

# A Collaborative Filtering Recommendation Algorithm Basing on Preference Navigation for Agricultural Science and Technology Information

Wenchao Wang<sup>1</sup>, Jinling Song<sup>1</sup>, Liming Huang<sup>2</sup>, Aiyong Liu<sup>1</sup>

<sup>1</sup>College of Mathematics and Information Technology, Hebei Normal University of Science & Technology, Qinhuangdao Hebei

<sup>2</sup>College of Business Administration, Hebei Normal University of Science & Technology, Qinhuangdao Hebei  
Email: [songjinling99@126.com](mailto:songjinling99@126.com)

Received: Jun. 2<sup>nd</sup>, 2015; accepted: Jun. 23<sup>rd</sup>, 2015; published: Jun. 26<sup>th</sup>, 2015

Copyright © 2015 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

Although the recommendation technology has been used in agricultural science and technology information service website, it is still obvious that the recommendation result lacks of personalization. In order to solve this problem, a collaborative filtering recommendation method basing on preference navigation for agricultural science and technology information is proposed. For the website with single user type, the candidate preference information are selected for the target user firstly according to the browsing quantity, then the recommendation information is determined according to the ratings of the other users. For the website with various types of users, user preference navigation information is chosen according to the browsing history and rating of the target user firstly, the navigation users are determined basing on user preference navigation information secondly, then similar users are determined through the ratings of navigation users, finally predict ratings of other items and recommend according to the similar users.

## Keywords

User Preference, Collaborative Filtering, Single User Type, Multiple User Type, Similarity

---

# 基于兴趣导航的农业科技信息协同过滤推荐方法

王文超<sup>1</sup>, 宋金玲<sup>1</sup>, 黄立明<sup>2</sup>, 刘爱勇<sup>1</sup>

<sup>1</sup>河北科技师范学院数信学院, 河北 秦皇岛

<sup>2</sup>河北科技师范学院工商管理学院, 河北 秦皇岛

Email: [songjinling99@126.com](mailto:songjinling99@126.com)

收稿日期: 2015年6月2日; 录用日期: 2015年6月23日; 发布日期: 2015年6月26日

## 摘要

由于推荐技术在农业科技信息服务领域的应用处于起步阶段, 而传统推荐技术的针对性又存在不足, 针对这些问题, 提出了基于兴趣导航的协同过滤农业科技信息推荐技术。对于用户类型比较单一的网站, 首先根据信息的浏览量选择目标用户的候选兴趣信息, 然后根据其他用户对候选信息的评分确定最终的推荐信息。对于用户类型多样化的网站, 首先根据目标用户的历史浏览信息及评分确定用户兴趣导航信息, 然后根据导航信息确定导航用户集, 再通过导航用户集的项目评分确定相似用户集, 最后根据相似用户对其他信息进行评分估计并进行推荐。

## 关键词

用户兴趣, 协同过滤, 单用户类型, 多用户类型, 相似度

## 1. 引言

随着农业信息化的快速发展, 农业科技信息的发展呈现出庞大且复杂的趋势, 据统计当前中国农业类相关网站数量已经超过了4万家(其中政府主持的也超过了4000家), 3000多个农业期刊, 涉农类综合报纸或专业报纸也到达数百种, 广播电视也都有农业类节目或栏目[1]。但是这种数量扩大的优势并没有给农民带来多少真正的实惠, 大多数涉农企业、偏僻的农村、渴望致富的9亿农民对信息的需求有近90%得不到满足[1]。究其原因在于面对如此庞大的信息量, 农民无法从中快速、准确地找到他们所需要的农业科技信息。

推荐技术作为解决信息过载的主要技术之一在电商等领域已经得到了越来越多的应用[2]-[7]。协同过滤推荐技术是最为传统的推荐技术, 也是目前最为成功有效的推荐技术[2]。其基本思想是基于用户——项目评分数据集, 通过收集相似用户的兴趣信息进而为目标用户进行推荐。因此, 将推荐技术应用到农业科技信息服务网站, 是解决农民的信息需求、提高农业信息服务质量的有效途径。

