

Ceramic Products Recommendation Based Improved Collaborative Filtering

Huiying Li Yilai Zhang Gongxun Cheng

School of Information Engineering

Jingdezhen Ceramic Institute

Jingdezhen, China

e-mail: lhyjdzty@126.com

Abstract—In order to achieve personalized recommendation in the e-commerce platform for ceramic products, and improve the quality of recommendation, an improved collaborative filtering recommendation algorithm based on collaborative filtering algorithm was presented. By analyzing the insufficient of traditional collaborative filtering recommendation algorithm, combined with the characteristics of ceramic products in similar neighbor selection, the user was classified, and classification results as a basis for similar neighbors selected to optimize the choice of similar neighbors. Finally, experiments verify the improved algorithm, experimental results show that the improved method was optimized.

Keywords- Ceramic Products; User Classifying; Collaborative Filtering

I. INTRODUCTION

With the rapid development of information technology, e-commerce platform is developing rapidly. The current e-commerce platform has gathered a lot of data, there has been the phenomenon of information overload. When consumers login the platform for consumption, and get a lot of confusion. However, along with the rapid growth of the number of the products and users, China Ceramic Mall for ceramic products also encounters this kind of problem (www.ccmall.cn). Therefore, in order to provide users with high-quality service in a China Ceramic Mall, combination of product features and user characteristics of the mall, improved collaborative Filtering algorithm for product recommendation is studied.

The research of recommended method based Internet has got a rapid development since the last century 90 years, and the two main types of the method were formatted, namely: the recommendation of content-based and collaborative-based. Since these two types algorithm have some shortcomings, including data sparsity, cold start and new items, thus affecting the recommended speed and quality. To solve these problems, many researchers integrated into the recommendation method of association rules, semantic analysis, clustering, statistical analysis and so on. These studies can be combined with the recommended object characteristics and recommended method of application areas to improve the quality of the recommended methods.

Collaborative filtering is the most successful recommendation technology, the traditional collaborative filtering is called user-based collaborative recommended, the basic idea is: The rated of target users not rated projects by

similar users to predict the ratings of these items [1]. This recommended method exists the problem of poor scalability, the recommended efficiency is significantly reduced with the increase in the number of users and the number of projects. To solve this problem, researchers used clustering to narrow the search range of similar users [2], or by singular value decomposition to reduce the number of dimensions of the project space [3]. Item-based collaborative filtering is a fundamental solution to problems caused by similar users real-time computing time complexity with the growth index rose [4]. The basic idea of item-based collaborative filtering: The item rated can be predicted by the similarity between the item and the user has rated. The similarity between items is relatively stable, can be updated once in a long time interval, and significantly improved the scalability of the system but item-based collaborative filtering prediction result has not been significantly improved.

The user preferences clustering based on the characteristics of ceramic products and collaborative filtering algorithm is combined and improved the user neighbors selection based collaborative filtering is presented and the data sparsity problem is improved. Experimental results show that the improved algorithm can improve the recommendation quality.

II. CERAMIC PRODUCTS RECOMMENDATION BASED IMPROVED COLLABORATIVE FILTERING

Collaborative filtering algorithm consists of four steps, namely: similarity computation, similar neighbors' selection, missing value prediction and product recommendation.

A. Similarity Computation

The similarity has the similarity of the two types of users and items. The similarity computation mainly has the method of cosine similarity and Pearson correlation coefficient, the Pearson correlation coefficient method is selected to calculate the similarity.

Given a ceramic products recommender system have the training set and scored ceramic products set, there are U training users in the training set, and there are P ceramic products, the relationship between U training users and P scored ceramic products is denoted by $UM=U \times P$ matrix, $v_{u,i}$ represents scored from the user u on the ceramic products in this scored matrix UM . If user u didn't buy ceramic products i , or ceramic products i did consumed but user u didn't score, then $v_{u,i}=0$.

Based on the assumption and PCC method, user-based similarity calculation method is as follows [5]:

$$sim(u_a, u) = \frac{\sum_{i \in I} (v_{u_a, i} - \bar{v}_{u_a})(v_{u, i} - \bar{v}_u)}{\sqrt{\sum_{i \in I} (v_{u_a, i} - \bar{v}_{u_a})^2} \sqrt{\sum_{i \in I} (v_{u, i} - \bar{v}_u)^2}} \quad (1)$$

Where user u_a and u represent active user in the China Ceramic Mall and user in the matrix UM , respectively. $I = I_{u_a} \cap I_u$ is the subset of ceramic products which both active user u_a and user u have scored previously, $v_{u_a, i}$ is the scored of active user u_a on the ceramic product i , $v_{u, i}$ is the scored of the user u on the ceramic products i in the matrix UM , \bar{v}_{u_a} and \bar{v}_u represent average scored of different ceramic products observed by user u_a and u , respectively. When two users have null ceramic products intersection ($I = NULL$), then the value of $sim(u_a, u) = 0$, namely we do not determined between users similarity.

