Exploring Social Annotations with the Application to Web Page Recommendation

ARTICLE *in* JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY · NOVEMBER 2009

Impact Factor: 0.67 · DOI: 10.1007/s11390-009-9292-6 · Source: DBLP

Daniel Dajun Zeng
Chinese Academy of Sciences

jj Wang
The Hong Kong University of Science and Tec…

Wenji Mao
Chinese Academy of Sciences

Li HQ, Xia F, Zeng D *et al.* Exploring social annotations with the application to Web page recommendation. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 24(6): 1028–1034 Nov. 2009

Exploring Social Annotations with the Application to Web Page Recommendation

Hui-Qian Li[1] (李慧倩), Fen Xia[1] (夏 粉), Daniel Zeng[1,2] (曾大军), *Senior Member*, *IEEE*
Fei-Yue Wang[1] (王飞跃), *Senior Member*, *CCF*, *Fellow*, *IEEE*
and Wen-Ji Mao[1] (毛文吉), *Senior Member*, *CCF*, *Member*, *ACM*

[1] *Key Laboratory of Complex Systems and Intelligence Science, Institute of Automation, Chinese Academy of Sciences Beijing 100190, China*

[2] *Department of Management Information Systems, University of Arizona, Tucson AZ 85721, U.S.A.*

E-mail: {huiqian.li, fen.xia, dajun.zeng, feiyue.wang, wenji.mao}@ia.ac.cn

Received June 9, 2009; revised July 17, 2009.

Abstract Collaborative social annotation systems allow users to record and share their original keywords or tag attachments to Web resources such as Web pages, photos, or videos. These annotations are a method for organizing and labeling information. They have the potential to help users navigate the Web and locate the needed resources. However, since annotations are posted by users under no central control, there exist problems such as spam and synonymous annotations. To efficiently use annotation information to facilitate knowledge discovery from the Web, it is advantageous if we organize social annotations from semantic perspective and embed them into algorithms for knowledge discovery. This inspires the Web page recommendation with annotations, in which users and Web pages are clustered so that semantically similar items can be related. In this paper we propose four graphic models which cluster users, Web pages and annotations and recommend Web pages for given users by assigning items to the right cluster first. The algorithms are then compared to the classical collaborative filtering recommendation method on a real-world data set. Our result indicates that the graphic models provide better recommendation performance and are robust to fit for the real applications.

Keywords graphic model, EM (expectation-maximization), social annotation, tag, recommendation

1 Introduction

Collaborative social annotation, which is also called social collaborative tag, is a classic representation of Web 2.0 application. An annotation, typically a single word, describes Web resources, such as Web pages, photos and videos. At the same time, annotations provide useful information about the users who annotate the resources, such as their interests and preferences. Both industry and academia consider social annotation as a potential useful measure for Web knowledge discovery, search and share.

In recent years, some researchers have showed great interests on a systematic study on the usefulness of social annotations[1−10]. Some pioneer researchers debated on the positive and negative properties of customer generated annotations. Macgregor and McCulloch[3] found that a large fraction of annotations on Flickr and del.icio.us, which are two of the largest social annotation systems, are erroneous, plural forms and various symbols. Kipp and Campbell[2] pointed out that many annotations are synonymous and highly related. In other words, it is difficult for people to find potential resources which may be annotated with synonyms or even erroneous tags. However, Paolillo and Penumarthy's study[4] indicates that annotations and users are highly clustered so that synonymous annotations can be classified into the same group. On the perspective of technology, one problem emerges: how to facilitate knowledge discovery over synonymous annotations and through similar users.

The current typical way to facilitate users in finding potential resources on Web is to recommend resources according to their fetching histories or more specifically, to their interests. The most studied recommendation method is collaborative filtering recommendation, which recommend items to users by mining the association information between similar users or items. The research work can be summarized as two primary perspectives: one is memory based

Regular Paper

This work is supported in part by the National Natural Science Foundation of China under Grant Nos. 60621001, 60875028, 60875049, and 70890084, the Chinese Ministry of Science and Technology under Grant No. 2006AA010106, and the Chinese Academy of Sciences under Grant Nos. 2F05N01, 2F08N03 and 2F07C01.

collaborative recommendation[11−13] and the other is model based collaborative recommendation[14−16].

