

## Research on Recommendation Model Optimization

Xiaofeng Li

Department of Information Science,  
Heilongjiang International University,  
Harbin 150025, China  
mberse@126.com

**Abstract.** The accuracy of the traditional online recommendation system, much depends on the collaborative filtering recommendation algorithm, however, recommend system aims to attract the interest of consumers and turn visitors into buyers, rather than accurately predict their score. Online recommendation system is the service version of social filtering process. Most previous studies emphasize the accuracy of the collaborative filtering algorithm. However, the effective recommendation system must be credible. It requires that the system logic be transparency and the system be able to provide consumers a new, inexperienced item.

**Keywords:** Recommend model, Diversity, Predictive scoring algorithm, Collaborative filtering recommendation algorithm

### 1 Introduction

With tremendous information provided by Internet, the efficiency and quality of information used by e-commerce system users are greatly declined because of information overload and disorientation. Recommendation systems can assist users in making choices from various alternative approaches [1-2]. The goal of such systems is to predict users' preference and provide prediction about appropriate information. Individualized recommendation technologies perform remarkably excellently in electronic commerce field. They can recommend according to consumers' interest and hobbies commodities which are consistent with their favors, substituting and surpassing the role of recommendation by traditional on-site stores' sales persons. The personalized suggestions are based on interaction or preferences, which thus offers suppliers a big opportunity [3-4].

Ganesh [5] pointed out that online selling has become retailers' important strategy. After talking with online retailers, we found more than a half of them have planned or will plan to carry out recommendation function on their websites. Recommendation systems include artificial recommendation and computer recommendation, with the former including cross-selling finished by manpower [6-7]. The automatic recommendation system finished by computers is collaborative

filtering, which utilizes data mining techniques to optimize interactive information with consumers for the purpose of personalized e-business. Automatic recommendation stresses real-time and machine learning algorithms rather than fixed data model [8-9]. The level of consumers' acceptance to recommendation systems has direct impacts on the effect of them. Being affected by many factors like personalization, source of recommended information and recommendation time, such systems play a dual role in consumers' decision-making process.

## 2 Recommendation model improvement based on freshness measurement

Express rating data with a two dimensional scoring matrix. Let  $U = \{u_1, u_2, \dots, u_m\}$  the collection of m users and the collection of n items is  $I = \{i_1, i_2, \dots, i_n\}$ . Users' rating data about one item is matrix R at  $m \times n$  dimension. Every item  $r_{u,i}$  in matrix R represents the score of item i by user u. With evaluation matrix, we can compute the similarity between any two users and any two items. Then, we fetch from user's scoring files a certain number of training sets and test sets for the related experiment.

### User similarity based on Pearson measurement

Pearson similarity is the commonest measuring method based on users. In the scoring matrix R of user against items, user x's rating for item i is  $r_{x,i}$ . The average score for user x and y is respectively  $\bar{r}_x$  and  $\bar{r}_y$ . Then the similarity of user x and y based on Pearson similarity can be expressed as (1).

$$sim(x, y) = \frac{\sum_{i \in I_{x,y}} (r_{x,i} - \bar{r}_x)(r_{y,i} - \bar{r}_y)}{\sqrt{\sum_{i \in I_{x,y}} (r_{x,i} - \bar{r}_x)^2 (r_{y,i} - \bar{r}_y)^2}} \quad (1)$$

### User similarity based on cosine method

The scoring data is expressed with n-dimensional vector. The vector value of those without evaluation value is 0. User data is expressed with m-dimensional vector. Through computation of cosine values between different angles, we can get user's similarity. The score of user x and y in n-dimensional space is vector x and vector y. Hence the relative similarity equation based on angle cosine method is shown in (2).

$$sim(x, y) = \cos(\bar{x}, \bar{y}) = \frac{\bar{x} \bullet \bar{y}}{\|\bar{x}\|_2 * \|\bar{y}\|_2} \quad (2)$$

Where, numerator equals to the inner product of two users' scoring vectors, while denominator is the product of two users' vector modules.

