Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI-10)

UserRec: A User Recommendation Framework in Social Tagging Systems

Tom Chao Zhou, Hao Ma, Michael R. Lyu and **Irwin King**
Department of Computer Science and Engineering
The Chinese University of Hong Kong, Shatin, N.T., Hong Kong
Email: {czhou, hma, lyu, king@cse.cuhk.edu.hk}

Abstract

Social tagging systems have emerged as an effective way for users to annotate and share objects on the Web. However, with the growth of social tagging systems, users are easily overwhelmed by the large amount of data and it is very difficult for users to dig out information that he/she is interested in. Though the tagging system has provided interest-based social network features to enable the user to keep track of other users' tagging activities, there is still no automatic and effective way for the user to discover other users with common interests. In this paper, we propose a *User Recommendation* (*UserRec*) framework for user interest modeling and interest-based user recommendation, aiming to boost information sharing among users with similar interests. Our work brings three major contributions to the research community: (1) we propose a tag-graph based community detection method to model the users' personal interests, which are further represented by discrete topic distributions; (2) the similarity values between users' topic distributions are measured by Kullback-Leibler divergence (KL-divergence), and the similarity values are further used to perform interest-based user recommendation; and (3) by analyzing users' roles in a tagging system, we find users' roles in a tagging system are similar to Web pages in the Internet. Experiments on tagging dataset of Web pages (Yahoo! Delicious) show that UserRec outperforms other state-of-the-art recommender system approaches.

Introduction

Social tagging systems have emerged as a popular way for users to annotate, organize and share resources on the Web, such as Yahoo! Delicious and Flickr. Social tagging systems enjoy the advantages that users can use free-form tags to annotate objects, which can ease sharing of objects despite vocabulary differences. As a form of users' individual behavior, tagging activity not only can represent users' judgments on the resources (Heymann, Koutrika, and Garcia-Molina 2008; Si et al. 2009), but also can indicate users' personal interests (Suchanek, Vojnovic, and Gunawardena 2008). However, due to the fast growth of social tagging systems, a user is easily overwhelmed by the large amount of data and it is very difficult for the user to dig out information that he/she is interested in. Several functions aiming at finding people with similar interests have been incorporated into tagging systems, such as *network* in Yahoo! Delicious and *contact* in Flickr. Take *network* in Yahoo! Delicious as an example, if a user Bob notices many of Jack's bookmarks as interesting, Bob can add Jack to his *network*. After that, when Jack updates his new bookmarks, they will also appear in Bob's bookmark pool to make it more convenient for Bob to browse resources he is interested in. However, no automatic interest-based user recommendation service is provided and it is not easy for a user to find other users with similar interest.

Solving the problem of modeling users' interest and performing interest-based user recommendation in social tagging systems achieve two benefits. At a fundamental level, we gain insights into utilizing information in social tagging systems to provide personalized service for each user. At a practical level, it can bring several enhancements. Firstly, it is more convenient for an active user to know the latest resources on particular topics he/she may be interested in because users with similar interests are recommended. Secondly, it can help users obtain high-quality results through social filtering. Thirdly, interest-based user recommendation can help build interest-based social relationships, and forming interest-based social groups, therefore increasing intra-group information flow on the corresponding topics. In this paper, we propose an effective two-phase *User Recommendation* (*UserRec*) framework for users' interest modeling and interest-based user recommendation, which can help information sharing among users with similar interests.

