

Collaborative Filtering Algorithm Based on the Similarity of User Ratings and Item Attributes

Aili Liu^{1, a}, Baoan Li^{1, b*}

¹Computer School, Beijing Information Science and Technology University, Beijing, China

^abjliuaili@163.com, ^bliba2010@139.com

*Corresponding author

Keywords: Collaborative Filtering; Personalized Recommendation; Data Sparsity; Item Attributes; Similarity

Abstract. Collaborative filtering recommendation algorithm is key technologies of personalized recommendation system, as the serious data sparsity of rated items, the traditional collaborative filtering algorithms only depending on users data cannot achieve satisfactory recommended quality, an improved collaborative filtering recommendation algorithm based on the similarity of user ratings and item attributes is proposed. The experimental results based on Movie Lens dataset show that the improved hybrid collaborative filtering recommendation algorithm obtains the better recommendation accuracy than traditional similarity calculation method.

Introduction

With the rapid development of the Internet and e-commerce systems, the problem of the overload information has cropped up. In order to solve the current problem of overload information on the Internet, many large e-commerce sites use the various forms of recommendation system in varying degrees [1]. The common recommendation algorithm contains: the recommendation based on content, collaborative filtering recommendation, the recommended based on knowledge, the recommendation based on network structure and the mixed recommend [2]. Collaborative filtering (CF) is considered one of the most successful and widely used personalized technologies.

Its core idea is based on the user rating to the item to find the similarity between users, to find the most similar neighbor user with the target user, and then use the score of the neighbor to the item to predict the preferences of target user to the item, and then find the produce item [3]. However, the traditional collaborative filtering technology has the problems such as sparse data, cold start and the ability of extension and other issues affecting the quality of recommendation.

For the recommended quality problems exist in the traditional user-based collaborative filtering algorithm, the researchers analysis it by the user and item information, from another point of view to reduce problems caused by spare data, the main methods contain multidimensional reduction method, the item-based collaborative filtering and intelligent Agent methods. Sarwar, who use the singular value decomposition (SVD) method to reduce the dimension of user-item rating matrix to obtain a relative dense data [4], in order to solve the sparsity problem of collaborative filtering, by using *concept hierarchy* and *community filtering* thought, new consumer profile is built, and the connotative relation between items is found [5]. It alleviates the sparseness of consumers' rating data, helps users find right neighbors and get better recommendations.

The above method can ease of data sparseness problem from different angles, but did not accurately get the user's interest, and so cannot be expected to recommend. Therefore, we propose an improved collaborative filtering recommendation algorithm based on the similarity of user ratings and item attributes.

The Adjusted Collaborative Filtering Algorithm

Item attributes similarity

As a result of the traditional collaborative filtering techniques only consider the user - item rating matrix data, it is easy to affect by sparse data, eventually lead to recommend the result is not accurate [6]. Therefore, this literature considering the influence of the item attributes to recommendation.

The information of item attributes is generally represented by a vector. For example, the attribute information about an item can be expressed as $P = \{P1, P2, \dots, Pk\}$, h_{ij} indicates whether the item has the property p , if there was 1, and 0 otherwise[7]. One item can have a variety of properties. It was showed as Table 1.

	A1	A2	A3	A4	A5
I1	1	1	0	0	0
I2	1	1	0	0	0
I3	1	1	1	0	0

Table 1. The matrix of item attributes

Calculated by the previous item attributes similarity formula finds I1 and I2 and I3 have the same attribute similarity, but, in fact, the attribute similarity of I1 and I2 is higher than I3 and I1. Therefore, this study considers the item similarity is affected by the uncommon attributes. Get the Eq. 1 of attribute similarity between the two items:

$$\text{sim}_c(I_a, I_b) = \frac{N^{I_a \cap I_b}}{N - N^{I_a \cup I_b}} \quad (1)$$

Therein Eq. 1, N represents all the number of item attributes, $N^{I_a \cap I_b}$ represents the number of attributes belong I_a, I_b together, $N^{I_a \cup I_b}$ represents the number of attributes neither belong to I_a , or I_b .

