

# A User-based Filtering Recommendation Algorithm for Health Service

Tong Wu <sup>a</sup>, Runtong Zhang <sup>b</sup>

School of Economics and Management, Beijing Jiao Tong University, Beijing, 100044, China

<sup>a</sup>14120638@bjtu.edu.cn, <sup>b</sup>rtzhang@bjtu.edu.cn

**Abstract.** Under the mobile environment, the traditional collaborative filtering algorithm can't combine with context information well, which result in the quality of recommendation decline. So this paper proposes an improved algorithm which combined with effective context rates with user-based collaborative filtering. The algorithm specifics to the problem of mobile information service recommendation and is based on context awareness. And use the public datasets from MovieLen to validate the algorithm. The result shows that the improved algorithm is more accurate compared with the traditional collaborative filtering recommendation algorithm and recommendations obtained from the results is in line with user preferences better.

**Keywords:** context aware; personalized recommendation; user-based collaborative filtering; health service.

## 1. Introduction

The research of recommendation system dedicated to ease the 'information overload' problem due to the growth of network information. The traditional personalized recommendation algorithm just use the relationship between user and resources, but neglect the environment of users and the status of users. However in the mobile Internet environment, the context information will influence the user's potential demand [1-6]. Domestic and foreign scholars have done a lot of research on how to use context information reasonably to improve the accuracy of the recommendation system. Kim, etc. [1] put forward a model of context awareness, and apply it to the user preferences in the process of collaborative filtering on the missing value extraction, to realize the combination of the collaborative filtering and context awareness. T. h. Dao, etc. [2] put forward a new model of the recommendation, which combine the -aware with genetic algorithm and collect situation data of mobile users to calculate similarity of the context to create a new context awareness model of collaborative filtering and use context awareness with genetic algorithm to construct a model of advertising recommended based on location.

Considering the complexity and diversity of context factors in the mobile Internet environment, this paper puts forward an improved algorithm combining effective context rates with user-based collaborative filtering, which make the traditional collaborative filtering algorithm with the context awareness. The more appropriate information resources can be recommended to the user in accordance with the current situation. Finally, Using Movie Len public data to test the effectiveness of the algorithm, the results show that the improved algorithm has certain advantages comparing with the traditional user-based collaborative filtering algorithm.

## 2. Related Research

### 2.1 A Collaborative Filtering Recommendation Algorithm Based on context awareness

Traditional collaborative filtering algorithm find the similar users through the users' ratings of the similar items, and then using this similarity for recommending information resources. In the collaborative filtering, the user ratings of information is static, constant, while in a dynamic environment the user's decision will be affected by context factors. Chen [3], puts forward a collaborative filtering algorithm based on the correlation coefficient of the situation of resources and user-context similarity; Linas [4] put forward an algorithm, combining the situation information with traditional user-item score matrix and adding the appropriate situation characteristics for each resource, then realizing context awareness recommendation by using matrix decomposition and

nearest neighbor collaborative filtering. Anyhow, collaborative filtering recommendation technology based on situational awareness is join the context information into the user, resource similarity and similarity model, in order to improve the recommendation accuracy.

## 2.2 Extracting Effective Situation Information

The primary function of effective situation information extraction is to extract useful context factors and update the user's personal context awareness model. For example, user is influenced by many situation information when see a doctor, such as location situation, state of illness situation and time, budget, etc. But not all context information will make great influence to the user preference, such as the user's job. So we need to extract effective situation from all kinds of context factors and use the effective situation information to make recommendation results easier to let the user satisfaction.

## 3. A User-Based Collaborative Filtering Algorithm Combining With Effective Context Rating.

### 3.1 Design Ideal

The algorithm based on 'user-context-item' model extended the user's preferences under the condition of a single dimension situation and gave weights of situation according to the effective impact on the preference, then improved the similarity calculation algorithm in the traditional algorithm. At the same time, it introduced the penalty function for reducing the uncertainty due to the sparse data. Finally it concluded the personalized recommendation algorithm which can deal with multidimensional context information.

### 3.2 Algorithm Process

#### (1)The User-based Collaborative Filtering Recommendation Algorithm

There are many kind of user similarity calculation in collaborative filtering algorithm. But the modified cosine similarity is most popular be used because it consider the dimension of the difference of each dimensions, solve the problem of the different users' rating scale. The user similarity calculation formula is as follows:

$$\text{sim}(U_a, U_b) = \frac{\sum_{i \in N(u) \cap N(v)} (r_{ai} - \bar{r}_a)(r_{bi} - \bar{r}_b)}{\sqrt{\sum_{i \in N(u) \cap N(v)} (r_{ai} - \bar{r}_a)^2 \times \sum_{i \in N(u) \cap N(v)} (r_{bi} - \bar{r}_b)^2}} \quad (1)$$

$\bar{r}_u, \bar{r}_v$  represent the average score of their own choice program of the user  $u$  and  $v$ .