In recent literature, some researchers modified the memory based collaborative filtering recommendation algorithms by associating them with annotations in order to facilitate active Web page searching. For example, Tso-Sutter *et al.*[17] proposed a generic method that allows annotations to be incorporated into standard collaborative filtering algorithms by reducing three-dimensional correlations to three two-dimensional correlations and then applying a fusion method to re-associate the adapted correlations. However, such methods only superficially use social annotations, leaving the problems such as synonymous annotations unsolved.

Some elementary work has also been done to incorporate social annotations into model-based methods. Wu *et al.*[18] proposed a method to develop ontologies statistically from collaborative social annotations by applying them to search for semantically similar Web pages. Inspired by this work, Plangprasopchok and Lerman[19] proposed the Interest-Topic Model, where users and Web resources are clustered separately indicating user interests and Web resources topics. This work achieves better performance than Wu's work since it is consistent with the observations by Paolillo and Penumarthy[4]: Web pages, users and annotations usually have different group patterns. But none of their work explains why the annotations can promote the performance of the model.

In this paper, we explore the usefulness of social annotations with its application to Web page recommendation. We mainly concern with three problems. First, we need to decide the pattern into which users, annotations and Web pages can be semantically grouped. Second, we need to justify how annotations can help locating potential resources. Last, we need to explore the different patterns of users online surfing and posting annotations, e.g., they do find potential resources through annotations or they just use annotations as descriptions after finding out the resources. Thus we propose four different graphic models representing four different patterns for user visit and annotation as well as clustering patterns. The models are then applied to Web page recommendation so that user annotation activities can be used to facilitate locating potential interesting Web pages. We compared the Web page recommendation performance of our models with that of the traditional collaborative filtering recommendation method. The real-data experiments show that our graphic models achieve far better recommendation quality than the collaborative filtering methods.

In Section 2, we introduce the Bayes based Web page recommendation and our graphic models. Section 3 presents our real-data experiments with comparison to the classical collaborative filtering methods. In Section 4, we discuss the potential underlying principles that lead to the experimental results. Section 5 summaries this paper and proposes our future works.

2 Exploring Social Annotations from Recommendation Perspective

Let us first introduce notions for formally describing the collaborative social annotation information. Let $X = \{x_1, x_2, \ldots, x_M\}$ be a set of Web pages, $Y = \{y_1, y_2, \ldots, y_N\}$ be a set of users, $T = \{t_1, t_2, \ldots, t_T\}$ be a set of annotations. Let $S = \{(x_{(1)}, y_{(1)}, t_{(1)}), \ldots, (x_{(L)}, y_{(L)}, t_{(L)})\}$ be the data set that consists of all tuples of Web pages, users and annotations, each of which denotes that a given user posts an annotation to a specific Web page. In a potential resource locating target, the goal is to predict the Web page x that a target user y would give to collect given the training data set including the tuples associated with the user.

Traditionally, from Bayesian view, the central task of Web page recommendation is to estimate the conditional probability $P(x|y)$, which represents the likelihood that a given user y collects the Web page x. Since we now have the information of annotations, we can predict the interest t of a given user y on a Web page x. So we need to estimate the probability $P(x, t|y)$ and further calculate the probability $P(x|y)$ over all the interests t, as $\Sigma_t P(x, t|y)$.

As reported by Paolillo and Penumarthy[4], there are apparent group patterns for users and annotations in collaborative annotation systems. This indicates that the users who have interest in some specific pages may have interest in the collections of other users who also have interest in the same specific pages. This mechanism is adopted to the user-based collaborative filtering recommendation methods, one of memory-based method which recommend an item to a user according to the similar users. However, it is difficult for the memory-based collaborative filtering method to consider similar users, items and even annotations, or the combination is trivial.

