

基于数据挖掘的社交电商助力农产品上行的实证分析

费威¹, 李诗允¹, 吕欣阳²

(1.东北财经大学 经济学院, 辽宁 大连 116025; 2.中国人民大学 农业与农村发展学院, 北京 100872)

摘要:为促进社交电商有效助力农产品上行,以社交电商拼多多为例,通过对拼多多平台的农产品销售数据进行挖掘,利用LDA主题模型对农产品销售数据进行分类分析。研究表明:消费者对品牌溢价和高附加值商品的购买意愿较高;蔬菜类商户的区域品牌意识和蛋类、其他类商户的高附加值意识都较为薄弱;果蔬类商品区域品牌的市场价值有待增加;蛋类和其他类商品附加值有待提高。据此,应充分发挥社交电商平台的优势,加速形成头部效应,提升流量与销量之间的转化率。

关键词:社交电商;农产品上行;LDA主题模型;数据挖掘

中图分类号: F323.7

文献标志码: A

文章编号: 1009-2013(2022)04-0104-11

Empirical analysis of boosting online sales of agricultural products with E-Commerce based on data mining

FEI Wei¹, LI Shiyun¹, LYU Xinyang²

(1.School of Economics, Dongbei University of Finance and Economics, Dalian 116025, China; 2.School of Agricultural Economics and Rural Development, Renmin University of China, Beijing 100872, China)

Abstract: In order to boost effective online sales of agricultural products with social e-commerce, the paper takes social e-commerce Pinduoduo as an example, adopts the LDA topic model to classify and analyze the agricultural product sales data by data mining of the agricultural product sales data on Pinduoduo platform. The study shows that consumers are much more willing to purchase brand premium and high value-added commodities. The vegetable merchants are weak in regional brand awareness while egg merchants and other types of merchants are weak in high value-added awareness. The market value of regional brands of fruits and vegetables needs to be increased whereas the added value of eggs and other commodities needs to be raised. Hence, it is proposed that the advantages of social e-commerce platforms should be fully utilized, the formation of head effects should be accelerated, and the conversion rate between traffic and sales should be improved.

Keywords: social e-commerce; online sales of agricultural products; LDA topic model; data mining

一、问题的提出

随着电商的快速发展,农产品上行成为必然的选择。从2012年淘宝与遂昌县合作衍生的遂昌模式,再到阿里巴巴集团“千县万村”农村战略,京东建立“新通路”事业部设立乡村自提点等,电商

企业为促进农产品上行在不断努力^[1]。农业农村部、中央网络安全和信息化委员会发布的《数字农业农村发展规划(2019—2025年)》,对农产品网络零售额占农产品总交易额比重提出由2018年的9.8%提升至2025年15%的目标要求。然而,目前农产品上行尚存在生产经营分散化、农产品标准化实现难^[2]、仓储和物流成本高^[3]、冷链系统不完善等诸多阻碍^[4]。

传统电商销售农产品的功能主要体现在线上支付,其线下对接复制了通常的流通模式,而社交电商为农产品上行提供了新路径。近年来社交电商正

收稿日期: 2022-07-22

基金项目: 国家自然科学基金青年项目(71703014); 教育部人文社会科学研究规划基金项目(21YJA790015); 辽宁省教育厅基础科研项目(LN2020J07); 东北财经大学提升社会服务能力建设项目(SF-Z202205)

作者简介: 费威(1982—),女,辽宁鞍山人,教授,博士,主要研究方向为经济优化、食品安全管理研究。

处于蓬勃发展时期。Wang 等将社交电商定义为一种融合了线上线下环境的由社交媒体作为媒介的商务形式,通过社交媒体将购物和社交网络活动结合起来进行潜在销售^[5]。目前国内外学者主要聚焦于从消费者和商户两类主体研究社交电商促进农产品上行的影响因素。就消费者视角而言,Ernst 等指出农产品质量的不确定性是影响消费者购买意愿的主要因素,农产品质量信号传递是促进农产品电商发展的关键^[6];吴自强实证分析表明产品种类认知、品牌认知、食品安全与健康支付意愿等因素,对生鲜农产品网购意愿有正向影响^[7];王克喜等构建二元 logit 选择模型分析认为消费者自身属性、农产品和网购平台特征对生鲜农产品购买意愿具有显著影响^[8];Doha 等实证分析表明消费者的社交电商行为倾向可由功利价值、经济价值和社会价值予以解释^[9];Li 等基于 PPM 模型分析了消费者从传统电商转换到社交电商的影响因素^[10];巩永华等利用熵权法计算得出消费者的熟悉度和偏好相似度对用户信任的显著影响^[11]。从商户的角度,田刚等采用 Tobit 模型实证分析得出 IT 人才占比、与供应商的关系和物流设施水平对生鲜电商企业技术效率有显著正向影响^[12];曾亿武等实证分析认为先前工作经历抑制农户使用电商销售农产品,创业经历和培训经历、社会资本对此发挥促进作用^[13];侯振兴采用扎根理论研究发现基础设施、网络安全、第三方电商服务等因素会影响农户和企业采用电商销售农产品^[14];汪兴东等实证分析指出农村电商政策对农户电商销售行为有正向影响^[15];王婵等通过实证分析得出农户的社交软件使用数和社交软件好友数等都会促进农户参与社交电商^[16]。此外,利用大数据研究农产品网购消费行为的研究日益增加。相关研究主要集中在消费者在线购买的影响因素分析,利用数据分析,讨论在线评论特征以及网络口碑等对消费者购买决策的影响。胡雅淇等通过爬取淘宝平台生鲜羊肉产品的月度截面数据,分析得出在线评论数量和可视化评论能够显著提高生鲜农产品销量,差评数量会明显降低其销量^[17];王英等利用爬虫技术和统计方法分析评论数据,探讨了品牌、在线评论和农产品销量之间的关系^[18];雷兵等利用大数据分析地方特色、网络口碑与农产品上行之间的作用机制^[19];王翠翠等通过构建农户可持续生计评价指标体系,以湖北

