E-Commerce Letters 电子商务评论, 2025, 14(1), 1579-1586
Published Online January 2025 in Hans. https://www.hanspub.org/journal/ecl
https://doi.org/10.12677/ecl.2025.141195

Hans汉斯

基于协同过滤的电商平台蔬菜产品信息智能推荐方法

丁云鹤，柴　洪*

甘肃农业大学财经学院，甘肃　兰州

收稿日期：2024年11月5日；录用日期：2024年11月19日；发布日期：2025年1月15日

摘　要

由于电商平台蔬菜产品信息繁杂且用户偏好多变，导致推荐匹配度较低。为了解决上述问题，提出基于协同过滤的电商平台蔬菜产品信息智能推荐方法。通过分析电商平台用户的购买历史和浏览行为，利用slope one方法和熵权法构建用户兴趣模型，深入理解了用户的蔬菜偏好和需求。采用协同过滤算法，结合用户对产品的评分数据和行为数据，计算蔬菜产品之间的相似度，得到更精准的产品关联信息。基于Spark平台，实现电商并行化协同过滤推荐技术，高效地生成个性化的蔬菜产品信息智能推荐。实验结果表明，研究方法能够在多种实验场景下显著提升推荐匹配度，具有实际应用价值。

关键词

协同过滤算法，电商平台，蔬菜产品信息，智能推荐，用户偏好预测

Intelligent Recommendation Method of Vegetable Product Information on E-Commerce Platform Based on Collaborative Filtering

Yunhe Ding, Hong Chai*

College of Finance and Economics, Gansu Agricultural University, Lanzhou Gansu

Received: Nov. 5th, 2024; accepted: Nov. 19th, 2024; published: Jan. 15th, 2025

Abstract

Because the vegetable product information of e-commerce platform is complicated and the user's

*通讯作者。

文章引用: 丁云鹤, 柴洪. 基于协同过滤的电商平台蔬菜产品信息智能推荐方法[J]. 电子商务评论, 2025, 14(1): 1579-1586. DOI: 10.12677/ecl.2025.141195

preference is changeable, the recommendation matching degree is low. In order to solve the above problems, an intelligent recommendation method of vegetable product information on e-commerce platform based on collaborative filtering was proposed. By analyzing the purchase history and browsing behavior of e-commerce platform users, slope one method and entropy weight method are used to build user interest model, and the user's vegetable preference and demand are deeply understood. The collaborative filtering algorithm was used to calculate the similarity between vegetable products by combining the user's score data and behavior data to obtain more accurate product association information. Based on the Spark platform, it realizes the parallel collaborative filtering recommendation technology of e-commerce, and efficiently generates personalized intelligent recommendation of vegetable product information. The experimental results show that the proposed method can significantly improve the recommendation matching degree under various experimental scenarios, and has practical application value.

Keywords

Collaborative Filtering Algorithm, E-Commerce Platform, Vegetable Product Information, Intelligent Recommendation, User Preference Prediction

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

在电子商务领域，推荐系统已成为提升用户体验和销售业绩的重要手段。农产品电商作为近年来快速发展的细分市场，面临着用户信息过载和商品种类繁多的挑战，迫切需要有效的推荐算法来帮助用户快速找到符合其需求的商品。

基于细粒度特征交互选择网络的农产品推荐算法，通过引入细粒度交互层和特征交互选择层，利用内积和哈达玛积有效地学习特征交互[1]。在训练过程中，该模型能够自动识别重要的特征交互，并删除冗余的交互，最终将重要的特征交互和一阶特征输入到深度神经网络中，得到最终的点击通过率(CTR)预测值，从而实现农产品的精准推荐。另一种融合表征的农产品推荐算法，则通过融合用户行为数据和农产品特征，构建用户－农产品关联矩阵，并运用深度神经网络对矩阵进行分解和优化，从而为用户提供精准的个性化推荐服务[2]。

尽管上述方法在农产品推荐领域取得了一定的成果，但对于新用户或新产品，由于历史数据的缺乏，推荐算法的有效性仍受到限制。针对这一不足，本文提出了一种基于协同过滤的电商平台蔬菜产品信息智能推荐方法，旨在提升推荐的精确性和个性化程度，从而更好地满足用户需求。

