A Collaborative Filtering Recommendation Algorithm Based on User Interest

Zhenyu Chen[1, a], Wenye Yu[2, b]

[1]School of Information, Nanchang Institute of science & Technology, Nanchang, 330108, China

[2]School of Education, Nanchang Institute of science & Technology, Nanchang, 330108, China

[a]zhenyuchenpaper@163.com, [b]csuzju@163.com

Keywords: Collaborative Filtering, User Interest, Recommendation System.

Abstract. Collaborative filtering recommendation is one of the most successful technologies for applying in recommendation system of E-commerce websites. However, the traditional collaborative filtering recommendation system is under-powered for dealing with user interest and data sparsity. To solve this problem, this paper proposes a collaborative filtering recommendation algorithm based on user interest (CFRUI). The algorithm introduces user interest weights to establish user-item rating matrix, implements similarity calculation by clustering analysis, so as to get the final recommended result. The experiment results show the algorithm can make effectively utilization of user interest to improve recommended quality.

Introduction

The rapid development of web technology makes application of E-commerce gradually infiltrate into people's daily lives. Along with people having increasing demand for information, every internet service providers have gradually changed their study emphasis, and implemented interested resource recommendation for users in accordance with the user's existing network information. And then the design and selection of recommendation algorithm is even more important in recommendation system.

As present, there are some recommendation algorithms, such as recommendation algorithm based on demography, recommendation algorithm based on mixed recommendation, recommendation algorithm based on content, recommendation algorithm based on knowledge and recommendation algorithm based on collaborative filtering, etc [1]. The collaborative filtering algorithm of tradition has some drawback, which does not focus on the user interest but one-sided emphasis on similarity users' behaviors.

The traditional Collaborative Filtering Algorithm

At present, the collaborative filtering algorithm based on user is one of the most recommendation algorithms, which core is to find out the similarity rules by comparison of target user and "the nearest neighbour" item rating.

Building User Information.

Rating matrix R [2] is $m \times n$ matrix, m is the number of users, and n is the number of items. $R = [r_{ij}]$, the element r_{ij} represents the interest rating of user i to item j.

Table 1 User-item Interest Rating Matrix

User	I_i	...	I_j	...	I_R
$User_1$	$R_{1,1}$	...	$R_{1,j}$	...	$R_{1,R}$
$User_2$	$R_{2,1}$	...	$R_{2,j}$	...	$R_{2,R}$
...	...	...	...	...	...
...	...	...	...	...	...
$User_m$	$R_{m,1}$	...	$R_{m,j}$	...	$R_{m,R}$

Table 1 is a user-item interest rating matrix. Contrary to Mean Absolute Error(MAE), if $R_{i,j}$ value is bigger,the degree of users being interested in item is higher .

Calculation of User Similarity.

The steps of similarity calculation between user i *and* user j is follows:

Step1: Calculating and acquiring all items of user i and user j rating.

Step2: Calculating the similarity between user i and user j by similarity calculation, which is $sim(i, j)$.

There are three kinds of calculating similarity algorithm:

1) If taking rating of user i and user j in n dimensional items marked as I and J, and their similarity is expressed by formula (1):

$$sim(i, j) = \cos(I, J) = \frac{I \cdot J}{\|I\| \times \|J\|} \tag{1}$$

Where the numerator represents as evaluated product of user i and user j in the n dimensional items, the denominator represents as dot-product of vectors of user i and user j in the n dimensional items.

Person similarity

Training the item set, which had been evaluated commonly by user i and user j. Person similarity is expressed by P_{ij}, and the similarity between user i and user j is expressed by formula (2):

$$sim(i, j) = \frac{\sum_{a \in I_{ij}} (R_{i,a} - \overline{R_i})(R_{j,a} - \overline{R_j})}{\sqrt{\sum_{a \in I_{ij}} (R_{i,a} - \overline{R_i})^2} \sqrt{\sum_{a \in I_{ij}} (R_{j,a} - \overline{R_j})^2}} \tag{2}$$

The modified cosine correlation [3]

The CFRUI method is not concerned with issues of interest rating of users, which overcomes the above-mentioned shortcomings with modified cosine correlation, and subtracts the average rating of user to items. Let command rating items of user i and user j to be I_{ij}. I_i and I_j represent rating of user i and user j in n dimensional items respectively, and the similarity of user i and user j is calculated Ed by in formula (3):

$$sim(i, j) = \frac{\sum_{a \in I_{ij}} (R_{i,a} - \overline{R_i})(R_{j,a} - \overline{R_j})}{\sqrt{\sum_{a \in I_i} (R_{i,a} - \overline{R_i})^2} \sqrt{\sum_{a \in I_j} (R_{j,a} - \overline{R_j})^2}} \tag{3}$$

The producing of Recommendation.