省农户调查数据为例,评估了农村电商扶贫成效,认为是否参与农村电商是影响扶贫的直接因素^[20]。

综上所述,现有文献为促进农产品上行提供了丰富的研究参考。学者们以实证分析、问卷调查、扎根理论等方法为主,分析了农产品特征、消费者自身属性和农户历史经验等影响因素,还有部分学者采取爬虫技术,利用数据在线评论分析消费者网购行为,但针对社交电商农产品上行的销量数据挖掘研究农产品上行的相对较少。不同于传统计量方法,数据挖掘可以实现更全面更广泛的数据搜集,且在数据的采集、清洗和分析等方面均体现出新的特征,有助于进行更直观的定量分析^[21]。对此,笔者拟以社交电商拼多多为例,用数据挖掘方法分析社交电商平台农产品销售情况,提炼社交电商助力农产品上行存在的主要问题,据此剖析原因,提出相应的解决对策。

二、农产品销售数据挖掘的研究设计

(一) 研究方法

本研究使用潜在狄利克雷分布(Latent Dirichlet Allocation, LDA)进行主题建模,利用数据挖掘方法对农产品进行分类。LDA 是一种用于从大量未标记的文档中提取有意义主题的无监督算法^[22]。构成数据集的所有商品名称被视为语料库,在语料库中的每个商品名称被称为文档。LDA 是将每个文档看作是若干词的向量,每个词都可以表示某个主题的概率分布,其输出将是一个词集列表,表示语料库中每个文档的不同主题及主题分布。

文档生成过程如图 1 所示,参数 α 和参数 η 分别控制一个狄利克雷分布。参数 α 首先随机生成文档对应的主题多项式分布 θ ,再由 θ 随机生成一个主题 z 。参数 η 随机生成主题对应词语的多项式分布 β ,综合主题 z 和主题对应词语分布情况 β 生成词语 w ,如此循环生成一个文档,其中包含 N 个词语,最终生成 K 个主题下的 M 篇文档。在文档中,每个单词 w_i 都与一个潜在变量 z_i 有关,其中 $i=\{1, \dots, K\}$,从而表明单词 w_i 是由哪个主题生成的^[23]。单词 w_i 的概率可以表示为:

$$P(w_i) = \sum_{j=1}^k P(w_i \mid z_i = j)P(z = j) \quad (1)$$

式(1)中 $P(w_i \mid z_i = j) = \beta_{ij}$ 是单词 w_i 在主题 j 中的概率, $P(z = j) = \theta_j$ 是一个文档特定的混合

权重,表示主题在文档中的比例。

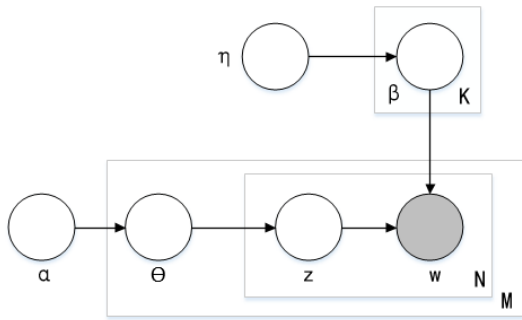


图1 LDA模型的文档生成过程

在主题建模中,依据相关研究^[24],令参数值 $\alpha=0.1$ 、 $\beta=0.01$,采用主题困惑度确定最佳主题数 K ,用主题一致性评估其可解释性。主题困惑度越低代表模型的预测能力越高,用软件Python3.8.8得到关于主题困惑度与主题数的折线图中会存在一个拐点,拐点处的主题数即为最佳主题数量。主题一致性通常假设如果某主题更易于解释,则该主题中的顶部单词对应语料库的文档中将更频繁地共同出现,主题一致性得分越高越有解释力度。

(二) 研究设计

目前拼多多已经成为销售额最高、参与销售商户规模最大的农产品社交电商平台。选择拼多多作为农产品上行的数据挖掘平台具有典型代表性。拼多多成立于2015年,是专注于拼团购物的第三方社交电商平台,平台用户通过发起和朋友、家人、邻居等社交群体的拼团,实现以更低的价格购买优质商品的目标。2020年拼多多农产品及农副产品订单总额超过2700亿元人民币,规模同比翻倍,2021年全年累计订单610亿件,同比增长59%^①。拼多多主要面向下沉市场,即把目标用户定位在三线及

以下城市的广大消费者。针对农产品上行的“最后一公里”问题,2018年拼多多率先提出“最初一公里”的战略,通过产地直销实现了消费端“最后一公里”和原产地“最初一公里”的直连,避免了采购入仓等中间环节造成的高成本。流通成本的降低使在不损害商户利益的情况下实现了消费者、平台和商户的共赢。

拼多多以相较于传统电商更宽松的资质条件进行招商:商户及相关企业在拼多多经销农产品,仅需上传食品经营许可证和缴纳保证金;平台对农产品实行零佣金;入驻商户在一个类目下可以同时开设两家店铺,入驻水果生鲜类目的新商户个人和企业分别需要缴纳不同额度的店铺保证金。拼多多的营销流程如图2所示。拼多多在营销方式上主要以拼单分享为主,通过低价引导其初始农产品消费者分享商品给好友进行拼团购买,建立用户链接,发展潜在消费者,同时采取微信等社交平台用户自发加入不同兴趣群体的方式,让消费者与熟人之间分享商品链接,通过信任吸引潜在消费者发生购买行为,即从传统电商的“人找货”转变为“货找人”。这些群体能够自发且较为精准地完成用户群体细分,容易产生重复购买行为,实现从关系到交易、生活到生意的转化。在交易完成后,消费者还会对商品质量、平台服务等做出评价,在社交圈广泛传播,逐渐形成商品的口碑。在这种营销模式下,“关键意见领袖(KOL)”应运而生,拥有流量的网红、社群群主等可以通过分享“好物”推销商品,随着商品口碑的形成与流传不断吸引新的用户关注,从而产生正反馈效应。