2. 分析电商平台用户偏好

由于电商平台积累了大量用户的购买历史和浏览行为，这些数据能够反映出用户的蔬菜偏好和需求[3]。为了更准确地理解用户的喜好，本文首先对这些数据进行深入分析。设电商平台拥有注册用户集合 $U=\{u_1,u_2,\cdots,u_n\}$ 和蔬菜产品信息集合 $I=\{I_1,I_2,\cdots,I_m\}$，通过构建评分矩阵 $V\in R^{n\times m}$，其中元素 v_{ij} 代表用户 u_i 对产品 I_j 的评分，评分可以是布尔型(如 1 表示喜欢，0 表示不喜欢)，若用户未对产品进行评分，则相应评分为 0。为了更精确地确定用户偏好度，本文引入 slope one 方法。该方法通过分析不同产品在用户中的评分数值之间的线性关系，来推断用户对产品的偏好度差异。设 D_{jk} 表示同时评价过产品 I_j 和 I_k 的用户集合的评分差均值，即：

$$D_{jk}=\frac{\sum_{u_i\in D_{jk}}\left(v_{ij}-\overline{v}_j\right)}{\left|D_{jk}\right|}-\frac{\sum_{u_i\in D_{jk}}\left(v_{ik}-\overline{v}_k\right)}{\left|D_{jk}\right|} \tag{1}$$

公式(1)中，v_{ik} 表示用户 u_i 对产品 I_k 的评分，$\overline{v}_j$ 和 $\overline{v}_k$ 分别表示产品 I_j 和 I_k 的平均评分，$\left|D_{jk}\right|$ 表示集合 D_{jk} 的大小。

对于用户 u_i 对未评分产品 I_t 的预测评分 $\hat{v}_{it}$，slope one 方法通过以下公式(2)进行预测：

$$\hat{v}_{it}=\overline{v}_t+\frac{\sum_{I_j\in R_i\cap N_t}D_{jt}\cdot W_{ij}}{\sum_{I_j\in R_i\cap N_t}W_{ij}} \tag{2}$$

式(2)中，R_i 表示用户 u_i 评价过的产品集合，N_t 表示与产品 I_t 有共同评分用户的产品集合，W_{ij} 表示用户 u_i 对产品 I_j 评分的权重。

同时，本文还深入剖析用户行为数据，发现其呈现出长尾分布特征。为了捕捉这种用户行为模式，本文采集了用户在阅读、收藏、购买以及加入购物车等过程中的行为数据，作为反映用户兴趣的关键指标[4]。设用户行为数据矩阵为 $B\in R^{n\times p}$，其中元素 b_{iq} 表示用户 u_i 在行为 q 上的强度。为了构建精确的用户兴趣模型，本文通过去除重复信息，利用熵权法科学分配不同兴趣特征的权重。设用户兴趣特征向量 $f_i=\left[f_{i1},f_{i2},\cdots,f_{ip}\right]$，其中，$f_{iq}$ 表示用户 u_i 在兴趣特征 q 上的得分。则兴趣特征 q 的熵 E_q 可以通过公式(3)表示为：

$$E_q=-\sum_{i=1}^{n}\frac{f_{iq}}{\sum_{i=1}^{n}f_{iq}}\log\frac{f_{iq}}{\sum_{i=1}^{n}f_{iq}} \tag{3}$$

基于熵值计算结果，兴趣特征 q 的权重 w_q 如公式(4)：

$$w_q=\frac{1-E_q}{\sum_{q=1}^{p}\left(1-E_q\right)} \tag{4}$$

最终，用户 u_i 的兴趣模型可以表示为加权后的兴趣特征向量 $F_i=\left[w_1\cdot f_{i1},w_2\cdot f_{i2},\cdots,w_p\cdot f_{ip}\right]$，从而为后续的蔬菜产品信息智能推荐提供了有力支持。