Hybrid Collaborative Filtering Algorithm

In order to combine the similarity with Rating and item attributes, the paper introduces the weighting factor w to control user ratings and the item attributes. The range of w is $[0, 1]$. In this study, the similarity of the score measured by Correlation of Pearson. Set U as users set of I_a and I_b scored together, then Get the Eq. 2 of attribute similarity between the two items:

$$\text{Sim}_R(I_a, I_b) = \frac{\sum_{c \in U} (R_{u, I_a} - \bar{R}_{I_a})(R_{u, I_b} - \bar{R}_{I_b})}{\sqrt{\sum_{c \in I_{ij}} (R_{u, I_a} - \bar{R}_{I_a})^2} \sqrt{\sum_{c \in I_{ij}} (R_{u, I_b} - \bar{R}_{I_b})^2}} \quad (2)$$

Therein, R_{u, I_a} is the score of user u to item I_a , R_{u, I_b} is the score of user u to item I_b . Then, the similarity between the two items can be expressed by the Eq. 3.

$$\text{Sim}(I_a, I_b) = w \times \text{sim}_c(I_a, I_b) + (1 - w) \times \text{Sim}_R(I_a, I_b) \quad (3)$$

Then set the highest similarity items as the neighbor sets M of the target item I_a . according to neighbors can predict the score which user u to ungraded item I_a are as the Eq. 4.

$$R_{u, I_a} = \bar{R}_{I_a} + \frac{\sum_{v \in M} \text{Sim}(I_a, D) \times (R_{v, I} - \bar{R}_I)}{\sum_{v \in M} |\text{Sim}(I_a, D)|} \quad (4)$$

Results and Discussion

Experimental Data

The data were provided by the United States Group Lens Team Movie Lens dataset, Movie Lens is an opportunity for study of Web recommendation system, for receiving user ratings of the film and

provide the recommendation list. Data set contains 943 users to 1682 films to make a total of 100000 scores, each user at least 20 films for the score, ratings range from 1 to 5. Each film has its own respective categories, used as the film's property to obtain project attribute similarity. Sparse level of the data set is: $1 - 100000 / (943 \times 1682) = 0.93695$, from that particular sparse. We select 80% of the experimental data as the training set, 20% of the test set.

Evaluation Index

The evaluation criteria of the quality of recommended system include statistical accuracy metrics and decision support precision measurement of two ways [8]. Usually used in the former method of Mean Absolute Error (Mean Absolute Error, MAE) measurement. MAE is forecast by the target user ratings and target users to compare the actual score, to the degree of deviation between the two to represent the effect of the accuracy of recommendation. MAE smaller value indicates higher prediction accuracy, the better the quality recommended. MAE calculation formula can be expressed as Eq. 5.

$$MAE = |E| = \frac{\sum_{i=1}^n |p_{i,j} - r_{i,j}|}{n} \quad (5)$$

Where n is the number of projects the target user ratings, $r_{i,j}$ is the user's actual score, $p_{i,j}$ is the prediction score of recommended system.

Experimental Design and Results Analysis

In this experiment, Set the w value from 0 to 1, every increase of 0.1, then observe the changes in the value of MAE, thus recommended that the impact on system performance. The experimental results can be shown in Fig. 1, w is 0.9 recommended best [9].

In order to verify the proposed hybrid similarity score based on user and project properties collaborative filtering algorithm is effective [10], with the user-based collaborative filtering (user-based CF) and Item-based collaborative filtering (Item-based CF) as contrast, comparing the MAE value; the number of neighbor set to from 10 to 50, interval of 10, the weight w is set to 0.9. The result of the experiment can be shown in table2, comparison chart shown in Fig. 1.

the number of neighbors Set	MAE		
	User-based CF	Item-based CF	Adjusted CF
10	0.931	0.893	0.816
20	0.912	0.867	0.798
30	0.893	0.882	0.772
40	0.889	0.887	0.768
50	0.882	0.871	0.773

Table 2. The experimental results

As can be seen from the below of figure 3, for different number of neighbors, the improved hybrid collaborative filtering algorithms achieved the smallest value of MAE. This can instruct that the instruction of the information of item attribute, to some extent alleviated the problems caused by data sparseness, then pushing the recommended quality.

