

基于偏好特征提取的农产品电商信息个性化推送方法

何雨

武汉科技大学马克思主义学院, 湖北 武汉

收稿日期: 2024年12月17日; 录用日期: 2024年12月31日; 发布日期: 2025年2月18日

摘要

由于现行方法在农产品电商信息个性化推送中存在用户偏好理解不够深入、推送算法不精准等, 导致应用效果不佳。具体表现为推送内容与用户实际需求存在偏差, 用户满意度不高, 同时, 由于推送算法的性能限制, 导致推送准确性较差, 无法达到预期的推送效果。针对农产品电商信息个性化推送中现行方法存在的用户偏好理解不足、推送算法精准度低等问题, 本文提出了一种基于偏好特征提取的个性化推送方法。该方法通过深入分析电商平台上的用户反馈信息, 区分正负维度并提取正样本, 进一步从正样本中提取用户偏好特征。基于这些偏好特征, 我们对农产品进行推送评分, 并生成个性化推送名单, 在电商平台上实施推送。实验结果显示, 该方法在MRR和HR两个评价指标上均不低于0.8, 表现出较高的推送精度, 有效解决了现有推送方法中的不足, 实现了农产品电商信息的个性化精准推送。

关键词

偏好特征提取, 电商信息, 个性化, 推送, 维度区分, 正样本

Personalized Push Method for Agricultural Product E-Commerce Information Based on Preference Feature Extraction

Yu He

School of Marxism, Wuhan University of Science and Technology, Wuhan Hubei

Received: Dec. 17th, 2024; accepted: Dec. 31st, 2024; published: Feb. 18th, 2025

Abstract

Due to the lack of in-depth understanding of user preferences and inaccurate push algorithms in

the current method of personalized information push for agricultural product e-commerce, the application effect is not satisfactory. Specifically, there is a deviation between the pushed content and the actual needs of users, resulting in low user satisfaction. At the same time, due to the performance limitations of the push algorithm, the accuracy of the push is poor, and the expected push effect cannot be achieved. This paper proposes a personalized push method based on preference feature extraction to address the problems of insufficient understanding of user preferences and low accuracy of push algorithms in the current methods of personalized push of agricultural product e-commerce information. This method analyzes user feedback information on e-commerce platforms in depth, distinguishes positive and negative dimensions, and extracts positive samples to further extract user preference features from the positive samples. Based on these preference characteristics, we conduct push ratings on agricultural products and generate personalized push lists for implementation on e-commerce platforms. The experimental results show that this method achieves a high push accuracy of not less than 0.8 on both MRR and HR evaluation indicators, effectively solving the shortcomings of existing push methods and realizing personalized and accurate push of agricultural product e-commerce information.

Keywords

Preference Feature Extraction, E-Commerce Information, Individualization, Push, Dimensional Differentiation, Positive Sample

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

随着互联网技术的飞速发展，电子商务已成为推动各行各业转型升级的重要力量。特别是在数字乡村振兴战略的引领下，农产品电商凭借其独特的优势，为农业经济的创新发展注入了新的活力。农产品电商不仅有效克服了传统线下交易的局限，拓宽了农产品的销售渠道，还极大地提升了消费者的购物体验，使农产品购买变得更加便捷和高效。

农产品电商的兴起，不仅促进农业产业的数字化转型，还引领农业发展模式的深刻变革。然而，在农产品电商领域，信息的海量性和多样性也给用户带来选择上的困扰。如何精准地推送用户感兴趣的内容，成为农产品电商提升用户体验、增强用户黏性的关键所在[1]。

个性化推荐技术作为当前信息技术领域的热门研究方向，已经在新闻、影视、电子商务等领域取得了显著的应用成效。然而，农产品电商信息的个性化推荐却面临着诸多挑战。农产品种类繁多，质量差异大，结构复杂，用户的购物偏好也呈现出多样化的特点，使现有的推荐方法在农产品电商领域的应用效果并不理想。文献[2]中提出的基于近邻传播聚类的推荐方法虽然在一定程度上成功实现了用户和商品的分类，但其局限性在于当用户兴趣发生变化时，无法迅速调整推荐内容，从而影响了推荐的准确性和效果。相比之下，文献[3]介绍的基于云计算的推荐方法借助云计算的强大存储和计算能力，致力于实现内容的个性化推荐。然而，该方法显著依赖于大规模用户数据，导致计算负担较重，并且在确保推荐精度方面存在不足。