For probability models, it is facilitated to explore user, Web page and annotation similarity through clustering, which also handles the data parse problem at the same time. Specifically, for graphic models, clusters are incorporated as hidden variables. Since we have tree types of data with our social annotation data set, we can generate three cluster sets: the Web page cluster set $Z_x = \{z_{x_1}, z_{x_2}, \ldots, z_{x_K}\}$, the user cluster set $Z_y = \{z_{y_1}, z_{y_2}, \ldots, z_{y_H}\}$, and the annotation cluster set $Z_t = \{z_{t_1}, z_{t_2}, \ldots, z_{t_D}\}$. The mechanism of the graphic

model with clusters as extra variables is as follows: for any given item, a Web page, user or annotation, a cluster should be assigned so that the item can be presented by the cluster that it is assigned to. For instance, for a given user, he or she should be assigned to a user cluster so that we can estimate the potential Web page the user would have interest in by estimating the potential interest of the cluster. Since the cluster which consists of similar users would provide more information than a single user, the estimation can be more accurate.

The probability graphic model not only explore similarity information more easily than memory based method, but also explicitly explain the annotation information. In graphic models, users, Web pages and annotations can be clustered separately, so that the pattern by which users, Web pages and annotations are coupled can be set to be different. The pattern they are coupled explicitly represent the activities that users follow when collecting and annotating in social annotation systems. Since graphic model can also be applied with a recommendation task, the user activity pattern is estimated by the performance of Web page recommendation.

Thus we map the different relationships between users, Web pages, annotations and the respective clusters into graphic models and conduct Web page recommendation to make further estimation, which we describe in the following subsection.

2.1 User Annotation Models

The issue we need to address in this subsection is how users, Web pages, annotations and their clusters are coupled to form a tuple of $P(x,t|y)$. More explicitly, what is the user's activity pattern when they make collections and annotations in social annotation systems.

As a common sense, users mostly follow one of the two patterns when visiting social annotation systems: pick up an interesting Web page and post it with annotations or search for the potential resources with specific annotations; and make selections from the feedbacks. The two patterns could be reflected by the two different couple patterns of how user clusters, Web page clusters and annotation clusters are clustered. Another difference can exist in whether a term should be clustered, for instance, whether to cluster annotations. As a result, we propose four graphic models that represent different couple patterns and whether to cluster annotations.

The first model is shown in Fig.1, where z_y denotes a cluster of user, z_x denotes a cluster of Web page and z_t denotes a cluster of annotation.

The generation of the tuple of (x,t,y) is modeled by the following process:

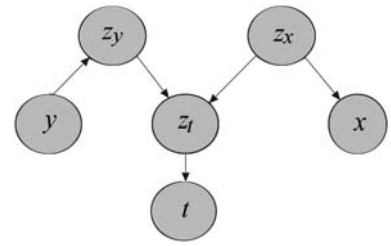

Fig.1. Graphic Model 1.

1) Given a user y, the probability that y belongs to a cluster z_y is estimated, represented by $P(z_y|y)$.

2) Then for any given Web page cluster z_x, the probability of collecting a Web page x from this cluster is estimated, represented by $P(x|z_x)$.

3) The cluster z_y and z_x determine the probability that a cluster of annotation z_t will be selected, which finally gives a estimation of annotation t.

The final probability estimation is given by the following equation:

$$P(x,t|y) = \Sigma_{Z_x,Z_y,Z_t} P(z_y|y)P(z_x) P(x|z_x)P(z_t|z_x,z_y)P(t|z_t). \quad (1)$$

This model is constructed as the supposed user surfing activities: a user saves an interesting Web page and gives it an annotation. All the terms, users, Web pages and annotations are clustered. But different users may have different surfing habits and clusters may not be necessary for some terms. So we propose three different models to present different habits and cluster patterns.

The second model is shown in Fig.2. This model is different from Model 1 as this model do not cluster for annotations. The annotation t is directly determined by a given cluster of user y and a given cluster of Web page x. So the final probability estimation is given by the following equation:

$$P(x,t|y) = \Sigma_{Z_x,Z_y,Z_t} P(z_y|y)P(z_x)P(x|z_x)P(t|z_x,z_y). \quad (2)$$

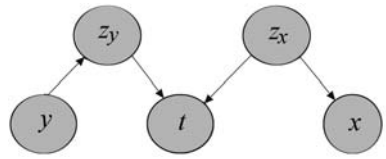

Fig.2. Graphic Model 2.

