Using Self-Defined Group Activities for Improving Recommendations in Collaborative Tagging Systems

Danielle H. Lee and Peter Brusilovsky
School of Information Sciences, University of Pittsburgh
135 N. Bellefield Ave, Pittsburgh, PA 15260, USA
hyl12@pitt.edu, peterb@pitt.edu

ABSTRACT
This paper aims to combine information about users' self-defined social connections with traditional collaborative filtering (CF) to improve recommendation quality. Specifically, in the following, the users' social connections in consideration were groups. Unlike other studies which utilized groups inferred by data mining technologies, we used the information about the groups in which each user explicitly participated. The group activities are centered on common interests. People join a group to share and acquire information about a topic as a form of community of interest or practice. The information of this group activity may be a good source of information for the members. We tested whether adding the information from the users' own groups or group members to the traditional CF-based recommendations can improve the recommendation quality or not. The information about groups was combined with CF using a mixed hybridization strategy. We evaluated our approach in two ways, using the *Citeulike* data set and a real user study.

Categories and Subject Descriptors
H.5.3. Information Interfaces and presentation: Group and Organization Interfaces – Collaborative Computing

General Terms
Design, Experimentation, Human Factors

Keywords
Social Network, Group, Hybrid Recommendations, *Citeulike*

1. INTRODUCTION
The power of collaborative filtering (CF) recommendations is based on a relatively simple idea: target users will like items favored by their likely-minded peer cohorts. That is to say, CF technology filters or evaluates target users' information based on the interests of the users' peers. While generations of CF systems have proven the effectiveness of this approach, it is known that CF is not performing well in situations where density of user ratings is relatively low. These situations occur, when a CF system just begins its operation (cold start) or when the number of items is very large and the number of users is relatively low (sparsity). A good example of the latter case is provided by popular collaborative tagging systems such as *Delicious* or *Citeulike*. Since items in these systems are submitted by the users themselves, the number of items is very large and an item can be bookmarked by just a few users. As a result, an approach utilizing only co-rating (i.e., co-bookmarking) does not work as a reliable approach to identify cohorts [9], and it results in a low quality of automated CF recommendations in these systems [2]. To solve this problem, more recent research efforts have focused on finding other sources of information to identify cohorts such as the peers who can be extracted from user tags or social connections [5, 18].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
RecSys'10, September 26–30, 2010, Barcelona, Spain.
Copyright 2010 ACM 978-1-60558-906-0/10/09...$10.00.

Among these approaches, the use of social connections might be a promising ways according to the existing research. Massa and Avesani [10] showed that users' trust networks can solve the malicious user problem, improve recommendation prediction and attenuate computational complexity using *Epinions* data set. In addition, for users who have rated few items (less than 5 ratings), trust network-based recommendations could make recommendations for 66% of the users, while CF could make recommendations for only 14% of the users with higher margin of error. Another study indicated that, for a user with unique tastes, his own social networks could increase his satisfaction in the recommendations, since he is able to know to where the information came from [16]. The recommendations made by friends were known to be frequently better and more useful than recommendations made by systems [15].

In our work, we utilize a user's own social connections as a source of recommendations, as in early-day "push" and "pull" collaborative filtering systems [14]. In a situation of so called *object-centered sociality* where social networks grow around common interests, the information about users' networks may be a good source for recommendations [6]. Herein, the social connections in considerations were users' groups. People join a group to share and acquire topic-related information as a form of community of interests or practice. For instance, when a group is about a Web programming language, the members share the language-related news or resources. If a group is made up of music fans, the members introduce good music to each other. Various social systems from social networking systems such as *Facebook*, *LinkedIn*, or *Friendster* to social tagging systems such as *Citeulike*, *Flickr*, or *Bibsonomy* support group activities.

The primary targets of existing group recommendations are, however, a group of users, not an individual user. The researchers suggested TV programs to watch [11], music to listen to together [4], or restaurants or venue to enjoy [1, 12] to a group of people by aggregating each person's preferences into one set. Few of these studies explored how to employ a user's group or community information to personalize his information. The study done by Yuan and colleagues [17] was one of the few examples. They exploited memberships of the same group and friendship-based social networks as a part of the recommendations. The hybridizations through weighted similarity calculations or through graph-based social features were tested against the traditional CF approach, using *Last.FM* data set. The results represented that fusing users' social networks with CF recommendations improved the recommendation quality. The hybridization in the study was done when the user similarities were calculated, and the similarities were based on users' ratings [17]. In our study, we focused on the recommendations in the situation when no rating is available and hybridization is done at the time of computing predictions. In the following section, we explained the recommendation algorithm and experimental setting.