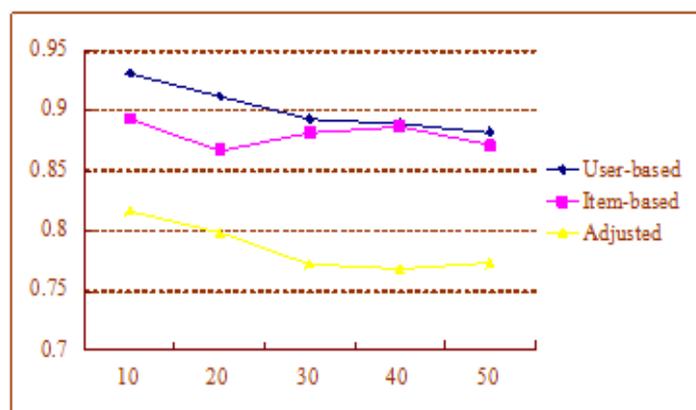


Fig. 1. The experimental contrast data

Conclusion

The rating of data sparsity cannot guarantee real-time and recommended data quality problems of collaborative filtering algorithm under circumstances [11]. This study introduces the Project properties, using the similarity score and the project attributes similarity weighted combination of approaches, to reduce the impact of the recommendation system of the sparsity of data, for improving the quality recommended and reducing the error. The experimental results show that the new algorithm can effectively improve the recommendation quality [12], and it can be extended in the recommended site.

Acknowledgement

The work was supported by Project of Construction of Innovative Teams and Teacher Career Development for Universities and Colleges under Beijing Municipality (Grant No. IDHT20130519), and also Project of Graduate Education Quality Engineering of Beijing Information Science and Technology University (Grant No. YJT201511).

References

- [1] Sarwar B, Karypis G, Konstan J, et al. Application of Dimensionality Reduction in Recommender System-A Case Study[C]. ACM Web KDD on Web Mining for E-commerce workshop (2000)
- [2] Xiaomin Zhang, Commerce personalized recommendation system technology and its application [D] Collaborative filtering, Chongqing University (2007)
- [3] Peiyong Xia. Research on personalized recommendation technology in collaborative filtering algorithm [D], China Ocean University (2011)
- [4] Sanchen S N et al, A feature mining based approach for the classification of text documents into disjoint classed information [J]. Information Processing and Management,38(4):583-604 (2002)
- [5] Adomavicius G, Tuzhilin A. Toward the next generation of recommender system: a survey of the state-of-the-art and possible extensions [J]. IEEE Trans on Knowledge and Data Engineering, 17(6):734-749 (2005)
- [6] Xiaosheng Ji, Yan Liu Bing, Laming Luo, Collaborative filtering method based on similarity measure of user interest [J], Journal of Computer Applications, 10: 2618-2620 (2010)
- [7] Yonglong Zhuang. Study filtration characteristics of the model-based collaborative project recommendation algorithm [D], Nanjing University of Technology (2008)

- [8] Zhimin Chen, Zhiqiang Li, Collaborative filtering algorithm based on user characteristics and item attributes [J], *Journal of Computer Applications*, 07: 1748-1750 (2011)
- [9] Tang, J., Zhang, Y., Sun J. et al., Quantitative study of individual emotional states in social networks, *IEEE Transactions on Affective Computing*, 3(2), pp: 132-144 (2012)
- [10] Li, B.A., Social Network Based on Context-aware and Component Technology, *Journal of Information & Computational Science*, 9(15), pp: 4279-4285 (2012)
- [11] DU, S.P., Masia B., Hu S. M., & Gutierrez D. A metric of visual comfort for stereoscopic motion, *Siggraph Asia 2013*, 32(6) (2013)
- [12] Gong, L., Liu, Y.J. and Guo, Z.J., Scenario Engine: The Core of the Next Generation Recommender System, *Communication of China Computer Federation*, 9(3), pp: 14-18 (2013)