This model is similar to the Interest-Topic Model proposed by Plangprasopchok and Lerman[19]. But they are different in two aspects. First, the estimation of the optimal solution is different. The reason for this is because for our model the Web pages are determined given a specific Web page cluster, whereas for

the Interest-Topic Model the Web page cluster is determined after a specific Web page is selected. Second, our model emphasizes on the user's activity by describing the order in which users generate a user-Web page-tag tuple, while the target of Interest-Topic Model is to combine different user interests and item topics so as to generate a user-item-tag tuple.

The third model is shown in Fig.3.

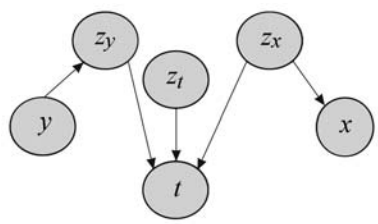

Fig.3. Graphic Model 3.

In this model, the annotation t is determined by all the clusters of users, Web pages and annotations, as z_y, z_x and z_t. The final probability estimation is given by the following equation:

$$P(x,t|y) = \Sigma_{Z_x,Z_y,Z_t} P(z_y|y)P(z_x)P(x|z_x) P(t|z_t,z_x,z_y)P(z_t). \quad (3)$$

The forth model is shown in Fig.4. This one is very different from the previous models in that the Web page cluster z_x is determined by the cluster of user z_y and the cluster of annotation z_t. This indicates that a given user firstly determines his/her interest, which is represented by annotation cluster z_t, and then chooses a Web page within the range of his/her interests. The final probability estimation is given by the following equation:

$$P(x,t|y) = \Sigma_{Z_x,Z_y,Z_t} P(z_y|y)P(z_t|z_y)P(t|z_t)P(z_x|z_t,z_y)P(x|z_x). \quad (4)$$

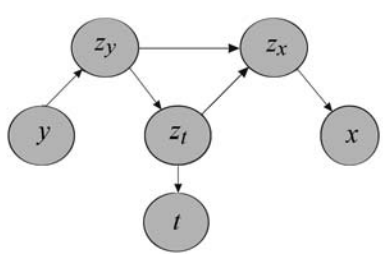

Fig.4. Graphic Model 4.

2.2 Probability Estimation

We adopt classical EM (Expectation-Maximization) algorithm to iteratively search for the optimal estimation. We take Model 1 as an instance. The log-likelihood is given below:

$$L = \log P(x,t|y) = \log \Sigma_{Z_x,Z_y,Z_t} P(z_y|y)P(z_x) P(x|z_x)P(z_t|z_x,z_y)P(t|z_t). \quad (5)$$

The parameters are $P(z_y|y)$, $P(z_x)$, $P(x|z_x)$, $P(z_t|z_x,z_y)$, $P(t|z_t)$, which should be fixed during the iteration. At the beginning of the algorithm, we initiate the parameters randomly. At step E, we estimate the joint probability of the hidden variable (z_x, z_y, z_t) using the current parameters, as shown by this equation:

$$P(z_x,z_y,z_t|x,y,t) = \frac{P(z_y|y)P(z_x)P(x|z_x)P(z_t|z_x,z_y)P(t|z_t)}{\Sigma_{Z_x,Z_y,Z_t} P(z_y|y)P(z_x)P(x|z_x)P(z_t|z_x,z_y)P(t|z_t)}. \quad (6)$$

At step M, the log-likelihood is maximized according to the current distribution of the hidden variables fixed in step E:

$$\arg\max_{\theta} \sum_{l=1}^{L} \sum_{Z_x,Z_y,Z_t} P(z_x,z_y,z_t|x,y,t) \lg P(z_y|y)P(z_x)P(x|z_x)P(z_t|z_x,z_y)P(t|z_t). \quad (7)$$