在农业科技信息服务过程中, 由于用户所经营的农业生产分类不同, 用户之间感兴趣的信息也会不同, 传统的协同过滤推荐技术单纯根据用户或项目之间的相似性对目标用户进行推荐, 却忽略了用户本身的兴趣, 使得推荐结果针对性不强、推荐精度较低, 而推荐技术在农业科技信息服务领域的应用研究[8]还处于起步阶段, 在这种情况下, 需要对传统的协同过滤推荐技术进行改进, 提出更加适用于农业科技信息服务的推荐技术。

为了给目标用户提供针对性的、感兴趣的农业科技信息,根据农业信息服务网站的用户种类对网站进行分类,针对不同类型的网站,以用户兴趣为导航,并结合传统的协同过滤推荐方法,分别提出了基于兴趣导航的农业科技信息协同过滤推荐方法。

## 2. 基于兴趣导航的农业科技信息协同过滤推荐方法

有的农业信息服务网站(比如中国黄骅冬枣网)虽然信息量很大,但是它的用户类型比较单一(均为冬枣种植者),该网站的农业信息也比较单一(与冬枣相关的农业信息)。对于这种用户类型单一的网站,由于所有用户所关注的信息基本相同,即用户兴趣基本相同,这种情况下无需显式表达目标用户的兴趣,网站上受欢迎程度高的信息也是目标用户的兴趣所在,而网站信息的浏览量和评分可以代表信息的受欢迎程度,某条信息的浏览量越大、评分越高,说明该条信息越受欢迎。因此可以根据各个信息的浏览记录,选择浏览量最高的  $N$  条信息作为候选信息,再根据各条信息的用户评分对信息进行综合测评,选择评分均值大于某阈值的信息推荐给目标用户。

有的农业信息服务网站(比如中国农业信息网),它的农业科技信息涉及农林牧副渔等各个生产部门,农业科技信息不仅数量众多而且类型复杂,这种网站的用户可能为果农、菜农、渔民等不同类型,不同类型用户所关注的信息也各不相同,即用户之间的兴趣差异较大。对于这类网站,只有目标用户的兴趣被明确表达出来,才能对相似用户的筛选起到真正的导航作用。因此,对于多用户类型的农业信息服务网站,我们首先构造用户兴趣导航信息集,然后根据用户兴趣导航信息集筛选导航用户集及相似用户集,最后将相似用户集评分最高的前  $N$  条信息推荐给目标用户。

### 2.1. 相关符号及定义

为了区分不同类型的农业信息服务网站,我们为网站增加了用户类型特征  $D$ ,  $D$  的取值范围为  $\{0,1\}$ ,其中  $D = 0$  表示该网站是单用户类型网站,  $D = 1$  表示该网站是多用户类型网站。设网站的用户集合为  $U = \{U_1, U_2, \dots, U_m\}$ , 农业科技信息(即项目)集合为  $I = \{\langle I_1, B_1 \rangle, \langle I_2, B_2 \rangle, \dots, \langle I_n, B_n \rangle\}$ , 其中  $B_j$  表示项目  $I_j$  的浏览量, 用户——项目评分数据集用  $m \times n$  阶矩阵  $R$  表示。 $R_{i,j}$  ( $1 \leq i \leq m, 1 \leq j \leq n$ ) 表示用户  $U_i$  对项目  $I_j$  的评分,  $R_{i,j}$  为 null 表示用户  $U_i$  对项目  $I_j$  没有进行评分。用户  $U_i$  的评分集合记为  $R_{U_i} = \{R_{i1}, R_{i2}, \dots, R_{in}\}$ , 农业科技信息  $I_j$  的评分集合记为  $R_{I_j} = \{R_{1j}, R_{2j}, \dots, R_{mj}\}$ 。

