

An Improved Graph-based Recommender System for Finding Novel Recommendations among Relevant Items

Ranran Liu^{1, a}, Zhengping Jin^{1, b}

¹State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, 100876, China;

^aranran_liu135@163.com, ^bzhpjin@bupt.edu.cn

Keywords: Recommender system, Novelty, Relevance, Social relationship.

Abstract. Recommender system has been extensively studied to provide the most relevant data to users in this era of information explosion. Among all kinds of recommendation algorithms, collaborative filtering (CF) algorithm is one of the most famous ones because of its high accuracy and simple implementation. Recently, scholars have proposed a new approach to find fresh and novel items, but the relevance of some novel items may be far from good which reduced system's precision accordingly. In this paper, we propose an improved recommender system to increase the relevance when finding out novel items. This approach is motivated by the fact that social relationships could reflect the similar interests between users in a recommender system. Thus, social relationship is taken into consideration when we build the profile graph of each user. We test the system on Last.fm data and the result shows that the improved graph-based recommender system could indeed provide fresh recommendations while the accuracy have increased by 0.7% on average at the same time.

Introduction

With the rapid development of the Internet, information overload has become one of the most pressing problems. People's requirements for personalized service stimulate the tremendous development of recommender systems (RS). Collaborative filtering (CF) [1, 2], as one of the most popular technique in RS, has been successfully employed in many applications, such as ecommerce, social networks and review sites. With the efforts of scholars around the world, CF's prediction precision steadily improved. Recent years, another increasingly crucial problem have aroused attention: how to recommend some novel items to users. This problem comes from the applications of RS that sometimes users don't want to see what they have already reviewed. According to CF algorithm, the high score items could be more accurate concerned about relevancy, they may be too popular that users are very likely to be familiar with them and feel the recommendation useless. This problem is a reflection of the famous long tail effect [3]. More and more studies realized that being accurate is far from enough [4] and turned their focus from precision to other factors, for example diversity and novelty. Scholars has proposed an ingenious RS called graph-based recommender system (GRS) to find out novel items from the accurate ones [6] using Shannon entropy [7] as the standard method. However, in these novel items there exist a degree of unexpected ones which are irrelevant recommendations. This is come from the risk of single use of entropy as the measure of novelty.

In this paper, we have studied GRS and came up with an improved graph-based recommender system for finding out novel recommendations from the relevant ones. We leverage social relationships to build user profile graph. Then we compute the entropy of items in the graph and choose the ones with higher entropy as the novel items. Finally we compose the novel recommendations with the ones produced by the classic CF algorithm as the final recommendations list.

The rest of the paper is organized as follows. Section 2 introduces PRS briefly and proposes the social relationship based PRS (SR-GRS). Section 3 demonstrates the data set and evaluations used

in this RS. Section 4 presents empirical results on one popular collaborative filtering data set and demonstrates the efficiency. Section 5 concludes the paper.

Social Relationship Weighted Graph-Based Recommender System (SR-GRS)

In this paper, we present a graph-based recommender system (GRS) which is based on Kibeom Lee and Kyogu Lee's proposal. As we all know that in a RS, social relationship always represents users' similar interests, so we improve GRS by leveraging social relationship to control the unpredictability of picking items based on entropy. In this section, we present a social relationship weighted graph-based recommendar system (SR-GRS) aimed to provide novel recommendations with higher precision as well. The algorithm can be divided into four main steps. Firstly, we convert the rating matrix into a adjacency matrix. Then social-based user profile graph will be established according to the adjacency matrix and social relationship. Thirdly, we calculate each node's entropy in the graph. Finally, the recommendation will be composed of two parts: novel ones and accurate ones. The details are as following.

Build adjacency matrix.

In order to rise the relevance of items, we use positively-related items to populate the adjacency matrix. These items are defined as the set of items in a user's profile that have ratings higher than the user's average rating. We assume that there are 5 items {1,2,3,4,5} and 4 users {A,B,C,D}. Each users' positively-related item set are as follows:

$$P(A) = \{1, 2\} \quad P(B) = \{2, 3\} \quad P(C) = \{1, 4, 5\} \quad P(D) = \{3, 4\}$$

With P(A), we can create the following weighted adjacency matrix (left) and processed weighted adjacency matrix M (right) after adding all users' matrix together:

$$\begin{array}{c} \begin{array}{ccccc} & 1 & 2 & 3 & 4 & 5 \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} & \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} & + & \begin{array}{ccccc} & 1 & 2 & 3 & 4 & 5 \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} & \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} & + \dots = & \begin{pmatrix} 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{pmatrix} \end{array} \end{array}$$

Fig.1. Process to get the matrix M

Establish social-based user profile graph.