So we get:

$$\begin{aligned} P(z_x) &= \sum_l \sum_{z_y,z_t} \frac{P(z_x,z_y,z_t|x_l,y_l,t_l)}{L} \\ P(x,|z_x) &= \sum_{l:x_l=x} \sum_{z_y,z_t} \frac{P(z_x,z_y,z_t|x_l,y_l,t_l)}{L \times P(z_x)} \\ P(z_y|y) &= \frac{\sum_{l:y_l=y} \sum_{z_x,z_t} P(z_x,z_y,z_t|x_l,y_l,t_l)}{\sum_{l:y_l=y} \sum_{z_x,z_y,z_t} P(z_x,z_y,z_t|x_l,y_l,t_l)} \\ P(t|z_t) &= \frac{\sum_{l:t_l=t} \sum_{z_x,z_y} P(z_x,z_y,z_t|x_l,y_l,t_l)}{\sum_l \sum_{z_x,z_y,z_t} P(z_x,z_y,z_t|x_l,y_l,t_l)} \\ P(z_t|z_x,z_y) &= \frac{\sum_l P(z_x,z_y,z_t|x_l,y_l,t_l)}{\sum_l \sum_{z_t} P(z_x,z_y,z_t|x_l,y_l,t_l)}. \end{aligned} \quad (8)$$

The steps E and M will iterate until the value of log-likelihood does not change any more.

3 Experimental Results

To evaluate each proposed models, in this section, we present our recommendation results by experiment. The data set was obtained by crawling the most popular social annotation Web site, del.icio.us. Crawlers followed a breadth-first strategy. First, a user was randomly selected as the seed. The crawlers fetched all

the Web pages and corresponding tags that the user had collected. Then the crawlers retrieved the user's friend network and followed the network connections to the user's friends. This iteration continued until a specified user collection size is matched. In our experiment, in order to reduce the amount of raw data, we filtered out users who saved much less than 100 pages and employed less than 10 annotations. For Web pages associated with the users after data filtration, we preserved only those pages collected by at least 10 users. Tags were evaluated in the same fashion. Table 1 summarizes the data sets.

Table 1. Summary of del.icio.us Data Sets

No. User	No. Page	No. Annotation	No. Transaction
612	1676	889	46843

We used a five-cross-validation to evaluate the algorithms, and generated the top-N recommendations based on the potential estimate values obtained from the algorithms. Four widely-accepted performance measurements: precision, recall, F-measure and rank score[20] were used to evaluate our recommendation results. For the rank score, we used $h = 10$ as the half life parameter. We calculated the performance levels for the top-5, top-10 and top-20 Web page recommendations. We have only reported our evaluation results for the precision of the top-10 recommendation as a representative example, and similar results were obtained for the top-5 and the top-20 recommendations.

The memory and time consuming for EM on large scale data set is known to be very high. In our experiments, we adjusted the scale of clusters to restrict the memory and time consuming within a tolerable range while still maintaining a considerable recommendation performance. The recommendation performance under variable cluster numbers denotes that when $|z_y| > 5$, $|z_x| > 5$ and $|z_t| > 10$, the recommendation precision and recall can be hardly improved. Consequently, our final experiments were implemented with the number of users set to $|z_y| = 5$, the number of Web pages set to $|z_x| = 5$ and the number of annotations set to $|z_t| = 10$.

The recommendation performance is compared with the two most widely used traditional collaborative filtering (CF in short) methods: the user-based method and the item-based method[21] and two modified CF methods[22] with annotation information involved as attributes.

Table 2 summarizes our experimental results of precision. We can see that the graphical model based recommendations listed on the right side of Table 2, achieve better performance than the classical CF methods on the left side. We set similar results for rank score, which are omitted here.

Table 2. Recommendation Performance of All Algorithms

Method	Precision	Method	Precision
Traditional user-based	0.0346	Model 1	0.0435
Annotation user-based	0.0284	Model 2	0.0701
Traditional item-based	0.0279	Model 3	0.0716
Annotation item-based	0.0363	Model 4	0.0662

4 Discussions

There are several potential underlying principles which can account for our experimental results. As shown in Table 2, Model 1 has a lower recommendation quality compared with the other models. Particularly, Model 1 is less effect than Model 2 which is simpler than Model 1. The difference between them is that Model 2 does not cluster for annotations. We have noticed that for the same Web page different users may give semantically completely different annotations, as shown in Fig.5.

