

A Study of User Preference-Based Collaborative Filtering Algorithm

Sumi Shin

S&T Information Center, Korea Institute of Science and Technology Information

ABSTRACT

A recommendation system is a program which attempts to predict items that users may be interested in, considering their preference and taste. Collaborative filtering is an algorithm which is most commonly used in the recommendation system. It recommends the items that users are interested in with similar preferences. The system finds similar users based on all users' purchase and rating data. However, user preferences can change, and this kind of change may slow down recommendation performances. To upgrade these recommendation performances, this study proposes a user preference-reflected recommendation technique. The precision and MAE of the proposed method were measured, using MovieLens data set. Compared to conventional collaborative filtering techniques, the proposed method revealed better results.

Keywords: Collaborative Filtering, Recommender System, Recommendation, User Preference

Date of Submission: 22 December 2016



Date of Accepted: 05 January 2017

I. INTRODUCTION

A recommender system is a program which attempts to predict items that users may be interested in, considering their preference and taste. Many recommendation services which help consumers make a decision have been developed and upgraded by online shopping malls and digital contents service providers. The most widely used technique in the recommendation system is collaborative filtering [1]. Collaborative filtering has no limitations in item or user attributes. It also offers an opportunity to make use of other users' preference information in a diverse manner. Collaborative filtering is designed based on the fact that when a certain consumer chooses a particular item, other consumers with a similar preference may like this item as well[2]. Under this concept, the program recommends this item to those with similar tastes. Therefore, collaborative filtering requires a process to find the users with similar interest and analyze the items based on neighboring users' ratings on these items. In general, people's preferences can change over time, and this kind of change may weaken recommendation performances. To improve these recommendation performances, this study proposes a user preference-reflected recommendation technique. In other words, conventional systems use a similarity measurement method based on total data to select similar users. In contrast, the proposed technique reflects user preferences with an addition of time. This study is structured as follows: In chapter 2, collaborative filtering is explained. In chapter 3, a user preference-added upgraded algorithm is proposed. In chapter 4, the proposed method is compared to conventional systems. In chapter 5, conclusion is given.

II. RELATED WORKS

The conventional studies which have adopted this kind of collaborative filtering approach include Group Lens[3], Ringo[4] and MovieLens[5]. Collaborative filtering is an algorithm which recommends useful items to target users based on the preferences on the items provided by neighboring users which have shown similar purchase patterns. In other words, this technique starts with the following assumption: 'People's interests or preferences are not just randomly distributed. There exist general trends and patterns among people's preferences.' In fact, people get recommendations from their acquaintances or through online communication with other users and refer to them when choosing an item. Collaborative filtering is a method which refers to other users' past preferences for the purpose of predicting a certain user's preference. The Information Tapestry built by the Xerox Palo Alto Research Center (Nichols and others) is the first system which adopted collaborative filtering. Collaborative filtering has no limitations in analysis of item attributes or user attributes. In addition, it offers a chance to utilize neighboring users' diverse information. This algorithm can be divided into user-based collaborative filtering and item-based collaborative filtering.

User-based collaborative filtering is performed as follows:

(i) Construct input data: A user-item matrix is constructed as stated in Table 2 based on the data just like Table 1 (past purchases or ratings by the users).

Table 1. Examples of the Item Rating Data

Alice: Shrek(Rating:4), SnowWhite(Rating:5), Superman(Rating:3)
Bob : SnowWhite(Rating:3), Superman(Rating:5), spiderman(Rating:5)
Chris: Superman(Rating:5), spiderman(Rating:5)
Toby: Superman(Rating:3)

Table 2. User-Item Matrix

	Shrek	Snow White	Superman	spiderman
Alice	5	5	5	-
Bob	4	3	5	4
Chris	3	-	5	-
Toby	-	-	3	-

(ii) Search for neighboring groups:

The ratings on the same item are compared to other users' ratings. Then, similarity among users is calculated, and neighboring groups are constructed. As stated in Table 2, once neighboring groups are selected against Toby, Alice, Bob and Chris assess the superman rated by Toby with the same scores. As a result, the three users are chosen as the neighbors with the same similarity.

(iii) Create recommendation list: A recommendation list is created after estimating preference on the items based on neighboring groups.

If a recommendation list is created based on Table 2, all films rated by Alice, Bob and Chris become the targets of recommendation. Since their average rating is the same with '4,' all films would be recommended. .

In case of Toby, for example, too many neighbors are chosen as similar ones during the search for neighboring groups. As a result, too many films are qualified to be recommended.

In item-based collaborative filtering, the ' ii ' and ' iii ' steps are modified as follows during user-based collaborative filtering[6]:

(ii) Search for similar items: The similarity among items is calculated according to the degree of purchase by other users and level of similar ratings.

