Improving Recommendation based on Implicit Trust Relationships from Tags

Hyunwoo Kim and Hyoung-Joo Kim,

Seoul National University, Seoul, Korea
hwkim@idb.snu.ac.kr, hjk@snu.ac.kr

Abstract. In this paper, we proposed an implicit trust relationship extraction approach to alleviate the sparsity problem in recommender systems. The recommender system cannot generate relevant items when a user-item matrix is sparse. It is a serious weakness of collaborative filtering based recommender systems. In social tagging system, tagging information is useful data source for recommendation. We investigate eliciting implicit trust relationships from the tagging information. The relationships are derived by Kullback-Leibler divergence of users' tagged items and tags. The experimental results show that proposed approach provides relevant items precisely and performs well in practice.

Keywords: Recommendation, Trust relationships, Tagging information, Recommender Systems.

1 Introduction

An item recommendation is a task that recommends highly relevant items with a given user. The correct recommendation is increasingly important because of information overload. It is impossible for a user to search all items to discover interesting items which are matched with the user's preference because the number of existing items is too large.

Recommender systems have proposed the methods with content analysis of items and user's ratings of items. Collaborative filtering is one of the most successful methods for recommendation. A collaborative filtering method is based on a user similarity. Given a user, the system searches similar users based on the user similarity and recommends the items that the similar users are interested in. The system utilizes rating information to searches similar users of the given user. A user similarity is a subjective, personal, and symmetric relationship. If user u_a is similar to user u_b, it necessarily means that u_b is similar to u_a. A sparsity problem is a serious weakness of the collaborative filtering method [1]. A user cannot rate all items in the recommender system because the number of items is too large. A user-item matrix has high sparsity and no overlap between users [2]. It causes the difficulty of searching similar users.

Trust is "a subjective expectation an agent has about another's future behavior based on the history of their encounters" [3]. In recommender systems, some users may trust a certain user but others may not trust the user. It depends on the user's personal interest or preference. Analogous to the user similarity, trust is also a subjective and personal relationship. However, trust is an asymmetric relationship. If u_a trusts u_b, it does not necessarily mean that u_b trusts u_a. The asymmetry is a significant feature of trust relationships and it facilitates the propagation of trust relationships. A trust relationship has direction and it is inherently transitive. Trust relationships are also dynamic. They are gradually built up and keep changing over time. By utilizing trust information, recommender systems relieves sparsity problem [4]. Nevertheless, most of the recommender systems provide no means of representing explicit trust relationships between users.

Social tagging gives a new opportunity to researchers who study recommendation. In social tagging environment, a user annotates an item with relevant keywords, i.e. tags, about the item. Annotated tags form folksonomy. Tags provide additional information about the item than a rating because tags also reflect the user's preference. Users generate folksonomy with their tags. Many research studies on recommendation already have been taken advantage of these factors of social tagging [5].

In this paper, to improve the performance of recommender systems, we proposed a recommendation method using implicit trust relationships that are derived from tagged items and tags in social tagging system.

2 Related Work

Recommender systems with trust information are divided into two kinds of approaches. First, a recommender system utilizes explicit trust relationships that are provided by users. In this case, users explicitly represent their trust information in the recommender system. Second, a recommender system elicits implicit trust relationships from users' data. In this approach, users do not provide their trust information. Thus, the recommender system derives the trust relationships from the user's information, e.g., profile, rated items, ratings, and tags. Papagelis et al. proposed a trust inference approach to alleviate sparsity problem of collaborative filtering [6]. O'Donovan and Smyth also proposed trust-based recommender systems [7]. Bhuiyan et al. proposed a method to develop trust networks from user tagging information [1]. They developed trust network from tags but it requires descriptions of items. Their approach cannot be applied to recommender systems which do not provide the descriptions of items.

3 Recommendation Model

We proposed a method which recommends items to a user with the user's implicit trust relationships that are derived from users' tagging information. A user's tagging information consists of triples which include users, items, and tags. A triple *<user, item, tag>* is a basic building block of tagging information. If user u annotates tag t to item i, then the triple $<u, i, t>$ is stored in the recommender systems dataset. We utilize these triples to elicit implicit trust relationships. Naïve approach conducts extraction using conditional probability between two users. Our proposed approach utilizes Kullback-Leibler divergence [8] to extract implicit trust relationships.