Related Work

Typically, recommender systems are based on Collaborative Filtering, which has been widely employed, such as in Amazon, MovieLens[1] and etc. Two trends have rised in collaborative filtering: one is memory-based algorithms (Herlocker et al. 1999), and the other is model-based algorithms (Hofmann 2004). Pearson Correlation Coefficient (PCC) (Resnick et al. 1994) and Vector Space Similarity (VSS) (Breese, Heckerman, and Kadie 1998) are applied in memory-based algorithms. Recently, several matrix factorization methods which focus on modeling the user-item rating matrix using low-rank approximations have been pro-

[1] http://movielens.umn.edu

Table 1: **An example of user-generated tags of a URL.**

URL	http://www.nba.com/
Tags of u_1	basketball, nba
Tags of u_2	sports, basketball, nba

posed for collaborative filtering (Salakhutdinov and Mnih 2008; Ma, King, and Lyu 2009). The above approaches ignore that user behaviors can provide semantic meanings, such as tagging activities, and their performances deteriorate when used for user recommendation in social tagging systems. Our proposed method shares similar goals with many of the above studies to perform recommendation, but includes several differences. Firstly, our proposed approach takes into account that tags can represent users' judgments about Web contents and represent users' interests. Secondly, the proposed framework discovers users' interests based on tag-graph, resulting in more semantic meanings. There are plenty of research efforts on social tagging systems. Several papers (Heymann, Koutrika, and Garcia-Molina 2008; Li, Guo, and Zhao 2008; Zhou et al. 2009) study the utility of tags, and find that tags are good at characterizing users' interests about Web contents and expressing concepts of resources. However, previous efforts focus on aggregating users' tags and perform data mining on the overall dataset. Our approach is different because we leverage each user's tagging activity to provide personalized service.

UserRec Framework

User Interest Modeling

A social tagging system consists of users, tags and resources (e.g. URLs, images, or videos), and we define the set of users $U = \{u_i\}_{i=1}^{I}$, the set of tags $T = \{t_k\}_{k=1}^{K}$, and the set of resources $R = \{r_j\}_{j=1}^{J}$. Users can use free-form tags to annotate resources. An annotation of a set of tags to a resource by a user is called a *post* or a *bookmark*. In order to facilitate discussions in the following sections, we define formulas related to *post* as follows:

$$\begin{aligned} R(u) &= \{r_i | r_i \text{ is a resource annotated by } u, r_i \in R\}, \\ S(u) &= \{t_j | t_j \text{ is a tag used by } u, t_j \in T\}, \\ T(u,r) &= \{t_k | t_k \text{ is a tag used by user } u \text{ to annotate the} \\ &\qquad \text{resource } r, t_k \in T\}. \end{aligned}$$

Users in social tagging systems may have many interests, and research efforts have shown that users' interests are reflected in their tagging activities. In addition, patterns of frequent co-occurrences of user tags can be used to characterize and capture users' interests (Li, Guo, and Zhao 2008). For example, it is very likely that for two tags t_k and t_m, if $t_k \in T(u_i, r_j)$ and $t_m \in T(u_i, r_j)$, t_k and t_m are semantically-related, and can reflect one kind of this user u_i's interests. Table 1 demonstrates one example.

The method for modeling users' interests consists of two stages. In the first stage, we generate an undirected weighted tag-graph for each user. The nodes in the graph are tags used by the user, the weighted edges between two nodes represent

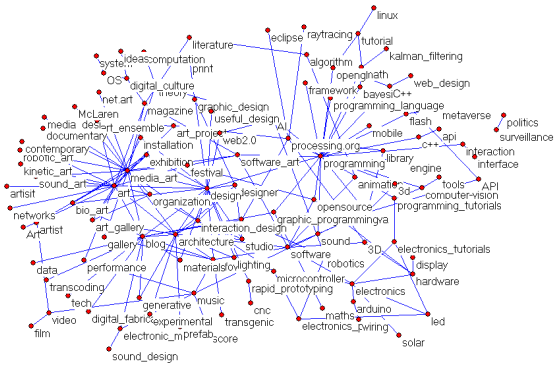


Figure 1: **Tag graph of one user.**

the strength of semantic relations between two tags, and the weights are calculated based on the user's tagging activities. Algorithm 1 shows the pseudo code of our method for generating a tag-graph for each user. The intuition of Algorithm 1 is the more often two tags occur together, the more semantically related these two tags are. The generated undirected weighted tag-graph is mapped to an undirected unweighted multigraph based on (Newman 2004). Figure 1 demonstrates one weighted tag-graph of a user generated by Algorithm 1, and to make the graph clear, the weights are not shown here. Intuitively we can find this user has interests on programming, art, etc. In addition, we can find co-occurrences of tags, such as *art* and *media_art*, *art* and *art_gallery*, can characterize a kind of this user's interests.

