International Journal of u- and e- Service, Science and Technology
Vol.9, No. 5 (2016), pp.299-308
http://dx.doi.org/10.14257/ijunesst.2016.9.5.27

A Recommendation Method basing on Synthetic Strategy for Agricultural Science and Technology Information

Song Jinling[1], Huang Liming[1], Gao Yan [2], Liu Aiyong[1] and Liu Haibin[1]

[1]*Hebei Normal University of Science & Technology, Qinhuangdao 066004, China*
[2]*Liaoning Institute of Science and Technology Benxi 117004, China*
songjinling99@126.com

Abstract

To ameliorate the limitations of traditional collaborative filtering technologies and enhance the recommendation quality of agricultural science and technology information, a collaborative filtering recommendation method based on synthetic strategy is proposed. Firstly, filter the user set and user-item rating matrix according to the location of target user, which can solve the regional problem. Then, predict ratings of items according to the similarity of users or item content, which can relax the impact of the sparse rating. In addition, add the rating time to the user-item rating matrix to distinguish the timeliness of the user preference, and add user preference shifting in the similarity formula as a factor which can express the similarity of users or item content better. Our method can not only guarantee the recommended information is local and suit to current season of agricultural production, but also ensure the recommending precision under sparse rating data.

Keywords: *rating sparseness; synthetic strategy; recommending precision; timeliness; user preference shifting*

1. Introduction

Recommendation technologies locate and push the preference contents to users initiatively from the overload information [1-2], which has been widely used in many fields such as electronic commerce [3-16]. Collaborative filtering is one of the most effective recommendation technology [3-6],[9-10] and [12,-15], it recommends the items which one similar user is interested in to the target user basing on the similarity of users (or items). In order to improve the service quality of agricultural information and supply specific and interesting information to farmers, recommendation technologies have been applied to the field of agricultural information service too [16].

However, since agricultural production is regional and seasonal [17], for the agricultural science and technology information service, the traditional collaborative filtering technologies have the following limitations: (1) the problem of sparseness of rating data [1], which is the important problem of collaborative filtering recommendation system itself. Rating sparseness will cause the lower recommendation precision. Although some researches have been focused it, such as user-based and item-based recommendation method [4], user-based(item-based) and trust-based recommendation method[3-6] and [16], soial relation based recommendation method [7-8],[10-14], and so on, all of them still exist some limitations. The user-based and item-based method only composes the user similarity formula and the item similarity formula by adding coefficient simply,but both the user and item similarity calculation in it depend on abundant rating data, which indicate that the method does not really solve the problem. The user-based(item-based) and trust-based recommendation method add trust mechanism when selecting similar users, it need plenty rating data too so it is difficult to solve the problem of data sparsity really. The soial relation based recommendation

ISSN: 2005-4246 IJUNESST

method can not be applied in the absence of social network data. Therefore, sparse rating data problem need to be solved in this papar according to the characteristics of the agricultural science and technology information itself. (2) The regional problem of recommendation information. The main crops in various regions are different, and users in different regions interest in diverse information naturally, so it is necessary to push the specific information needed by the production of target user's region. However the traditional collaborative filtering recommendation technologies didn't consider the regional problem. (3) The timeliness problem of user preference. Because the agricultural production is seasonal, the preference of the target user in a season will be markedly different to other seasons, but the traditional collaborative filtering recommendation technologies didn't care the timeliness of user preference.

To solve the above problems of traditional collaborative filtering technologies and improve the recommendation quality of agricultural science and technology information, we propose a collaborative filtering recommendation method based on synthetic strategy: Firstly, the user set and user-item rating matrix are filtered according to the location of target user to solve the regional problem; Then to resist the rating data sparseness, the items which haven't been rated by target user are predicted according to the similarity of users or item content; In addition, we add the rating time to the user-item rating matrix to distinguish the timeliness of the user preference, and user preference shifting degree is added into the similarity formula as a factor to express the similarity of users or item content better.

2. Definitions and Symbols

Since the service of agricultural science and technology information is regional, and in the recommendation system the agricultural science and technology information itself does not contain geographical features except users can represent the geographical features by registration address or IP address. An information is evaluated by a local user implicates that it must be needed for the local agricultural production, if the target user references other users in the same region in collaborative filtering recommendation, then the recommended information will be with regional characteristics natively. Accordingly, we select all local users with target user as the user set in our recommended system, the user set below refers to this too.