2. RECOMMENDATION DESCRIPTION

In this paper, we explored a hybrid approach, which combined CF-based recommendations and group-based recommendations. This version of social recommendations used information about users' self-defined groups. As the first step of the process, we used a CF approach to generate a cohort for each target. Our recommendation approach was designed for social tagging systems, where ratings are typically not present and users express their interests by bookmarking. Therefore, we used Jaccard similarity coefficient for co-bookmarked information instead of Pearson correlation for co-ratings to compute the similarity between two users (equation 1) [7]. The Jaccard similarity represented the fraction of shared information in the joint information space of both users. Here A and B were sets of items in the collection of two users being compared.

$$\textbf{Jaccard Similarity} = (A \cap B)/(A \cup B) \qquad \text{eq (1)}$$

Once we selected the peer cohorts based on the Jaccard similarity regardless of whether they were in the same group or not, we produced CF-based item ranking according to the aggregated prediction scores of the items in the top 10 cohorts' collections, as following equation 2.

$$\mathrm{CF_{ia}} = \sum_{j=1}^{n} \mathrm{sim_{ij}} \times \mathrm{w_{ja}} \qquad \text{eq (2)}$$

CF_{ia} was a CF recommendation prediction score for user i about item a. It was calculated for every item a, which belonged to the collection of at least one of the top 10 peers' collections (peer j). W_{ja} represented whether peer j had the item a in his collection. If he had the item, the value was 1, otherwise the value was set as zero.

The next step generated recommendations based on group information. This step utilized two kinds of group information: group members' collections and groups' collections. Group collections exist in a number of social systems such as *Citeulike*, where users are able not only to manage interesting information in their personal repository (a group member's collection), but also jointly build group collections (a group's collection). The recommendations based on group members' collection was calculated using Jaccard similarity and the same ranking approach as used in CF step. The only difference was using group members instead of the top 10 peer users. In equation 3, variable k represented the other member of user i's group(s) and W_{ka} showed whether the group member k had the item a or not.

$$\mathrm{GMem_{ia}} = \sum_{k=1}^{n} \mathrm{sim_{ik}} \times \mathrm{w_{ka}} \qquad \text{eq (3)}$$

After that, the recommendations based on group members' collection were fused with the recommendations based on group collections. For each item belonging to at least one of user's group collections or group member's collections, the prediction scores were computed based on the mixed aggregation of similarities of group and group members. To compute the recommendations based on group collections, we used the same Jaccard similarity between a target user and one of his group or his group by comparing individual and group collections. In eq. 4, W_{xa} was the weight to denote whether the group x has item a or not.

$$\mathrm{G_{ia}} = \sum_{k=1}^{n} \mathrm{sim_{ik}} \times \mathrm{w_{ka}} + \sum_{x=1}^{y} \mathrm{sim_{ix}} \times \mathrm{w_{xa}} \qquad \text{eq (4)}$$

In the final step, our hybrid approach combined both – CF-based recommendations ($\mathrm{CF_{ia}}$) and group-based recommendations ($\mathrm{G_{ia}}$), which in turn a fusion of group and group members-based recommendations. We used a simple version of *mixed* hybridization approach. Once the CF recommendations and two kinds of group-based recommendations selected a list of candidate items respectively, the items were combined together and summed up the recommendation scores for each item. According to the final scores, we ordered the items and pick the top N items [3]. Figure 1 explained the operation in the mixed strategy and equation 5 showed how to compute the hybrid recommendations.

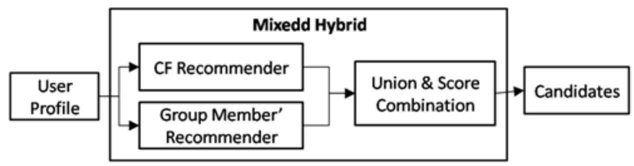

Figure 1.Mixed Hybrid Strategy

$$\mathrm{H_{ia}} = \sum_{j=1}^{n} \mathrm{sim_{ij}} \times \mathrm{w_{ja}} + \sum_{k=1}^{n} \mathrm{sim_{ik}} \times \mathrm{w_{ka}} + \sum_{x=1}^{y} \mathrm{sim_{ix}} \times \mathrm{w_{xa}} \qquad \text{eq (5)}$$

3. EXPERIMENTAL EVALUATION

To assess the prospects of group-based recommendations, we ran two studies – a traditional n-fold prediction study with a large data set and a real user study.

