A Trust-aware Recommender System Based on Implicit Trust Extraction

Zarli Htun
Department of Information Technology
University of Technology, Yatanarpon Cyber City, Republic of Union of Myanmar

Phyu Phyu Tar
Department of Information Technology
University of Technology, Yatanarpon Cyber City, Republic of Union of Myanmar

Abstract— Trustworthiness is used to solve cold-start problem in recommender systems. In such systems, user assigned explicit trust rating such as how much they trust each other and trust can be used for neighbor formation to generate automated recommendation. However, reliable explicit trust data is not always available. In this paper we propose a new method of deriving implicit trust relationships based on user's interest similarity without needing explicit trust data. To measure the users' interest similarity, user's personalized tagging information is used. The derived implicit trust values are used to generate resource recommendations for social bookmarking website users. Our experiment result shows that the proposed recommendation method outperforms the traditional collaborative filtering approach.

Keywords —Recommendation, Collaborative Filtering, Social Bookmarking, Tags, Trust

I. INTRODUCTION

With the rapidly growing amount of information available on the World Wide Web, it is important to have tools to help users to select the relevant part of online information. Recommender systems are a solution to information overload problem. Its task is to recommend 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 get the items the user interests because the number of existing items is too large. In this case, recommender system helps the user to find his interesting items. Typically, in a recommender system, we have a set of users and a set of items. Each user u rates a (small) subset of the set of all items with some numeric score, e.g. on a scale from 1 to 5. The recommender system has to predict the unknown rating for source user u on a non-rated target item based on the known ratings.

Collaborative Filtering (CF) is one of the most successful technologies among recommender systems [13]. CF is a process that the system uses the preference similarity from users in the system to sort out the relevant information from the non-relevant information. Most of the CF-based recommender systems use either the user-based CF or the item-based CF schemes for recommending resources to users. The approach of user-based CF calculates the similarity between neighbors and the target user for predicting the new proper items to the target user. On the other hand, the approach of item-based CF collects the interaction history of user-item from the system and selects the most similar items based on the similarity between items.

To overcome the well-known cold-start problem, recommender systems are extended with trust-awareness. Trust-aware recommender systems make recommendations based on the ratings of users that are directly or indirectly trusted by the target user. They assume that users tend to find trusted users when receiving recommendations. To say it formally, trust is the subjective probability that the trustier, u will have the same preferences and tastes as the trustee v. By utilizing trust information, recommender systems relieves sparsity problem [4]. However, most of the trust-aware recommender systems provide no means of representing implicit trust relationship between users. In the proposed system, trust values between users are tried to estimate implicitly without needing explicit information.

Social tagging systems have gained popularity on the web. They allow users to create tags that annotate and categorize content and share them with other users. People tag resources for future retrieval and sharing [10]. Tags can convey information about the content and creation of a resource [11]. Tags identify what the resource is about and the characteristics of a resource [12]. This additional metadata could also be used to support the recommendation process. The tags assigned to items by the user describe user's personal interest and preferences who added them. In

this paper, how the information represented by the folksonomy data can be used to improve the performance of traditional collaborative filtering algorithms. This paper presents a resource recommendation method which is based on implicit trust relationships which are obtained from users' tagged resources and tags.
The rest of the paper is organized in the following ways. In section 2, other related works are presented. Section 3 describes the traditional approach of making recommendation. Section 4 presents the framework of proposed system. Section 5 explains the dataset and experimental setup. Section 6 presents the experimental results and the paper concludes in section 7.

II. Related Work

Trust-awareness is used to overcome cold-start problem in recommender systems. Trust could be used as supplementary or replacement method of widely used collaborative filtering [1, 2]. Reference [4] introduced a trust-aware recommender system which relies on user's explicitly stating trust values for other users. Reference [3] conducted analysis on the correlation between interpersonal trust and interest similarity. They suggested that the relationship between users' trust and similarity is positive. Reference [1] presents a survey on the relationship between trust and interest similarity in a social network. Reference [6] proposed a method of developing trust network based on user tagging information but they required item description information while defining the semantic meaning of tags. This work differs from the above research in several ways, most notably, in the estimation of trust values. Rather than having the user explicitly declare trust vales, the recommender system will implicitly estimate trust values between users. Users may not want to expend effort on assigning trust values to other users.
There is not so much research work done on the item recommendation. Reference [5] discussed about using the tag information to do item recommendation. In [5], the tag information was converted into two 2-dimensional relationships, user-tag and tag-item, and was used as a supplementary source to extend the rating data. A tf-idf weighted tag-based item profiles have been used for web page recommendation in [7] and [8].
In this paper, user's tagging behavior in social bookmarking website will be explored to know the user's interest and preferences. After determining user's preferences and interest based on user's tag usage, user-item rating matrix is derived which is not explicitly available in social bookmarking website based on user's tagging behavior. And, implicit trust relationships between users are derived in order to be used in proposed resource recommender system.

