道 个头 灰枣 感觉 口感 枣核 生产日期 评价 甜度 价格便宜 保质期
2	体验与性价比	18.20	价格 物流 实惠 性价比 质量 商家服务 活动 售后 购物 优惠 品质 预期 总体
3	物流与客户体验	20.80	质量 速度 发货 购物 物流 品质 态度 卖家 体验 自营 客服 产品 态度
4	特性与用途	17.11	红枣 口感 产品 煲汤 肉质 甜度 果肉 品质 泡茶 泡水 食用 零食 色泽 材质
5	健康与养生	22.68	核桃 松鼠 红枣 营养 口感 大枣 美味 味道 孩子 礼盒 枣仁 包装 老人 坚果

（二）高频词分析

高频词图如图3。高频词汇分析可快速揭示网络评论的核心议题。在预处理后的在线评论中，名词常刻画产品特性（如“味道”“质量”“包装”“口感”“物流”“价格”），形容词多表达消费者对购物经历的情感反应及整体评价（如“香甜”“美味”“物有所值”“满意”），其中正面评价词汇出现频繁，说明整体情感倾向较为积极。动词则描绘消费者的购买行为模式（如“喜爱”“购买”“复购”“信赖”“推荐”），体现消费者的行为意向。

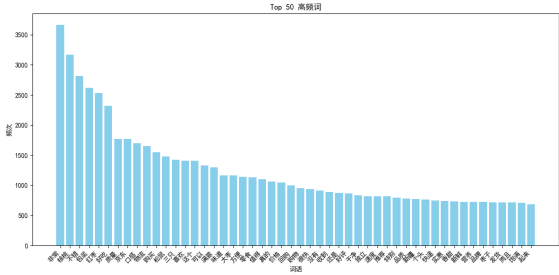


图3：高频词图



图4：评论词云图

评论词云图如图4所示，评论词云图以图形化方式展示文本数据中的高频词汇，直观反映了评论内容的关键词分布和主题倾向：“购买、购物、零食、实惠、推荐”等词语频率较高，证明了消费者关注购买过程和购买体验的便捷性和实惠性；“品牌、香甜、好吃、比较”等表明消费者在购买产品时考虑产品的品牌和口味；“包装、物流、快递、服务、态度”等说明消费者重视商品包装和电商平台的服务质量；“健康、性价比、优惠、回购、继续”反映了消费者希望购买健康并且物美价廉的商品。

（三）在线评论情感分析

在线评论所蕴含的情感通常可分为“积极”“中性”“消极”三类。借助ROST Content Mining软件，对消费者在京东平台针对助农产品的评论进行了情感分析，采用更加多级的情绪强度：一般积极（0< 赋值≤10）、中度积极（10< 赋值≤20）、重度积极（赋值>20）、一般消极（-10≤赋值<0）、中度消极（-20≤赋值<-10）、重度消极（赋值<-20）。不同主题下助农产品评论的情感分析详细结果，如表2所示。

表2: 不同主题的情感分析结果

分类	总评论数量 / 条	类型	数量 / 条	百分比
总体	10432	积极情绪	7958	66.26%
		中性情绪	3335	27.77%
		消极情绪	717	5.97%
品质与口感	3543	积极情绪	2869	80.98%
		中性情绪	610	17.22%
		消极情绪	64	1.81%
体验与性价比	1946	积极情绪	1593	81.86%
		中性情绪	319	16.39%
		消极情绪	34	1.75%
物流与客户体验	1821	积极情绪	1347	73.97%
		中性情绪	440	24.16%
		消极情绪	34	1.87%
特性与用途	2298	积极情绪	2173	94.56%
		中性情绪	122	5.31%
		消极情绪	3	0.13%
健康与养生	3153	积极情绪	2866	90.90%
		中性情绪	206	6.53%
		消极情绪	81	2.57%

积极情感最为较大的类别是产品特性与功能、健康益处以及养生价值、以及消费体验与价格效益，它们的正面评价比例分别是94.56%、90.90%以及81.86%，这些数据说明了消费者对于助农产品有着十分突出的偏好，更愿意依靠购买行为来帮助农产品销售难题的解决。消费者对产品特性与功能的积极反馈特别突出，这可追溯到农村电子商务的兴起和发展，农产品得到了广泛的推广，凭借物美价廉的优势迅速赢得了众多消费者的认可。此外，平台已构建起较为成熟的销售体系，不仅能够更有效地预防农产品滞销情况的发生，还让更多人认识并喜爱上农产品，进而提升了购买体验。