3.1 The Data Source

As the data source, our study used a social tagging system, *Citeulike*. *Citeulike* is one of the leading systems for managing and sharing bibliographic references. *Citeulike* has a well-developed group mechanism. Users can create a group, join existing groups or be invited to join a group. When group members find interesting references, they are able to add them not only to their personal repositories, but also to the group collection. The tags assigned to the item will also be added to the group record. The up-to-date group collection is available to all group members. The group members are also able to easily copy references from the group collection to their personal repositories and back.

In order to crawl the data, the site was visited in October and December, 2008. At that time, there was a page with a list of all groups. We collected all group names that were shown on the page at the time of the visit. Then, the group collections, the group members and each member's personal collection were collected. The collected information included the bibliography (article title, list of authors, journal name, publication year, etc.), the tags and the posted date and time. In total, information for more than 700 groups was collected. After filtering out single-member groups and group members who had insufficient collection ($n < 5$), the total number of groups was reduced to 619 and then total number of users to 2643 with 3528 memberships (i.e., user-group connections). Each user was a member of 1.34 groups and each group had 5.7 members on average. To diversify the data set, we also selected other users, who posted new items at the time of visit, and collected their bibliography, tags and the posted date and time. Table 1 is the description of the whole data set. It shows the statistics not only about the group members but also about other users who were not in any group.

Table 1. Descriptive Statistics of *Citeulike* Dataset

Total no. of users (including group members)	19958
Total no. of distinct items (papers)	1070389
Average no. of items per user	64.75
Total no. of groups	619
Total no. of group members	2643
Total no. of group memberships	3528
Average no. of items per group	445.89
Average no. of items per group members	188.24

3.2 The Formal Evaluation

We used the collected *Citeulike* data set to perform a traditional formal *N* fold cross validation (*N*=5). The idea of the cross-validation approach is to assess the quality of a recommendation approach by its ability to predict ratings (or bookmarks) already made by the users. In our 5-fold cross validation, the information collection of each test user was randomly partitioned into 5 equally sized subsets. Each subset was then used as the test set and the other subsets were used as the training set to predict recommendations in the test set. This procedure was run 5 times, and the results of 5 runs were averaged to produce the final accuracy. The final results were assessed through evaluation metrics for information retrieval: precision and recall. Precision aims to measure how precise the prediction is and recall aims to measure how complete the prediction is. More specifically, precision at point N (precision@N) is the ratio of the number of correctly predicted items in the top-N list to N (eq. 6). Recall at point N (recall@N) is the ratio of the number of correctly predicted items in the top-N list to the total number of relevant items (eq. 7) [13, 17].

$$\text{precision@N} = \frac{\text{No. of correct prediction}}{\text{N of top N set}} = \frac{\text{test} \cap \text{top N}}{\text{N}} \qquad \text{eq (6)}$$

$$\text{recall@N} = \frac{\text{No. of correct prediction}}{\text{size of test set}} = \frac{\text{test} \cap \text{top N}}{\text{test}} \qquad \text{eq (7)}$$

The 5-fold cross validation was used to compare the quality of four approaches discussed in the previous section including CF-based recommendations as a baseline. We compare precision and recall at five points taking into account the top 1, 3, 5, 10 and 20 recommendations. To select the test users, we randomly chose 630 users who were members of groups. First, Table 2 shows the results of precisions. Even though our algorithm was quite simple, we were able to predict more than one of third of the test set correctly in the top 1 and top 3 recommendations. The results of the precision showed that the recommendations made based on group members' information ($GMem_{ui}$) were the worst, and the group- and group members-based (G_{ui}) and hybrid approaches (H_{ui}) were better than other two approaches. These differences were all statistically significant. We reckoned that two users joined the same group since their interests are similar. In one of our previous studies [8], however, we found that the interest overlaps between the members of a group were too trivial to infer one user's interests from other members in his group. On the other hand, the group's collection was made up of several members' contributions, and the aggregated set of group's collection items was largely overlapped with their members. Therefore, in group-based recommendations, rather than the members' collections, the group collections which were assembled by the members were more helpful to get useful information. In the comparison of recall, we found the same results. Among four recommendations, the recall of hybrid recommendations was the highest. The results were significant, as well.