III. Overview of Collaborative Filtering Recommender System

In this section, the basic work of recommender system is described. The goal of a CF algorithm is to recommend products to a target user based on the opinions of other users. A recommender system assumes a set of users U= $\{u_1,\ldots u_N\}$ and a set of items I= $\{i_1,\ldots\ldots i_M\}$. Each user u rates a set of items by $r_{u,i}$. $r_{u,i}$ can be any real number, but more often ratings are integers, e.g., in the range [1,5]. The basic task of recommendation is as follows. Given a user u U and an item i I for which ru,i is unknown, predict a rating for u on item i.
For each user u_i we have a list of items I_{ui} for which the user has expressed an opinion about. These opinions can be either explicitly given by the user as a rating score (as is the case with Netflix) or can be implicitly derived from the user's purchase records. Under this setting we consider a distinguished user $u_a \in U$ called the active user for whom the task of a collaborative filtering algorithm is to suggest other items that the active user might like. This suggestion can take either of the following two forms:

- Prediction: Provide a numerical value, $P_{a,j}$ expressing the predicted likeliness of item i j / $\in$ I_{ua} for the active user u_a. The predicted value should be within the same scale (e.g., from 1to 5) as the opinion values provided by u_a in the past.
- Recommendation: Provide a list of N items, Ir I, that the active user will like the most. Obviously the recommended list should only contain items not contained in I_{ua}. This kind of suggestion is also known as Top-N recommendation.

The proposed system is a top-N collaborative filtering recommender system that provides users with a ranked list of top-N items.

III. Proposed trust-based Recommender system

In social bookmarking system, the following data exists:
$U=\{u_1, u_2,\ldots\ldots\ldots,u_m\}$ is a set of n users,
$T=\{t_1,t_2,\ldots\ldots\ldots\ldots,t_l\}$ is a set of tags annotated by users to describe bookmarked resources,

$I=\{i_1,i_2,\ldots\ldots\ldots\ldots,i_n\}$ is the set of resource items tagged by users.

Trustworthiness between users is useful for recommender systems. However, the trust information is not always available, and even available, it may change over time. In this research, we propose to automatically construct the trustworthiness between users based on users' online information and online behavior. Users with similar tagging behavior and resource items will trust each other. Therefore, user's tagging information will be used to measure trustworthiness between users.

In social bookmarking systems, users annotate the bookmark resources using their own keywords called tags. Therefore, tags show user's interest and preferences on the bookmarked resources. The proposed system will analyze user's tag usage and find out user's interest on resource items. Therefore, as the first step the proposed system finds the interest factor of a user for a particular tag. Then the resulted interest factor is used in calculating the interest factor of user for a particular item. A user will be more interested in tags which are highly used by the user than tags which are rarely used. And he will be interested in related resources in the same way.

Each user u_k E U has a set of personal tags: $T_k= \{t_{kl},t_{k2},\ldots\ldots\ldots,t_{kl}\}$ used to annotate the particular resources. Therefore, as the first step, the tag score of a user for a particular tag, $ts(u_i,t_j)$, is defined as follows,

(3)

where $freq(t_j)$ is the total frequency of tag i used by user to annotate his resource items. Freq (t_l) is total frequency of all tags used by user i. Therefore, how much each tag important to the user in all the tags used by this user is measured in this step. Then the interest Factor of a user to a resource item 'i' is the sum of interest factor of all tags used by the user to annotate the resource item 'i'. It is written in formula as follows:

$$IN(u,i) = \sum_{j \in T_r} ts(u,t_j) \qquad (4)$$

Bookmarked items annotated with important tags by the user are important to user and will have highest interest factor values.

After finding user's interest on tags and resource items, the next step is to measure trust values between users based on their interests. Prior work on trust-enhanced recommender systems worked on the explicit trust relationship of users. They need users to provide explicit trust rating for other users. But it is not always possible to obtain explicit trust information from users. In our research, measuring trust values between users base on three similarity measures: tag usage similarity, resource item similarity and interest similarity on resource items.