### User similarity based on amended angle cosine method

Instead of using traditional cosine methods which don't consider the problem with scoring magnitude of different users, lots of researchers applied the amended cosine method, based on which, the above problem is solved through deduction of mean scores. The user similarity based on the amended cosine method is like (3).

$$sim(x, y) = \frac{\sum_{i \in I_{x,y}} (r_{x,i} - \bar{r}_x)(r_{y,i} - \bar{r}_y)}{\sqrt{\sum_{i \in I_{x,y}} (r_{x,i} - \bar{r}_x)^2 \sum_{i \in I_{x,y}} (r_{y,i} - \bar{r}_y)^2}} \quad (3)$$

Set matrix S the similarity matrix between  $n \times n$  dimensional users. Every  $S_{x,y}$  in the matrix refers to the similarity between user x and y. The above is the estimation of similarity generated based on users' historic rating records. If the number of products being definitely rated by a user is quite less, then the similarity got by that method is not accurate, as it's calculated based on the itemset which are appraised by both user x and y. If the number of users increases, the computation amount grows linearly and performance gets worse.

## 3 Experiment Design and Discussion

### 3.1 Classification of movie items in the dataset

According to features of movie items in the set, we classify movies into 19 types (Table 1).

Based on item type, the item class matrix G is generated at  $1683 \times 19$  dimension. When i item belongs to class  $g_i$ ,  $G_{i,j} = 1$ ; or,  $G_{i,j} = 0$ . Now we make clustering analysis for the item. Some items' class matrix is listed in Table 3.

For extreme sparseness of rating matrix, item class matrix needs to rely on simple categorization of item classes for higher clustering quality.

### 3.2 Similarity calculation method

For the clustering dataset, we need calculate the similarity between users and that between items. As for different datasets, the different results exist between item-based collaborative filtering similarity computation and user-based collaborative filtering similarity computation. We compared the effect of respectively user similarity method based on Pearson measuring, cosine similarity method and the modified cosine

method. Learning that different similarity results could affect the accuracy of recommendation, we got the mean value of MAE after cross tests for 5 iterations for the chosen dataset. The modified cosine similarity method based on item performs the best in terms of recommendation accuracy. Therefore it is chosen for the comparison test in the next step. In the experiment, because it's required to pre-determine the value of initial clustering quantity Kcluster, the preset value of Kcluster has big influences on test results. From recommendation results got with different values of Kcluster, the recommendation accuracy is not consistent due to different K values. Here we chose Kcluster's best value (K=15) for the following control experiment.

**Table 1.** Movie classification

Item ID	Movie classification
0	unknown
1	Action
2	Adventure
3	Animation
4	Children's
5	Comedy
6	Crime
7	Documentary
8	Drama
9	Fantasy
10	Film-Noir
11	Horror
12	Musical
13	Mystery
14	Romance
15	Sci-Fi
16	Thriller
17	War
18	Western

## 4 Conclusion

It outlined the progress that collaborative filtering recommendation method made from the mere method to user experience. Then it proposed the recommendation model based on variety. The recommendation quality was enhanced by adjusting the percentage of long-tail items through augmentation of freshness parameter when the candidate recommendation item set was calculated. The simulation experiment was designed to evaluate the diversity-based recommendation algorithm. The nearest neighbor quantity, clustering quantity and parameter value were appropriately

selected. The result confirmed that the proposed method is superior in diversity and coverage rate while keeping certain accuracy rate.

## References

1. Li Wenhai, Xu Shu. The design and implementation of e-commerce recommendation system based on Hadoop. *Computer engineering and design*, **2014**, 01:130-136.
2. Wang Binbin, Zhou Zuojian, Jie, Pan Jingui. Iterative training Web Service hybrid collaborative filtering recommendation model based on computer. *Research and development*, **2013**, S2:153-162.
3. Hu Jiming, a Mandi. On the socialization of minor user resource association recommendation model based on. *Information theory and practice*, **2014**, 04:123-126.
4. Wang Zhisheng, Li Qi, Su Weijie. Large scale personalized recommendation based on implicit feedback. *Journal of software*, **2014**, 09:1953-1966.
5. Ganesh J, Reynolds K E, Lockett M, et al. Online shopper motivations, and e-store attributes: an examination of online patronage behavior and shopper typologies. *Journal of Retailing*, **2010**, 86(1): 106-115.
6. Yang Xingyao, in Turgen Ibra sound, Jiong, Qian Yurong, Sun Hua. Consider the collaborative filtering recommendation model item property. *Computer applications*, **2013**, 11:3062-3066.
7. Ju Chunhua, Bao Fuguang. The situation and the main features of multi dimension of research into personalized recommendation model based on. *Journal of communication*, **2012**, S1:17-27.
8. Guo Yingguang, Chen Kai buy. Recommendation model current news concerned products. *Computer applications and software*, **2009**, 10:162-165.
9. Ji Lianghao, Wang state. Research on Collaborative Filtering Recommendation Based on user information. *Computer engineering and design*, **2008**, 08:2047-2051.