In the second stage, we adopt a fast greedy algorithm for community discovery in networks (Clauset, Newman, and Moore 2004), which optimizes the modularity, Q, of a network by connecting the two vertices at each step, leading to the largest increase of modularity. For a network of n vertices, after $n-1$ such joins we are left with a single community, and the algorithm stops. The complexity of the community discovery algorithm is $O(n \log^2 n)$, and n is the number of vertices in the graph. The concept of modularity of a network is widely recognized as a good measure for the strength of the community structure. Modularity is defined in Eq. (1):

$$Q = \frac{1}{2m} \sum_{ij} [A_{ij} - \frac{k_i k_j}{2m}] \delta(c_i, c_j), \tag{1}$$

where k_i is degree of node i, and is defined in Eq. (2),

$$k_i = \sum_k A_{ik}, \tag{2}$$

and A_{ij} is the weight between node i and node j, $\delta(c_i, c_j)$ is 1 if node i and node j belong to the same community after partition; otherwise, $\delta(c_i, c_j)$ is 0. $m = \frac{1}{2}\sum_{ij} A_{ij}$ is the total weights of all edges in this tag-graph. The idea of modularity is that if the fraction of within-community edges is no different from what we would expect for the randomized network, then modularity will be zero. Nonzero values represent deviations from randomness. After detecting communities in previously generated unweighted multigraph, we

Algorithm 1 Generate a tag-graph for user u_i.

procedure GenTagGraph(user u_i)
Input
$\mathbf{R(u_i)}$, the set of resources annotated by the user u_i
$\mathbf{S(u_i)}$, the set of tags used by the user u_i
$\forall \mathbf{T(u_i, r_j)}, \mathbf{where\ r_j} \in \mathbf{R(u_i)}$, the set of tags used by the u_i to annotate resource r_j
$\mathbf{G(u_i)} = (\mathbf{V}, \mathbf{E})$, V are nodes in G, E are weighted edges in G
$\mathbf{V} = \emptyset, \mathbf{E} = \emptyset$

```
1:  for all r_j ∈ R(u_i) do
2:    for all t_k ∈ S(u_i) do
3:      for all t_m ∈ S(u_i) do
4:        if t_k ∈ T(u_i, r_j) and t_m ∈ T(u_i, r_j) then
5:          if w(t_k, t_m) not exists in E then
6:            Add w(t_k, t_m) = 1 to E
7:          else
8:            w(t_k, t_m) = w(t_k, t_m) + 1
9:          end if
10:         Add t_k and t_m to V if they not exist in V
11:       end if
12:     end for
13:   end for
14: end for
```

Output
Tag-graph $\mathbf{G(u_i)}$

Table 2: **Sample topics of two users.**

u_a	sound_art, networks, artist, art, art_gallery kinetic_art, contemporary, artisit, Art
	programming_tutorials, programming_language programming, computer-vision, opengl
u_b	citations, bibliography, Research
	privacy, phishing, myspace, internetsafety cyberbullying, InternetSafety, bullying

can find *topics* of a user. A topic, which is represented by a set of tags used by a user in our framework, can show the user's interests. Thus, each community indicates one topic of the user. The set of topics of all the users is named C here. We define all the topics of a user u in Eq. (3):

$$UC(u) = \{c_m^u | c_m^u \text{ is a topic of the user u, } c_m^u \in C\}, \quad (3)$$

where c_m^u is a topic of the user u, and is defined as follows:

$$\begin{aligned} c_m^u \quad = \quad & \{t_k | t_k \text{ is a tag belonging to the corresponding} \\ & \text{community of topic } c_m^u \text{ through the} \\ & \text{community detecting algorithm, } t_k \in T\}. \quad (4) \end{aligned}$$

Through our proposed two-stage method, we can model users' interests with several topics, which consist one or more tags. Table 2 demonstrates sample topics of two users.