In this step, we will establish a social based user profile graph. Firstly, we create a sub-matrix of M for a target user, for instance user A. By selecting not only the rows corresponding to items in P(A) but also rows related to user A's friends which reflect user's social relationship. We assume that user C and D are A's friends and the most-liked items of C and D is item 4. At the same time, user A hasn't rated item 4, which means that A may be not familiar with item 4 to a large extent and his friends are interested in it. Intuitively, we can conclude that A may be interested in item 4 and we have enough reason to adding item 4 into user profile graph. So we expend the user profile set as P'(A) = {1,2,4} The sub-matrix of M is shown in Fig.2.

$$\begin{array}{ccccc} & 1 & 2 & 3 & 4 & 5 \\ \begin{array}{c} 1 \\ 2 \\ 4 \end{array} & \begin{pmatrix} 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{pmatrix} \end{array}$$

Fig.2. User A's sub-matrix

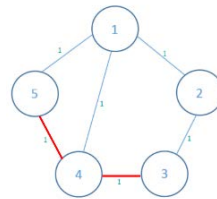


Fig.3. User's profile graph G

Secondly, based on user's sub-matrix, we establish user profile graph G shown in Fig.3. Nodes in G are the items related to the sub-matrix and the edges in G are (1,2) (1,4) (1,5) (2,3) (3,4) (4,5) and the weights of these edges are the corresponded elements in the sub-matrix, such as the weight

of edge (1,2) is 1. It should be noted that the two red edges shown in the picture are added when we take A's friends' favorite item 4 into consideration.

Compute entropy.

In this step, we calculate the Shannon entropy of each item in G with formula (1). Entropy is a measure of the uncertainty of the signal emitted by a source, and the formula is shown in Fig 1. In the formula, $H(X)$ is the entropy of a discrete random variable X with possible values $\{x_1, \dots, x_n\}$ and $P(x_i)$ is the probability of x_i 's occurrence. Here, we use the frequency to approximate the probability. In this approach, X is the item we want to calculate the entropy and x_i represents the item relevant to X and the frequency is the number of two items interested users simultaneously, which is exactly the weight of the edge in graph G corresponding two items.

In order to be fair that the number of values related to each item is different, we normalized entropy $H(X)$ by using $\ln z$ where z is the number of values related to item i .

$$H(X) = -\sum_{i=1}^n P(x_i) \log_b P(x_i) \quad (1)$$

Entropy is used to measure the uncertainty of a variable's value in the system, higher entropy represent higher uncertainty. Here in this approach we calculate each item's entropy of a certain user. As we all know that the relevance matrix of a RS is extremely sparse matrix, a large proportion of the elements are zero. In this case, if user's profile set just has a few items, a lot of entropies may be zero and we can be sure that the items user have already be interested in have non-zero entropies. Thus, this may lead to a bad result that the so called novel items are familiar to users.

Compose recommendations.

The last step is to give the final result which is relevant and novel at the same time. In order to filter out less relevant items in novel ones, we take use of tags users give to the items. Firstly, we find out the interests of every user through the most frequently used tags marked as T-set. Secondly, we sort the result by the entropies of items in descending order and use the T-set to filter the irrelevant items in the sorted recommendations. So far, we could be confident to a certain degree about the leftover novel recommendations belong to the regions or areas user is interested in which means the relevance of these items may be not too bad. Finally, we compose the novel recommendations with the relevant ones produced by classic CF algorithm as the final result.

Evaluation

In order to compare the performance of the proposed approach comprehensively, we take the PSG algorithm and traditional CF as the base-line algorithm and evaluate their precision and novelty by the use of evaluations below.

Items' average degree: The RS's novelty is evaluated by the use of items' average degree [12]. Recommendations with lower average degree are more novel for users because lower degree means less popular. Here is the average degree $N(L)$:

$$N(L) = \frac{1}{ML} \sum_u \sum_{\alpha \in R(u)} k_\alpha \quad (2)$$

Where M is the number of test users, L is the length of recommendation list, $R(u)$ is the set of user u 's recommendations, k_α is the degree of recommendation α . We get k_α by calculating the number of users who have listened to this recommendation or artist α in the training data set.

Precision: Regarded as the most intuitive way to present the accuracy of a RS, its definition is as follows:

$$Precision(L) = \frac{\sum_u |R(u) \cap S_{test}(u)|}{\sum_u |R(u)|} \quad (3)$$

Where $S_{\text{test}}(u)$ is the test data set of user u .

Results

The proposed RS is experimented on the data set collected from Last.fm. We compute the average playcount for each user and gain the items which have higher playcounts than the average. For each user, we divided the data into a training set which covers the first 80% of each user-listened period, and the remaining 20% as a testing set. In the SR-GRS algorithm, we set p as the number of items chosen from user's friends' favorite items, which is equal to 5, 10, 15 and 20, respectively.