中性情绪的评论数量达3335条, 占总评论数的27.77%; 与之相对, 消极情绪的评论数量为717条, 仅占总评论数的5.97%。这一数据对比表明, 即便存在消极情绪的情况下, 消费者的满意度依然存在较大的提升潜力, 并且在各个主题下的消极评论中, 也呈现出与之相近的比例特征。这意味着, 在消极情绪表达中, 消费者并未采取完全否定的态度, 而是以一种较为理性的方式进行批评。

（四）共现网络分析

共现网络分析法以一种更为直观、系统的方式揭示关键词的中心地位（即重要性）以及它们之间的复杂联系（刘自强等，2019）^[16]。词语的视觉呈现尺寸与其在文本中出现的频次成正比，即词语越大表示其出现得越频繁；词与词之间的空间距离反映了它们之间的关联强度，距离越近的词语，表明它们之间的联系或共现关系越紧密。图5展示了不同主题领域内关键词之间的共现现象，进而全面把握消费者对京东平台农产品助销项目评价的总体特征与趋势（丁宇诚和李浩，2023）^[17]。

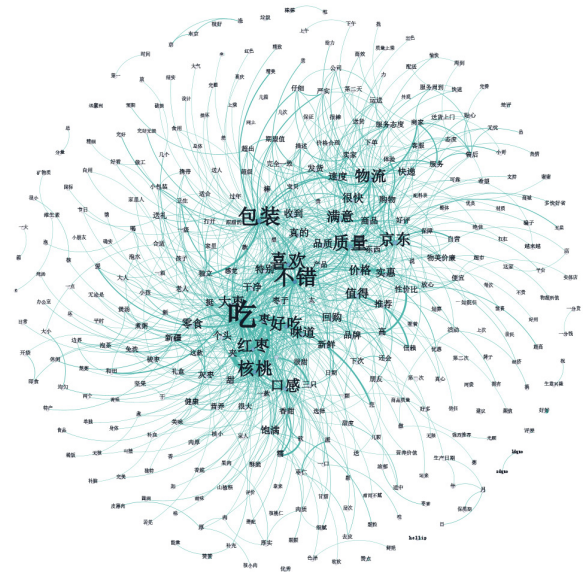


图5: 共现网络分析

共现网络分析图的结果与LDA主题分析相互印证,得出以下结论:(1)商家服务与物流配送协同。共现网络词云分析说明“商家、服务、速递、态度、体验、满意度、物流、时效、发货、快速”这些关键词之间存在密切联系,它们相互穿插,一起构建出一幅生动的关联画面,凸显出两者在助农产品购买流程中的协同作用。作为支撑消费体验的核心环节,客服团队与物流人员的专业素养及服务效率,均对消费者的满意度产生显著影响,友善的服务态度与高效的响应机制共同构筑起优质的服务体验。

(2)产品质量与价格的内在联系。依靠对“商品、特性、质地、性能价格比”等核心要素进行关联性探讨,可以发现消费者在购买行为中有着理智取向,也就是他们倾向于在保证质量可靠的前提下才会去考虑性价比,更倾向于购买性价比高、物美价廉的产品。这一发现印证价格与质量维度在消费选择中的相互制约关系,消费者在权衡产品价值时展现出明确的偏好倾向(李杰等,2023)^[18]。

深入分析共现数据，甄别出新兴关联模式：第一，“再来、下次、光顾、再买”等词汇组合体现显著的复购倾向，说明高质量产品和服务能激发持续消费意愿；第二，“包装、完好、严实”等节点的关联性，反映消费者对包装质量的高度认可，印证包装设计对提升消费体验的积极作用；第三，“分量、足、少”等对立词汇揭示消费者对产品分量感知存在显著差异，可能源于产品规格说明模糊或个体认知偏差，为产品优化提供了方向；第四，“满意、购物、产品”等节点的关联性，表明消费者对京东平台产品及购物流程的整体满意度较高，体现了京东在产品运营与用户体验管理方面的成效。

四、结论与建议

借助 LDA 模型,采用京东电商平台里的助农商品的在线评价数据以深入分析消费者偏好,主要结论如下:

（一）性价比是首要期望

“价格与性价比”是最受关注的议题之一，“口感与味道”次之，消费者决策时倾向于综合考量。情感分析表明，在性价比讨论中中性情绪占比较大，意味着消费者契合感有提升空间。因此，京东平台的成功策略应着重于持续提供高性价比助农产品，并始终以消费者需求为核心。

（二）服务质量是关键影响因素

在众多因素中，物流与商家服务的综合影响力次于产品自身品质。情感分析结果显示，物流配送和售后服务的负面评价占比

相对最低，证实了消费者对服务的总体满意度尚可。

（三）感知价值直接影响满意度

当消费者感知到的功能性效益（口感、外观、份量等）、质量性效益及体验性效益超过支付成本时，满意度提升。情感分析数据显示，该维度（体验与性价比主题）的积极和消极情绪占在主题中排名居中，表明其在消费者评价体系中处于中等关注度和满意度等级。因此，提升助农产品质量是平台增进消费者满意度的关键策略。

参考文献

- [1] 徐向梅. 农村电商迈入兴农新阶段 [N]. 经济日报, 2022-07-08(011).
- [2] 王雅丽, 杨雪云. 农村电商发展对城乡消费差距的影响 [J]. 科学决策, 2024(11): 6379.
- [3] 李宏兵, 王爽, 赵春明. 农村电子商务发展的收入分配效应研究：来自“淘宝村”的经验证据 [J]. 经济经纬, 2021, 38(01): 3747.
- [4] 刘明辉, 卢飞. 平台型电商普惠能否提振农村居民消费?: 来自中国淘宝村的证据 [J]. 管理评论, 2023, 35(09): 155168, 193.
- [5] 田野, 夏杰长. 农村电商、农村居民就业与共同富裕 [J]. 中国流通经济, 2024, 38(06): 6676.
- [6] 曹增栋. 电子商务促进了农民农村共同富裕吗?: 基于收入增长与差距缩小的双重视角 [J]. 西部论坛, 2024, 34(03): 95110.
- [7] 陈华帅, 谢可琴. 农村电商与农村家庭收入包容性增长 [J]. 经济评论, 2024(05): 92107.
- [8] 林莉, 杨承亮. 后疫情时代基于农产品直播平台消费者购买意愿及影响因素研究 [J]. 延边大学农学学报, 2022, 44(02): 7785.
- [9] 程谦益, 姜娟. 基于论坛文本的新能源汽车用户评价分析：以特斯拉和蔚来为例 [J]. 物流工程与管理, 2023, 45(03): 29-34.
- [10] 周婷玮. 基于共现网络与情感分析的多平台消费者评论主题比较研究 [J]. 知识管理论坛, 2023, 8(02): 79-91.
- [11] 杨梦琳, 卢益清. 基于在线评论的生鲜电商消费者满意度分析研究 [J]. 中国物流与采购, 2022(06): 44-46.
- [12] 王二朋, 倪郑宇. 农产品线上消费者的偏好特征研究：基于“京东”销售苹果在线评论数据的分析 [J]. 价格理论与实践, 2020(02): 120-123.
- [13] 殷彤, 于力超, 鲍伟. 精准扶贫背景下电商扶贫产品消费者满意度及其影响因素研究 [J]. 数理统计与管理, 2022, 41(04): 599-609.
- [14] 池毛毛, 潘美钰, 王伟军. 共享住宿与酒店用户评论文本的跨平台比较研究：基于 LDA 的主题社会网络和情感分析 [J]. 图书情报工作, 2021, 65(02): 107-116.
- [15] Bleidm, Ngay, Jordanmi. Latent Dirichlet allocation [J]. The Journal of Machine Learning Research, 2003, 3(04/05): 993-1022.
- [16] 刘自强, 许海云, 岳丽欣, 等. 基于 Chunk-LDAvis 的核心技术主题识别方法研究 [J]. 图书情报工作, 2019, 63(09): 73-84.
- [17] 丁宇诚, 李浩. 基于文本挖掘的医用防护服消费者评论分析 [J]. 丝绸, 2023, 60(06): 47-57.
- [18] 李杰, 陶源, 冯良珠. 俄乌冲突背景下俄罗斯顶级智库涉华评论主题和情感分析 [J]. 情报杂志, 2023, 42(08): 29-35, 48.