Interest-based User Recommendation

Based on the topics of each user generated by our two-stage method for modeling users' interests, we further propose a two-stage method to perform interest-based user recommendation. In the first stage of our interest-based user recommendation method, we represent the topics of each user with a discrete random variable. A probability value is calculated for each topic of a user according to the impact of this topic on the user. Here we introduce how to measure the impact of each topic to a user. In Eq. (3) we have defined the formula to express all the topics of a user, and in Eq. (4) we have defined the formula to express one topic of a user. $N(t_k, u_i, c_m^{u_i})$ is the number of times tag t_k is used by user u_i, where $t_k \in S(u_i)$, and $t_k \in c_m^u$. We define the impact of a topic $c_m^{u_i}$ to a user u_i in Eq. (5):

$$TN(u_i, c_m^{u_i}) = \sum_{t_k \in c_m^{u_i}} N(t_k, u_i, c_m^{u_i}). \quad (5)$$

We formulate Eq. (5) based on the idea that, if a user uses tags of a topic $c_m^{u_i}$ more often than his or her tags of another topic $c_n^{u_i}$, it is very likely that this user is more interested in the topic $c_m^{u_i}$ than the topic $c_n^{u_i}$. After defining the impact of a topic to a user, we define the total impacts of all the topics on a user in Eq. (7). The formula for calculating the probability value of each topic of a user is defined in Eq. (6), which shares similar idea with the maximum likelihood estimation method. Through the first stage of our method for performing interest-based user recommendation, we can get users' topic distributions.

$$Pr(u_i, c_m^{u_i}) = \frac{TN(u_i, c_m^{u_i})}{TTN(u_i)}, \quad (6)$$

where

$$TTN(u_i) = \sum_{c_m^{u_i} \in UC(u_i)} TN(u_i, c_m^{u_i}). \quad (7)$$

In the second stage, we propose a Kullback-Leibler divergence (KL-divergence) based method to calculate the similarity between two users according to their topic distributions. In information theory, the KL-divergence is a measure between two probability distributions. The formula to calculate the similarity value of a user u_j for a user u_i is defined in Eq. (8):

$$KL(u_i|u_j) = \sum_{c_m^{u_i} \in UC(u_i)} Pr(u_i, c_m^{u_i}) \log \frac{Pr(u_i, c_m^{u_i})}{Pr(u_j, c_m^{u_j})}. \quad (8)$$

Algorithm 2 shows the details of how to calculate the KL-divergence based similarity value of user u_j for user u_i. In line 2 of Algorithm 2, all the tags t_k belong to the same topic $c_m^{u_i}$ are sorted in a descending order according to their used frequencies $N(t_k, u_i, c_m^{u_i})$. The reason for the sorting is that, the more often a tag $t_k, t_k \in c_m^{u_i}$, is used by user u_i to express the topic $c_m^{u_i}$, the more representative this tag t_k is for the topic $c_m^{u_i}$. In other words, different tags may carry different weights to a topic just as different topics may carry different weights to a user. Line 2 to line 7 mean if topic $c_m^{u_i}$ of u_i has a corresponding topic $c_m^{u_j}$ in u_j, the value is calculated and added to the KL-divergence value. Because one topic may contain several tags, the corresponding topic exists if both topics have at least one tag in common. Line

8 to line 12 are used to avoid divide-by-zero problem if no corresponding topic exists, and it is a common way used in calculating the KL-divergence.