In consideration of precision, the result in Fig.4 shows that the SR-GRS algorithm is inferior to the baseline algorithm CF and superior to GRS. We can observe that SR-GRS performed better than GRS which means SR-GRS could reduce the unexpectedness by the use of social relationship and increase the accuracy of recommender system. Here note that when the parameter p equals to 10, the precision is the highest of all. Besides, because the parameter p has nothing to do with the other two algorithms, the precision results vibrate slightly.

Table 1. Items' average degree

Parameter p	CF	GRS	SR-GRS
5	0.361	0.720	0.416
10	0.368	0.724	0.642
15	0.357	0.693	0.511
20	0.348	0.688	0.653

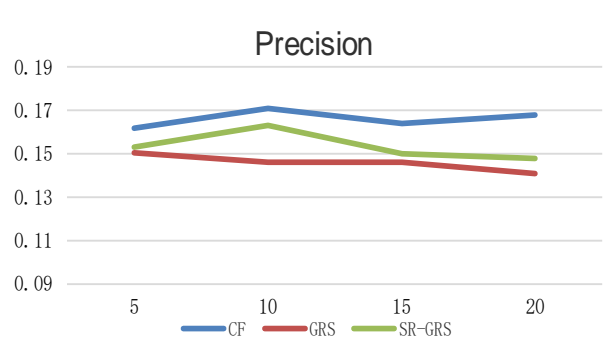


Fig.4 Precision of different algorithms

In Table 1, we calculate items' average degree of each algorithm to measure RS's novelty. The classic algorithm CF didn't perform well according to this assessment, and GRS is the best one. We noted that, SR-GRS has significant higher average degree than CF which is slight lower than GRS's. This demonstrates that SR-GRS has the ability to find out fresh items for users.

From above experiment, we can see that our proposed system has higher precision than GRS while items' average degree didn't reduce much and were still significantly higher than CF.

Conclusion

In this paper, we have proposed an improved graph-based recommender system (SR-GRS) by using social relationship. When considering both accuracy and novelty, we can draw a conclusion that the proposed algorithm has improved the baseline algorithm by overcoming the disadvantage of entropy partly when finding novel items. In other words, social relationship can help to filter some extremely irrelevant items recommended by high entropy. In short, our improved graph-based recommender system can find novel recommendations among relevant items. Nevertheless, the proposed system does come with its limitation because not each RS has social relationship between users. We can come up with other methods to improve the novelty while maintaining a relatively high accuracy, such as filtering irrelevant items by using tags and so on.

Acknowledgements

This work is supported by NSFC (Grant Nos. 61300181, 61502044), the Fundamental Research Funds for the Central Universities (Grant No. 2015RC23).

References

- [1] Adomavicius G, Tuzhilin A. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions, *IEEE Trans. Knowled. Data Eng.* 17 (6), 2005, p. 734–749.
- [2] Su X, Khoshgoftaar T M. A Survey of Collaborative Filtering Techniques[J]. *Advances in Artificial Intelligence*, 2009, 2009(2009). A survey of collaborative filtering techniques, *Adv. Artif. Intel.* 2009, 2009, p. 4:2–4:2.
- [3] Anderson, Chris. The long tail: Why the future of business is selling less of more. Hyperion. 2008
- [4] Mcnee S M, Riedl J, Konstan J A. Being accurate is not enough: How accuracy metrics have hurt recommender systems. In *CHI '06: CHI '06 Extended Abstracts on Human Factors in Computing Systems*. USA: ACM. New York, NY, 2006, p. 1097–1101.
- [5] Herlocker J L, Konstan J A, Terveen L G, et al. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, 2004, 22, p. 5–53.
- [6] Lee K, Lee K. Escaping your comfort zone: A graph-based recommender system for finding novel recommendations among relevant items. *Expert Systems with Applications* 42. 2015, p. 4851–4858
- [7] Shannon, C. E. A mathematical theory of communication. *ACM SIGMOBILE*. ACM New York, NY, USA, 2001, p. 3-55
- [8] Sun Z, Han L, Huang W, et al. Recommender systems based on social networks. *Journal of Systems and Software* , Volume 99, 2015, p. 109-119.
- [9] Sun Z, Han L, Huang W, et al. Recommender systems based on social networks. *Journal of Systems and Software*, 2015, p. 109–119
- [10] Domingues, M., Gouyon, F., Jorge, A., et al. Combining usage and content in an online recommendation system for music in the long tail. *International Journal of Multimedia Information Retrieval*, 2012, 2, p. 3–13.
- [11] Zhu Y X. Evaluation Metrics for Recommender Systems. *Journal of University of Electronic Science and Technology of China*. Vol.41(2012) No.2