Resolving Data Sparsity and Cold Start in Recommender Systems

Guibing Guo

School of Computer Engineering
Nanyang Technological University, Singapore
gguo1@e.ntu.edu.sg

1 Research Problems

Recommender systems (RSs) are heavily used in e-commerce to provide users with high quality, personalized recommendations from a large number of choices. Collaborative filtering (CF) is a widely used technique to generate recommendations [1]. The main research problems we desire to address are the two severe issues that original CF inherently suffers from:

- *Data sparsity* arises from the phenomenon that users in general rate only a limited number of items;
- *Cold start* refers to the difficulty in bootstrapping the RSs for new users or new items.

The principle of CF is to aggregate the ratings of like-minded users. However, the reported matrix of user-item ratings is usually very sparse (up to 99%) due to users' lack of knowledge or incentives to rate items. In addition, for the new users or new items, in general, they report or receive only a few or no ratings. Both issues will prevent the CF from providing effective recommendations, because users' preference is hard to extract. Although many algorithms have been proposed to date, these issues have not been well-addressed yet.

2 Progress to Date

Due to the popularity of social networks such as Facebook, more and more researchers turn to incorporate the social relationships (e.g. trust[1]) of users to help complement users' preference in addition to item ratings. However, trust has not been well utilized so far since the improved performance of trust-based approaches is marginal. Therefore, we would like to propose a new method that incorporates trust with CF to achieve better performance.

On the other hand, with the advent of virtual worlds such as Second Life, e-commerce in virtual reality (VR) is believed to have a promising future. Since the VR accommodates richer item (product) related information than the traditional online environment, a few works (e.g. [5]) attempt to apply traditional

[1] Trust reflects the extent to which users' opinions are valuable in decision makings.

J. Masthoff et al. (Eds.): UMAP 2012, LNCS 7379, pp. 361–364, 2012.

recommendation methods based on the additional information. However, none of them makes good use of the features of VR, such as that users can effectively interact with 3D virtual products and human beings. We believe that VR is more likely and meaningful to be a medium that embraces users' pro-active interactions rather than a mere information source of products. Thus we propose a way to take advantage of these features to enrich the user-item matrix itself.

2.1 Incorporating Trust with Collaborative Filtering

To begin with, we seek a trust-based solution to the concerned issues of CF. Trust relationships of users are often employed in order to correlate more potential raters for the active users who require recommendations, in addition to users' ratings on items. The active user can report his evaluations of trustworthiness towards other users, and each trusted user (neighbor) can also have their own trusted neighbors. Hence a web of trust (WOT) can be constructed. It is reported that trust is positively correlated with similarity [6]. Trust-aware recommender systems [4] attract researchers' attention because trust is able to propagate through the WOT. However, trust propagation is time-consuming and the best length of propagation can only be determined empirically.

Motivated by the previous usage of trust, we propose a simple but effective method to utilize trust to find more similar users whose ratings can be aggregated to generate recommendations [2]. More specifically, the ratings of trusted neighbors will be merged to represent the preference of the active user. Note that the active user is also regarded as a trusted neighbor of himself. Formally, the weighted average of trusted neighbors' ratings on a certain item i will be calculated as follows, used as the new rating for the active user u on this item.

$$\tilde{r}_{u,i} = \frac{\sum_{v \in TN_u} t_{u,v} r_{v,i}}{\sum_{v \in TN_u} t_{u,v}} \tag{1}$$

where $\tilde{r}_{u,i}$ is the merged rating, $t_{u,v}$ is the trust user u has toward user v, and TN_u denotes the set of u's trusted neighbors (including himself).

After that, similar users will be probed using the newly formed rating profile of the active user, and traditional CF is applied to generate recommendations. At the present, only explicit trust information specified by the users themselves is made use of because explicit trust is more reliable than implicit trust in general.

Our method is demonstrated to be effective using three real-life datasets. The results show that our method achieves the best performance in terms of both accuracy and coverage comparing with other benchmarks, especially in the case of cold-start users where the greatest improvement in performance is reached. We also analyze that theoretically, our method is able to function well even in two extreme cases: 1) users have only specified trusted neighbors but not rated any items; and 2) users have only rated items but not identified any trusted neighbors. It will fail to work if and only if both trust and rating information are out of reach. In that case, information other than trust or item ratings is required to model user preference. Furthermore, trust propagation does not provide significant benefits and hence is not necessary for our method, considering the cost and issues it brings. Overall, the data sparsity and cold start problems are largely alleviated.