Let the user set is $U=\{U_1,U_2,\ldots,U_m\}$, the set of agricultural science and technology information (*i.e.*, items) is $I=\{I_1,I_2,\ldots,I_n\}$, the user-item rating dataset is a $m\times n$ matrix R, where $R(i,j)$ $(1\le i\le m, 1\le j\le n)$ is the rating of user U_i to item I_j, and $R(i,j)$ is null when user U_i doesn't rate the item I_j. The rating set of user U_i is denoted as $R(U_i)=\{R_{i1}, R_{i2},\ldots, R_{in}\}$, and the rating set of item I_j is denoted as $R(I_j)=\{R_{1j}, R_{2j},\ldots, R_{mj}\}$.

Definition 1 (Candidate Neighbors of User) Let user-item rating matrix is R, target user is $U_i\in U$, the item to predict for user U_i is I_k. If there is a rating $R(j,k)\neq$null of user $U_j \in U(j\neq i)$ in the rating set $R(I_k)$, then U_j is a candidate neighbor of target user U_i for item I_k, and the candidate neighbor set of target user U_i can be expressed as:

$$C_{Ik}(U_i)=\{U_j|U_j\in U, j\neq I \text{ and } R(j,k)\in R(I_k), R(j,k)\neq \text{null}\}.$$

Definition 2 (User Similarity) Let the user set in one region is $U=\{U_1, U_2,\ldots, U_m\}$, the target user is $U_i\in U$, the similarity of any user $U_j\in C_{Ik}(U_i)$ with user U_i is denoteded as $SIMU_{i,j}$.

Definition 3 (Similar Preference Neighbors) Let the users in one region is $U=\{U_1, U_2,\ldots, U_m\}$, the target user is $U_i\in U$, if the similarity $SIMU_{i,j}$ of any user $U_j\in C_{Ik}(U_i)$ with target user U_i is higher than T_{simU} (*i.e.*, $SIMU_{i,j}> T_{simU}$), where T_{simU} is the threshold of user similarity, then the user U_j is the similar preference neighbor of the target user U_i, and the similar preference neighbor set of the target user U_i can be expressed as:

$$SIMU(U_i) = \{U_j | U_j \in C_{Ik}(U_i) \text{ and } SIMU_{i,j} > T_{simU}\}.$$

Definition 4 (Candidate Neighbors of Item) Let user-item rating matrix is R, target user is $U_i \in U$, the item to predict for user U_i is I_k. If there is a rating $R(i,j) \neq null$ of the item $I_j \in I (j \neq k)$ in the rating set $R(U_i)$, then the item I_j is a candidate neighbor of item I_k, and the candidate neighbor set of item I_k can be expressed as:

$$C_{Ui}(I_k) = \{ I_j | I_j \in I, j \neq k \text{ and } R(i,j) \in R(U_i), R(i,j) \neq null\}.$$

Definition 5 (Item Similarity) Let the item set in recommendation system is $I = \{I_1, I_2, \ldots, I_n\}$, the target item is $I_k \in I$, the similarity of any item $I_j \in C_{Ui}(I_k)$ with item I_k is denoted as $SIMI_{k,j}$.

Definition 6 (Similar Content Neighbors) Let the item set in recommendation system is $I = \{I_1, I_2, \ldots, I_n\}$, the target item is $I_k \in I$. For any item $I_j \in C_{Ui}(I_k)$, if the similarity $SIMI_{k,j}$ of I_j with the target item I_k is higher than T_{simI} (*i.e.*, $SIMI_{k,j} > T_{simI}$), where T_{simI} is the threshold of item similarity, then the item I_j is the similar content neighbor of the item I_k, and the similar content neighbor set of the target item I_k can be expressed as:

$$SIMI(I_k) = \{I_j | I_j \in C_{Ui}(I_k), \text{ and } SIMI_{k,j} > T_{simI}\}.$$

3. Period Based User Preference and User Preference Shifting

The agricultural production changes in different seasons, for example, spring is the planting time of soybean, corn, sorghum and other crops, summer is the growing time of various crops, autumn is not only the harvest time of various crops but also the planting time of winter wheat, winter is the growing time of winter wheat. So, the agricultural science and technology information needed by a user in different season will be distinct too, *i.e.*, the user preference will transform from various seasons greatly. To realize the timeliness of recommendation result, we need to understand the user preference on various periods (seasons) and the user preference shifting between different seasons, so that it can produce precision recommendation result.