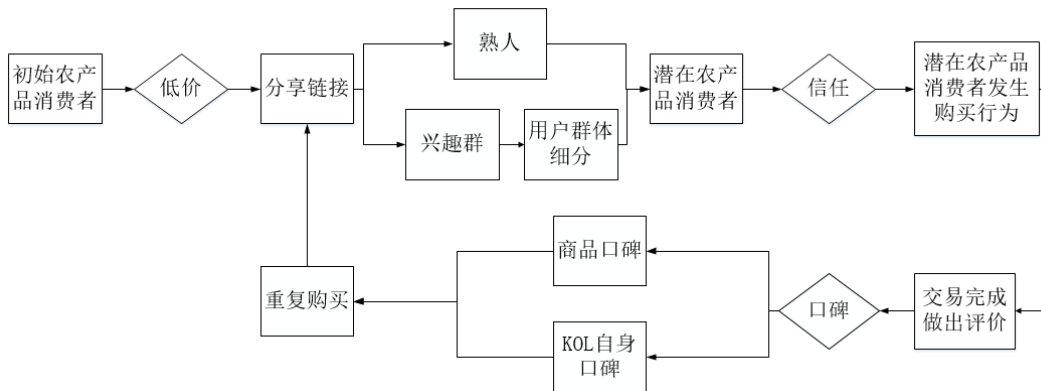


图2 拼多多营销流程图

(三) 数据来源与分析框架

1. 数据来源

本研究通过在拼多多网页对“农产品”“农产品大全”“农产品直销”等关键词进行检索，利用软件 XPath Helper 2.0.2 对商品名称、价格和销量^②数据进行爬取。整个数据获取过程在 2021 年 10 月 29 日至 11 月 4 日一周内完成，最终爬取到 13866

条商品信息。原始数据示例如表 1 所示。商品分为蛋类、果蔬类和其他三类。为简化且便于说明，每类选取两个商品作为例子在表 1 中予以展示。其中，商品名称中包括很多修饰词凸显商品特点，如售卖方式包括 10 斤、5 斤等，商品包装类型包括罐装、精品礼盒装等，商品产地包括内蒙古、福建平和和琯溪等。

表 1 原始数据示例

商品名称	价格(单位:元)	销量(单位:万单)
无穷盐焗鸡蛋整箱 600 克爱辣乡巴佬香卤鸡蛋休闲即食零食营养早餐	10.34	10
现捡土鸡蛋农家散养新鲜柴鸡蛋现捡草鸡蛋笨鸡蛋月子蛋整箱包邮	19.8	0.0014
官方旗舰店 35 年老树红心柚子福建平和琯溪红心蜜柚精品礼盒装	5.5	10
正宗突尼斯软籽石榴超甜新鲜应季水果红石榴无籽 5/3/1 斤整箱批发	6.8	10
枫正宗内蒙古手撕牛肉干西藏特产风干牦牛肉麻辣五香零食熟食	24.9	9.7
香辣蟹罐装熟食小螃蟹即食全母蟹一元蟹大闸蟹钳切蟹麻辣海鲜零食	11.9	10

2. 数据分析框架

通过将数据挖掘应用于农产品电商的研究，一方面能够利用实时数据，对农产品电商现状进行更为直观的定量分析；另一方面，基于经济学理论的

相关性分析能够挖掘数据间的联系，探索农产品电商的主要影响因素^[25]。因此，本研究通过数据清洗、分词处理和主题分类的步骤进行数据处理。具体如图 3 所示。

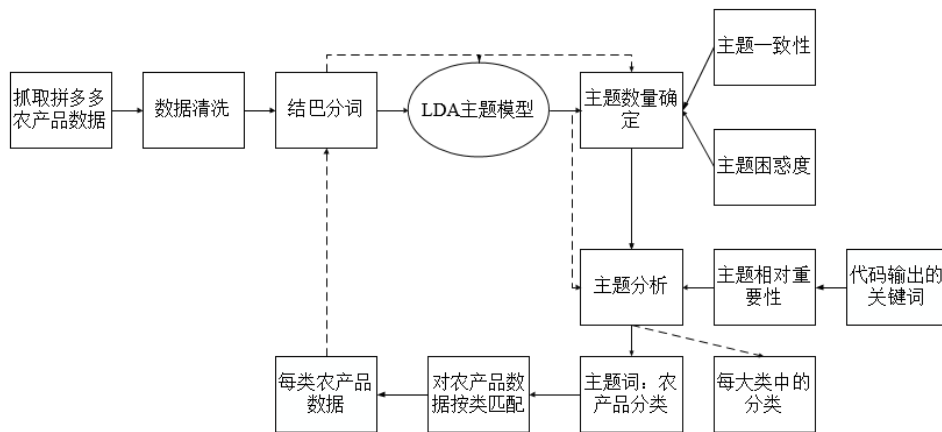


图 3 农产品销售数据分析框架

(1) 数据清洗。针对多次少量爬取数据存在的部分商品被多次爬取的情况，需要删除爬取的重复信息，并且剔除与农产品无关的商品数据。经过上述数据清洗处理后，留存 10471 条商品有效信息。

(2) 分词处理。目前中文分词根据实现原理和特点分为基于词典分词算法和基于统计的机器学习算法。词典分词算法的原理是对前缀词典实现高效的词图扫描，生成句子中汉字所有可能成词情况构成的有向无环图，采用动态规划查找最大概率路径，找出基于词频的最大切分组合^[26]。本研究对商品名称进行文本挖掘，由于商品名称大部分由形

容词和名词构成且不易产生歧义，因此采用基于词典分词的开源中文结巴分词工具。根据商品名称特征设计专门的停用词表^③，例如在停用词表上增加分辨商品种类无意义的词，如包邮、顺丰、整箱、批发等，以更好地对商品类别进行主题提取。

(3) 主题分类。利用 Python 中的机器学习库 scikit-learn(简称 sklearn)进行 LDA 主题模型分析。针对爬取到的农产品商品名称确定主题数量，再用主题一致性和主题困惑度测度主题数量是否合理，得到科学合理的主题数量。通过代码输出每个主题的关键词并进行主题分析以概括出主题词，主题词

即为农产品数据的初步分类。然后根据得到的农产品分类对商品数据进行匹配和归类,经过以上步骤分别对挖掘的每类农产品进行处理再分析^④。

三、社交电商助力农产品上行的实证分析

(一) 主题确定

本研究通过构建不同主题数量的 LDA 模型以确定最佳主题数量。由于进行主题分类的对象不是常见的逻辑连贯的文本(评论、政策等),而是由形容词和名词组成的商品名称,需要结合模型的可视化图、主题困惑度和主题一致性综合衡量进行最佳主题数(K)确定。

分别输出 $K=3$ 至 $K=8$ 的 6 个可视化图。例如,图 4 是主题数为 3 时的可视化结果,直观展示了主题分布情况,3 个圆分别在 3 个不同象限内并且没有交叉重叠部分,说明主题数量为 3 时分类清晰。

