

# Collaborative Filtering Recommendation Based on Preference Order

Li Yu and Xiaoping Yang

School of Information, Renmin University of China, Beijing 100872, P.R. China  
[buaayuli@ruc.edu.cn](mailto:buaayuli@ruc.edu.cn) [yang@ruc.edu.cn](mailto:yang@ruc.edu.cn)

**Abstract.** Collaborative filtering is an important personalized method in recommender systems in E-commerce. It is infeasible that traditional collaborative filtering is based on absolute rating for items since users are difficult to accurately make an absolute rating for items, and also different users give different rating distribution. In this paper, an improved collaborative filtering algorithm based on preference order of items is proposed, which user only give rating order of items. The presented method enhances collaborative filtering by improving accuracy and adeptness. An example is provided to verify the method.

**Keywords:** *Customer relationship management, Electronic commerce, Collaborative filtering*

## 1. INTRODUCTION

Collaborative filtering is the most important personalized recommendation used widely in recommender system in E-commerce [1]. But traditional collaborative filtering is based on user's absolute rating for items [2], that is to say, it is based on semantic differential. Usually, user give high rating for preferred product and low rating for dislike product, as shown in the following.

Above the rating scheme have two supposing:

Like 5 4 3 2 1 Dislike
------------------------

First, all users have same understanding for rating which mean same valuing standard. It is often that the preprime is hard to satisfy. One hand, users give the highest rating for same item, which do not means they have equal preference for the item. On the other hand, rating distribution given by different user is different. Some users prefer to give high rating and some prefer to give low rating. Table 1 shows experience statistical rating distributing. Seeing from table 1, ratio of high rating and low rating is not same. Most users prefer to give high rating and some users prefer to give low rating. That is to say, rating information is skewed.

**Table 1. Ratio of Different Rating for Items by Users**

Rating	1 Dislike	2 ...	3 adiaphorous	4 ...	5 Like
Ratio	0.082	0.095	0.226	0.224	0.372

Second, often users are hard to give the absolute rating. This means absolute rating cannot reflect accurately users' preference. Collaborative filtering based on them is not accurate. But it is easier to give order preference of items.

According to above analysis, collaborative filtering based on preference order is better method while absolute rating is hard to given. Once user give preference order of items, collaborative filtering will give recommendation result. But one must be pointed that it is also difficult if user need to give preference order for a lot of items. In fact, the method is more difficult than absolute rating. In practice, preference order can be given by users or be got by computing. For example, it is given according to time of browsing web pages in system based web; it is given according to buying cost on a certain product in supermarket. In addition, it can be transferred to order rating when absolute rating is not enough accurate, then collaborative filtering based order rating is employed.

## 2. PREFERENCE ORDER

### 2.1 Order

Define: Order is a sequence of objects according to certain attribute<sup>[3]</sup>, denoting as  $O$ . That is to say that order  $O$  is order set of objects.

For example, there are three objects:  $x_1, x_2, x_3$ , they can be order according to certain user' preference for them.

$$x_3 \succ x_1 \succ x_2$$

Denoting  $r(O, x)$  as order number of object  $x$  in Order  $O$ , for example, For Order  $O$ ,

$$O = x_3 \succ x_1 \succ x_2,$$

Then

$$r(O, x_1) = 2, r(O, x_2) = 3, r(O, x_3) = 1$$

As shown in Figure 1, there three items, A, B and C. Order number B in Order sequence is 3.

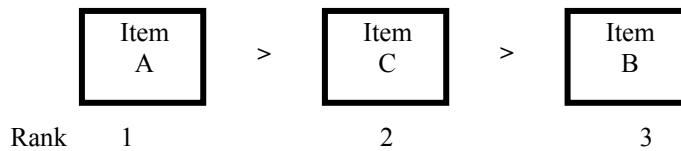


Figure 1. Preference Order

## 2.2 Require of Preference Order

Requiring accurately of use's preference order for items is the first step in application of collaborative filtering based it. Generally, preference order can be gotten by the following methods.

(1) Explicit preference order by user

When number of items is few, users can directly give the preference order for all items according to their interests. Because of the limit of human ability, it is difficult while number of items is large. The following two methods are often employed.

(2) Implicit preference order based on transaction data

Transaction data show user preference. More transaction on a certain product shows more users' preference for it. In the case, preference order can be gotten according to the buying cost on product or to quality of transaction on product.

(3) Implicit preference order based on Web mining

Web mining developed in recent years is an important method to get the users' preference by analyzing browsing pattern. For example, preference for product can be measured by time of browsing web pages by mining log information. The method has especial effect in E-commerce.