3. 基于协同过滤算法计算蔬菜产品相似度

在对电商平台用户偏好的深入分析基础上，本文进一步利用协同过滤算法来计算蔬菜产品之间的相似度，以便为智能推荐系统提供更精准的产品关联信息。产品相似度的计算主要依赖于用户对产品的评分数据以及用户行为数据，本文利用用户对产品的评分数据来计算产品之间的皮尔逊相关系数和余弦相似度。设蔬菜产品 I_j 和 I_k 的评分向量分别为 v_j 和 v_k，同时评价过这两个产品的用户集合为 U_{jk}。则产品 I_j 和 I_k 的皮尔逊相关系数 sim_{jk} 可以通过公式(5)进行表示：

$$sim_{jk}=\frac{\sum_{u_i\in U_{jk}}\left(v_{ij}-\overline{v}_j\right)\left(v_{ik}-\overline{v}_k\right)}{\sqrt{\sum_{u_i\in U_{jk}}\left(v_{ij}-\overline{v}_j\right)^2\sum_{u_i\in U_{jk}}\left(v_{ik}-\overline{v}_k\right)^2}} \tag{5}$$

除了皮尔逊相关系数外，本文结合余弦相似度作为评估产品间相似性的另一指标[5]。蔬菜产品 I_j 和 I_k 的余弦相似度 $\cos sim_{jk}$ 如公式(6)所示：

$$\cos sim_{jk} = \frac{\sum_{u_i \in U_{jk}} v_{ij} v_{ik}}{\sqrt{\sum_{u_i \in U_{jk}} v_{ij}^2} \sqrt{\sum_{u_i \in U_{jk}} v_{ik}^2}} \tag{6}$$

为了进一步计算产品间的相似度，本文利用用户行为数据来计算蔬菜产品之间的相似度。对于每一对蔬菜产品 I_j 和 I_k，本文计算同时与这两个产品有过行为的用户集合 U_{jk}，然后应用余弦相似度度量方法来量化它们之间的相似性，这一计算过程遵循公式(7)：

$$\cos sim_{jk}^{bh} = \frac{\sum_{u_i \in U_{jk}} b_{ij} b_{ik}}{\sqrt{\sum_{u_i \in U_{jk}} b_{ij}^2} \sqrt{\sum_{u_i \in U_{jk}} b_{ik}^2}} \tag{7}$$

式中，$\cos sim_{jk}^{bh}$ 表示基于用户行为的余弦相似度，它反映了同时与产品 I_j 和 I_k 有过行为的用户集合中，用户行为强度的相似性。

最后，为了得到更精确的产品相似度，本文将基于评分的相似度和基于用户行为的相似度进行融合，得到综合相似度。设产品 I_j 和 I_k 的综合相似度为 $\cos sim_{jk}^{cb}$，则它可以通过公式(8)进行计算：

$$\cos sim_{jk}^{cb} = a \cdot sim_{jk} + (1-a) \cdot \cos sim_{jk}^{bh} \tag{8}$$

其中，$\cos sim_{jk}^{cb}$ 表示蔬菜产品 I_j 和 I_k 的综合相似度，a 表示权重系数，用于调节基于评分的相似度和基于用户行为的相似度在综合相似度中的贡献程度[6]。通过调整 a 的值，可以得到更符合实际需求的综合相似度结果。

4. 生成蔬菜产品信息智能推荐

为了优化用户的购物体验，本文在深入分析电商平台用户偏好和产品相似度的基础上，采用基于 Spark 平台的电子商务并行化协同过滤推荐技术，以高效地生成蔬菜产品信息智能推荐。在用户并行化推荐阶段，本文首先依据用户的历史行为数据，筛选出可能符合其兴趣的蔬菜产品信息集合。并利用协同过滤算法，识别出与目标用户兴趣高度匹配的 k 个用户。针对这组用户，本文计算他们对目标蔬菜产品的喜好评分，该评分是通过将这些用户已评分的蔬菜产品与目标产品的相似度进行加权平均得出的。这一步骤中，用户喜好评分的计算公式为：

$$tR(u,s) = \frac{\sum_{v \in U_{u,k}} \cos sim(u,v) \times R(v,s)}{\sum_{v \in U_{u,k}} \cos sim(u,v)} \tag{9}$$

公式(9)中，$\cos sim(u,v)$ 表示用户 u 和用户 v 之间的相似度，$R(v,s)$ 表示用户 v 对与 $s_{\max}$ 相似的产品 s' 的喜好评分，该评分是通过计算用户 v 已评分的蔬菜产品与 s' 的相似度加权平均值得出的[7]。