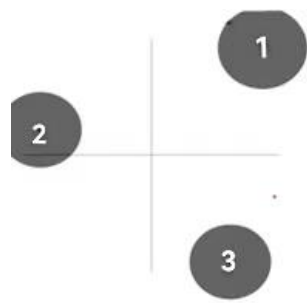


图 4 主题数为 3 时基于整体文本的可视化结果

困惑度分析结果如图 5 所示。主题困惑度随着主题数量的增加而递减,并且主题数 K 为 3 时存在拐点,即时曲线急剧下降, K 在 3~4 时曲线基本趋于平稳,因此,确定 LDA 模型的最佳主题数量为 3。

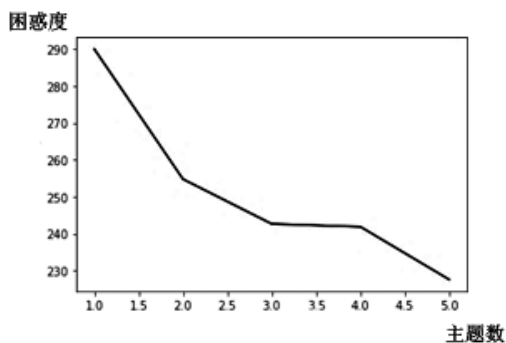


图 5 主题困惑度得分

由于主题模型是一种无监督学习方法,不能保证输出主题的可解释性,本研究用一致性评估模型的可解释性,结果如图 6 所示。 K 在 1~3 时主题一致性得分不断上升,在 3~4 时趋于平稳。综合上

述分析,最佳主题数量 K 选定为 3。

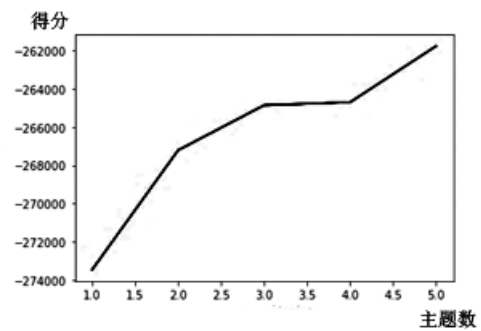


图 6 主题一致性得分

根据 LDA 模型得到主题和关键词表,具体如表 2 所示。表 2 展示了 3 个主题及其排列在前 8 个的关键词。输出的主题关键词只有顺序而没有主题名称,根据关键词及其之间的逻辑关系对每个主题进行命名,例如以“鸡蛋”“红心”“鸭蛋”和“皮蛋”为关键词的主题被命名为“蛋类”。其中,“农村”“特产”“网红”等均为对商品的描述,“冰糖”“五香”分别是以冰糖橙子、冰糖心苹果、五香牛肉等形式出现,属于无关词汇,与农产品本身的分类无直接关系。

表 2 主题和关键词表

主题	主要关键词
蛋类	鸡蛋 农家 红心 鸭蛋 板栗 农村 特产 皮蛋
果蔬	水果 蔬菜 干货 农家 苹果 特产 冰糖 橙子
其他类	零食 特产 小吃 水产 五香 网红 牛肉 食品 (含肉制品及水产)

(二) 实证分析结果

基于拼多多的商品信息数据挖掘,本研究剖析了社交电商平台农产品销售现状及其特点,对具有代表性的各类农产品进行分析,包括商品信息中有无产地标识对销售的影响、商品是否经过加工对销售的影响等。

1. 农产品整体销售情况分析

对爬取的农产品数据按类别进行匹配,通过筛选、分类、计算后的结果如表 3 和表 4 所示。为简化且不失一般性,参照相关研究将农产品销量数据进行分级处理^[26],共分为 4 个等级,销量 1 级至销量 4 级分别为累计销量(销售计数单位)小于 1000、1000 至 1 万、1 万至 10 万和超过 10 万。

根据表 3 和表 4 数据易见:在总体销售情况上,销售蛋类的商户占比不足 20%,且蛋类销量占比低于其商品数量占比的 6.53%,蛋类销量明显低于果

蔬和其他类；果蔬是平台上销量最高的农产品，但由于单价较其他类偏低等原因使销售额占比有所下降；其他类的销量占比 41.19%，由于包括牛肉和水产等单价较高的商品，其销售额占农产品整体销售额的 50% 以上。

表 3 农产品销售总体情况

	蛋类	果蔬	其他类
商品数量 (单位: 个)	1980	4031	4205
销量 (单位: 万单)	5705.94	20407.38	18286.7
销售额 (单位: 万元)	97618.62	232028.2	411895.4

表 4 农产品销售比率

	蛋类/%	果蔬/%	其他类/%
商品数量占总数量的比率	19.38	39.46	41.16
销量占总销量的比率	12.85	45.96	41.19
销售额占总销售额的比率	13.16	31.29	55.55
销量 1 级的商户比率	21.77	14.31	13.34
销量 2 级的商户比率	31.82	17.84	24.16
销量 3 级的商户比率	28.74	29.82	31.99
销量 4 级的商户比率	17.68	38.03	30.51

从单个商户的农产品销售情况来看，超过一半

的果蔬和其他类农产品的销量达到 1 万以上，蛋类销量达到 10 万以上的占比最少，并且销量小于 1000 的占比最多，从而可以推断购买蛋类农产品的消费者相对较少，社交电商平台中“头部效应”尚不明显。商户更应该注重提升商品品质及售后服务水平，加速正面口碑在消费者之间的传播，实现扩大销售规模的目标。其中，“头部效应”^⑤中“头部”可以理解为每类农产品中销量 4 级的商户，这些商户由于累计销量足够多、口碑足够好、知名度足够高等更易吸引消费者购买，相应商户的商品销量持续增加形成正反馈的效应即“头部效应”。

本研究对三大类农产品再分别进行 LDA 主题分类，提取与农产品名称相关的关键词，依据频率降序罗列，并将表示一类商品的名称合并整理，如脐橙、甜橙、橙子可以合并成橙子，得到农产品数据三级分类。通过农产品的三级分类发现，蛋类按常识可分为未加工和经初加工的二级分类，果蔬类和其他类可以显著区分为水果和蔬菜、肉和水产，最终得到表 5 所示的农产品分类。