To compute tag usage similarity for the two users u_a and u_b, let T_a and T_b be the sets of tags for each user u_a and u_b respectively. Tag usage similarity is measured based on the common tags used by the two users, u_a and u_b. It is calculated as follows:

$$sim_T(u_a,u_b) = \frac{(T_a \cap T_b)}{|T_a|} \qquad (5)$$

To compute the resource item similarity between two users u_a and u_b, both of their resource items are considered as two sets, and the Jaccard Index is applied between these sets. The Jaccard index is a well known statistic, widely used to compare the similarity between two sets. This formula is presented below in equation 1, where I_a represents the item set of user u_a and I_b represents the item set of u_b.

$$sim_R(u_a,u_b) = J(u_a,u_b) = \frac{(I_a \cap I_b)}{|I_a|} \qquad (6)$$

To calculate interest similarity on resource items, we use interest factors values calculated from equation (4). Interest factor values calculated at equation (4) indicate the users' interest on resource items. These values can be taken as user's implicit ratings on resources items. Then, the similarity between two users' rating behavior can be by using various similarity measures. In the proposed system, to measure item interest similarity called $sim_I(u_a,u_b)$, Pearson Correlation similarity method is used. Let a and b be two users, $R_{a,p}$ be the rating of user a for item p and P be the set of items, rated both by a and b. Then Pearson correlation coefficient is defined as follows:

(7)

Finally, level of trust 'st' is the sum of three similarity measures stated above,

$$st(u1, u2) = sim_A(u$$ (8)

The proposed social bookmark recommender utilizes trust values calculated in the above step to filter most trusted users in neighborhood selection process. After neighborhood selection, the resource recommender subsequently chooses which resource items of neighbor users to be recommended to user. In order to generate a top-N list of items, the ranking score of an item is computed according to the following equation,

$$RScore\ (u,i) = \sum_{n \in Nei\ (u)} st(u,n)$$ (9)

Where Nei(u) is neighbors of user u produced from neighborhood formation phase. st(u,n)=trust value of user u and his neighbor n.

IV. EXPERIMENTS

This paper uses the version of the Delicious.com dataset published for the HetRec 2011(hetrec2011-delicious-2k). This dataset was obtained from Delicious social bookmarking system. Its users are interconnected in a social network generated from Delicious "mutual fan" relations. Each user has bookmarks, tag assignments, i.e. Triples [user, tag, bookmark] and contact relations within the dataset social network. Each bookmark has a title and URL. Table 1 shows some statistics about the dataset.

Table I. DATA STATISTICS OF HETREC-DELICIOUS-2K DATASET

1867 users
69226 bookmarked URLs 38581 bookmarked principal URLs(e.g., www.delicious.com for http://www.delicious.com, http://www.delicious.com/
104799 bookmarks (avg. 56.1 bookmarked URLs user; avg. 1.5 users/bookmark)
53388 tags 437593 tag assignments(avg. 234.4 tags/user;avg. 6.3 tags/URL
7668 bi-directional user relations (avg. 8.22 relations/user)

An experimental study was performed to evaluate the performance of our resource recommender. The recommender was tested with derived implicit trust relationships for various numbers of trustful users. In our experiments, the dataset is divided into two parts – training (80%) and testing datasets (20%).

The "Precision and Recall" method is used to evaluate the recommendation performance. Precision and Recall for an item list recommended to user u_i is computed based on the following equations:

$$\text{Precision} = \frac{|T_i \cap P_i|}{|P_i|}$$ (10)

$$\text{Recall} = \frac{|T_i \cap P_i|}{|T_i|}$$ (11)

Where T_i is the set of all items preferred by user u_i and P_i is the set of all recommended items generated by the recommender system.

The proposed trust-aware recommender system (ST) was compared with traditional user-based collaborative filtering recommender system. Both methods ST and Traditional CF used the same recommendation technique

except the neighborhood formation phase. With proposed ST algorithm, top N most trusted users are selected as neighbors. For neighborhood formation in traditional user based CF, cosine similarity method is used. In the performed experiment, the number N of recommended items is set to 100. The dataset contains significantly more items than users which is different from many other data sets, so using a small value for N will produce generally poor results for all compared methods. And then the system is tested with various numbers of neighbors from 10 to 50 by an interval of 10.

Fig 1 shows the average recall values of our proposed system and traditional CF. When the recommender system deploys the implicit trust relationships, the performance of the system is higher than that of the traditional system without trust information. If the user has the small number of trustful users, the derived implicit trust relationships are important to improve the quality of recommendation results.

According to the figure, the proposed implicit trust relationship extraction can find more hit items than the traditional approach.