Let the period t_1 is the current season, then the items which target user browses and rates on t_1 period can express his truly preference on current period, and the his preference will shift in some degrees on other period (such as t_2) relative to period t_1. In order to get user preference on various periods, we need to get the user-item rating matrix on various periods firstly. We can save the rating time along with the user's rating and then get the matrix on any period by filling the rating matrix according to the rating time on this period. Let the user-item rating matrix on period t_1 is R_{t1}, and $R_{t1}(i,j)$ $(1 \leq i \leq m, 1 \leq j \leq n)$ is the rating of the user U_i to the item I_j on period t_1. Let the user-item rating matrix on period t_2 is R_{t2}, and $R_{t2}(i,j)$ $(1 \leq i \leq m, 1 \leq j \leq n)$ is the rating of the user U_i to the item I_j on period t_2.

Let the item set which the target user U_i has been rated on the period t_1 is $I_{t1}(U_i) = \{I_{c'}, \ldots, I_{d'}\}$, we extract a number of keywords from each item of $I_{t1}(U_i)$ respectively and construct the preference documents $D_{t1}(U_i)$ of user U_i, that is the user preference of U_i on the period t_1. Analogously, let the item set rated on the period t_2 is $I_{t2}(U_i) = \{I_{c''}, \ldots, I_{d''}\}$, the preference documents constructed by above method is $D_{t2}(U_i)$, which is the user preference of U_i on the period t_2.

For the user preference shifting between another period and period t_1, we measure it synthetically by the similarity of the content and rating of user preference document on two periods. Let the common keyword set of $D_{t1}(U_i)$ and $D_{t2}(U_i)$ is $W = \{W_1, W_2, \ldots, W_g\}$, and the frequency of any keyword $W_i (1 \leq i \leq g)$ appears in it's item is $|W_i|$, the rating set of each item containing one keyword in $D_{t1}(U_i)$ is $R_{t1}(W) = \{R_{t1}(i, I(W_1)), R_{t1}(i, I(W_2)), \ldots, R_{t1}(i, I(W_g))\}$, the rating set of each item containing one keyword in $D_{t2}(U_i)$ is $R_{t2}(W) = \{R_{t2}(i, I(W_1)), R_{t2}(i, I(W_2)), \ldots, R_{t2}(i, I(W_g))\}$, then the preference shifting of the target user U_i between the period t_1 and t_2 is Formula (1).

$$UPS_{t1,t2} = \frac{\sum_{j=1}^{g} (|w_j| R_{t1}(i, I(W_j)))(|w_j| R_{t2}(i, I(W_j)))}{\sqrt{\sum_{j=1}^{g} (|w_j| R_{t1}(i, I(W_j)))^2} \sqrt{\sum_{j=1}^{g} (|w_j| R_{t2}(i, I(W_j)))^2}} \tag{1}$$

In this formula we represent user preference by combining the feature and rating of items, and calculate the cosine similarity of two preference documents by taking the product of keyword frequency and the item rating as factor, which can express the user preference shifting more accurately. It can be known that $UPS_{t1,t2}$ =1 when t_2=t_1.

The calculation algorithm of user preference shifting(UPSA) is described as following:

UPSA (U_i, t_1,t_2,R)

Input: the target user U_i, period t_1, period t_2, user-item matrix R;

Output: the user preference shifting $UPS_{t1,t2}$;