#### (2)Effective Context Rating

Using the ideas of research the result of changing a variable, this paper change the situation information from Multidimensional  $C$  to a single dimension  $C_t$ , then use the Relative Standard Deviation(RSD) to calculate user preferences' change under the condition of  $C_t$ , defined as influence coefficient  $Icf$ .  $R_{ijc_t}$  Represent the degree of change of user's preference of item  $i_j$  under the situation  $C_t$ , among  $R_{ijc_t} \in (0,5)$  other context factors' scores are total of  $N$  when calculate  $R_{ijc_t}$ .

$$R_{ijc_t} = \frac{\sum_{c_t \in C} R_{ijc_t}}{n} \quad (2)$$

The formula of the volatility of user preferences under Single dimension situation information is as follow:

$$Icf_t = \frac{SD}{\bar{X}} = \sum_{u_i \in U} \frac{\sqrt{\frac{1}{|c_t|-1} \sum_{c_t \in C_t} \left( R_{ijc_t} - \overline{R_{ijc_t}} \right)^2}}{\overline{R_{ijc_t}}} \quad (3)$$

This paper define  $\theta_t$  represent the impact of user preference under the useful context factor  $c_t$ :

$$\theta_t = \frac{Icf_t}{\sum_{i=1}^k Icf_i} \quad (4)$$

In the formula (4),  $k$  represent the  $k$  useful context factors.

(3) An Algorithm Combining User-Based Collaborative Filtering Recommendation with Effective Context Rating

According to the result of user preferences under the condition of single  $C_t$  situation, this paper improve the formula of calculate the user similarity. And then calculate the user similarity under the multidimensional situation.

Based on the users similarity calculation of single dimension situation information  $C_t$ , to improve the correction of cosine similarity:

$$\text{sim}(U_a, U_b)_{-C_t} = \frac{\sum_{(j,t) \in S(x,y)} (r_{ajt} - \bar{r}_a)(r_{bjt} - \bar{r}_b)}{\sqrt{\sum_{(j,t) \in S(x,y)} r_{ajt}^2 \times \sum_{j \in N(u) \cap N(v)} r_{bjt}^2}} \quad (5)$$

The similarity between two users is more reliable only when they have more common rating items, so this paper introduce the penalty function, as follows:

$$P(u, v) = \frac{\min\left(\sum_{u, v \in U} \{u \neq 0, v \neq 0\}, \partial\right)}{\partial} \quad (6)$$

The formula which combine the penalty function with modified cosine similarity is:

$$\text{sim}(U_a, U_b)_{-C_t} = P(u, v) \text{sim}(U_a, U_b)_{-C_t} \quad (7)$$

So the situation similarity can be written to:

$$\text{Sim}_{-context}(U_a, U_b) = \sum_{t=1}^k \theta_t \text{sim}(U_a, U_b)_{-C_t} \quad (8)$$

As a result, user similarity calculate formula of multidimensional context information is as follows:

$$\text{SimFinal}(U_a, U_b) = \beta \text{sim}_{-context}(U_a, U_b) + (1 - \beta) \text{sim}_{-Nocontext}(U_a, U_b) \quad (9)$$

So user rating prediction formula for the items under the condition of multidimensional context is:

$$\text{MultiContext}_{-P} = \bar{r}_v + \frac{\sum_{v \in U} (\text{SimFinal}(u, v)(r_{v,i,c} - \bar{r}_v))}{\sum_{v \in U} \text{sim}_{-Final}(u, v)} \quad (10)$$

## 4. Simulation and Competition

### 4.1 Data Source

This paper demonstrates results by the typical MovieLens-100k dates which derived from the University of Minnesota of the Group Lens Research item, including 943 users' evaluations to 1682 films of 100000 times (the film's score between 1 to 5 points). Each user evaluate at least 20 films. At the same time, the data set provides five to 80% / 20% of the training set and test set to help us to predict on the training set, and compared the test set general prediction results with the actual results .

The movie ratings contain less situation information, While the records of users' evaluations include the timestamp when the user given by film score, we can extract the timestamp to rate translates into user behavior when user time situation. In addition, we will the user's gender, age and as simulation scenarios.