鉴于现有方法在农产品电商信息个性化推荐中存在的不足，本文提出基于偏好特征提取的农产品电商信息个性化推送方法，通过挖掘用户的购物偏好，实现更加精准、个性化的信息推送，从而提升农产品电商的用户体验和竞争力。

2. 用户反馈正负维度区分

在农产品电商平台上收集用户反馈信息，反馈信息存在以下三种情况：

第一，用户对农产品的操作频次大于 1 时，视为正样本。因为它代表用户对农产品进行了不止一次的购买、浏览操作，这反映出用户对该农产品具有较高的偏好度，并具有较高的可信度。

第二，用户对农产品的操作频次为 0 时，表示用户对农产品没有进行过任何操作。然而，这并不意味着这些农产品就一定为负样本[4]。由于缺乏足够的信息，只能做出一些合理的推断，而无法确定其确切的正负属性。

第三，用户对农产品的操作频次为 1 时，即用户对农产品仅进行过一次操作，一次点击行为并不能充分反映用户的偏好，它可能仅仅是一个偶然的为行为，也可能代表用户的真实兴趣。因此，无法准确判断其正负属性。

为了更全面地考虑用户行为的复杂性和多样性，并灵活地处理用户对农产品的操作频次为 0 的情况，引入参数 δ 以及一个额外的用户活跃度参数 α 来细化正负样本的区分。参数 α 可以基于用户在平台上的整体活跃程度(如登录频率、浏览时长、购买历史等)来设定，以反映用户对平台的整体兴趣和参与度[5]。根据收集的用户反馈信息样本正负维度区分，其用公式表示为：

$$K = \begin{cases} 1, & \text{if } p \geq 1 \\ \alpha \cdot \delta & \text{if } p = 0 \text{ and } \alpha > 0 \\ 0, & \text{if } p = 0 \text{ and } (\alpha = 0 \text{ or } \delta = 0) \end{cases} \quad (1)$$

式中， K 表示用户反馈信息样本正负维度区分结果；1 表示正样本； p 表示用户对农产品的操作频次；0 表示负样本； α 表示用户活跃度参数，取值范围是[0, 1]，其中 1 表示非常活跃，0 表示完全不活跃； δ 表示处理缺失值的参数，取值范围是[0, 1]。当 $\delta = 1$ 时，倾向于将缺失值视为负反馈；当 $\delta = 0$ 时，则不将缺失值视为明确的负反馈。通过以上公式区分用户反馈信息的正负维度。

3. 用户偏好特征提取

选取区分为正的样本(即用户与农产品的有效交互行为，如购买、收藏、浏览等)，用于提取用户偏好特征。当前阶段，用户的种种行为所透露出的兴趣倾向称为短期偏好。具体来说，如果在某个时间周期 T 内，用户购买、收藏、浏览某种农产品次数有所增加，那么可以认为用户对此类产品特征的偏好程度较高，或者至少对与这类特征相关的产品有较高的兴趣[6]。为了量化这种短期偏好，使用公式(1)来表示：

$$P_d = \phi \sum_{k=1}^N k + \beta w_T \quad (2)$$

式中， P_d 表示用户短期偏好值； ϕ 、 β 表示行为次数和直接评分对用户短期偏好的贡献权重，两者之和为 1。两个系数的选择基于实际数据的分析结果进行调整，以最优地反映用户行为对短期偏好的影响； k 是行为次数，表示在指定时间周期内，用户对某类农产品的各种交互行为的次数之和； w_T 表示时间周期 T 内用户成功交互农产品的直接评分[7]。 ϕ 、 β 的敏感性分析：通过调整这两个参数的值，观察用户短期偏好值 P_d 的变化情况。如果 P_d 对这两个参数的变化非常敏感，说明用户行为评分和时间周期内的直接评分对用户偏好特征的影响较大。短期偏好并不能代表用户的真实偏好，因此根据多个用户短期偏好，计算用户特征的长期偏好值：