R_{t1}= user-item matrix R on period t_1;
R_{t2}= user-item matrix R on period t_2;
/*the item set rated by U_i on period t_1*/
for each(R(i,j)∈ R_{t1} (U_i)) ***do***/* the item set rated by U_i on period t_1 */
    if(R(i,j) ≠null) ***then***
     $I_{t1}(U_i)= I_{t1}(U_i) \cup I_j$;
endfor
for each(R (i,j)∈ R_{t2} (U_i)) ***do*** /* the item set rated by U_i on period t_2 */
    if(R(i,j) ≠null) ***then***
     $I_{t2}(U_i)= I_{t2}(U_i) \cup I_j$;
endfor
$D_{t1}(U_i)$=keyword set of $I_{t1}(U_i)$;
$D_{t2}(U_i)$= keyword set of $I_{t2}(U_i)$;
W= $D_{t1}(U_i) \cap D_{t2}(U_i)$;
for each(W_i∈ W) ***do***
    Fz= Fz +$|W_i|*R_{t1}(i,I(W_i))*|W_i|*R_{t2}(i,I(W_i))$;
    Fm1= Fm1+$(|W_i|*R_{t1}(i,I(W_i)))^2$;
    Fm2= Fm2+$(|W_i|*R_{t2}(i,I(W_i)))^2$;
endfor
$UPS_{t1,t2}$ = Fz/ $(Fm1*Fm2)^{0.5}$;
return $UPS_{t1,t2}$;

4. User Based Rating Prediction

The user based rating prediction algorithm predicts the rating of the unevaluated item of the target user basing on the similarity between users. The main steps are: Let the item to rate for target user U_i is I_k, ① Search the candidate neighbor set C_{Ik} (U_i) of the target user U_i for item I_k; ② To each user in C_{Ik} (U_i), calculate his similarity with the target user U_i; ③Calculate the similarity threshold T_{simU} of all users in C_{Ik} (U_i) with the target user U_i; ④Choose user whose similarity is greater than T_{simU} as the similar user and construct similar neighbor set SIMU (U_i); ⑤Calculate the prediction rating of I_k by taking each similarity of users in SIMU (U_i) as weight.

To improve the calculated accuracy of the similarity between one user in C_{Ik} (U_i) and the target user U_i, we choose the period with most common rating items to calculate the similarity. However the selected period may be not the current period, so it is ought to take into account the preference shifting of target user on this period. We measure the user similarity by the product of the modified cosine similarity formula and the preference shifting of target user. Let the period with most common items between user U_i and U_j is

t_i, and the common item set of user U_i and U_j on period t_i is $I_{ti}(U_i, U_j)$, then the similarity $SIMU_{i,j}$ between users U_i and U_j is as Formula (2), where $\overline{R_{ti}(U_i)}$ and $\overline{R_{ti}(U_j)}$ are the average ratings on common items $I_{ti}(U_i, U_j)$ of user U_i or U_j on period t_i, $UPS_{t1,ti}$ is the preference shifting of target user U_i on period t_i.

$$SIMU_{i,j} = \frac{\sum_{I_k \in I_{ti}(U_i,U_j)} (R_{ti}(i,k) - \overline{R_{ti}(U_i)})(R(j,k) - \overline{R_{ti}(U_j)})}{\sqrt{\sum_{I_k \in I_{ti}(U_i,U_j)} (R_{ti}(i,k) - \overline{R_{ti}(U_i)})^2} \sqrt{\sum_{I_k \in I_{ti}(U_i,U_j)} (R_{ti}(j,k) - \overline{R_{ti}(U_j)})^2}} . UPS_{t1,ti} \tag{2}$$

To the user similarity threshold T_{simU}, we use the average similarity of all users in $C_{Ik}(U_i)$, which is as Formula (3).

$$T_{simU} = \frac{\sum_{U_j \in C_{Ik}(U_i)} SIMU_{i,j}}{|C_{Ik}(U_i)|} \tag{3}$$

The algorithm of user based rating prediction (URPA) is described as following:

URPA (U_i, I_k, U, R, t_1)

Input: the target user U_i, the item I_k to be rated, user set U, user-item matrix R, current period t_1;

Output: prediction rating R(i,k) of item I_k;