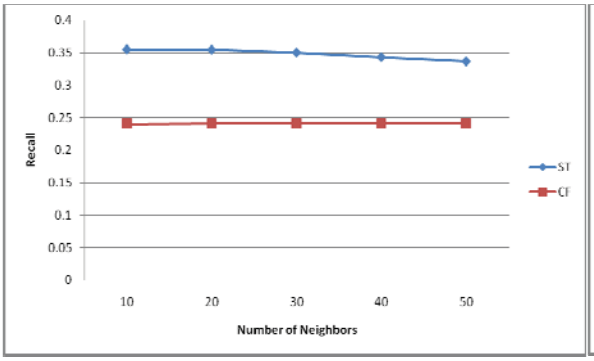

Fig 1 Recall comparison of the two models for resource recommendation by varying number of neighbors N=10,….,50 and fixing top N =100.

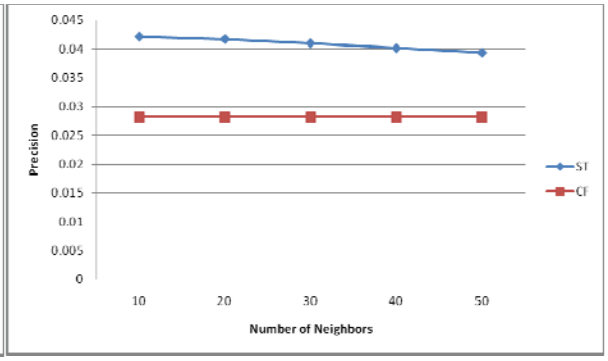

Fig 2 Precision comparison of the two models

V. CONCLUSION

In this paper, a recommendation method based on implicit trust relationships is presented. Implicit trust relationships are derived by analyzing the users' tagging information and preferences. The implicit trust metric is incorporated into top n recommendation system to provide more accurate and reliable recommendations to users. Moreover, user's interest and preferences are captured by analyzing user's tagging information and generated the user-item rating matrix to be used in top-N recommender system even though there is no explicit rating matrix available. For future work, we will evaluate our recommendation approach with other datasets, such as CiteULike.

REFERENCES

[1] T. Bhuiyan, "A survey on the relationship between trust and interest similarity in online social networks", The Journal of Emerging Technologies in Web Intelligence, Vol 2, No. 4, 2010.

[2] J. O'Donovan, "Capturing trust in social web applications", In Computing with social Trust, volume 3, 2009.

[3] C. N. Ziegler and J. Golbeck, "Investigating interactions of trust and interest similarity", In Decision Support Systems, Vol. 43,Issue 2, March, *2007*, pp 460-475.

[4] P. Massa and P. Avesani (2007), "Trust-aware recommender systems", In ACM Recommender Systems Conference , Minneapolis, MN, USA.

[5] K.H.L. Tso-Sutter, L.B. Marinho, and L.Schmidt-Thieme. "Tag-aware recommender systems fusion of collaborative filtering algorithms", Proceedings of the 2008 ACM symposium on Applied computing, ACM, USA, 2008.

[6] T. Bhuiyan, Y. Xu and A. Josang, "SimTrust:a new method of trust network generation", IEEE/IFIP 8th International Conference on Embedded and Ubiquitous Computing(EUC),2010,IEEE,Hong Kong, China, pp. 718-722.

[7] S.Niwa,T. Doi and S. Hon'Iden, "Web page recommender system based on folksonomy mining", Transactions of Information Processing Society of Japan, 47(5):2006.

[8] A. Shepitsen, J. Gemmell, B. Mobasher and R. Burke, "Personalized recommendation in social tagging systems using hierarchical clustering", In Proc. Of the 2008 ACM conference on Recommender systems, 2008.

[9] M. R. McLaughlin and J. L. Herlocker. "A collaborative filtering algorithm and evaluation metric that accurately model the user experience", In SIGIR '04: Proceedings of the 27th international ACM SIGIR conference on Information Retrieval, New York, NY, USA, 2004.

[10] C. Marlow, M. Naaman, D. Boyd and M. Davis, "HT06, tagging paper, taxonomy, flickr, academic article, to read" in: Proceedings of Hypertext, New York: ACM Press, 2006.

[11] M.hbcgt Memmel, M. Kockler and R. Schirru, "Providing multi source tag recommendations in a social resource sharing platform," Journal of Universal Computer Science, vol. 15, no. 3, 2009.

[12] S.A. Golder and B. A. Huberman, "The structure of collaborative tagging systems," Journal of Information Science, vol. 32, no. 2, 2006.

[13] G. Adomavicius and A. Tuzhilin , "Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extentions. IEEE Transactions on Knowledge and Data Engineering, Vol.17, No.6, pp.734-749.