定义 1 (候选兴趣项目集合)对于类型  $D_i = 0$  的网站, 设该网站的用户集合为  $U = \{U_1, U_2, \dots, U_m\}$ , 项目集合为  $I = \{\langle I_1, B_1 \rangle, \langle I_2, B_2 \rangle, \dots, \langle I_n, B_n \rangle\}$ , 设目标用户为  $U_i \in U$ , 则目标用户  $U_i$  的候选兴趣项目集合为  $I$  中浏览量最大的前  $N$  条信息, 候选兴趣项目集合  $CPI(U_i)$  可表示为:  $CPI(U_i) = \{\langle I_j, B_j \rangle \mid \langle I_j, B_j \rangle \in I \text{ 且 } B_j \text{ 至少大于其他 } |I|-N \text{ 条信息}\}$ 。

定义 2 (综合评分)给定农业科技信息推荐系统的用户集合  $U = \{U_1, U_2, \dots, U_m\}$ , 设目标用户为  $U_i \in U$ , 目标用户的候选项目集合为  $CPI(U_i)$ , 对任意的项目  $\langle I_j, B_j \rangle \in CPI(U_i)$ , 根据  $R_{I_j}$  中所有评分得到的该项目的综合评分记为  $SR_j$ 。

定义 3 (用户兴趣导航信息集)给定农业科技信息推荐系统的用户集合  $U = \{U_1, U_2, \dots, U_m\}$ , 用户——农业科技信息评分数据集为  $R$ , 设目标用户  $U_i \in U$ , 用户  $U_i$  的评分项目记为  $I_{U_i} = \{\langle I_c, B_c \rangle, \dots, \langle I_d, B_d \rangle\}$ , 其中项目  $\langle I_j, B_j \rangle$  ( $c \leq j \leq d$ ) 的评分  $R_{i,j} \neq null$ , 则用户  $U_i$  的兴趣导航信息集为  $I_{U_i}$  中评分最高的  $N$  ( $1 < N < n$ ) 条信息, 用户  $U_i$  的兴趣导航信息集  $UPNI(U_i)$  表示如下:

$UPNI(U_i) = \{\langle I_l, B_l \rangle, \dots, \langle I_{l+N-1}, B_{l+N-1} \rangle\}$ , 其中  $I_{l+x} \in I_{U_i}$  ( $0 \leq x \leq N-1$ ) 且  $R_{i,l+x}$  大于  $I_{U_i}$  中其他项目的评分。

定义 4 (导航用户集)对目标用户  $U_i$  的导航信息集  $UPNI(U_i) = \{\langle I_l, B_l \rangle, \dots, \langle I_{l+N-1}, B_{l+N-1} \rangle\}$ , 如果某用户  $U_j$  在这些信息上都作出过评分, 则用户  $U_j$  为目标用户  $U_i$  的导航用户, 所有导航用户构成的集合记为

$$UN(u_i) = \{U_j | U_j \in U \text{ 且 } R_{j,i+x} \neq \text{null} (0 \leq x \leq N-1)\}。$$

定义 5 (用户相似度) 给定农业科技信息推荐系统的用户集合  $U = \{U_1, U_2, \dots, U_m\}$ , 设目标用户  $U_i \in U$ , 任意用户  $U_j \in UN(u_i)$  与用户  $U_i$  的相似性表示为  $SIM_{i,j}$ 。

定义 6 (相似用户集) 给定农业科技信息推荐系统的用户集合  $U = \{U_1, U_2, \dots, U_m\}$ , 设目标用户  $U_i \in U$ , 对任意用户  $U_j \in UN(u_i)$ , 如果他与目标用户的相似度  $SIM_{i,j} > T_{sim}$ , 其中  $T_{sim}$  为用户相似度阈值, 则用户  $U_j$  为目标用户  $U_i$  的相似用户, 目标用户  $U_i$  的相似用户集  $SIM(u_i)$  可表示为:

$$SIM(u_i) = \{U_j | U_j \in UN(u_i), \text{ 且 } SIM_{a,b} > T_{sim}\}。$$

## 2.2. 基于单用户类型的兴趣导航协同过滤推荐算法

若网站的用户类型比较单一(网站类型根据网站开发目的即可确定), 尽管农业科技信息数据量很大, 但是网站的信息构成并不复杂, 根据前文分析, 这种情况下的农业科技信息的推荐较容易。推荐的主要过程如下: 首先根据信息的浏览量选择目标用户的候选兴趣信息, 然后根据其他用户对候选信息的评分确定每条信息的综合评分, 最终选取综合评分大于某阈值的信息作为推荐信息。