在项目并行化推荐阶段，本文进一步寻找与用户喜好评分值最高的蔬菜产品 $s_{\max}$ 相似的 e 个产品集合 $I_{s_{\max,e}}$。产品 s' 的喜好评分可以通过公式(10)进行表示：

$$tR(u,s') = \frac{\sum_{s \in S} \cos sim(s,s') \times R(u,s)}{\sum_{s \in S} \cos sim(s,s')} \tag{10}$$

为了生成最终的推荐列表，本文还考虑推荐列表的实时更新以及个性化推荐的实现。在个性化推荐的实现中，本文为目标用户 u 未评分的蔬菜产品集合 I_{ur} 中的每个产品生成预测评分。未评分蔬菜产品 i_{ur}

的评分预测可以通过公式(11)进行表示：

$$tPR\left(u,i_{ur}\right)=\bar{R}_u+\frac{\sum_{v\in U_{u,k}}\cos sim\left(u,v\right)\times\left(R\left(v,i_{ur}\right)-\bar{R}_v\right)}{\sum_{v\in U_{u,k}}\cos sim\left(u,v\right)} \tag{11}$$

公式中，$\bar{R}_u$ 和 $\bar{R}_v$ 分别表示用户 u 和用户 v 对所有已评分蔬菜产品的平均评分。这一预测评分综合考虑了近邻用户的相似度、各自对所有已评分蔬菜产品的平均评分以及近邻用户对特定产品的具体评分[8]。最终，本文将预测评分不为 0 的产品纳入推荐列表 L_e 中，从而生成更加精准和个性化的蔬菜产品信息智能推荐。推荐过程如图 1 所示。

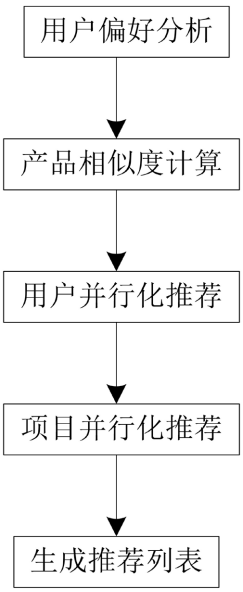

Figure 1. Collaborative filtering recommendation process
图 1. 协同过滤推荐过程

这一过程不仅能够提升推荐的精准度，还能极大地丰富用户的购物体验。由此，完成了本文基于协同过滤的电商平台蔬菜产品信息智能推荐方法的研究。

5. 实验

5.1. 实验准备

在实验正式开始之前，本文首先搭建一个稳定且高效的实验环境，以支持从数据采集、存储、处理到推荐算法实现的全过程。表 1 是实验环境的详细配置：

Table 1. Experimental environment configuration table
表 1. 实验环境配置表

序号	组件	具体参数
(1)	服务器	CPU：Intel Xeon E5-2680 v4 (28 核 56 线程) 内存：256GB DDR4 ECC 存储设备：2TB SSD 硬盘

续表

(2)	存储设备	容量：2 TB 接口类型：SATA/SAS 读写速度：≥500 MB/s
(3)	操作系统	Ubuntu 20.04
(4)	大数据处理平台	Apache Spark 3.2
(5)	数据库	MySQL 8.0
(6)	数据处理库	NumPy, Pandas, SciPy
(7)	网络带宽	≥1 Gbps
(8)	网络安全	配置防火墙

通过以上配置，能够确保实验环境的高性能、稳定性和安全性，从而满足电商平台蔬菜产品信息智能推荐实验的需求。接下来，实验将遵循既定步骤，开展数据的全面收集、预处理、算法的具体实现以及结果的精确评估等一系列工作，以期验证智能推荐算法在电商平台蔬菜产品信息推荐中的实际应用效果。

5.2. 实验场景设计

为了验证推荐算法在不同情境下的效果，本文精心设计了多个实验场景，具体如表 2 所示。

Table 2. Design the experiment scene
表 2. 设计实验场景

场景编号	实验场景名称	目标
T1	新用户冷启动场景	验证推荐算法在新用户无历史行为数据情况下的推荐效果
T2	个性化推荐场景	验证推荐算法根据用户历史行为数据提供的个性化推荐效果
T3	时令蔬菜推荐场景	验证推荐算法根据季节和天气因素提供的时令蔬菜推荐效果
T4	促销活动推荐场景	验证推荐算法在促销活动期间提供的商品推荐效果
T5	跨品类推荐场景	验证推荐算法在蔬菜产品与其他品类之间的跨品类推荐效果