R_{t1}= user-item matrix R on period t_1;
for each($R_{t1}(j,k) \in R_{t1}(I_k)$) ***do***
  if($R_{t1}(j,k) \neq$null) ***then***
    $C_{Ik}(U_i) = C_{Ik}(U_i) \cup U_j$;
endfor
for each($U_j \in C_{Ik}(U_i)$) ***do***
  t_i=the period with most common rating items of user U_i and U_j;
  $I_{ti}(U_i, U_j)$= the common item set of user U_i and U_j;
  $\overline{R_{ti}(U_i)}$ =the average rating of user U_i on items $I_{ti}(U_i, U_j)$;
  $\overline{R_{ti}(U_j)}$ = the average rating of user U_j on items $I_{ti}(U_i, U_j)$;
  for each($I_k \in I_{ti}(U_i, U_j)$) ***do***
    Fz= Fz+$(R_{ti}(i,k) - \overline{R_{ti}(U_i)})(R(j,k) - \overline{R_{ti}(U_j)})$;
    Fm1= Fm1+$(R_{ti}(i,k) - \overline{R_{ti}(U_i)})^2$;
    Fm2= Fm2+$(R_{ti}(j,k) - \overline{R_{ti}(U_j)})^2$;
  endfor
  $SIMU_{i,j}$=Fz/(Fm1*Fm2)$^{0.5}$* UPS(U_i, t_1,t_i,R);
  x=x+1;
  SUM= SUM+ $SIMU_{i,j}$;
endfor
T_{simU}= SUM/x;
for each($U_j \in C_{Ik}(U_i)$) ***do***
  if($SIMU_{i,j} > T_{simU}$) ***then***
    SIMU (U_i)= SIMU (U_i)$\cup U_j$;
endfor
l=0; Rating=0;
for each($U_j \in$ SIMU (U_i)) ***do***
  l=*l*+$SIMU_{i,j}$;
  Rating = Rating + R(j,k) * $SIMU_{i,j}$;
endfor
R(i,k) = Rating /*l*;

5. Item Content Based Rating Prediction

Because of the sparseness of rating data, to one unevaluated item of the target user, there may be few or no other ratings on it. Under this case, it is impossible to predict the item by the similarity between users (similar users can't be found or untrustworthy), but if there are other items containing approximate content with the item, then the rating can be predicted according to the similar items. The method can relax the disadvantage of the rating sparseness on a certain extent and improve the diversity of recommended result.

The item content-based rating prediction algorithm predicts the rating of the unevaluated item basing on the similarity between of item content. The main steps are as below: Let the item to rate for target user U_i is I_k, ①Search the candidate neighbor set $C_{Ui}(I_k)$ of the target item I_k; ②To each item in $C_{Ui}(I_k)$, calculate it's similarity with the target item I_k; ③Calculate the similarity threshold T_{simI} of all items in $C_{Ui}(I_k)$; ④Choose item whose similarity is greater than T_{simI} as the similar item and construct similar content neighbor set SIMI (I_k); ⑤Calculate the prediction rating of I_k by taking the similarities of items in SIMI (I_k) as weight.

To calculate the similarity of one item in $C_{Ui}(I_k)$, firstly calculate the similarity of the common keywords between the item and target item, in addition it need to combine the preference shifting of target user as a factor because the item maybe not belong to the current period which indicate that the item deviates from the current user preference. So the similarity between items measures by the product of cosine similarity of common keywords and user preference shifting. Let the target item is I_k, the current period is t_1, the period containing the candidate neighbor item I_j is t_i, the common keyword set of item I_k and I_j is $W=\{W_1, W_2,\ldots, W_g\}$, the frequency of any common keyword $W_l(1\le l\le g)$ appears in its item is $|W_l|$, then the similarity $SIMI_{k,j}$ between items of I_k and I_j is as Formula (4).

$$SIMI_{k,j} = \frac{\sum_{l=1}^{g} (|w_l|\mathrm{R}(i,k))(|w_l|\mathrm{R}(i,j))}{\sqrt{\sum_{l=1}^{g} (|w_l|\mathrm{R}(i,k))^2}\sqrt{\sum_{l=1}^{g} (|w_l|\mathrm{R}(i,j))^2}}.UPS_{t1,ti} \tag{4}$$

To the item similarity threshold T_{simI}, we use the average similarity of all items in $C_{Ui}(I_k)$, which is as Formula (5).

$$T_{simI} = \frac{\sum_{I_j\in C_{Ui}(I_k)} SIMI_{k,j}}{|C_{Ui}(I_k)|} \tag{5}$$

The algorithm of item content based rating prediction (ICRPA) is described as following:

ICRPA (U_i, I_k, U, R, t_1)
Input: the target user U_i, the item I_k to be rated, user set U, user-item matrix R period t_1;
Output: prediction rating R(i,k) of item I_k;
for each(R(i,j)∈R(u_i)) ***do***
  if(R(i,j) ≠null) ***then***
    $C_{Ui}(I_k) = C_{Ui}(I_k) \cup I_j$;
endfor
for each($I_j \in C_{Ui}(I_k)$) ***do***
  t_i= the rating period of item I_j;
  W=the common keyword set of item I_k and I_j;
  for each($W_l \in$ W) ***do***

Fz= Fz+$(|w_l|R(i,k))(|w_l|R(i,j))$;

Fm1= Fm1+$(|w_l|R(i,k))^2$;

Fm2= Fm2+$(|w_l|R(i,j))^2$;

endfor

$SIMI_{k,j}$=Fz/ $(Fm1*Fm2)^{0.5}$* UPS (U_i, t_1,t_i,R);

x=x+1;

SUM= SUM+ $SIMI_{k,j}$;

endfor

T_{simI} = SUM/x;

for each($I_j \in C_{ui}(I_k)$) ***do***

if($SIMI_{k,j} > T_{simI}$) ***then***

SIMI (I_k)= SIMI (I_k) $\cup I_j$;

endfor

l=0; Rating=0;

for each($I_j \in$ SIMI (I_k)) ***do***

l=*l*+ $SIMI_{k,j}$;

Rating= Rating +R_{ij}* $SIMI_{k,j}$;

endfor

R(i,k) = Rating /*l*;

6. Collaborative Filtering Recommendation Algorithm Basing on Comprehensive Strategies for Agricultural Science and Technology Information

The main steps of collaborative filtering recommendation algorithm basing on comprehensive strategies for agricultural science and technology information is: ①Choose all the users in the same region with the target user U_i as U, and select the rating matrix R for user set U; ② Find out all the items $I'(U_i)$ which the target user U_i hasn't rated; ③To each item I_k in $I'(U_i)$, count the users which have rated it on current period; ④If the rating users of item I_k is great than the given threshold T_{count}, then call the URPA algorithm to predict the rating on I_k; ⑤Otherwise call the ICRPA algorithm to predict the rating on I_k; ⑥When all the items have been rated, select N items with the highest rating in $I'(U_i)$ as the recommendation result.

The algorithm of SSCFRA is described as following:

SSCFRA (U_i, U′,R′, t1, Tcount, N)

Input: the target user U_i, the initial user set U′, the initial user-item matrix R′, the current period t_1, the rating user count threshold T_{count}, the count N of recommendation items;

Output: the recommendation result set RI;

U=all the users of U′ in same region with target user U_i;

R= the user-item matrix for use set U in R′;

for each($R(i,j) \in R(U_i)$) ***do*** /*Search the items unevaluated by U_i */

if(R(i,j) =null) ***then***

$I'(U_i)= I'(U_i) \cup I_j$

endfor

for each($I_k \in I'(U_i)$) ***do***

Count=0;

R_{t1}= user-item matrix R on period t_1;

for each($R_{t1}(i,k) \in R_{t1}(I_k)$) ***do***

if($R_{t1}(i,k) \neq$ null) ***then***

Count = Count+1;

```
    endfor
    if(Count > T_count) then
        URPA (U_i, I_k, U, R, t_1);
    else
        ICRPA (U_i, I_k, U, R, t_1);
  endfor
  RI= the N items with the highest rating in I′(U_i);
  return RI;
```

7. Experiment

To test the recommendation performance of SSCFRA algorithm, the following two experiments are designed: (1) Under various rating sparseness, the recommendation precision of the SSCFRA algorithm compares with other collaborative filtering methods; (2) Whether the SSCFRA algorithm can recommend the information of current season ahead. The first experiment aims at the recommendation precision problem, and the second experiment aims at the timeliness of SSCFRA algorithm.

The data set we select contains ratings of 1000 user to 1200 items (agricultural science and technology information), where the rating is from 1 to 5 and each user rates 20 pieces of information at least. In the data set there is a field to show the rating time of each item, and the rating time is generated randomly to distinguish various seasons. In our experiments, 80% data set is taken as basic data to predict the ratings of other items and the remaining 20% is taken as test data to evaluate the recommendation results.

The methods which we selected to compare with SSCFRA algorithm are UCF (user-based collaborative filtering) [4] and ICF (item-based collaborative filtering) [5].