1) The users' preference for any items

Let the nearest target neighbors of use i to be U_i, and the predicted rating of use i *to* item j can be expressed by formula(4):

$$P_{i,j} = \overline{R_i} + \frac{\sum_{n \in U_i} sim(i,n) * (R_{i,j} - \overline{R_n})}{\sum_{n \in U_i} sim(i,n)} \tag{4}$$

2) The Top-N Recommendation

After calculating the interest of use i *to* different items respectively, selecting the highest weight value of the interest and N items of no in users' interest of the rated items set as the Top-N recommendation[4].

In conclusion, owing to the traditional similarity algorithms have some defects, which don't make a decision effectively and calculate the similarity of users when the user rating data sizes are sparse. But the modified cosine similarity measure algorithm can solve the problem. Compared with the traditional algorithms, the interest rating of users does not affect the similarity calculation, instead, it can more effectively and more precisely get results.

The Collaborative Filtering Recommendation Algorithm based on User Interest

In the CFRUI algorithm, let i and j to be user and item respectively, rating *of* user i to item j r and item to be $P(i, j)$, U_i to be user set of i, U_j to be user set of j. If threshold of the ratio of i and U_i is over normal average, and it means that use i *is* h interest in item j, not vice versa. Therefore, $P(i, j)$ can be measured interested degree of user i to item j, that is $P(i, j)$ can be calculated by formula (5) [5].

$$P_{i,j} = \frac{\sum_{n \in U_j} sim(i, j)}{|U_j|} \tag{5}$$

In the CFRUI algorithm, the user interest weights were introduced to record and distinguish user interest. The algorithm sorts user interest by analyzing the record file of user, and calculates the proportion of single interest of user, assigns weight value.

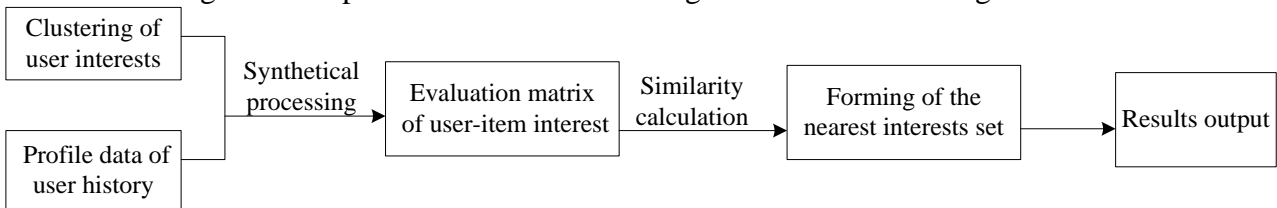The flow diagram of improved recommendation algorithm is shown in Fig.1.

Fig. 1The flow diagram of collaborative filtering algorithm based on user interest

The steps of improved algorithm are as follows:

Step1: Calculating the similarity weighting of user interest by using the modified cosine similarity.

Step2: Calculating the Top-N recommendation of the user in accordance with number of user neighbor;

Step3: Calculating the users' interesting degree $P(i, j)$ and interesting values of clustering users;

Step4: Calculating the User-Item interest matrix with XQ by analyzing the user neighborhood;

Step5: Inputting matrix R, and defining the user and neighbor as I, J respectively;

Step6: Switching the matrix to database;

Step7: Modifying the similarity by formula (1);

Step8: Calculating the similarity of user in accordance with step 5, and getting the rating by contra- simplifying;

Step9: Predicting the rating of i to j by the formula (6):

$$P_{i,j} = \overline{R_i} + \frac{\sum_{n \in U_i} sim(i, n) * (R_{i,j} - \overline{R_n})}{\sum_{n \in U_i} sim(i, n)} \tag{6}$$

Step10: Predicting the rating of user i to item j in accordance with rating predict formula of the step 7, and implementing the Top-N recommendation.