通过表 2 设计的场景，旨在全面覆盖电商平台蔬菜产品信息推荐中的各类实际情况。对于新用户冷启动场景，当新用户刚注册电商平台且尚未产生任何购买或浏览记录时，平台可以基于新用户注册时填写的信息以及已有物品的属性信息进行推荐，同时采用基于内容的推荐算法，根据蔬菜产品的属性特征为用户推荐相似的产品，以帮助新用户快速了解平台上的蔬菜产品，提高用户的购物体验和满意度。对于已有购买或浏览记录的用户，平台则可以根据其历史行为数据为其提供个性化的推荐，通过分析用户之间的相似性或物品之间的关联性，为用户推荐与其兴趣相似的蔬菜产品，以提高推荐的准确性和匹配度，满足用户的个性化需求，增加用户的购物频次和忠诚度。此外，随着季节的变化，蔬菜的种类和价格也会发生变化，平台需要根据时令蔬菜的情况为用户提供合适的推荐，引入时序加权和热门物品惩罚等方法，以更好地反映用户当前的兴趣和物品的当前热度，同时结合蔬菜产品的季节性特征进行推荐，帮助用户及时了解到当前季节的热门蔬菜产品，提高用户的购物体验和满意度，同时也有助于促进时令蔬菜的销售。电商平台经常会举办各种促销活动，如打折、满减、赠品等，平台需要根据用户的购物习惯和偏好，为其推荐合适的促销活动，通过分析用户的历史购买记录和评价记录等数据，挖掘用户的购

物习惯和偏好，然后结合当前的促销活动进行推荐，同时采用基于内容的推荐算法，根据促销活动的特点为用户推荐相似的产品，以提高促销活动的参与度和转化率，增加平台的销售额和利润。在接下来的实验中，将分别在这些场景下对推荐算法进行验证和评估。

5.3. 实验结果及分析

为了验证本文方法的优越性，按照上述设计的多个实验场景进行实验，并将本文方法与两种对比方法进行比较。对比方法 1 是基于细粒度特征交互选择网络的农产品推荐算法，对比方法 2 是一种融合表征的农产品推荐算法。以下是各实验场景下不同推荐方法的推荐匹配度对比结果：