Algorithm 2 KL-divergence based similarity measure for user u_j to user u_i.

procedure KL-sim(user u_i, user u_j)
Input
$\forall \mathbf{Pr}(\mathbf{u_i}, \mathbf{c_m^{u_i}})$, $where\ c_m^{u_i} \in UC(u_i)$
$\forall \mathbf{Pr}(\mathbf{u_j}, \mathbf{c_m^{u_j}})$, $where\ c_m^{u_j} \in UC(u_j)$
$\mathbf{KL}(\mathbf{u_i}|\mathbf{u_j}) = \mathbf{0}$
1: **for all** $c_m^{u_i} \in UC(u_i)$ **do**
2: **for** $t_k \in c_{u_j}^{u_i}$ **do**
3: **if** $t_k \in c_m^{u_j}$ **then**
4: $KL(u_i|u_j) = KL(u_i|u_j)$
5: $+Pr(u_i, c_m^{u_i}) \log \frac{Pr(u_i, c_m^{u_i})}{Pr(u_j, c_m^{u_j})}$, **BREAK**
6: **end if**
7: **end for**
8: **if** $\forall t_k \in c_m^{u_i}, not\ \exists c_m^{u_j}\ \ that\ \ t_k \in c_m^{u_j}$ **then**
9: $KL(u_i|u_j) = KL(u_i|u_j)$
10: $+Pr(u_i, c_m^{u_i}) \log \frac{Pr(u_i, c_m^{u_i})}{\epsilon}$,
11: where ϵ is a very small real value
12: **end if**
13: **end for**
Output
$\mathbf{KL}(\mathbf{u_i}|\mathbf{u_j})$

Experimental Analysis

Dataset Description and Analysis

The dataset is crawled from Yahoo! Delicious, and in Yahoo! Delicious, users use free-form tags to annotate URLs that they are interested in. In addition, a user can add other users who share similar interests to their personal *network*. Users are informed the latest interesting resources added by people from his or her *network*. In addition, a user is informed the list of users who have added him or her to their personal *network*, and a list of *fans* appears in this user's profile. In our crawling, we crawl users' bookmarks, and here a bookmark consists of a *user*, a *URL*, and one or several *tags* annotated by this user to this URL. In addition, we crawl users' *network* information and *fans* information. Our crawling lasts one month during year 2009. Table 3 shows the statistics of our whole dataset. Figure 2(a) shows the distribution of the number of users in a user's network which follows a Power Law distribution. Figure 2(b) shows the distribution of the number of fans of a user. It is surprised to see that this distribution also follows a Power Law distribution. As we know, the number of fans of a user cannot be determined by the user himself or herself. However, it seems that certain users are well known by other users, and it is interesting to investigate the characteristics of the well known users. Roughly, *expertise* of a user in Yahoo! Delicious can be interpreted from two aspects: the first is the quality of bookmarked resources and the second is the number of bookmarks. We measure the *expertise* of a user through the second aspect. Figure 2(c) demonstrates the relation between a user's number of bookmarks and his/her number of fans, and we can find there is a positive relationship. The reason why this happens is similar to why the Web portals become very popular and have plenty of visits every day. Note the role of users with extremely large number of bookmarks is very similar to the role of Web portals on the Internet, or called hubs (Kleinberg 1999).

Table 3: **Statistics of the crawled dataset.**

Users	Bookmarks	Network*	Fans**
366,827	49,692,497	425,069	395,415

* This is the total number of users in all users' personal networks.
** This is the total number of fans of all users.

Experimental Results

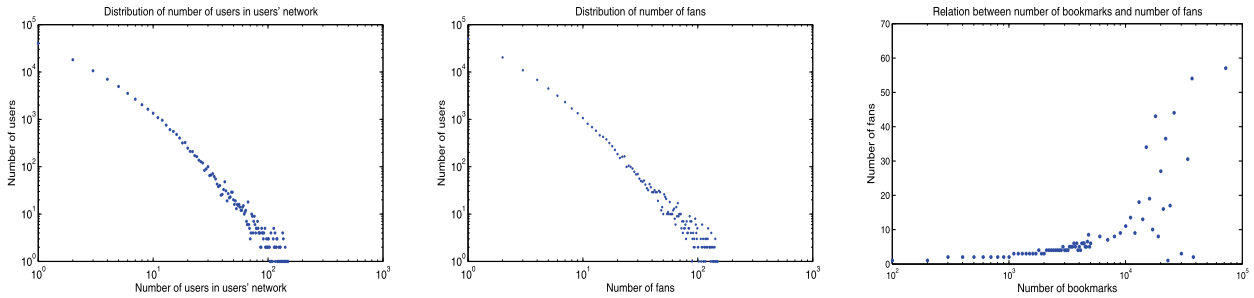
Two research questions are presented to give an idea of the highlights of our experimental analysis:

RQ1 Whether using tags is more effective than using URLs for recommender system approaches?

RQ2 How is our approach compared with the state-of-the-art recommender system approaches?

In order to investigate whether using tags is more effective than using URLs, we employ memory-based approaches and model-based approaches on both URLs and tags, and compare their performances. There are two memory-based approaches we employ, one is the Pearson Correlation Coefficient (PCC) method. The other memory-based approach we compare is the algorithm proposed by (Ma, King, and Lyu 2007). This is an effective PCC-based similarity calculation method with significance weighting, and we refer to it as PCCW. We set the parameter of PCCW to be 30 in our experiments. Two top-performing model-based recommendation algorithms are also employed, including Probabilistic Matrix Factorization (PMF) proposed by (Salakhutdinov and Mnih 2008), and Singular Value Decomposition (SVD) proposed by (Funk 2006). Both PMF and SVD employ matrix factorization approach to learn high quality low-dimensional feature matrices. After deriving the latent feature matrices, we still need to use memory-based approaches on derived latent feature matrices to perform the user recommendation task, and we employ both PCC and PCCW on latent feature matrices of SVD and PMF. We refer to them as SVD-PCC, SVD-PCCW, PMF-PCC, and PMF-PCCW respectively. We tune the dimension of latent matrices and set the optimal dimension value 10, and use five-folder cross-validation to learn the latent matrices for SVD and PMF.

It is shown that spreading interests within the *network* of Yahoo! Delicious users contribute a lot to the increase of popularity of resources (Wetzker, Zimmermann, and Bauckhage 2008). Thus, by crawling users' *network*, for each user in the test data, we consider users in his/her network share similar interests with him/her. In other words, users in a user's network is considered as relevant results in the user recommendation task. We randomly sample 400 users



(a) Distribution of Network. (b) Distribution of Fans. (c) Relation of # Bookmarks & # Fans.

Figure 2: **Statistics of users.**

Table 4: **Comparison With Approaches based on URLs (A Larger Value Means A Better Performance For Each Metric.)**

Metrics	Memory-Based Approaches		Model-Based Approaches				UserRec
	PCC	PCCW	SVD-PCC	SVD-PCCW	PMF-PCC	PMF-PCCW	
Precision@R	0.0717	0.1490	0.0886	0.0907	0.1136	0.1322	**0.3272**
MAP	0.1049	0.1874	0.1218	0.1245	0.1491	0.1745	**0.3752**
Bpref	0.0465	0.1148	0.0568	0.0582	0.0765	0.1029	**0.2913**
MMVRR	0.0626	0.1154	0.0710	0.0736	0.0858	0.1088	**0.2345**

whose number of users in their network is between 3 and 10, and further collect all the users in these 400 users' network resulting in 2,376 users in total. Then we crawl all these 2,376 users' bookmarks, and there are total 1,190,762 unique URLs and 139,707 unique tags.