$$P_c = \frac{1}{n} \sum_{n=1} P_{dn} \quad (3)$$

式中, P_c 表示用户对某类农产品的长期偏好特征值; n 是周期数量, 表示用户行为数据的时间跨度。选择依据是用户行为数据的可用性和研究目的; P_{dn} 第 n 个时间周期内的用户短期偏好值。 n 的敏感性分析: 通过改变时间周期的数量 n , 观察用户对某类农产品的长期偏好特征值 P_c 的变化情况。如果 P_c 对 n 的变化非常敏感, 说明时间跨度对用户长期偏好的影响较大。结合用户对某类农产品的短期偏好和长期偏好, 确定用户偏好特征值:

$$p = xP_d + yP_c \quad (4)$$

式中, p 表示用户对农产品偏好特征; x 、 y 表示短期偏好和长期偏好的权重系数。选择依据是用户行为数据的稳定性和研究目的。 x 、 y 的敏感性分析: 通过调整这两个参数的值, 观察用户对农产品偏好特征 p 的变化情况。如果 p 对这两个参数的变化非常敏感, 说明短期偏好和长期偏好对用户偏好特征的影响较大。

为了处理农产品的类别和品牌信息, 采用 one-hot 编码方法。对于同类别农产品的价格信息, 采用十分位数原理进行区间划分, 以便更好地反映用户对价格的感知差异, 采用三元组表征用户对农产品特征的偏好程度, 其用公式表示为:

$$\begin{cases} C = \langle P_{Ti,G}, P_{Ti,P}, P_{Ti,S} \rangle \\ P_{Ti,G} = x \left(\frac{n+1-T}{n} p \right) \\ P_{Ti,P} = y_1 \left(\frac{n+1+T}{n} P_{c,J} \right) \\ P_{Ti,S} = y_2 \left(\frac{n+1+T}{n} P_{c,S} \right) \end{cases} \quad (5)$$

式中, C 表示用户对农产品特征的偏好程度; $P_{Ti,G}$ 表示用户在周期 T 内对电商平台上农产品 i 的偏好; $P_{Ti,P}$ 表示用户对电商平台上农产品品牌的偏好; $P_{Ti,S}$ 表示用户对电商平台上农产品价格的偏好; $P_{c,J}$ 、 $P_{c,S}$ 分别表示用户对某品牌、某价格区间农产品操作频率; y_1 、 y_2 表示长期偏好特征系数, 用户对品牌和价格的长期偏好程度; n 表示时间周期的划分数量, 用于计算用户对农产品特征的偏好程度。 y_1 、 y_2 的敏感性分析: 通过调整两个参数的值, 观察用户对农产品特征的偏好程度 C 的变化情况。如果 C 对参数的变化非常敏感, 说明用户对品牌和价格的长期偏好对用户偏好特征的影响较大。采用熵权法计算用户长期偏好特征系数, 熵权法是基于信息熵的概念确定各个相应指标的系数, 信息熵的值越大, 就意味着信源中的不确定性越高, 反之则越低。基于信息熵的这一特性, 将用户长期的感知价格偏好和品牌偏好系数进行量化表示:

$$\begin{cases} y_1 = (1-x) \left(\frac{X_{c,J}}{X_{c,J} + X_{c,S}} \right) \\ y_2 = (1-x) \left(\frac{X_{c,S}}{X_{c,S} + X_{c,J}} \right) \end{cases} \quad (6)$$

式中, $X_{c,J}$ 、 $X_{c,S}$ 分别表示用户长期感知价格偏好、品牌偏好信息熵。将公式(6)计算的偏好系数代入到公式(5)中, 提取到用户对电商平台农产品的偏好特征。

4. 农产品电商信息个性化推送

基于以上提取的用户对电商平台上农产品的偏好特征, 对电商平台上农产品进行评分, 其用公式表示为:

$$F_i = \frac{\int_{-\infty}^{+\infty} (P_{Ti,G} + P_{Ti,P} + P_{Ti,S})}{C} \quad (7)$$

式中, F_i 表示电商平台上农产品 i 的用户推荐评分。评分越高, 则表示用户对农产品的感兴趣程度越高, 推送价值越高[8]。因此根据实际情况设定一个阈值, 将推荐评分 F_i 大于阈值的农产品信息列入到推荐列表, 以此生成农产品电商个性化推荐名单, 将列表上农产品信息其发送到电商客户端上进行可视化展示, 以此实现基于偏好特征提取的农产品电商信息个性化推送。