Item-based similarity computation is to calculate similarity between ceramic products i scored by active user u_a and ceramic products j scored by other user in the matrix, the formula is as follows [5]:

$$sim(i, j) = \frac{\sum_{u \in U} (v_{u, i} - \bar{v}_i)(v_{u, j} - \bar{v}_j)}{\sqrt{\sum_{u \in U} (v_{u, i} - \bar{v}_i)^2} \sqrt{\sum_{u \in U} (v_{u, j} - \bar{v}_j)^2}} \quad (2)$$

Where $U = U_i \cap U_j$ is the subset of users who have invoked both ceramic products i and j previously, $v_{u, i}$ and $v_{u, j}$ is scored values of ceramic products i and j scored by user u , and \bar{v}_i and \bar{v}_j represents the average values of the ceramic products i and j scored by different users. When $U = NULL$, then $sim(i, j) = 0$, namely we do not determined between cloud service item similarity.

B. Similar Neighbors Selection and Optimization

The similar neighbors' selection classic method of user and item is Top-K. In the application of the method to select similar neighbors, the similarity value of the selected users or items are considered, and the top k most similar users or items are selected for the similarity computation according to the similar value sorting, then the problem that some users or items have a few neighbors is ignored, where the show is data sparsity problem of the recommended method[6-7]. To overcome this problem, the introduction of the user clustering participation Similar neighbor selection. That is, when the user by the clustering method results is in the same category, regardless of similar value size, the user is regarded as a member of the K similar neighbors. Namely the selection neighbors users, the formula is as follows:

$$S(u) = \begin{cases} \{ua \mid ua \in T(u), \text{Top-K}\} \\ \{ua \mid ua \in T(u), (ua, u) \subset R_k\} \end{cases} \quad (3)$$

The selection neighbors ceramic products based on the formula as shown:

$$S(i) = \begin{cases} \{ik \mid ik \in T(i), \text{Top-K}\} \\ \{ik \mid ik \in T(i), (ik, i) \subset I_k\} \end{cases} \quad (4)$$

Where the formula 3 and 4, the selection similarity neighbors users, when the user and the predicted user is a category the R_k , the users similarity neighbors selection cannot affected by the size of the between user similarity values, and ceramic products similarity neighbors selection, when ceramic products and the predicted item is a category the I_k , the selection similarity neighbors items cannot affected by the size the between item similarity values.

Where the application method of user clustering is KMeans algorithm. According to the characteristics of the ceramic products, the user clustering R_k based product attributes: product shape, product design and color, product processing, the texture of the product, product usage and product packaging. Products classification I_k is based on product usage.

C. Rated Prediction for Ceramic Products

On the basis of the above on collaborative filtering based on user and ceramic products similarity computation method and improved the Top-K neighbors selection method, user-based CF apply similar users to predict the unrated ceramic products values by the following equation[4]:

$$UP(v_{u_a, i}) = \bar{u_a} + \frac{\sum_{u \in S(u)} sim(u, u_a)(v_{u, i} - \bar{v}_u)}{\sum_{u \in S(u)} sim(u, u_a)} \quad (5)$$

In the formula 5, $UP(v_{u_a, i})$ is predicted the missing value $v_{u_a, i}$ in the UM matrix. $\bar{u_a}$ is average values of different ceramic products rated by the active user u_a , and \bar{v}_u is average values of different ceramic products rated by the similar neighbors user u .

Item-based collaborative filtering methods apply similar items to predict the unrated ceramic products values by the following equation [4]:

$$IP(v_{u_a, i}) = \bar{ik} + \frac{\sum_{i \in S(i)} sim(i, ik)(v_{u, i} - \bar{i})}{\sum_{i \in S(i)} sim(i, ik)} \quad (6)$$

In the formula 6, $IP(v_{u_a, i})$ is predicted the missing values $v_{u_a, i}$ in the UM matrix. \bar{ik} and \bar{i} is average values of rated ceramic products I and j by different users, respectively.

In the rating of prediction unrated ceramic products, when the similar neighbors and similar neighbors are not empty, the big similar value is selected to predict unrated ceramic products. If one is empty, not empty similar neighbor is selected to predict unrated ceramic products, when both is empty, the rating of unrated ceramic products is

not predicted, but the average rating of ceramic product rated by different users are regarded as prediction results.

III. CERAMIC PRODUCT RECOMMENDATION

According to prediction results by improved collaborative filtering, the rating of ceramic products is sorted. The recommendation system of China Ceramic Mall push the user preference product based on the sorting results.

The preferences of prediction active users and recommended that may be of interest to the ceramic products, which can not only improve the user experience, but also improve user loyalty in the China Ceramic Mall.