Supposed there are attribute set  $R = \{A_1, A_2, \dots, A_m\}$  in a E-commerce, And corresponding, defining event  $e = (a_1, a_2, \dots, a_m, t), a_i \in D_{A_i}, t$  means time when event  $e$  happen.

An event sequence  $S$  is a set of all events. In WWW lot, an event can include visited URL, host IP and visiting time, etc, as shown in the following,

*(athene.cs.ncu.edu.cn,07/Aug/1996 15:37:11,/manila/data-mining.html,200,12134)*

In above example, the first attribute is host name, the second one is visiting time, the third is visited page, the last two attributes show requesting result.

Generally, users' preference order can be presented as,

$$S = \{(e_1, t_1), (e_2, t_2), \dots, (e_n, t_n)\}$$

For web page recommendation,  $e_i$  means web page,  $t_i$  represents spending time on visiting web page.

## 3. COLLABORATIVE FILTERING BASED ON PREFERENCE ORDER

Input: User's preference order for items

Output: Recommendation for Current User  $U^*$

Algorithm:

- (1) Getting the preference order according method introduced in 2.2;
- (2) Computing similarity of user  $U^*$  and other users---similarity  $(u^*, u_k)$  ( $k=1,2,\dots,n$ );
- (3) Determining neighbor uses by ranking the similarity  $(u^*, u_k)$ ;  
The step is similar to traditional collaborative filtering.
- (4) Predicting the preference for object item,

$$\begin{aligned}
 & \text{Score}(u^*, \text{item } A) \\
 &= \sum_{u \in \text{Neig}(u^*)} \text{Rank}(u, \text{item } A) \times \text{Weight}(u^*, u) \tag{1}
 \end{aligned}$$

(5) Rank the predicting value for all item that not be rated, and make the former n items as recommendation result.

In the above algorithm, step (2) is the most key step, which decided on recommendation result. In the next part, similarity of two orders is discussed in detail.

#### 4. SIMILARITY BASED ON PREFERENCE ORDER

The most difference between collaborative filtering based in preference order and traditional collaborative filtering is different user preference information. The former employ preference order and the later employ absolute rating, as shown in Table 2. The difference decides that different method must be used in computing the similarity between users.

**Table 2. User/rating Based Preference Order**

User\Item	A	B	C	D	E
U1	3	1	4	5	2
U2	4	2	5	1	3
U3	2	1	4	5	?

In computing the similarity, traditional *Cosine* or *Spearman* method cannot be used because user information is rank of item. And *Spearman's Rank Correlation* [3] is employed to measure the similarity of users. If two order have same rank number, then the following equation be employed,

$$\rho = \frac{\sum_{x \in X_1} (r(O_1, x) - \bar{r}_1)(r(O_2, x) - \bar{r}_2)}{\sqrt{\sum_{x \in X_1} (r(O_1, x) - \bar{r}_1)^2} \sqrt{\sum_{x \in X_2} (r(O_2, x) - \bar{r}_2)^2}} \tag{2}$$

and  $\bar{r}_i = (1/|X_i|) \sum_{x \in X_i} r(O_i, x)$ . Otherwise, the following equation be employed,

$$\rho = 1 - \frac{6 \times \sum_{x \in X_1} (r(O_1, x) - r(O_2, x))^2}{|X_1|^3 - |X_1|} \tag{3}$$

According to the above equation,  $\rho=1$  if the two order is same, otherwise  $\rho=-1$ .

If two orders have different items, two orders must be transferred by dropping the items that is not another order, and then similarity is computed. For example, having two order  $O_1$  and  $O_2$ ,

$$Q = x^1 \succ x^3 \succ x^4 \succ x^6 \quad Q_2 = x^5 \succ x^4 \succ x^3 \succ x^2 \succ x^6$$

Then

$$O'_1 = x^3 \succ x^4 \succ x^6 \quad O'_2 = x^4 \succ x^3 \succ x^6$$

$$r(O'_1, x^3) = 1 \quad r(O'_1, x^4) = 2 \quad r(O'_1, x^6) = 3$$

$$r(O'_2, x^3) = 2 \quad r(O'_2, x^4) = 1 \quad r(O'_2, x^6) = 3$$

According to the equation 2, the similarity of the two orders is

$$\rho = 1 - \frac{6((1-2)^2 + (2-1)^2 + (3-3)^2)}{3^2 - 3} = 0.5$$

If two orders have no relationship, then  $\rho=0$ .

## 5. SIMILARITY BASED ON BROWSING SEQUENCE

### 5.1 General Method

General similarity of two orders is discussed above. In practice, preference order often been got by mining users' browsing sequence. In the case, similarity based on browsing sequence is important.