5. 实验论证

5.1. 实验设置及数据集

为了验证本文提出的基于偏好特征提取的农产品电商信息个性化推送方法的可行性与可靠性, 以下将在 GYSADF、FAHHKS 两个数据集上对本文方法进行对比实验, 对照组设置为文献[2]提出的基于近邻传播聚类的推荐方法、文献[3]提出的基于云计算的推荐方法, 以下用对照组 1 和对照组 2 表述。

GYSADF、FAHHKS 数据集均来自某农产品电商平台, 具体见表 1 所示。

Table 1. Experimental dataset

表 1. 实验数据集

数据集	用户数	产品数	评分记录
GYSADF	2000	2451	32,546
FAHHKS	2500	4162	56,841

实验分别利用两个数据集检验三种方法的两个指标。实验环境设计为: 采用 windowsUO 操作系统, 搭配 Inter Core i8 处理器, 128 G 固态硬盘和 32 GB 内存。采用 1.11 版本 Python 编辑推送程序。

5.2. 实验流程及评价指标

基于以上实验环境开展对比实验, 将 GYSADF、FAHHKS 两个数据集中数据代入到公式(1)中进行正负维度区分; 利用公式(2)~(6)提取用户农产品偏好特征; 通过公式(7)计算用户推荐评分, 生成推荐名单进行个性化推送。

本次实验选择 MRR (平均倒数排名)和 HR (命中率)作为方法推荐准确性评价指标, MRR 反映的是推荐的信息是否摆在用户更明显的位置, 强调位置关系、顺序性。在理想情况下, 即推荐的项目总是位于列表的第一位时, MRR 为 1; 在最坏情况下, 即推荐的项目不在列表中时, MRR 趋近于 0。MRR 的值越高, 表示推荐方法越能将用户感兴趣的农产品电商信息排在推荐列表的前面, 从而反映出用户体验和满意度;

HR 反映的是推荐序列中是否包含了用户真正点击的农产品电商信息, 强调推荐的“准确性”。通过计算在推荐列表的前 N 个项目中包含用户真正感兴趣产品的比例来得到。由于它是一个比例数值, 因此是介于 0 和 1 之间的一个数。HR 的值越高, 表示推荐方法越能准确推荐用户感兴趣的农产品电商信息。