### 4.2 Standard of Evaluation

This paper take MAE (Mean Absolute Error) to represent the accuracy of prediction. Computation formula is as follows:

$$\text{MAE} = \frac{\sum_{i \in N(u) \cap N(v)} r'_{u,i,context} - r_{u,i,context}}{T} \quad (11)$$

In the formula (11),  $\hat{r}_{ij}$  represents the predict score,  $r_{ij}$  represents the real score. T is the number of test dataset.

### 4.3 Results Analysis

#### (1)The Traditional User-based Collaborative Filtering Algorithm

The traditional algorithm only consider two dimensions of ‘user-item’ information. It predict users’ score of items according to calculate similarity between users, according to the score value sequence and then predict the user's preferences. The calculation results of five training datasets as shown in table 1:

Table 1 the mean absolute error of the Traditional User-based Collaborative Filtering Algorithm

MAE	Data1	Data2	Data3	Data4	Data5
Traditional collaborative filtering algorithm based on user	0.772	0.781	0.775	0.776	0.763

#### (2)The Improved Algorithm

In order to verify the effectiveness of the proposed model, the first step is extracting the effective situation for user preferences. The results of five dataset such as table 2 and 3.

Table 2 situational factors and factors coefficient

Situation	$lcf(1)$	$lcf(2)$	$lcf(3)$	$lcf(4)$	$lcf(5)$
time $C_1$	0.00209	0.00224	0.00221	0.00231	0.00233
age $C_2$	0.00258	0.00249	0.00215	0.00261	0.00217
gender $C_3$	0.00196	0.00203	0.00169	0.00187	0.00201

Table 3 situational factors and factors coefficient

Situation	$\theta_t(1)$	$\theta_t(2)$	$\theta_t(3)$	$\theta_t(4)$	$\theta_t(5)$
time $C_1$	0.315234	0.316188	0.352445	0.318113	0.333333
age $C_2$	0.38914	0.376702	0.362563	0.39726	0.346093
gender $C_3$	0.295626	0.30711	0.284992	0.284627	0.320574

So, in the similarity calculation, the original similarity weight 70 percent, the similarity under the situation of information distribution weight of 30 percent. Finally it is concluded that the result of the user preferences MAE of the improved algorithm under the condition of multidimensional situation the result of the user preferences MAE as shown in table 4.

Table 4 the mean absolute error of improved algorithm

MAE	Data1	Data2	Data3	Data4	Data5
Improved algorithm	0.762	0.769	0.768	0.769	0.752

#### (3)Comparison

The mean Absolute error of two algorithms as shown in table 5:

Table5 the mean Absolute error

MAE	Data1	Data2	Data3	Data4	Data5
Traditional collaborative filtering algorithm based on user	0.772	0.781	0.775	0.776	0.763
Improved algorithm	0.762	0.769	0.768	0.769	0.752

Compared with the traditional algorithm, this article puts forward the prediction more accurate than traditional collaborative filtering recommendation algorithm, so the above experimental data can prove that situational information affects users' preferences. And recommendation algorithm added situation information makes the prediction of user preferences more accurate.

## 5. Summary

In summary, this paper researches the most popular recommendation algorithm and puts forward an improved algorithm combining with user-based collaborative filtering with effective context rating. The improved algorithm has been proved more effective than the traditional user-based collaborative filtering recommendation algorithm through the famous MovieLens datasets. And it has certain guidance and help for solving the practical.

## Acknowledgments

Thank you for the project 'The Wise Medical Health Management Innovation Driven by Big-Data' (Project No.71532002) which is supported by the NSFC (National Nature Science Foundation of China)

## References

- [1] Kim J, Lee D, Chung K Y. Item recommendation based on context awareness model for personalized u-healthcare service [J]. *Multimedia Tools and Applications*, 2014, 71(2): 855-872.
- [2] Dao T H, Jeong S R, Ahn H. A novel recommendation model of location-based advertising: context awareness Collaborative Filtering using GA approach [J]. *Expert Systems with Applications*, 2012, 39(3): 3731-3739.
- [3] Dey A K. Providing architectural support for building context awareness applications [D]. Georgia Institute of Technology, 2000.
- [4] Chen A. context awareness collaborative filtering system: Predicting the user's preference in the ubiquitous computing environment [M]//*Location-and Context-Awareness*. Springer Berlin Heidelberg, 2005: 244-253.
- [5] Adomavicius G, Tuzhilin A. context awareness recommender systems [M]//*Recommender systems handbook*. Springer US, 2011: 217-253.