IV. EXPERIMENTS EVALUATION

A. Experiments data set

300 users on the score of 150 ceramic products rated are selected as the training set in the China Ceramic Mall (www.ccmall.cn).Data set in the process of experiment is divided into two parts, namely: part of the training set as a prediction and the other part as a prediction of user data set.

B. Prediction results evaluation method

In order to analysis prediction results and measure the prediction quality of improved method in comparison with the tradition collaborative filtering, Mean Absolute Error (MAE) method is employed. MAE formula as follows:

$$MAE = \frac{\sum_{i,j} |v_{i,j} - pv_{i,j}|}{N} \quad (7)$$

Where the formula 7, $pv_{i,j}$ is the predicted value of the user i for ceramic products j , $v_{i,j}$ is the actual score value of the user i for ceramic products j , N is the number of predicted values.

C. Analysis of experimental results

In order to test improved collaborative filtering algorithm,250 users for the training set randomly selected, the set density is given 10%,20%, and 30%,respectively,the rated products number of active user is given 20,40 ,and 60,respectively.The Top-K value is given 10.the 100 times experiment are done based on 9 kinds setting. MAE is taken the average of 100 times as the experimental results. Finally, the experimental results of improved and tradition collaborative filtering are compared, the results are shown in the table below:

TABLE I. MAE COMPARISON WITH TRADITION COLLABORATIVE FILTERING (A SMALLER MAE VALUE MEANS A BETTER PERFORMANCE)

Evaluation Method	Training Set Density	Algorithm	Training Users (250)		
			20	40	60
MAE	10%	CF	186	170	153
		ICF	184	168	151
	20%	CF	183	172	148
		ICF	181	170	145
	30%	CF	175	168	148
		ICF	170	159	143

According to MAE results in the Table 1, we can learn: With the increase of the number of initialization active user rating, the value of MAE is gradually reduced. With the increase of the density of initialization rating in the training set, the value of MAE is also reduced. This declare that the more the items are rated, the higher the accuracy are predicted, and the more items of active user in the training set can also get a more accurate prediction results. The results of experiment can basically simulate the actual situation. And the comparisons with results of two algorithms also show that improved collaborative filtering algorithm is better than the tradition collaborative filtering.

V. CONCLUSION

An improved recommendation algorithm for e-commerce platform of ceramic products is presented. Compared with the tradition collaborative filtering, experiment results show the effectiveness of the method and the quality of recommendation is improved. We will further explore the factors of affecting the quality of recommendation, such as the relationship between the products and associated group, friendships and membership, the user content and so on. With the explosion of SNS (Social Network Service) size, we will also focus on scalability and effectiveness of the algorithm.

ACKNOWLEDGMENT

This work was financially supported by the National Science-technology support Program, China (2012BAH25F02, 2013BAF02B01), the Province Natural Science Foundation, Jiangxi, China (2009GZS0065).

REFERENCES

- [1] Joseph A. Konstan, Bradley N. Miller, David Maltz, Jonathan L. Herlocker, Lee R.Gordon, and John Riedl, "GroupLens: applying collaborative filtering to usenet news," Communications of the ACM, vol. 40, Mar. 1997, pp. 77-87, doi:10.1145.
- [2] Al Mamunur Rashid, Shyong K. Lam, Adam LaPitz, George Karypis, and John Riedl, "Towards a Scalable kNN CF Algorithm: Exploring Effective Applications of Clustering," In: Advances in Web Mining and Web Usage Analysis (WebKDD 2006), Springer-Verlag Berlin Heidelberg, 2007, pp. 147-166, doi: 10.1007.
- [3] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl, "Incremental singular value decomposition algorithms for highly scalable recommender systems," Fifth International Conference on Computer and Information Technology, IEEE Press, 2002, pp. 399-404, doi: 10.1.1.3.7894.
- [4] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl, "Item-based collaborative filtering recommendation algorithms," Proceeding WWW '01 Proceedings of the 10th international conference on World Wide Web, ACM Press, 2001, pp. 285-295, doi: 10.1145.
- [5] Zhang Fuzhi, and Zhang Qifeng, "An Improved Collaborative Filtering Algorithm Based on Manifold Alignments," Pattern Recognition & Artificial Intelligence, vol. 22, April. 2009, pp. 614-618.
- [6] Zibin Zheng, Hao Ma, Michael R. Lyu, and Irwin King, "Qos-aware web service recommendation by collaborative filtering," IEEE TRANSACTIONS ON SERVICES COMPUTING, vol. 4, April. 2011, pp. 1-12, doi: 10.1109.

- [7] Cheng Gongxun, Liu Lilan, Lei Huahuan, Lin Zhiqi, Li Manping, and Yu Tao, "QoS-prediction cloud service recommendation by collaborative filtering in cloud manufacturing platform," *Advances in Information Sciences and Service Sciences*, vol. 5, Mar. 2013, pp. 295-304, doi: 10.4156.