表 5 基于数据挖掘的农产品分类

一级品类	二级品类	三级品类
蛋类	蛋 (未加工)	鸭蛋, 种蛋, 鹅蛋, 鸽蛋, 鹌鹑蛋, 鸡蛋, 胚蛋 (喜蛋、毛蛋)
	蛋 (经初加工)	皮蛋, 茶叶蛋, 咸鸭蛋, 卤蛋
果蔬类	水果	苹果 (蛇果), 橙子 (脐橙、甜橙), 柚子 (蜜柚、葡萄柚), 枣子 (冬枣、红枣、青枣、大枣), 猕猴桃 (奇异果), 石榴
	蔬菜	地瓜 (红薯、番薯、蜜薯、香薯), 柿子 (番茄), 青菜 (白菜、菠菜、生菜、大白菜、酸菜、野菜、咸菜、榨菜、酱菜), 萝卜 (红萝卜、青萝卜), 生姜 (鲜姜、洋姜), 黄瓜 (青瓜), 豆角, 大蒜 (干蒜), 芋头 (山芋、香芋)
其他类	肉	牛肉 (酱牛肉、牛筋、蹄筋、牛排)
	水产	螃蟹 (大闸蟹、梭子蟹、河蟹), 海参 (刺参), 大虾 (虾仁、虾皮、虾米), 鱿鱼

2. 果蔬类农产品销售情况分析

对果蔬类中三级分类商品的价格做描述性统计得到表 6，表中按二级分类排列，即前 6 行是水果，其余行是蔬菜。由表 6 可以看出，平台上水果占比更高，苹果和橙子占比已达 13.84%，可见水果商户在平台上的发展趋于成熟；水果相比于蔬菜的

商品价格差异性更大，离散程度更高，水果价格的标准差均在 8 以上，67% 以上的蔬菜价格标准差在 4~7 之间；水果的商品附加值较蔬菜更高，如苹果和猕猴桃的价格最高分别可达 114 元和 125 元，而蔬菜中番茄的最高价格为 88.80 元，其他蔬菜的最高价格均在 15~40 元之间。

表 6 果蔬类价格的描述性统计

	平均值	标准差	最小值	最大值	中位数	商品数量	占比/%
苹果	7.89	10.89	1.00	114.00	9.30	291	7.24
橙子	13.82	13.32	1.00	93.00	9.90	266	6.60
柚子	10.35	8.78	0.50	66.90	7.90	145	3.60
枣子	10.69	8.03	0.78	37.63	7.90	97	2.41
猕猴桃	11.54	14.06	1.90	125.00	7.80	86	2.13
石榴	14.42	11.02	2.00	59.80	11.80	58	1.44

表6(续)

	平均值	标准差	最小值	最大值	中位数	商品数量	占比/%
红薯	7.70	5.24	0.60	40.50	6.80	179	4.44
番茄	14.88	12.70	0.99	88.80	10.90	161	3.99
青菜	10.94	7.54	0.87	29.74	9.90	122	3.03
萝卜	10.11	6.92	0.50	39.90	8.75	115	2.85
生姜	5.48	4.58	0.50	32.80	4.68	107	2.65
黄瓜	13.17	5.12	2.38	25.80	13.89	80	1.98
豆角	12.96	15.07	3.96	28.88	9.83	74	1.84
大蒜	8.28	3.61	0.80	15.78	7.90	49	1.22
芋头	6.71	4.09	0.60	15.8	5.70	43	1.07

进一步分析果蔬类中三级分类商品的销售情况,如表7所示。水果的成交量相较于蔬菜保持在更高水平。具体表现为:水果的销量均值显著超过蔬菜,水果的销量均值在4万至7万单之间,56%以上的蔬菜销量均值在3万至4万单之间;水果的累计销量显著高于蔬菜,水果(除石榴)的累计销量均超过500万单,78%以上的蔬菜累计销量低于500万单;果蔬类达到销量4级和销量3级的商户(以下简称4级商户,3级商户)占总商户比率全

部超过了50%,说明果蔬类商品在平台上的头部效应已初步形成;水果的平台头部效应更强,蔬菜商户可发展的空间相对较大,水果4级商户比率高于蔬菜4级商户比率,水果1级商户比率远远低于蔬菜1级商户比率;蔬菜商户的销量差异显著,如红薯和芋头4级商户占比均超过45%,而青菜和黄瓜4级商户占比均不足20%,青菜和生姜1级商户比率接近30%,而大蒜和芋头1级商户的比率则低于7%。

表7 果蔬类销售情况

	销量均值 (单位:元)	累计销量 (单位:万单)	销量4级的 商户比率/%	销量3级的 商户比率/%	销量2级的 商户比率/%	销量1级的 商户比率/%
苹果	5.58	1629.20	43.64	30.24	14.78	11.34
橙子	4.08	1084.92	31.95	20.68	26.69	20.68
柚子	5.85	848.76	44.83	31.03	20.00	4.14
枣子	6.00	582.42	40.21	42.27	13.40	4.12
猕猴桃	6.96	598.82	56.98	26.74	8.14	8.14
石榴	6.42	372.56	50.00	25.86	15.52	8.62
红薯	5.69	1018.13	46.93	20.67	14.53	17.88
番茄	4.72	759.71	32.30	36.65	13.66	17.39
青菜	3.11	382.71	19.67	20.49	31.97	27.87
萝卜	3.83	440.99	24.35	33.91	21.74	20.00
生姜	3.55	379.65	24.30	26.17	19.63	29.91
黄瓜	3.44	275.47	17.50	33.75	31.25	17.50
豆角	3.59	269.25	25.68	28.38	32.43	13.51
大蒜	7.65	374.67	67.35	22.45	10.20	0.00
芋头	6.44	276.83	48.84	30.23	13.95	6.98

根据表6,选择水果和蔬菜中最具代表性的、商品数量占比最高的苹果和红薯进行对比分析,进一步得出水果和蔬菜的平台销售现状差异可能存在的原因,通过观察商品名称发现果蔬类农产品中最明显的区别是有无产地标识,并据此予以区分,结果如图7和图8所示。

观察苹果和红薯的商品数量占总数量的比率,

可以看出红薯有产地标识的比率远低于苹果有产地标识的比率,如图7、图8所示。有产地标识的红薯数量占红薯总数量的比例不足30%,而有产地标识的苹果占比超过60%,表明水果的区域品牌建设领先蔬菜;有产地标识的苹果和红薯商品数量占比、销量占比和销售额占比均呈递增趋势,说明有产地标识的商品能够以更高的价格实现更多的成