2.2 Recommender Systems in Virtual Reality

At present we turn to the emerging VR environment in which even richer information is available compared to 2D online environments. For example, the actual products can be represented in the 3D models which enable users to interact with the virtual products, such as viewing from different angles, zooming in and out, touching the surface, customizing the components and even trying them on. It has been reported that the virtual product experience can help users judge the product value and make a better decision prior to purchase [3]. Other features such as the sense of presence, can also enhance user's evaluation of products.

Inspired by those features, we define two new concepts called *virtual rating* and *physical rating*. The former refers to the rating that is reported during or after users' interaction with 3D virtual products prior to purchase whereas the latter corresponds to the rating that is given after users' purchase and experience the actual products. Hence, virtual ratings are the outcomes of product evaluations according to users' virtual product experience. Inversely, physical ratings are the evaluative outcomes of products in the light of real product experience, which in fact are the user-item ratings in traditional RSs. We propose a new RS that makes use of both virtual and physical ratings for recommendations.

For virtual ratings, there could be two approaches for the proposed RS to collect data: manual and automatic. Firstly, the system could provide a user interface to accept users' manual inputs of ratings. Our ongoing research is mainly focused on this approach. Secondly, the system could automatically gather users' ratings according to their emotional responses. More specifically, the positive emotions are captured in the form of electroencephalogram (EEG) signals while users are interacting with 3D virtual products prior to purchase, with the help of a real-time convenient wireless EEG headset. The virtual ratings are then calculated from the averaged relative power of the collected EEG signals.

Recommender systems can benefit from introducing virtual ratings in at least two aspects. Firstly, the data sparsity is reduced because more user-item ratings (virtual and physical) are available. Users can form concrete opinions towards the products they interact with in VR, though they may not have the intention to purchase, which may depend on many other factors in addition to product value, such as consumption need or goal, income or salary, time constraints and so on. Users may interact with lots of products, but only purchase a small portion of them, e.g., users may experience different alternatives before choosing one of them. In most cases, users are not willing to express their opinions about the purchased products unless they are well motivated by good incentive mechanisms. In a word, physical ratings occupy only a small proportion of all products whereas virtual ratings may cover a wider range. The proposed RS will have a sound rating richness and reduce its reliance on incentive mechanisms.

Secondly, the cold start problem is largely alleviated. For users who have virtual and/or physical ratings, their preference can be extracted based on which similar users can be found and recommendations can be generated. Since the preference is more complete, the recommendations generated will be more accurate. Similarly, for the new products, as long as there are some virtual ratings (no need to be purchased) available, they are made possible for recommendations.

3 Future Research

For our trust-based recommender systems, majority strategy is another way to merge the ratings of trusted neighbors, which works well if the ratings are diverse. Thus we would like to investigate how the majority strategy can possibly improve our previous method, especially when the item receives many ratings.

The main future work is to design an effective CF that incorporates both virtual and physical ratings to make better recommendations. Basically, virtual ratings can be utilized at least two ways: 1) separately to form another user-item rating matrix to generate recommendations; or 2) transformed into physical ratings somehow and integrated into the original user-item rating matrix to fill in the missing values. In order to determine the best way to utilize virtual ratings, we need to understand them more deeply. In principle, there are two fundamental research concerns. The first is the method that users use to evaluate product values in VR and the factors that are taken into considerations. The other is the connections and distinctions between virtual and physical ratings. To validate the soundness and effectiveness of the proposed method, we may need to build up a prototype of virtual malls and recruit real users for ratings collection.

Since VR also supports users' interaction with other users in real time and face to face, another line of future research focuses on the social (trust) relationships among users. Although we have successfully employed explicit trust, it is not clear whether the method is applicable to VR. Furthermore, now that virtual ratings are introduced, it is worth investigating whether trust can be built upon virtual ratings. In that case, trust usage will be even more complex.

References

1. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions, pp. 734–749 (2005)
2. Guo, G., Zhang, J., Thalmann, D.: A Simple But Effective Method to Incorporate Trusted Neighbors in Recommender Systems. In: Masthoff, J., et al. (eds.) UMAP 2012. LNCS, vol. 7379. Springer, Heidelberg (2012)
3. Jiang, Z., Benbasat, I.: Virtual product experience: Effects of visual and functional control of products on perceived diagnosticity and flow in electronic shopping. Journal of Management Information Systems 21(3), 111–147 (2004)
4. Massa, P., Avesani, P.: Trust-aware recommender systems. In: Proceedings of the ACM Conference on Recommender Systems, pp. 17–24 (2007)
5. Shah, F., Bell, P., Sukthankar, G.: A destination recommendation system for virtual worlds. In: Proceedings of the Twenty-First International Florida Artificial Intelligence Research Society Conference (2010)
6. Ziegler, C., Golbeck, J.: Investigating interactions of trust and interest similarity. Decision Support Systems 43(2), 460–475 (2007)