(iii) Create recommendation list: A recommendation list is created based on similarity with the items bought or experienced by a user in the past

III. PROPOSED ALGORITHM

In general, human preferences can change over time. In case of the items which are sensitive to trends such as film and music, in particular, user preferences can rapidly change. However, a collaborative filtering algorithm targets all data regardless of changes in user preferences. In this sense, the rating performances of the items which are sensitive to trends could drop. This study proposes a new similarity calculation approach which measures user preferences based on similarity changes between two users. The proposed technique calculates similarity with the items rated by Users A and B at a certain point of time when trend should be reflected among total rating values. This method can be user similarity at a certain point of time. However, the two users may rate the items at a different point of time so that the number of items rated by both users at a certain point of time would be relatively low. If there are only few items for calculation, it wouldn't be able to estimate appropriate similarity. To take care of this problem, the items rated by at least one of the two users at a certain time zone were included during the calculation of similarity. For this, if a set of the items rated by Users A and B from the beginning to the time 'T' is AB(t), a set of the items rated by the two users from T-1 to T would be as follows:

$$\text{SetAB}(T) = \text{AB}(T) - \text{AB}(T-1)$$

Here, the similarity of two users is defined with the cosine similarity of SetAB(T). This can be summarized with pseudo code as follows Fig.1:

```

for (t=0 to T-1)
  if (item of A == item of B)
    add item to AB(T-1)
for (t=0 to T)
  if (item of A == item of B)
    add item to AB(T)

SetAB = AB(T)-AB(T-1)
Calculate similarity
    
```

Figure.1 Pseudo code of proposed algorithm

IV. EXPERIMENTS AND ASSESSMENTS

This study attempted to measure performances with the recommendation technique proposed based on the data collected through actual users. Among the MovieLens data collected as a part of GroupLens Research Project in the University of Minnesota, 1 million data having 1 million ratings were used. The MovieLens 1 million data consist of 1,000,209 ratings on 3,952 films performed by 6,040 users. In this experiment, recommendation performances on 604 users (10% of total users) were measured. The experiment results were compared to user-based collaborative filtering and item-based collaborate filtering with two different indicators.

First of all, performances were assessed with precision which tells how much the recommendation list matches the items actually bought. ‘Precision’ is a rating method which reveals how much a customer has actually spent to get some items among the recommendation list. When a set of the items used by ‘User a’ (V_a) and set (R_a) defined with a set of the items recommended to ‘User a,’ the precision (PR_a) on ‘User a’ can be defined equation 1 below:

$$PR_a = \frac{|V_a \cap R_a|}{|R_a|} \quad (1)$$

Next, rating-predicted performances were measured based on similarity For estimation on performances, the MAE grating scale on the performances of rating prediction was used on rating prediction performances. It can be expressed as follows: MAE is an indicator commonly used for prediction. Here, the recommendation performances are assessed based on absolute mean of differences between users’ actual rating and prediction by recommendation system. It can be stated as equation 2 below:

$$MAE = \frac{\sum_{i=1}^n |p_i - a_i|}{n} \quad (2)$$

At the first experiment, a fifty film recommendation list by technique, and precision was compared. Table 3 below represents the results of precision on the recommendation list. As shown in this test, the proposed method is greater than the conventional techniques in terms of precision.

Table 3. Precision Comparison

Algorithm	User-based Collaborative Filtering	Item-based Collaborative Filtering	Proposed Technique
Precision	23.4%	17.2%	25.1%

Table 4 reveals error measurements on prediction errors. As confirmed in the results, the proposed method is higher than the conventional techniques in terms of 7% or more.

Table 4. MAE Comparison

Algorithm	User-based Collaborative Filtering	Item-based Collaborative Filtering	Proposed Technique
MAE	0.64	0.81	0.61

V. CONCLUSION

As online business environment rapidly expands, a recommendation system has become more important. This study investigated collaborative filtering algorithm mostly used in a recommendation system and proposed a brand-new recommendation system considering user preferences. Under the proposed method, time is reflected on user preferences by transforming a similarity calculation method between two users. According to MovieLens data, the proposed method revealed upgraded performances (5-7% improved, compared to conventional collaborative filtering). Compared to MAE performance improvement, the accuracy of performance improvements were significantly increased, confirming usefulness in the creation of a recommendation list. Then, it was confirmed that it was about film-related data in which user preferences can be classified by time, and that the proposed algorithm would be useful in the fields where preferences change over time.

REFERENCES

- [1]. Goldberg David, Nichols David, Oki Brian M. and Terry Douglas, Using collaborative filtering to weave an information tapestry , *Communications of the ACM, Vol. 35*, 1992, 61-70.
- [2]. J. Bobadilla, F. Ortega, A. Hernando and A. Gutierrez, Recommender systems survey, *Knowledge-Based Systems*, vol.46, 2013, 109-132.
- [3]. Konstan, Miller, and Herlocker, GroupLens: Applying Collaborative Filtering to Usenet News, *Comm. ACM, Vol.40*, 1997, 77-87.
- [4]. U. Shardanand and P. Maes, Social Information Filtering: Algorithms for Automating 'Word of Mouth', *Proc. Conf. Human Factors in Computing Systems*, 1995.
- [5]. L. Getoor and M. Sahami, Using Probabilistic Relational Models for Collaborative Filtering, *Proc. Workshop Web Usage Analysis and User Profiling (WEBKDD '99), August*. 1999.
- [6]. B. Sarwar, G. Karypis, J. Konstan and J. Riedl, Item-based Collaborative Filtering Recommendation Algorithms, *Processing of the 10th International World Wide Web Conference, ACM Press*, 2001. 285-295.



S.M Shin received the B.S. and M.S. degrees in Computer Engineering from Hongik University, Korea, in 1997 and 2005, respectively. She is a Ph.D. student in Dept. of Computer Engineering, Hongik University, Korea. And she is currently a senior researcher in the Department of Information Service, KISTI, Korea. She is interested in recommender system, semantic web, big data analysis, databases, data retrieval and Internet of Things.