对于综合评分与评分阈值的确定, 我们都选择加权平均的度量方法。其中综合评分  $SR_j$  和综合评分阈值  $T_{SR}$  的公式分别如公式(1)和公式(2)所示。

$$SR_j = \frac{\sum_{U_i \in U} R_{i,j}}{|U|} \quad (1)$$

$$T_{SR} = \frac{\sum_{1 \leq j \leq N} SR_j}{N} \quad (2)$$

基于单用户类型的兴趣导航农业科技信息协同过滤推荐算法(SUCFRA)描述如下:

SUCFRA(D,  $U_i$ , U, I, R, N)

**输入:** 网站类型 D, 目标用户  $U_i$ , 用户集合 U, 项目集合 I, 用户项目评分矩阵 R, 常数 N;

**输出:** 推荐项目集合 RI;

**If** D=0 **then**

/\* 根据信息的浏览量选择目标用户的候选兴趣信息\*/

CPI( $U_i$ )= $\emptyset$ ;

**for**(i=1 to N)

CPI( $U_i$ )=CPI( $U_i$ ) $\cup$ {I 中剩余的  $B_i$  值最大的项目};

**endfor**

**for each**( $\langle I_j, B_j \rangle \in$  CPI( $U_i$ )) **do** /\* 计算 CPI( $U_i$ )各个项目的综合评分\*/

y=0;

**for each**( $R_{s,j} \in R_{i,j}$ ) **do**

**if**  $R_{s,j} \neq \text{null}$  **then**

SR = SR +  $R_{s,j}$ ;

y=y+1;

**endfor**

```

    SRx = SR/y;
    T = T+ SRx;
endfor
TSR = T/N; /* 计算 CPI 项目的综合评分阈值*/
for each(<Ix,Bx>∈ CPI(Ui)) do /* 选择 CPI 中大于综合评分阈值的项目作为推荐项目*/
    If SRx > TSR then
        RI= RI ∪ Ix;
    endfor
return RI;

```

### 2.3. 基于多用户类型的兴趣导航协同过滤推荐算法

对于用户类型多样化的网站，协同过滤推荐算法是以用户兴趣为核心进行设计，旨在通过发掘目标用户兴趣所在，为目标用户产生一个推荐列表。算法的主要步骤如下：1) 首先根据目标用户 $U_i$ 的历史浏览信息及评分确定用户兴趣导航信息 $UPNI(u_i)$ ；2) 选择对 $UPNI(u_i)$ 中所有信息作出评分的其他用户构成目标用户 $U_i$ 的导航用户集 $UN(u_i)$ ；3) 计算 $UN(u_i)$ 中每个用户与目标用户 $U_i$ 的相似度；4) 计算 $UN(u_i)$ 中所有用户的相似度阈值 $T_{sim}$ ；5) 选取 $UN(u_i)$ 中相似度超过阈值 $T_{sim}$ 的用户构成目标用户 $U_i$ 的相似用户集 $SIM(u_i)$ ；6) 选择 $SIM(u_i)$ 中所有用户都评价的项目构造相似项目集 $I_s$ ；7) 根据 $SIM(u_i)$ 中所有用户的相似度作为权重计算 $I_s$ 中各项目的预测评分；8) 选取 $I_s$ 中评分最高的 $k$ 条信息构成推荐列表。由于本推荐系统中为用户推荐的信息不宜太多，以使用户在短时间内看完，因此 $k$ 选取较小的数(如10左右)即可。