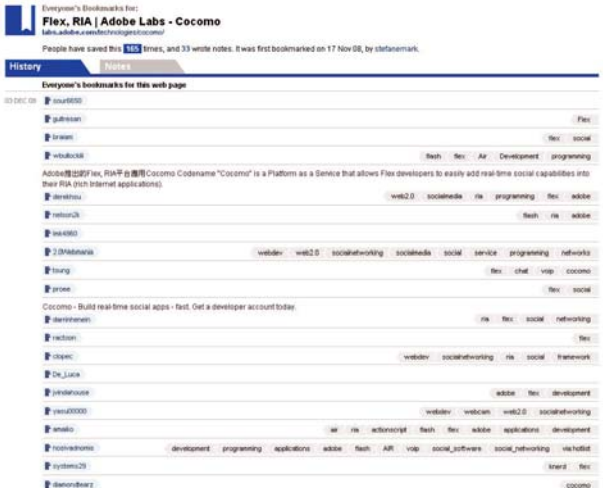

Fig.5. A snapshot of del.icio.us.

Thus a given cluster of annotations may consist of semantically different annotations because they have been posted to the same Web pages. As a result, if we use a single cluster of annotation, we may assign a wrong annotation to a specific Web page given a specific user because it has co-occurred with the right annotation on another Web page by other users. So if we directly choose the annotation instead of a cluster of them, as what done by Model 2, we get a better performance. Moreover, if we combine the information of the user, Web page and the annotation clusters to determine an annotation, we get the best performance. In other words, we should use the transactional information of all the clusters to determine a user, Web page and the annotation tuple.

We also find that the cluster number of users and Web pages has little effect on the recommendation performance. We have conducted experiments with the fixed number of annotations and the number of clusters of users and Web pages being 5, 10 and 20. The results only show little differences. That is to say that we can set the cluster numbers of users and Web pages considerably small and at the same time still keep relatively high performance. This makes the models very robust and easy to be adjusted.

We assume that if a given model is consistent with real user surfing and annotating activities, it is likely to achieve good recommendation performance. Under this assumption, our experimental results show that Model 3 can make more accurate recommendation. This indicates that a user selects a specific Web page from a given cluster and then gives it an annotation from a cluster of potential available annotations. This pattern also indicates that users treat annotations more as self-organized descriptions than a measure of resource discovery as in Model 4, where users first determine the annotations and then choose the Web pages connected with the considered annotations. But this model does not outperform Model 2 and Model 3, where annotations are assumed as self-descriptions of user collections. This observation also provides insight into system design: how the collaborative annotation systems could be designed in order to encourage users to trace Web resources following annotations, which may include user interface and background mechanisms.

5 Conclusion and Future Work

In this paper, we conducted an empirical study on the collaborative social annotations and applied it to the Web page recommendation. We generate semantic clusters for social annotations and also clusters for users and Web pages. Based on the transactional information of users, Web pages and annotations and the cluster information, we propose four graphic models which represent different mechanisms for a given user to choose an interested Web page and post it with an annotation.

We adopted these four models to generate Web page recommendation for each user. The experimental result shows that all the models achieve better performance than the classical CF recommendation method. We also find that using all the transactional information of the clusters will achieve the best recommendation quality. These models are quite robust in that the number of user clusters and the number of Web page clusters have little effect on the final recommendation.

We will further explore relationships among users based on the concept of semantic clustering of social annotations in the graphic models. We expect to gain better understanding of the user groups using their annotation information.

References

[1] George W. Furnas, Caterina Fake, Luis von Ahn, Joshua Schachter, Scott Golder, Kevin Fox, Marc Davis, Cameron Marlow, Mor Naaman. Why do tagging systems work? In *Proceedings of Conference on Human Factors in Computing Systems*, Montreal, Canada, April 22–27, 2006, pp.36–39.