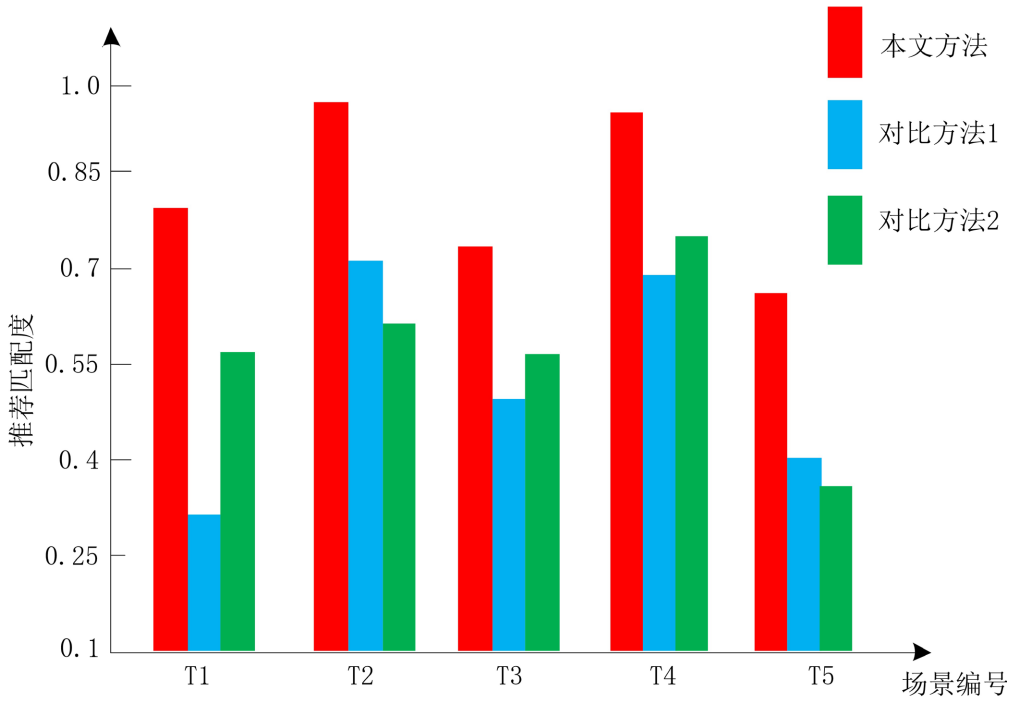

Figure 2. Recommendation matching degree of different recommendation methods in different experimental scenarios
图 2. 各实验场景下不同推荐方法的推荐匹配度

根据图 2 中的实验结果可以看出，本文方法在各个实验场景下均表现出色。特别是在新用户冷启动和个性化推荐场景中，其推荐匹配度远超对比方法 1 和对比方法 2，在数值上可达到 0.79，凸显了本文方法在处理新用户无历史行为数据及基于用户历史行为的个性化推荐上的优势。在时令蔬菜推荐场景中，本文方法的推荐匹配度同样领先，在数值上可达到 0.73，说明本文方法能够准确捕捉季节和天气因素，提供贴合时令的蔬菜推荐。在促销活动推荐场景中，其推荐匹配度高达 0.93，充分展现了对促销活动信息的有效利用，为用户提供更精准的商品推荐。尤为突出的是，在跨品类推荐场景中，本文方法依然保持高推荐匹配度，远超对比方法，进一步验证了其在处理复杂推荐场景时的稳定性和可靠性。

6. 结束语

本文研究了一种基于协同过滤的电商平台蔬菜产品信息智能推荐方法。该方法深入分析了用户历史行为数据和蔬菜产品特征，构建了精准的用户－产品关联模型，并运用协同过滤技术为用户提供个性化的蔬菜产品推荐。实验结果表明，该方法能够显著提升推荐的准确性和用户满意度，有效缓解了农产品

电商中的信息过载问题。特别是在促销活动推荐场景中，其推荐匹配度高达 0.93，充分证明了该方法对促销活动信息的有效整合与利用，进而为用户提供了更为精确的商品推荐。展望未来，将继续探索结合深度学习等先进技术，进一步提升推荐算法对新用户和新产品的推荐效果，并深入开展用户行为分析，深入挖掘用户的潜在需求，以期为用户提供更加精准和个性化的推荐服务。

基金项目

甘肃省人文社会科学项目：数字技术赋能甘肃农业现代化的实践路径研究(24ZZ13)。

参考文献

[1] 白雪, 王霞光, 金继鑫, 等. 基于细粒度特征交互选择网络的农产品推荐算法[J]. 计算机系统应用, 2024, 33(5): 271-279.

[2] 黄英来, 冀宇超, 刘镇波. 一种融合表征的农产品推荐算法[J]. 哈尔滨理工大学学报, 2024, 29(3): 20-27.

[3] 倪伟竣, 纪淑娟, 梁永全. 利用图神经网络的互补产品推荐[J]. 计算机工程与应用, 2024, 60(10): 292-300.

[4] 耿杰, 刘春丽, 魏雪梅, 等. 基于用户重购行为的产品推荐方法[J]. 计算机研究与发展, 2023, 60(8): 1795-1807.

[5] 高云梅, 张淑慧. 基于信息检索与 K 均值聚类的化工产品精准推荐算法研究[J]. 粘接, 2023, 50(3): 132-135.

[6] 罗远远. 协同过滤算法在农产品智能推荐系统中的应用研究[J]. 南方农机, 2023, 54(3): 33-35.

[7] 高云. 基于协同过滤算法的电商平台产品信息智能推荐方法[J]. 互联网周刊, 2022(13): 54-56.

[8] 王梦茹, 纪淑娟, 梁永全, 等. 融合注意力机制的互补产品推荐方法[J]. 计算机应用研究, 2022, 39(11): 3281-3286.