Table 2. Precision Comparison

	Top 1	*Top 3*	*Top 5*	*Top 10*	*Top 20*
CF_{ui}	19.62%	14.14%	11.66%	08.34%	06.04%
$GMem_{ui}$	7.10%	05.84%	04.98%	03.88%	02.94%
G_{ui}	22.59%	21.29%	19.92%	17.95%	15.28%
H_{ui}	**35.13%**	**30.04%**	**27.09%**	**22.88%**	**18.28%**

Table 3. Recall Comparison

	Top 1	*Top 3*	*Top 5*	*Top 10*	*Top 20*
CF_{ui}	1.63%	2.70%	3.26%	3.97%	4.76%
$GMem_{ui}$	0.38%	0.77%	1.01%	1.35%	1.73%
G_{ui}	1.42%	3.72%	5.35%	7.90%	10.61%
H_{ui}	**2.56%**	**5.53%**	**7.51%**	**10.41%**	**13.27%**

We also counted which recommendations produced the best results for each test user among the four approaches. This was to check whether the highest mean values of hybrid recommendations were caused by a small number of very high values or not. We used the precision results of the top 5. After excluding 122 test users (19.3% of the entire test users) whose all precisions were zero, the Table 5 shows the number of test users for whom each approach produced the best results. In this case, the hybrid recommendations mostly produced the best results (58.4% of the entire test users), while the results of CF recommendations were the second best (13.8%). We interpreted these results to mean that the recommendations based only on users' social connections could not beat the CF approach, but could make the CF recommendations better by hybridizing them together.

Table 4. The Number of Test Users for Whom Each Approach Produces the Best Results (the precision of the Top 5)

CF_{ui}	$GMem_{ui}$	G_{ui}	H_{ui}
87	9	44	368

As the final analysis, we tested whether fusing users' own social connections with CF recommendations can help the cold-start problem. Out of 630 test users, 90 users had too small number of items to generate appropriate CF-based recommendations ($n < 4$). Among them, 39 users (43.3%) did not have any CF-based recommendations but had group-based recommendations. On average, these 39 users had 1.91 items ($\sigma = 1.05$).

3.3 The User Study

As the second experiment, we evaluated group-based recommendations in a user study with 8 human subjects. The subjects were all members of the same group in *Citeulike* and had a considerable number of items in our data set (M = 191.0). For each subject we produced 30 personal recommendations by randomly mixing top 10 items generated for these users using three compared approaches: typical CF-based (CF_{ui}), group and group members-based (G_{ui}) and hybrid recommendation (H_{ui}). The participants were asked to rate the recommended items using 1 – 3 scale (1 – negative, 2 – neutral & 3 – positive). Figure 2 shows the actual ratings given by participants. Except one user

(user3), most of them rated recommendations powered by their group information more positively. We also evaluated whether these visual differences in ratings are significant or not by using the one-way ANOVA test. Table 5 represents the average ratings of three approaches. As expected, the hybrid recommendations were rated the higher than the CF and group information based recommendation. These mean values among the three approaches were significantly different (F = 3.73, p = .026), but the mean difference between CF and group based recommendations was not significant, according to the Scheffe post-hoc test.

Table 5. Comparison of Users' Rating

CF_{ui}	G_{ui}	H_{ui}
2.33 (σ=0.77)	2.31 (σ=0.76)	2.63 (σ=0.45)

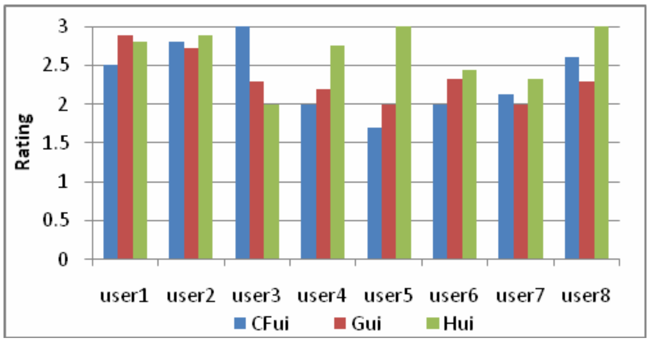

Figure 2. Average Ratings of Each User

4. CONCLUSION AND DISCUSSION

In the context of social tagging systems, the traditional collaborative filtering approach does not perform well due to the sparsity of user feedback (i.e., bookmarking). This paper explored one of the solutions to this problem based on users' social networks: more specifically user self-defined group memberships. We compared the quality of a group-based recommendation approach, which employed the information about group collections and group members' collections, with traditional CF recommendations. We also compared the quality of a hybrid approach, which combined the traditional CF and the group-based approaches. The approaches were compared using 5-fold cross validation on a crawled *Citeulike* data set, as well as a user study. From the results of precision and recall, the hybrid recommendations produced the best results. The results of the user study were similar. Most of the participants perceived that the hybrid recommendations and recommendations contributed by group information were better than those produced using the CF-based approach. We interpreted this result to mean that users' own social connections, which were mainly based on similar interests, were important source to acquire good information and should be a part of personalizing their information. They can also reinforce the existing CF-based recommendation quality.