We adopt four well known metrics that capture different aspects of the performance for the evaluation of the task, namely Precision at rank n (P@n), Precision at rank R(P@R), Mean Average Precision (MAP) and Bpref (Buckley and Voorhees 2004). In addition, we propose three novel metrics to help further evaluate the effectiveness of the proposed UserRec framework, namely Mean Multi-valued Reciprocal Rank (MMVRR), Top-K accuracy and Top-K recall. Multi-valued Reciprocal Rank (MVRR) is revised from the measure *reciprocal rank*. In our experimental scenario, the input of each measure is a user, and there are several relevant results for each user. We define $MVRR(u) = \sum_{i=1}^{N} \frac{1}{u_{r_i}}/N$, where u_{r_i} is the rank of a relevant result of user u, and N is total number of relevant results of user u. Mean Multi-valued Reciprocal Rank (MMVRR) is the mean value of MVRR in the test set. Top-K accuracy measures percentage of users who actually add at least one of the Top K-th recommended user in his/her network. Top-K recall measures percentage of people in users' network covered by top K recommended users.

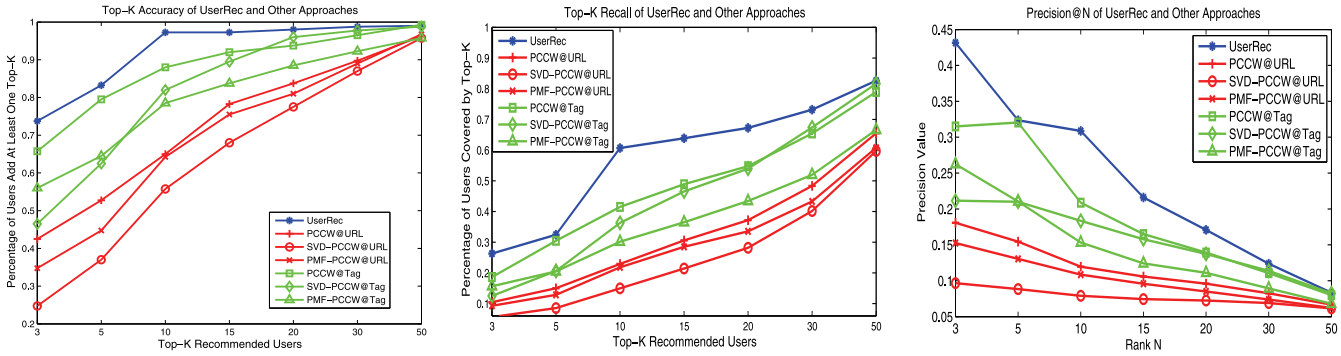
Table 4 demonstrates the results of metrics Precision@R, MAP, Bpref, and MMVRR of our method and other approaches when employing URLs. Table 5 shows the results of these metrics of our method and other approaches when employing tags. From Table 4 and Table 5, we can see that the proposed UserRec consistently outperforms other approaches on all these metrics. In addition, comparing results in Table 5 with results in Table 4, we can see that the same approaches achieve better performances when employing tags' information than when employing URLs' information, and this further confirms that tags can capture users' interests. In order to further compare the effectiveness of the proposed method with state-of-the-art approaches, and to further investigate whether employing tags can achieve better performances than employing URLs, we show Top-K accuracy, Top-K recall and Precision@N results of our method, and results of PCCW, SVD-PCCW, PMF-PCCW when employing on tags and URLs respectively. The results are shown in Fig. 3. From the results of Fig. 3, we can see the proposed UserRec method still outperforms other approaches in each metric, which is quite encouraging. In addition, we can find that the results of PCCW@Tag, SVD-PCCW@Tag, and PMF-PCCW@Tag are better than PCCW@URL, SVD-PCCW@URL and PMF-PCCW@URL respectively. From these three metrics, we can again confirm that tags are quite good resources to characterize users' interests.