Experiment Result and Analysis

Experiment data set and Rating Standard.
The experiment calculates the recommendation quality by the formula (7) [6].

$$MAE = \frac{\sum_{i=1}^{N} |P_i - Q_i|}{N} \tag{7}$$

The calculation results of formula (8) is inverse correlation with recommendation, the higher the result, the poorer the recommendation quality. MAE is highly correlated with number of user neighbor. According to the formula, the larger the N, the more number of user neighbor, the smaller

the MAE, the more obvious result. But the number of neighbor is larger than threshold, it will have an adverse impact on recommendation result. So amount selection of neighbor is very important. The number of data selection range from 30 to 40 in the experiment.

We take the web explorations record of family user as test dataset. In ordered to acquire the degree of user interest and construct the user interest matrix, the training set was taken as user interest matrix. In the experiment, we analyze and process the web log files of users in one quarter, which includes 1800 users browse records, 500 browse websites, 400 comments.

The Experiment Result Analysis.

Analyzing and calculating the experiment data with the improved the recommendation algorithm, and building the interest model, implementing the Top-N recommendation. We use two kinds of algorithms (the collaborative filtering recommendation algorithm based on user interest and the collaborative filtering recommendation algorithm) to simulate experiment, compared the number of user recommendation and satisfaction levels. The results of the comparison result are shown in Fig.2.
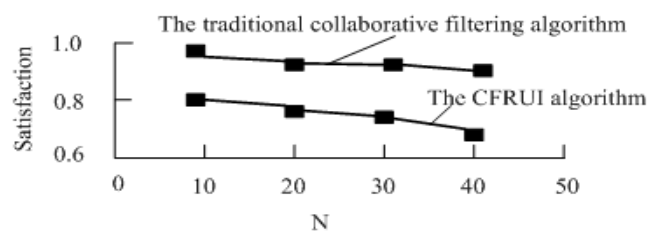

Fig. 2 The experiment results of algorithm

Figure 2 shows the MAE of the collaborative filtering recommendation algorithm based on user interest is much smaller than the collaborative filtering recommendation algorithm. Therefore, based on the above, one can draw a conclusion that the improved recommendation algorithm improves the accuracy rate of recommendation.

Summary

Aiming at the difficulty of fact that the traditional collaborative filtering technology can't reflect the user interest and solve problem of the data sparsity, this paper proposes the collaborative filtering recommendation algorithm based on user interest (CFRUI). The CFRUI algorithm introduces user interest weights to establish user-item rating matrix, implements similarity calculation by clustering analysis, so as to get the final recommended result. The experiment results show the CFRUI algorithm can improve the accuracy of recommendation system, and solve the problem of data sparsity.

References

[1]. XU H, PENG L, GUO A D. Study on Improved Strategy of Collaborative Filtering Based on User Multi-interest. Computer Technology and Development, Vol. 4(2011) No.21: 75-78.

[2]. PENG Y. Multiple Contents-Items Collaborative Filtering recommendation Based on personal characteristics of User. Chongqing, 2007, 12--126.

[3]. Sarwar B, Karypis G, Konstan, et al. Item-based collaborative filtering recommendation algorithm[C].//Proceedings of the tenth International World Wide Web Conference, [s.l.]: Springer, 2001:285-295.

[4]. YAN D M, NU C H. The Optimal Collaborative Filtering Recommendation Based on User Interest and Character. Application Research of Computers, Vol. 2(2012) No.29: 497-499.

[5]. LIU X D, CHEN D R, WANG H M. An Improved Collaborative Filtering Recommendation Algorithm. Journal of Wuhan University of Technology: Information and Management Engineering Edition. Vol, 32(2010) No. 4: 31-33.

[6]. ZHANG B Q. A Collaborative Filtering Recommendation Algorithm Based on Domain Knowledge. Computer Engineering. Vol. 21(2005) No 3:7-10.