利用 GYSADF 数据集检验三种方法 MRR, 利用 FAHHKS 数据集检验三种方法 HR。

5.3. 实验结果与讨论

以推送次数为变量, 图 1 给出了三种方法在不同推送次数下 MRR; 以农产品数量为变量, 图 2 给出了三种方法在不同农产品数量下 HR。

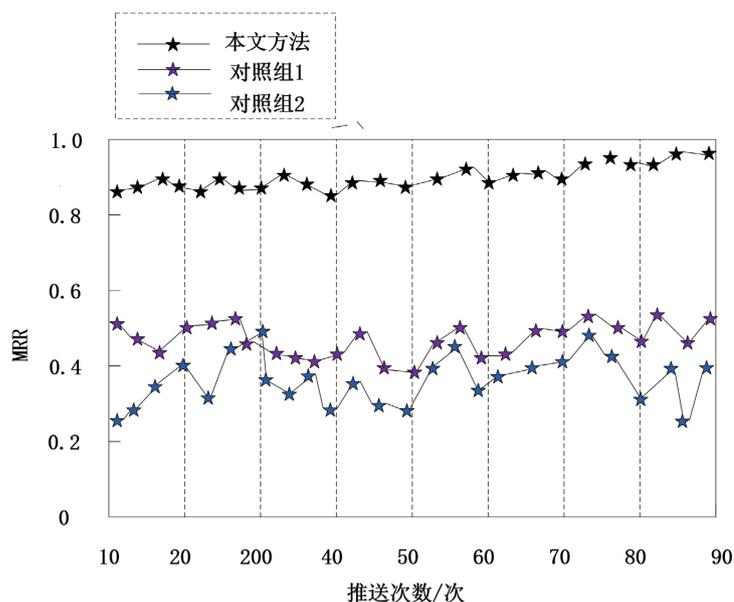


Figure 1. MRR comparison of three methods

图 1. 三种方法的 MRR 对比

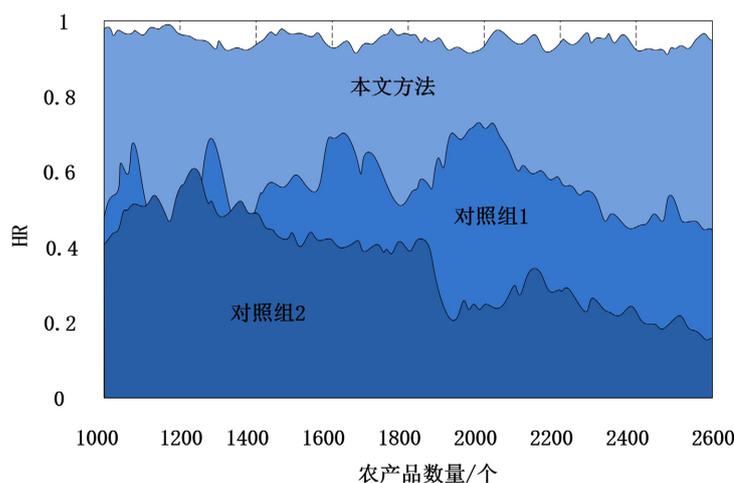


Figure 2. HR comparison of three methods

图 2. 三种方法的 HR 对比

从图 1 可以看出,在农产品电商信息个性化推送场景中,本文方法 MRR 在 0.8 以上,远高于其余两种方法,推送的农产品电商信息位置比较靠前;从图 2 可以看出,本文方法 HR 非常接近 1,HR 值高于对照组 1 和对照组 2,推荐的农产品电商信息命中率较高。这是因为本文采用了偏好特征提取技术,能够从用户和农产品的海量数据中提取出具有代表性的特征,这些特征能够更准确地反映用户的兴趣和农产品的属性。通过对特征进行深入分析和处理,能够更好地理解用户的需求和偏好,从而为用户提供更加个性化的推荐结果,进而保证了农产品电商信息个性化推送精度。因此通过以上对比证明,本文方法更适用于农产品电商信息个性化推送,具有良好的可行性与可靠性。

6. 结束语

本研究基于偏好特征提取的农产品电商信息个性化推送方法,通过深度挖掘用户的历史行为、消费

习惯及偏好特征，实现了对用户需求的精准捕捉和个性化推送。研究结果显示，该方法能够显著提高用户满意度，为农产品电商平台的运营效率和用户体验带来了显著提升。然而，本研究仍存在一定的局限性。随着用户需求的不断变化和市场环境的快速发展，如何持续优化推荐算法，提高推荐的准确性和实时性，是未来研究的重要方向。此外，农产品电商平台的多样性和复杂性也对推荐系统的通用性和适应性提出了更高的要求。未来将进一步探索推荐方法，以及融合多种数据源和推荐策略的混合推荐系统，以应对不断变化的用户需求和市场环境。同时，加强用户隐私保护和数据安全也是未来研究中不可忽视的重要议题。

参考文献

- [1] 谢梦怡. 基于近邻传播聚类的多源异构数据信息个性化推送方法[J]. 信息技术与信息化, 2024(7): 165-169.
- [2] 张博君. 基于近邻传播聚类的电商商品信息个性化推送研究[J]. 中国信息界, 2024(2): 246-248.
- [3] 王南. 基于云计算的短视频媒体资源个性化推送方法[J]. 兵工自动化, 2024, 43(2): 16-22.
- [4] 张迅. 基于大数据平台的电力营销信息个性化推送方法[J]. 信息与电脑(理论版), 2023, 35(24): 97-99.
- [5] 王珍, 许继平, 王小艺. 数字信息技术在个性化服装智能推送中的应用[J]. 丝网印刷, 2023(22): 93-95.
- [6] 张志. 基于用户画像对互联网用户个性化推荐与引导[J]. 电脑编程技巧与维护, 2022(12): 155-158.
- [7] 王金威. 基于大数据分析的高校云招聘信息个性化推送研究[J]. 安徽电子信息职业技术学院学报, 2022, 21(4): 25-31.
- [8] 张莉. 基于用户画像的常德地区休闲农业信息个性化推荐系统研究[J]. 软件, 2022, 43(2): 1-3.