Conclusions and Future Work

In this paper, we propose an effective framework for users' interest modeling and interest-based user recommendation in social tagging systems, which can help information sharing among users with similar interests. Specifically, we analyze the *network* and *fans* properties, and we observe an interesting finding that the role of users have similar properties with Web pages on the Internet. Experiments on a real world dataset show encouraging results of UserRec compared with the state-of-the-art recommendation algorithms. In addition,

Table 5: **Comparison With Approaches based on Tags (A Larger Value Means A Better Performance For Each Metric.)**

Metrics	Memory-Based Approaches		Model-Based Approaches				UserRec
	PCC	PCCW	SVD-PCC	SVD-PCCW	PMF-PCC	PMF-PCCW	
Precision@R	0.1495	0.3168	0.1540	0.2042	0.1875	0.2084	**0.3272**
MAP	0.1816	0.3444	0.1898	0.2469	0.2084	0.2440	**0.3752**
Bpref	0.1132	0.2395	0.1170	0.1479	0.1376	0.1707	**0.2913**
MMVRR	0.1129	0.1943	0.1151	0.1397	0.1300	0.1550	**0.2345**



(a) Top-K Accuracy. (b) Top-K Recall. (c) Precision@N.

Figure 3: **Performance comparison of UserRec and Other Approaches.**

experimental results also confirm that tags are good at capturing users' interests. In future work we would like to extend our methods and develop a more robust framework that can handle the *tag ambiguity* problem. Moreover, we plan to investigate how information, such as URLs and tags, is propagated in the social tagging systems.

Acknowledgments

The work described in this paper was fully supported by two grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CUHK4128/08E and CUHK4154/09E).

References

Breese, J. S.; Heckerman, D.; and Kadie, C. 1998. Empirical analysis of predictive algorithms for collaborative filtering. In *UAI*, 43–52.

Buckley, C., and Voorhees, E. 2004. Retrieval evaluation with incomplete information. In *SIGIR*, 25–32.

Clauset, A.; Newman, M. E. J.; and Moore, C. 2004. Finding community structure in very large networks. *Physical Review E* 70(6):066111+.

Funk, S. 2006. Netflix update: Try this at home. Technical report, sifter.org/simon/journal/ 20061211.html.

Herlocker, J. L.; Konstan, J. A.; Borchers, A.; and Riedl, J. 1999. An algorithmic framework for performing collaborative filtering. In *SIGIR*, 230–237.

Heymann, P.; Koutrika, G.; and Garcia-Molina, H. 2008. Can social bookmarking improve web search? In *WSDM*, 195–206.

Hofmann, T. 2004. Latent semantic models for collaborative filtering. *TOIS* 22(1):89–115.

Kleinberg, J. 1999. Authoritative sources in a hyperlinked environment. *JACM* 46(5):604–632.

Li, X.; Guo, L.; and Zhao, Y. E. 2008. Tag-based social interest discovery. In *WWW*, 675–684.

Ma, H.; King, I.; and Lyu, M. R. 2007. Effective missing data prediction for collaborative filtering. In *SIGIR*, 39–46.

Ma, H.; King, I.; and Lyu, M. R. 2009. Learning to recommend with social trust ensemble. In *SIGIR*, 203–210.

Newman, M. E. J. 2004. Analysis of weighted networks. *Physical Review E* 70(5):056131+.

Resnick, P.; Iacovou, N.; Suchak, M.; Bergstrom, P.; and Riedl, J. 1994. Grouplens: An open architecture for collaborative filtering of netnews. In *CSCW*, 175–186.

Salakhutdinov, R., and Mnih, A. 2008. Probabilistic matrix factorization. *NIPS* 20:1257–1264.

Si, X.; Liu, Z.; Li, P.; Jiang, Q.; and Sun, M. 2009. Content-based and graph-based tag suggestion. In *Proceedings of the ECML PKDD Discovery Challenge*, 243–260.

Suchanek, F. M.; Vojnovic, M.; and Gunawardena, D. 2008. Social tags: meaning and suggestions. In *CIKM*, 223–232.

Wetzker, R.; Zimmermann, C.; and Bauckhage, C. 2008. Analyzing social bookmarking systems: A del.icio.us cookbook. In *Proceedings of Mining Social Data Workshop on ECAI*, 26–30.

Zhou, T. C.; Ma, H.; King, I.; and Lyu, M. R. 2009. Tagrec: Leveraging tagging wisdom for recommendation. In *Proceedings of IEEE International Symposium on Social Intelligence and Networking*.