交易，同时也在一定程度上体现了消费者偏好购买有产地标识的商品，愿意为其支付更高的价格，这与张耘堂等的研究结论^[27]有一致性，即农产品原产地的知名度会提升消费者信任，且长此以往会上升为情感性的支持；再观察商品销量达到 4 级和 3 级的商户比率，销售有产地标识商品的 4 级和 3 级商户比率均高于销售无产地标识的比率，很多农产品以产地命名，因此产地成了农产品品牌，则可认为社交电商平台和农产品区域品牌相辅相成，且有产地标识的商户更可能形成头部效应，这同已有研究指出电商驱动农产品品牌价值创造的目的和结果是促进农产品品牌成长的结论相一致^[28]。销售有产地标识商品的 1 级和 2 级商户占比相对销售无产地标识的占比更少，说明区域品牌化可以在一定程度上提升以散户形式经营的商户数量。



图 7 苹果的销售情况

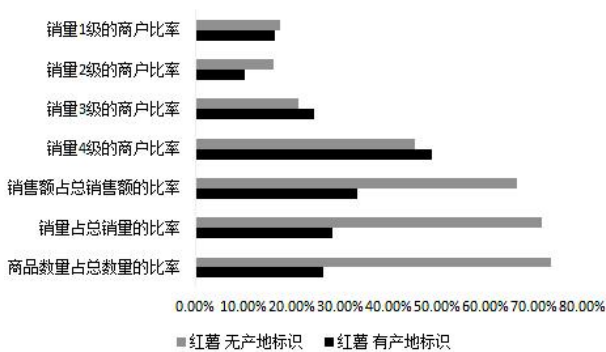


图 8 红薯的销售情况

综上，果蔬类农产品销售存在以下主要问题：第一，蔬菜类农产品区域品牌较水果类有更大的建设空间，蔬菜价格的标准差相对较低，表明商品同质化严重，市场趋于完全竞争，商户的区域品牌意识有待加强；第二，我国消费市场潜力巨大，消费者对品牌的平均溢价购买意愿较强，而我国很多农产品价值创造力不足，只能占据低端市场^[29]，因此

果蔬类农产品应注重提升区域品牌的市场价值，打造在国际市场上具有突出竞争力的高端农产品品牌；第三，社交电商平台的优势有待进一步发挥，有研究指出，农产品的微笑价值链曲线与工业品类类似，即在农产品的整个产业价值链中，品牌营销、物流、加工等环节会比单纯的种植或养殖环节创造更多的附加价值^[30]，因此商户需建立并维护好商品口碑，尤其是销售有产地标识农产品的商户，更应避免以劣充优、“搭便车”等损害区域品牌口碑现象的发生。只有保证商品质量，才能加强消费者品牌推崇意愿，更充分发挥社交电商平台正面口碑“一传多”的优势，从而形成正反馈效应。

3. 蛋类与其他类农产品销售情况分析

观察商品名称可得，蛋类和其他类（含肉制品和水产）农产品最明显的区别为是否经过加工，且这两类商品都具有在销售环节对物流、储藏条件高要求的特征，因此将这两类农产品合并分析。蛋类商品无法就任意一个三级品类（已按未加工和经初加工分类）对比分析，因此选取蛋类全部商品，以及肉类和水产中商品数量占比最高的牛肉和螃蟹，分析蛋类和其他类农产品的平台销售现状（具体销售情况见图 9、10、11）。

与果蔬类中苹果和红薯销售现状比较，蛋类数量占全部农产品数量的比率较低，且蛋类和其他类的 3、4 级商户比率不显著高于 1、2 级商户比率；蛋类与其他类农产品 4 级商户占比显著低于果蔬类 4 级商户，除经初加工牛肉的 4 级商户比率达到 40%，其余商品的 4 级商户比率均不足 30%，而果蔬类 4 级商户比率均在 40% 以上；三类农产品的 3 级用户比率基本持平，均维持在 30% 左右；蛋类和其他类 1 级、2 级商户比率显著高于果蔬类，其中未经加工蛋类 1 级和 2 级商户的比率之和接近 50%，果蔬类 1 级和 2 级商户的比率之和均低于 30%。

结合图 9、10、11，还可以观察到蛋类、牛肉和螃蟹经初加工商品数量占总商品数量的比率。蛋类和螃蟹的经初加工商品数量占比分别在 40% 和 20% 左右，牛肉的经初加工商品比率高达 80%，表明蛋类和水产的加工普遍处于初级阶段；以牛肉为代表的肉制品加工已经较为成熟，经初加工的蛋类、牛肉和螃蟹商品数量占比、销量占比和销售额

占比均呈先升后降的趋势,销量占比的上升说明消费者更偏好经初加工的商品,这是由于未经加工的商品有易腐性、非标准化、体验性强等特点,缺乏能以较低信息成本进行甄别的物理特性,消费者难以建立有效的信任机制^[31];而销售额较销量占比的降低是由于经初加工的商品每单成交价较低,如经初加工牛肉以每单售价不超过40元的牛肉丸、牛肉酱等形式销售,未经加工牛肉按斤销售,每单售价明显高于经初加工牛肉商品,表明经初加工商品的附加值有较大提升空间;经初加工的商品4级商户比率占比均高于未经加工商户占比,其中经初加工蛋类和牛肉4级商户比率显著高于未经加工的,表明未经加工的商品有更充分利用社交电商平台提升消费者信任的空间。