这里的用户相似度计算方法采用相关相似性方法，根据目标用户 $U_i$ 和用户 $U_j$ 对用户兴趣导航信息集上的评分的 Pearson 相关系数度量，用户相似性表示如公式(3)所示。

$$SIM_{i,j} = \frac{\sum_{I_c \in UPNI(u_i)} ((R_{i,c} - \bar{R}_i)(R_{j,c} - \bar{R}_j))}{\sqrt{\sum_{I_c \in UPNI(u_i)} (R_{i,c} - \bar{R}_i)^2} \sqrt{\sum_{I_c \in UPNI(u_j)} (R_{j,c} - \bar{R}_j)^2}} \quad (3)$$

用户相似度阈值则采用导航用户集 $UN(u_i)$ 中所有用户与目标用户 $U_i$ 的相似度均值表示，如公式(4)所示。

$$T_{sim} = \frac{\sum_{U_j \in UN(u_i)} SIM_{i,j}}{|UN(u_i)|} \quad (4)$$

基于多用户类型的兴趣导航农业科技信息协同过滤推荐算法(MUCFRA)的详细描述如下：

MUCFRA(D, U<sub>i</sub>, U, I, R, N,k)

**输入：** 网站类型 D, 目标用户 U<sub>i</sub>, 用户集合 U, 项目集合 I, 用户项目评分矩阵 R, 常数 N, 常数 k;

**输出：** 推荐项目集合 RI;

**If** D=1 **then**

    I<sub>U<sub>i</sub></sub>=∅; /\* 根据目标用户的浏览历史构建用户兴趣导航信息集\*/

**for each**(I<sub>j</sub>∈I) **do**

**If** R<sub>i,j</sub>≠null **then**

```

     $I_{U_i} = I_{U_i} \cup I_j;$ 
endfor
UPNI( $u_i$ )=  $I_{U_i}$  中评分最高的 N 个项目;
UN( $u_i$ )=对 UPNI( $u_i$ )中所有项目都有评分的用户集; /* 根据用户兴趣导航信息集构建导航用户集*/
x=0;
for each ( $U_j \in UN(u_i)$ ) do /* 计算导航用户集中各个用户与目标用户的相似度*/
    x= x+1;
     $SIM_x = SIM_{i,j};$ 
    SUM= SUM+  $SIM_x;$ 
endfor
 $T_{sim} = SUM/x;$  /* 计算导航用户集中各个用户与目标用户的相似度阈值*/
for each( $U_j \in UN(u_i)$ ) do/* 在导航用户集 UN( $u_i$ )中选择相似度大于阈值的用户构造相似用户集*/
    If( $SIM_{i,j} > T_{sim}$ ) then
         $SIM(u_i) = SIM(u_i) \cup U_j;$ 
    endifor
 $I_s = SIM(u_i)$ 中所有用户均有评价的项目集合; /* 构造相似项目集  $I_s$ */
l=0; CRI= $\emptyset$ ;
for each( $I_c \in I_s$ ) do/*计算  $I_s$  中各项目的预测评分*/
    l=l+1;
    for each( $U_j \in SIM(u_i)$ ) do
        Score= Score + $R_{j,c} * SIM_{i,j};$ 
        SIM = SIM +  $SIM_{i,j};$ 
    endfor
    Score $_l$ = Score/ SIM;
    CRI= CRI  $\cup \langle I_c, score_l \rangle;$ 
endfor
M= |CRI|; /*选取  $I_s$  中预测评分最高的 k 条信息构造推荐信息集合*/
RI= {  $I_c | \langle I_c, score_l \rangle \in CRI, score_l$  至少大于其他 M-k 个项目的预测评分值};
return RI;

```

### 3. 实验

为验证所提算法的推荐效果，本文对所提算法在不同的数据稀疏情况下与其他协同过滤方法的推荐精度进行了对比实验。本文选择的数据集中包含 1000 个用户对 1200 条农业科技信息的评分，分值为 1~5 分，其中每个用户至少评价 20 条信息，在数据集中用一个字段表示各个项目的评价时间，评价时间随机



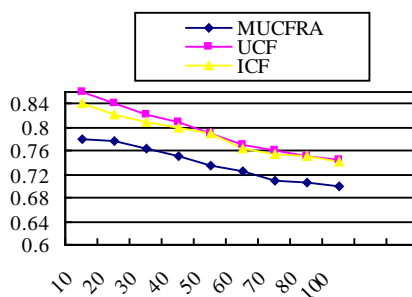


Figure 1. Impact of recommendation accuracy on various sparse rating  
图 1. 数据稀疏度对推荐精度的影响