Here, similarity of event sequence based on edit distance is employed [3-4]. Supposing the following sequence,

$$S = \{(e_1, t_1), (e_2, t_2), \dots, (e_n, t_n)\}$$

Any sequence can be transferred into another sequence by the following operations,

- ① Ins(e, t), Inserting event e when time t;
- ② Del(e, t), Deleting event e when time t;
- ③ Move(e, t, t'), Moving event e from time t to time t';

And every operation o has cost C(o).

For event sequence S is transferred into event sequence T though operation set  $OP = \{o_1, o_1, \dots, o_k\}$ , then the cost of operations is

$$C(OP) = \sum_{i=1}^k c(o_i) \quad (4)$$

Obviously, many transferring can realize from event sequence S event sequence T, selecting the min cost of transferring as distance between event sequence S and T. Noting  $d(S, T)$  as distance between event sequence S and T, defining

$$D(S, T) = \min\{C(OP_j) | j=1, 2, 3, \dots\} \quad (5)$$

### 5.2 Improved Method

Considering about the affect of length of sequence on edit distance, an improved edit distance is employed, as following,

$$D(S, T) = \frac{\min\{C(OP)\}}{\max\{|S|, |T|\}} \quad (6)$$

And  $|S|$  and  $|T|$  is respectively length of sequence  $S$  and  $T$ .

Once edit distance of two sequence is computed, their similarity can be computed as following:

$$\overline{Similarity}(S_i, S_j) = \frac{1}{D(S_i, S_j)} \quad (7)$$

Then following transferring is done,

$$Similarity(S_i, S_j) = \frac{\overline{Similarity}(S_i, S_j)}{\max_{\forall i, j} (\overline{Similarity}(S_i, S_j))} \quad (8)$$

### 5.3 Example

Supposing having four sequence  $S_1$ ,  $S_2$ ,  $S_3$  and  $S_4$ :

$$S_1 = \langle (A, 1), (A, 6), (D, 10), (E, 11), (F, 12), (G, 13) \rangle$$

$$S_2 = \langle (A, 3), (C, 8), (D, 10), (E, 11), (F, 12), (G, 13) \rangle$$

$$S_3 = \langle (A, 1), (A, 6) \rangle$$

$$S_4 = \langle (A, 3), (C, 8) \rangle$$

According to the equation,

$$d(S_1, S_2) = \frac{2.4}{6} = 0.4 \quad d(S_3, S_4) = \frac{2.4}{2} = 1.2$$

Then

$$\overline{Similarity}(S_1, S_2) = \frac{1}{0.4} \quad \overline{Similarity}(S_3, S_4) = \frac{1}{1.2}$$

$$Similarity(S_1, S_2) = \frac{1}{0.4} / \frac{1}{0.4} = 1 \quad Similarity(S_3, S_4) = \frac{1}{1.2} / \frac{1}{0.4} = 0.33$$

## 6. CONCLUSIONS

Collaborative filtering is most important recommendation in E-commerce. Traditional collaborative filtering is based on absolute rating which is hard to get or is not accurate because human has limit ability to understand a lot of objects accurately. Sometime, user's preference order is easy to get by mining usage data, analyzing transaction data, etc. Collaborative filtering algorithm based on preference order is presented. In the future, the algorithm should further being verified in a real database.

## ACKNOWLEDGMENTS

This research was supported by Open Foundation of Key Laboratory of Data Engineering and Knowledge Engineering, Ministry of Education P. R. China, Open Foundation of Key Laboratory of Information Management and Economy, Ministry

of Information Industry, P. R. China under grants F0607-31, F0607-42, and Research Foundation of Renmin Univ. of China.

## REFERENCES

1. P. Resnick and H.R. Varian, Recommender systems, *Communications of the ACM*. Volume 40, Number 3, pp.56-58, (1997).
2. B.M. Sarwar, G. Karypis, J.A. Konstan, and J. Riedl, Analysis of Recommendation Algorithms for E-Commerce, in *Proceedings of the ACM EC'00 Conference* (Minneapolis: MN, 2000), pp.158-167.
3. M.V.M. Halkidi and Y. Batistakis, On clustering validation techniques, *Journal of Intelligent Information Systems*. Volume 17, Number 2-3, pp.107-145, (2001).
4. Banerjee and J. Ghosh, Clickstream clustering using weighted longest common subsequences, in *Proc. of Workshop on Web Mining in First International SIAM Conference on Data Mining* (Chicago, 2001), pp.33-40.
5. L. Yu and L. Liu, Research on Similarity of event sequence, *Computer Engineering*. Volume 29, Number15, pp.13-14, (2003).