In the future, we will evaluate the group-based recommendations through other view points, for example, novelty, serendipity, or diversity. One of the participants pointed out that some recommendations were good but others were too obvious. We also plan to develop a more elaborated recommendation algorithm. According to our previous study [8], we found that, rather than item-unit based similarity comparison, more semantically rich information, for instance, tags or metadata, is a better measure of similarity. Therefore, metadata and tag information based algorithm will be considered.

5. REFERENCES

[1] Ardissono, L., et al., *INTRIGUE: Personalized recommendation of tourist attractions for desktop and handset devices*. Applied Artificial Intelligence, 2003. **17**(8): p. 687-714.

[2] Bogers, T. and A.v.d. Bosch, *Recommending scientific articles using citeulike*, in *Proceedings of the 2008 ACM conference on Recommender systems*., Lausanne, Switzerland. p. 287-290.

[3] Burke, R., *Hybrid web recommender systems*, in *The adaptive web: methods and strategies of web personalization*. 2007, Springer-Verlag. p. 377-408.

[4] Crossen, A., J. Budzik, and K.J. Hammond, *Flytrap: intelligent group music recommendation*, in *Proceedings of the 7th international conference on Intelligent user interfaces*. 2002, ACM: San Francisco, California, USA. p. 184-185.

[5] Guy, I., et al., *Personalized recommendation of social software items based on social relations*, in *Proceedings of the third ACM conference on Recommender systems*. 2009, ACM: New York, New York, USA. p. 53-60.

[6] John, B., *The Future of Social Networks on the Internet: The Need for Semantics*, D. Stefan, Editor. 2007. p. 86-90.

[7] Kostov, V., E. Naito, and J. Ozawa. Cellular phone ringing tone recommendation system based on collaborative filtering method. in *Computational Intelligence in Robotics and Automation, 2003. Proceedings. 2003 IEEE International Symposium on*. 2003.

[8] Lee, D.H. and P. Brusilovsky. Interest Similarity of Group and the Members: the Case Study of Citeulike. in *The Proceedings of the 2nd Web Science Conference*. 2010. Raleigh, NC, USA.

[9] Lee, D.H. and P. Brusilovsky. Social Networks and Interest Similarity: The Case of CiteULike. in *Proceeding of the 21st ACM Conference on Hypertext & Hypermedia*. 2010. Toronto, Canada.

[10] Massa, P. and P. Avesani, *Trust-Aware Collaborative Filtering for Recommender Systems*. Vol. 3290. 2004. 492-508.

[11] O'Connor, M., et al., *PolyLens: a recommender system for groups of users*, in *Proceedings of the 7th European Conference on Computer Supported Cooperative Work*. 2001, Kluwer Academic Publishers: Bonn, Germany. p. 199-218.

[12] Park, M.-H., H.-S. Park, and S.-B. Cho, *Restaurant Recommendation for Group of People in Mobile Environments Using Probabilistic Multi-criteria Decision Making*, in *Proceedings of the 8th Asia-Pacific conference on Computer-Human Interaction*. 2008, Springer-Verlag: Seoul, Korea. p. 114-122.

[13] Sarwar, B., et al. Application of dimensionality reduction in recommender systems a case study. in *In ACM WebKDD Workshop*. 2000.

[14] Schaefer, J.B., et al., *Collaborative Filtering Recommender Systems*, in *The Adaptive Web: Methods and Strateiges of Web Personalization*, P. Brusilovsky, A. Kobsa, and W. Nejdl, Editors. 2007, Springer: Berlin, Germany. p. 291-324.

[15] Sinha, R. and K. Swearingen. Comparing Recommendations Made by Online Systems and Friends. in *In Proceedings of the DELOS-NSF Workshop on Personalization and Recommender Systems in Digital Libraries*. 2001.

[16] Tintarev, N. and J. Masthoff, *Effective explanations of recommendations: user-centered design*, in *Proceedings of the 2007 ACM conference on Recommender systems*. 2007, ACM: Minneapolis, MN, USA. p. 153-156.

[17] Yuan, Q., et al. Augmenting Collaborative Recommender by Fusing Explicit Social Relationships. in *Workshop on Recommender Systems and the Social Web, Recsys 2009*. 2009. New York, NY

[18] Zhen, Y., W.-J. Li, and D.-Y. Yeung, *TagiCoFi: tag informed collaborative filtering*, in *Proceedings of the third ACM conference on Recommender systems*. 2009, ACM: New York, New York, USA. p. 69-76.