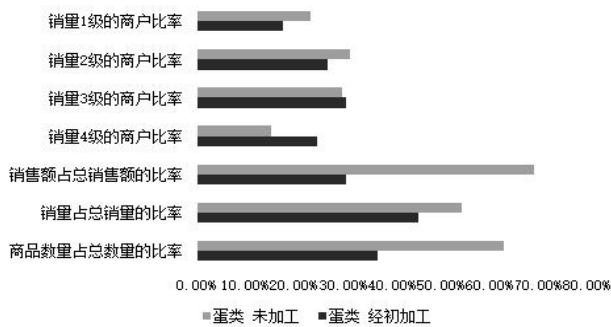


图9 蛋类的销售情况

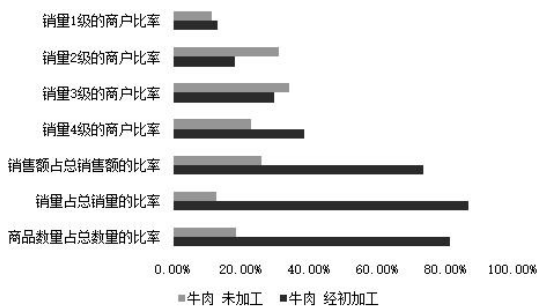


图10 牛肉的销售情况

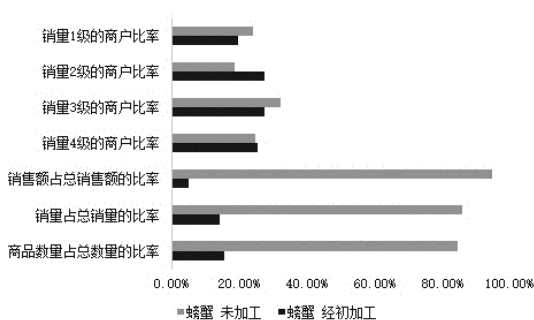


图11 螃蟹的销售情况

综上,蛋类和其他类农产品销售存在以下主要

问题:第一,蛋类和其他类农产品电商的发展处于初级阶段,其销售需要冷链技术和高效物流作为强大支撑,由于目前我国生鲜农产品冷链物流发展仍面临着“小、散、乱、弱”的现实困境^[32],消费者会对未经加工的蛋类、肉制品的新鲜度等品质存在顾虑,此类生鲜农产品的经销商户应提供有效的客户保障措施(如直接便利的售后退换货、补偿机制等)降低消费者的感知风险,提升消费者网购生鲜农产品品质的信任水平,从而较大程度地提高消费者网购生鲜农产品的意愿与实际购买频率^[33];第二,蛋类和其他类商品附加值有待提高,我国农业生产成本“地板”的持续上升与农产品价格“天花板”的不断触顶^[34],经初加工蛋类和其他类商品经销的商户需要合理提高其供给的农产品附加值,从而提升议价能力,实现增收;第三,社交电商平台的头部效应尚未成熟,商户除了对农产品本身进行深加工外,还应考虑从原料产地、加工的产品类型等角度实现加工农产品的附加值增值^[35],加快形成一条高质高效的农产品生产加工销售链条,累积销量和传播正面口碑,提升流量与销量的转化率。

四、主要结论及启示

本研究以社交电商拼多多为例,用数据挖掘方法爬取平台数据,基于LDA模型对农产品数据分类,分析平台中农产品销售的总体情况及各类农产品销售情况,从商户角度提炼社交电商助力农产品上行存在的主要问题并剖析原因,得出以下主要结论:一是消费者对品牌溢价和高附加值商品的购买意愿较高,蔬菜类商户的区域品牌意识和蛋类、其他类商户的高附加值意识较为薄弱,果蔬类商品区域品牌的市场价值有待提升,蛋类和其他类商品附加值有待提高;二是蛋类和其他类农产品的经销商户的发展仍处于初级阶段,需要配套完善的冷链物流运输以支撑和保障农产品电商的健康发展,进而提高消费者感知信任水平;三是应进一步发挥社交电商平台的优势,加速形成头部效应,提升流量与销量的有效转化率。

据此,可以得到以下启示:

第一,通过推进农产品区域品牌建设,提高农产品附加值。从农户、合作社到商户等都应增强农产品区域品牌建设意识,从农产品生产到营销确保

品牌商品的品质,以此为基础,利用社交电商平台的低价拼团和熟人信任等优势进一步提升品牌知名度。同时,随着人们收入水平的不断提高与健康安全认识的不断增强,消费者对初、深加工农产品的需求大量增加。加工农产品的附加值较高,有利于农户等农产品供给主体的整体效益提升。对此,基于消费偏好的异质性精准定位细分市场,实施差异化的市场销售策略至关重要,尤其要依据社交电商具有的熟人优势与信任特征,进行初深加工农产品的分类营销,丰富农产品上行市场,促进其繁荣发展。

第二,通过加强物流等基础保障设施建设,全面提高消费者感知信任水平。针对生鲜农产品,仓储、运输等物流基础保障设施和设备的完善尤为重要,对此根据各地发展现状,结合新基建进程,强化农产品流通环节的软硬件建设。以平台与数据技术研发企业为主导,政府部门助推的方式,加大区块链等数字技术的推广应用,提高电商平台经销农产品的全程可追溯性,全面提升物流配送效率,加快形成更加高效高质的农产品供应链条,为电商助力农产品上行提供可靠的基础保障体系。

第三,通过提升社交电商营销规范性,促进农产品上行良性发展。以拼多多为代表的社交电商平台在提供相对低廉商品的同时,由于商品质量、等级等参差不齐给平台的定位与形象造成了负面影响。对此,社交电商平台应在明确自身定位的基础上,确立平台发展的基本门槛条件,面向不同需求消费者的商品供给与营销都应建立在商品质量保障的基本前提下。商户进入平台后,对自身经销商品应有精准的分类定位,对于生鲜农产品、特色农产品,在保质的情况下进行分类分级销售,加强正面口碑传播,保持与维护社交电商的信任优势,实现可持续的良性发展循环。

注释:

- ① 数据来源: 网易. <https://www.163.com/dy/article/H387ARB60537MGX9.html>
- ② 销量是从商品链接建立至今的累计销量。
- ③ 停用词表是由在文本处理过程中为提高检索效率设置的停用词组成的。停用词是指如果遇到会立即停止处理和忽略掉的词。
- ④ 图 3 中实线为农产品总体数据的分析过程,虚线为二次分类时每类农产品的数据分析过程。

- ⑤ “头部效应”的原意是指在一个领域中,第一名往往会获得更多的关注,拥有更多的资源,“头部”就是所在赛道里的高价值并且有优势的领域。参考出处具体网址 <https://baike.baidu.com/item/%E5%A4%B4%E9%83%A8%E6%95%88%E5%BA%94/50884988>

参考文献:

- [1] 郝国强. 特色农产品电商营销模式及技术支持研究[J]. 广西民族大学学报(哲学社会科学版), 2019, 41(1): 77-84.
- [2] 刘欢. 线上线下融合发展模式下农产品流通的困境及出路[J]. 商业经济研究, 2016(18): 147-149.
- [3] 徐志闵. 线上线下融合发展模式下农产品流通困境与对策[J]. 农业经济, 2018(9): 136-138.
- [4] 王海南, 宁爱照, 马九杰. 疫情后我国生鲜农产品供应链的优化路径与策略[J]. 农村经济, 2020(10): 107-113.
- [5] WANG C, PING Z. The evolution of social commerce: the people, management, technology, and information dimensions[J]. Communications of the Association for Information Systems, 2012, 31(5): 105-127.
- [6] ERNST S, HOOKER N H. Signaling quality in an e-commerce environment: the case of an emerging e-grocery sector[J]. Journal of Food Products Marketing, 2006, 12(4): 11-25.
- [7] 吴自强. 生鲜农产品网购意愿影响因素的实证分析[J]. 统计与决策, 2015(20): 100-103.
- [8] 王克喜, 戴安娜. 基于 Logit 模型的绿色生鲜农产品网购意愿的影响因素分析[J]. 湖南科技大学学报(社会科学版), 2017, 20(2): 87-93.
- [9] DOHA A, ELNAHLA N, MCSHANE L. Social commerce as social networking[J]. Journal of Retailing and Consumer Services, 2019, 47: 307-321.
- [10] Li C Y, KU Y C. The power of a thumbs-up: Will e-commerce switch to social commerce?[J]. Information & Management, 2018, 55(3): 340-357.
- [11] 巩永华, 陈磊, 黄卫东. 基于关系强度和偏好相似度的社交电商用户信任融合模型[J]. 管理现代化, 2019, 39(6): 96-98.
- [12] 田刚, 张蒙, 李治文. 生鲜农产品电商企业技术效率及其影响因素分析——基于改进 DEA 方法与 Tobit 模型[J]. 湖南农业大学学报(社会科学版), 2018, 19(5): 80-87.
- [13] 曾亿武, 陈永富, 郭红东. 先前经验、社会资本与农户电商采纳行为[J]. 农业技术经济, 2019(3): 38-48.
- [14] 侯振兴. 区域农户农企采纳农产品电子商务的影响因素[J]. 西北农林科技大学学报(社会科学版), 2018, 18(1): 66-74.
- [15] 汪兴东, 刘雨虹. 农户电商销售意愿与行为偏差分析——基于江西省 54 个县的调查数据[J]. 农林经济管理

- 学报, 2021, 20(3): 316-325.
- [16] 王婵, 陈廷贵, 刘增金. 虚拟嵌入视角下农户社交电商参与行为及其影响研究——以陕西设施冬枣为例[J]. 农业现代化研究, 2022, 43(3): 431-443.
- [17] 胡雅淇, 林海. 在线评论特征对生鲜电商农产品销量的影响——来自淘宝羊肉大数据的证据[J]. 中国农业大学学报, 2021, 26(6): 206-218.
- [18] 王英, 杨丹, 王乐, 等. 基于品牌效应的在线评论对稻米销量影响的实证研究[J]. 中国农业大学学报, 2018, 23(1): 178-187.
- [19] 雷兵, 王巧霞, 刘小. 地方特色、网络口碑与农产品上行[J]. 中国软科学, 2021(2): 34-45.
- [20] 王翠翠, 夏春萍, 蔡轶. 农业电商扶贫可以提升农户的可持续生计吗?——基于农产品上行视角[J]. 浙江农业学报, 2022, 34(3): 636-651.
- [21] 李华杰, 史丹, 马丽梅. 基于大数据方法的经济研究: 前沿进展与研究综述[J]. 经济学家, 2018(6): 96-104.
- [22] BLEI D M, NG A Y, JORDAN M I. Latent dirichlet allocation [J]. the Journal of machine Learning research, 2003, 3: 993-1022.
- [23] 陈嘉钰, 李艳. 基于 LDA 主题模型的社交媒体倦怠研究——以微信为例[J]. 情报科学, 2019, 37(12): 78-86.
- [24] HAO H, ZHANG K, WANG W, et al. A tale of two countries: International comparison of online doctor reviews between China and the United States[J]. International journal of medical informatics, 2017, 99: 37-44.
- [25] 林海, 胡雅淇. 大数据剖析生鲜电商平台销量的影响因素——基于支持向量机方法的实证研究[J]. 贵州社会科学, 2021(3): 129-138.
- [26] 罗刚. 解密搜索引擎技术实战: Lucene&Java 精华版 [M]. 北京: 电子工业出版社, 2011: 144.
- [27] 张耘堂, 李东. 原产地形象对农产品电商品牌化的影响路径研究[J]. 中国软科学, 2016(5): 43-54.
- [28] 慕静, 东海芳, 刘莉. 电商驱动农产品品牌价值创造的机制——基于京东生鲜的扎根理论分析[J]. 中国流通经济, 2021, 35(1): 36-46.
- [29] 郑琼娥, 许安心, 范水生. 福建农产品区域品牌发展的对策研究[J]. 福建论坛(人文社会科学版), 2018(10): 197-202.
- [30] 涂传清. 农户介入农产品流通中高附加值活动的影响因素分析——基于赣南果农的实证研究[J]. 商业经济与管理, 2014(5): 12-23.
- [31] 杨晓娜, 李玉峰, 范丹丹. 基于消费者感知价值的生鲜农产品渠道选择策略[J]. 世界农业, 2021(11): 56-65+90.
- [32] 罗千峰, 张利庠. 农产品冷链物流高质量发展的理论阐释与实现路径[J]. 中国流通经济, 2021, 35(11): 3-11.
- [33] ZhENG Q, CHEN J, ZHANG R, et al. What factors affect Chinese consumers' online grocery shopping? Product attributes, e-vendor characteristics and consumer perceptions[J]. China Agricultural Economic Review, 2020, 12(2): 193-213.
- [34] 冷功业, 杨建利, 邢骄阳, 等. 我国农业高质量发展的机遇、问题及对策研究[J]. 中国农业资源与区划, 2021, 42(5): 1-11.
- [35] 王二朋, 高志峰, 耿献辉. 加工农产品的附加值来自哪里——消费偏好视角的分析[J]. 农业技术经济, 2020(1): 80-91.

责任编辑: 曾凡盛