产生。在实验时，将数据集的 80% 作为计算预测评分的基本数据，将剩余的 20% 作为测试数据，用来对预测推荐结果进行评估。本文所选取的对比协同过滤方法主要包括基于用户相似度的协同过滤方法(user-based collaborative filtering, UCF) [3]、基于项目相似度的协同过滤方法(item-based collaborative filtering, ICF) [4]，由于这两种方法均为多用户类型，我们用 MUCFRA 算法进行实验。其中我们采用相似用户数表示数据集中用户评价的稀疏性，相似用户数分别选取 10, 20, 30, 40, ..., 100 等值，用目标项目预测评分的平均绝对误差表示推荐精度。推荐精度随数据稀疏性的变化如图 1 所示，其中横坐标为相似用户数，纵坐标为预测评分的平均绝对误差。

由图 1 可知，随着相似用户数的减少，UCF 和 ICF 算法预测评分的平均绝对误差增长较快，说明稀疏性对 UCF 和 ICF 算法推荐结果的影响很大，MUCFRA 算法的平均绝对误差相对增长较慢，说明 MUCFRA 算法的推荐结果受稀疏性影响相对较小。随着相似用户数的逐渐增大，三种算法的平均绝对误差均显著减小，说明项目的预测评分精度均有明显提高，而且 MUCFRA 算法相对 UCF 和 ICF 算法的平均绝对误差较低，说明在任意数据稀疏情况下，本文的 MUCFRA 算法在预测精确度方面均优于其他两种方法。

#### 4. 结论

基于用户兴趣导航的协同过滤推荐算法的提出主要是为了解决目前农业科技信息服务的针对性问题，通过认定目标用户的兴趣大大减少候选推荐信息的数量，不仅可以大大降低寻找推荐信息的工作量，而且可以在一定程度上提高农业科技信息的推荐精度，满足农民对科技信息的针对性需求。

#### 基金项目

国家自然科学基金(No. 60773100, 61070032)，河北省自然科学基金(No. F2015407039)，河北省科技厅项目(No. 13227427)，河北省教育厅项目(No. sz132075, 2012~112)，河北科技师范学院博士基金(No. 2013YB007, 2013YB001)，河北科技师范学院科研创新团队基金(No. CXTD2012-08)。

#### 参考文献 (References)

- [1] 中国农业网 (2014) 2014 年中国农业互联网行业分析报告. <http://www.zgny.com.cn/ifm/consultation/2014-10-10/271777.shtml>
- [2] 杨博, 赵鹏飞 (2011) 推荐技术综述. *山西大学学报(自然科学版)*, **34**, 337-350.
- [3] Wang, J., De Vries, A.P. and Reinders, M.J.T. (2006) Unifying user-based and item-based collaborative filtering approaches by similarity fusion. *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, New York, 9 August 2006, 501-508.
- [4] Sarwar, B., Karypis, G., Konstan, J., et al. (2001) Item-based collaborative filtering recommendation algorithms. *Pro-*

*ceedings of the 10th International Conference on World Wide Web*, New York, 1-5 May 2001, 285-295.

- [5] 孙光福, 吴乐, 刘淇, 朱琛, 陈恩红 (2013) 基于时序行为的协同过滤推荐算法. *软件学报*, **24**, 2721-2732.
- [6] 贾冬艳, 张付志 (2013) 基于双重邻居选取策略的协同过滤推荐算法. *计算机研究与发展*, **50**, 1076-1084.
- [7] 顾梁, 杨鹏, 罗军舟 (2015) 一种播存网络环境下的 UCL 协同过滤推荐方法. *计算机研究与发展*, **52**, 474-486.
- [8] 陈诚, 廖桂平, 史晓慧, 等 (2011) 农业农村信息服务个性化推送技术. *中国农学通报*, **27**, 151-156.