3.1 Naïve approach

Collaborative filtering based recommender systems calculate the similarity between two users and deploy the user similarity to recommend items. Basically, user similarity is a symmetric relationship. For example, when user u_a is interested in item i_m and user u_b is interested in both item i_m and i_n, the similarity between u_a and u_b is 0.5 using Jaccard similarity. We examine the user similarity in depth. From u_a's point of view, u_b is interested in all items, i.e., i_m, which the u_a is interested in. Thus, u_a may prefer u_b's other items, i.e., i_n, which u_a is not aware of. However, from u_b's point of view, u_a is interested in partial items which the user u_b is interested in. u_b may not be interested in u_a's items or u_b may has seen all u_a's items already. We can calculate trust information from the asymmetric relationships.

Conditional probability is a measure to calculate an asymmetric relationship. Using tagged items, trust from u_a to u_b is defined as:

$$trust^{i}_{u_a \to u_b} = \frac{P_i(u_a \cap u_b)}{P_i(u_a)}$$

$P_i(u_a)$ is the probability of user u_a's items and $P_i(u_a \cap u_b)$ is the probability of items that both u_a and u_b are interested in. Trust is also calculated by tags. Using tags, trust from u_a to u_b is defined as:

$$trust^{t}_{u_a \to u_b} = \frac{P_t(u_a \cap u_b)}{P_t(u_a)}$$

$P_t(u_a)$ is the probability of user u_a's tags and $P_t(u_a \cap u_b)$ is the probability of tags that both u_a and u_b annotate.

3.2 Kullback-Leibler divergence approach

Kullback-Leibler divergence is an asymmetric measure of the difference between two probability distributions. Previously mentioned naïve approach only considers the existence of tagged items or tags but KL divergence takes into account the frequency of tagged items or tags. This aspect enables the detailed analysis of the user's preference. This approach is analogous to the weighted version of naïve approach. Using tagged items or tags, trust from u_a to u_b is defined as:

$$D^{i}_{KL}(u_b || u_a) = \sum_{k \in I(u_a)} f_i(u_b, k) \log \frac{f_i(u_b, k)}{f_i(u_a, k)}$$

$$D^{t}_{KL}(u_b || u_a) = \sum_{s \in T(u_a)} f_t(u_b, s) \log \frac{f_t(u_b, s)}{f_t(u_a, s)}$$

$I(u_a)$ is a set of tagged items that u_a is interested in and $T(u_a)$ is a set of tags that u_a annotates. $f_i(u_b, k)$ and $f_t(u_b, s)$ are probability mass functions and they are defined as:

$$f_i(u_b, k) = \frac{n_i(u_b, k)}{\sum_{l \in I(u_b)} n_i(u_b, l)}$$

$$f_t(u_b, s) = \frac{n_t(u_b, s)}{\sum_{r \in T(u_b)} n_t(u_b, r)}$$

$n_i(u_b, l)$ is u_b's distribution of tagged item l and $n_t(u_b, r)$ is u_b's distribution of tag r. Proposed recommender system calculates a trust relationship between two users based on the equations.

3.3 Item recommendation

We proposed a method that recommends items to a user with the user's tagging information. Given a user's tagging information, the system elicits implicit trust relationships from the user's dataset. To recommend appropriate items to the user, the system searches the user's trustful users who have high trust scores with the given user. Then, items of trustful users are aggregated to generate a final recommendation result. The system recommends relevant items that are matched the user's preference in order of relatedness.

4 Evaluation

Dataset. We conducted a series of experiments on *last.fm*[1] dataset. *Last.fm* is a music website and provides social tagging service for users. In *last.fm*, a user annotates tags to artists by the user's interest. Users manage and discover artists with tagging.

From the *last.fm* dataset, we evaluate the recommendation accuracy of derived implicit trust relationships of the different number of trustful users. Implicit trust relationships are extracted by two methods. Given a user, first method searches trustful users by the conditional probability between other users. Second method searches trustful users by the Kullback-Leibler divergence between two users.

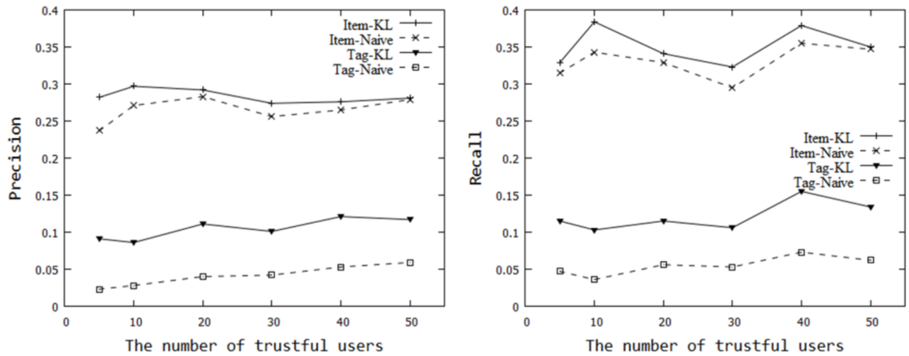

Fig. 1. Recommendation accuracy of the naïve and KL divergence approach

Fig. 1 shows the average precision and recall values of naïve and KL divergence approach. The values are averaged over randomly selected 100 users from the dataset. In Fig. 1, solid lines are our proposed method's results and dashed lines are naïve