Firstly, we test the prediction precision of SSCFRA, UCF, and ICF with various data sparseness. Due to T_{count} is the threshold of candidate neighbors in SSCFRA algorithm, it can represent the rating sparseness. Therefore, we select 10, 20, 30, 40... 100 to T_{count} as the rating sparseness respectively, and use the mean absolute error of prediction rating of target items to express the recommendation precision. The recommendation precision changes with the rating sparseness are shown in Fig. 1, where the horizontal ordinate is T_{count} value and the ordinate is the mean absolute error of prediction rating.

It can be known from Fig. 1 that the mean absolute error of prediction rating increase gradually along with the decrease of T_{count}, which indicates that the prediction rating of items depends mainly user-based method, so the rating sparseness influence the recommendation result greatly. Following T_{count} increases, the mean absolute error of prediction rating diminish observably, which indicates that the prediction precision of items is improved significantly under the synthesized strategy of user-based method and item content-based method. In addition, the prediction accuracy of our SSCFRA excelled other methods under any data sparseness.

Then, under various T_{count} value, we test the timeliness of the recommendation result of SSCFRA algorithm and other methods. The experimental result is shown in Figure 2, where the horizontal ordinate is T_{count} value and the ordinate is the mean absolute error between the period of the predicted item and the current period. Figure 2 indicates that the information recommended by SSCFRA algorithm suit for current period better than other methods.

Synthesize all experimental results, we get a conclusion: With sparse rating data, the recommendation precision and timeliness of algorithm SSCFRA are all superior to the traditional recommendation methods.

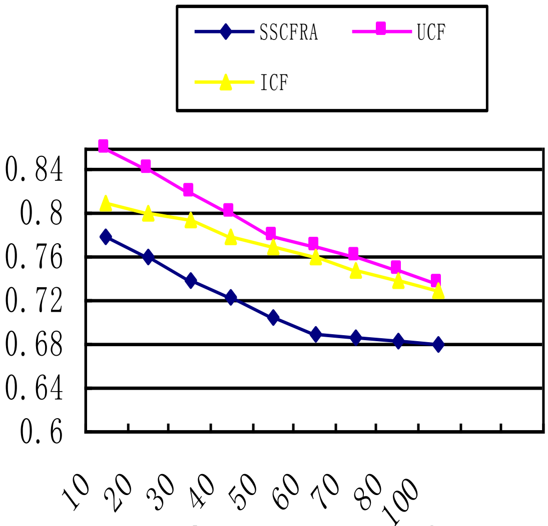

Figure 1. Impact of Recommendation Precision on Various Sparse Rating

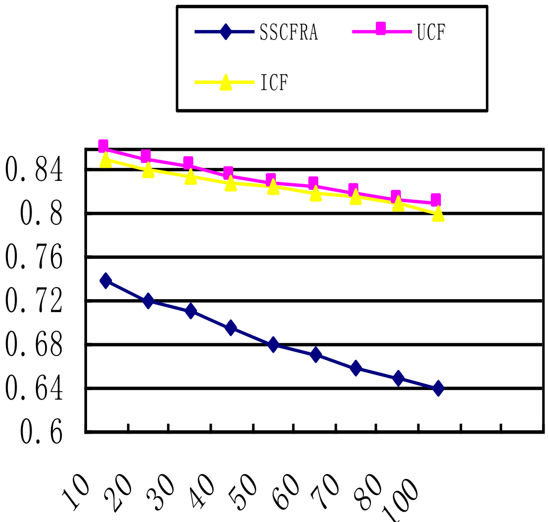

Figure 2. Impact of Recommendation Timeliness on Various Sparse Rating

8. Conclusion

In order to improve the service for agricultural science and technology information, we propose a collaborative filtering recommendation algorithm basing on comprehensive strategy. The algorithm can not only resist the influence of sparse rating and improve the recommendation precision greatly, but also can satisfy the regional characteristics and timeliness of recommendation result.

Acknowledgments

This work is supported by The Natural Science youth Foundation of Hebei Province (No. F2015407039), Hebei Province Department of science and technology Fund (NO.13227427), Hebei Province Department of Education Fund (NO.sz132075, sz141097, 2012-112), the Doctoral Foundation of Hebei Normal University of Science & Technology under Grant (No. 2013YB007, 2013YB001), and the Creative team Foundation of Hebei Normal University of Science & Technology (No. CXTD2012-08).