[1] http://www.last.fm

approach's results. Precision and recall values of KL divergence based methods are higher than naïve methods. Naïve approach captures the existence of tagged items or tags. Additionally, KL divergence based method captures the count of tagged items or tags. This feature causes the better performance of KL divergence based method. Irrelevant to the extraction methods, tagged item based method shows better accuracy than tag based method. Because the process is recommending not tags but items, recommender system which utilizes only tags cannot easily produce better performance than tagged item-based recommender system.

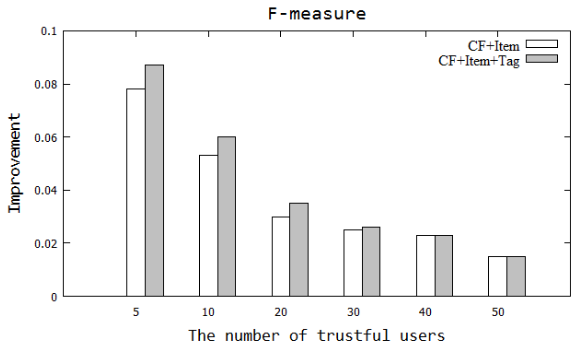

Fig. 2. Improvement over collaborative filtering

Fig. 2 shows that the recommender system's improvement over the collaborative filtering system. In Fig. 2, *CF+Item* is the result of utilizing tagged item-based implicit trust relationships with the collaborative filtering method. *CF+Item+Tag* is the result of utilizing both tagged item-based and tag-based implicit trust relationships with the collaborative filtering method. When the recommender system deploys implicit trust relationships, the performance of the system is higher than that of the system without trust information. In Fig. 2, when the number of trustful users is small, the improvement is larger than others. If the user has the small number of trustful users, our implicit trust relationships are critical feature to improve the quality of recommendation results. This result indicates the proposed approach alleviates the sparsity problem of the traditional recommender systems.

5 Conclusion

In this paper, we presented a recommendation method in social tagging system. Our approach elicits implicit trust relationships to alleviate sparsity problem. The system generates efficient and accurate recommendation with trust information. To derive trust information, we reflect asymmetric nature of users trust relationships. We utilize conditional probability as a naïve approach and Kullback-Leibler divergence as more sophisticated approach. Experimental result shows that the proposed approach performs well in practice. As future work, we plan to provide hybrid recommendation method, tag abstraction, and trust propagation.

Acknowledgments. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 20110030812).

References

1. Bhuiyan, T., Xu, Y., Jøsang, A., Liang, H., Cox, C.: Developing Trust Networks based on User Tagging Information for Recommendation Making. Web Information Systems Engineering 2010, pp. 357-364 (2010)
2. Massa, P., Avesani, P.: Trust-aware Recommender Systems. RecSys 2007, pp. 17-24 (2007)
3. Mui, L., Mohtashemi, M., Halberstadt, A.: A Computational Model of Trust and Reputation. In: Preceedings of the 35th Hawaii International Conference on System Science, pp. 2431-2439 (2002)
4. Massa, P., Bhattacharjee, B.: Using Trust in Recommender Systems: an Experimental Analysis. Trust management, pp. 221-235 (2004)
5. Milicevic, A.K., Nanopoulos, A., Ivanovic, M.: Social Tagging in Recommender Systems: a Survey of the State-of-the-art and Possible Extensions. Artificial Intelligence Review 33, pp. 187-209 (2010)
6. Papagelis, M., Plexousakis, D., Kutsuras, T.: Alleviating the Sparsity Problem of Collaborative Filtering using Trust Inferences. Trust management, pp. 125-140 (2005)
7. O'Donovan, J., Smyth, B.: Trust in Recommender Systems. Intelligent User Interfaces, pp. 167-174 (2005)
8. Kullback, S., Leibler, R.A.: On information and sufficiency. The Annals of Mathematical Statistics 22, pp. 79-86 (1951)