References

[1] F. Ricci, L. Rokach and B. Shapira, "Recommender Systems Handbook", Sringer, Berilin, **(2011)**.

[2] X. Hailing, W. Xiao and L. Xiaodong, "Comparison study of Internet recommendation system", Chinese Journal of Software, vol. 20, no. 2, **(2009)**, pp.350-362.

[3] J. Dongyan and Z. Fuzhi, "A Collaborative Filtering Recommendation Algorithm Based on Double Neighbor Choosing Strategy",Chinese Journal of Computer Research and Development, vol. 50, no. 5, **(2013)**, pp.1076-1084.

[4] J. Wang, A. P. D. Vries and M. J. T. Reinders, "Unifying user-based and item-based collaborative filtering approaches by similariity fusion", Proceedings of the 29th Annual Int ACM SIGIR Conf on Research and Development in Information Retrieval, Seattle, Washington, USA, **(2006)** August 6-11.

[5] B. Sarwar, G. Karypis and J. Konstan, "Item-based collaborative filtering recommendation algorithms. Proceedings of the 10^{th} Int Conf on World Wide Web, HongKong, China, **(2001)** May 1-5;.

[6] M. Jamali and M. Ester, "TrustWalker: A random walk model for combining trust-based and item-based recommendation, "Proceedings of the 15th ACM SIGKDD Int Conf on Knowledge Discoverry and Data Mining, Paris, France. **(2009)** June 28 - July 1.

[7] H. Ma, T. C. Zhou and M. R. Lyu, "Improving recommender systems by incorporating social contextual information, ACM Transaction on Information Systems, Article 9, vol. 29, no. 2, **(2011)**.

[8] H. Ma, I. King and M. R. Lyu, "Learning to recommend with explicit and implicit social relations", ACM Transaction on Intelligent System and Technology, Article 29, vol. 2, no. 3, **(2011)**,

[9] W. S. Guangfu and L. Qi, "Recommendations Based on Collaborative Filtering by Exploiting Sequential Behaviors", Chinese Journal of Software, vol. 24, no. 11, **(2013)**, pp.2721-2732.

[10] L. Guo, J. Ma, Z. M. Chen and H. R. Jiang, "Learning to recommend with social relation ensemble", Proceedings of the 21st ACM International Conference on Information and Knowledge Management, Maui, USA, **(2012)** October 29-November 2.

[11] L. Wu, E. H. Chen, Q. Liu, L. L. Xu, T. F. Bao and L. Zhang, "Leveraging tagging for neighborhood-aware probabilistic matrix factorization", Proceedings of the 21st ACM International Conference on Information and Knowledge Management, Maui, USA, **(2012)** October 29-November 2.

[12] Q. Liu, E. H. Chen, H. Xiong, C. H. Q. Ding and J. Chen "Enhancing collaborative filtering by user interests expansion via personalized ranking", IEEE Trans. on Systems, Man and Cybernetics-B., vol. 42, no. 1, **(2012)**, pp.218-233.

[13] L. Yanen, Z. Chengxiang and C. Ye, "Exploiting rich user information for one-class collaborative filtering", Knowledge and Information Systems, vol. 38, no. 2, **(2014)**, pp. 277-301.

[14] Y. Bhawsar, G. S. Thakur and R. S.Thakur, "User recommendation system using Markov model in social networks', International Journal of Computer Applications, vol. 86, no. 9, **(2014),** pp. 33-39.

[15] G. Liang, Y. Peng and L. Junzhou, "A Collaborative Filtering Recommendation Method for UCL in Broadcast-Storage Network", Chinese Journal of Computer Research and Development, vol. 52, no. 2, **(2015**), pp.474-486.

[16] S.-G. Deng, L.-T. Huang,J. Wu and Z.-H. Wu, "Trust-Based Personalized Service Recommendation: A Network Perspective", Journal of Computer Science & Technology. vol. 29, no. 1, **(2015)**, pp. 69-80.

[17] C. Cheng, L.Guiping, S. Xiaohui, "The Personalized Push Technology of Agriculture and Rural Information Service", Chinese Agricultural Science Bulletin, vol. 27, no. 29, **(2011)**, pp. 151-156.