[2] Margaret E I Kipp, D Grant Campbell. Patterns and inconsistencies in collaborative tagging systems: An examination of tagging practices. In *Proceedings of American Society for Information Science and Technology*, Austin, USA, Nov. 5–9, 2007.

[3] George Macgregor, Emma McCulloch. Collaborative tagging as a knowledge organization and resource discovery tool. *Library Review*, 2006, 55(5): 291–300.

[4] John C Paolillo, Shashikant Penumarthy. The social structure of tagging Internet video on Del.icio.us. In *Proc. Hawaii International Conference on System Sciences*, Hawaii, USA, Jan. 3–6, 2007, pp.85b.

[5] Hotho A, Jaschke R, Schmitz C, Stumme G. Trend detection in folksonomies. In *Proc. the First International Conference on Semantics And Digital Media Technology*, Athens, Greece, Dec. 6–8, 2006, pp.56–77.

[6] Mika P. Ontologies are us: A unified model of social networks and semantics. In *Proc. International Semantic Web Conference*, Galway, Ireland, Nov. 6–10, 2005, pp.5–15.

[7] Noruzi A. Editorial. Folksonomies: Why do we need controlled vocabulary? *Webology*, 2007, 4(2).

[8] Macgregor G, McCulloch E. Collaborative tagging as a knowledge organisation and resource discovery tool. *Library Review*, 2006, 55(5): 291–300.

[9] Paul Heymann, Georgia Koutrika, Hector Garcia-Molina. Can social bookmarking improve Web search? In *Proc. the International Conference on Web Search and Web Data Mining*, Palo Alto, USA, Feb. 11–12, 2008, pp.195–206.

[10] Ames M, Naaman M. Why we tag: Motivations for annotation in mobile and online media. In *Proc. the SIGCHI Conference on Human Factors in Computing Systems*, San Jose, USA, April 28–May 3, 2007, pp.971–980.

[11] Breese J S, Heckerman D, Kadie C. Empirical analysis of predictive algorithms for collaborative filtering. In *Proc. UAI*, Madison, USA, July 24–26, 1998, pp.43–52.

[12] Herlocker J L, Konstan J A, Riedl B J. An algorithmic framework for performing collaborative filtering. In *Proc. the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Berkeley, USA, Aug. 15–19, 1999, pp.230–237.

[13] Wang J, de Vries A P, Reinders M J T. Unifying user-based and item-based collaborative filtering approaches by similarity fusion. In *Proc. the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Seattle, USA, Aug. 6–11, 2006, pp.501–508.

[14] Zhang T, Iyengar V S, Kaelbling P. Recommender systems using linear classifiers. *Journal of Machine Learning Research*, 2002, 2(2): 313–334.

[15] Jin R, Si L, Zhai C. A study of mixture models for collaborative filtering. *Information Retrieval*, 2006, 9(3): 357–382.

[16] Hofmann T. Latent semantic models for collaborative filtering. *ACM Transactions on Information Systems*, 2004, 22(1): 89–115.

[17] Karen H L Tso-Sutter, Leandro Balby, Marinho, Lars, Schmidt-Thieme. Tag-aware recommender systems by fusion of collaborative filtering algorithms. In *Proc. 2008 ACM*

Symposium on Applied Computing, Fortaleza, Brazil, March 16–20, 2008, pp.1995–1999.

[18] Xian Wu, Lei Zhang, Yong Yu. Exploring social annotations for the Semantic Web. In *Proc. the 15th International Conference on World Wide Web*, Edinburgh, Scotland, May 23–26, 2006, pp.417–426.

[19] Anon Plangprasopchok, Kristina Lerman. Exploiting social annotation for automatic resource discovery. In *AAAI Workshop on Information Integration from The Web*, Vancouver, Canada, July 22–23, 2007, pp.86–91.

[20] Breese J S, Heckerman D, Kadie C. Empirical analysis of predictive algorithms for collaborative filtering. In *Proc. UAI 1998*, Madison, USA, July 24–26, 1998, pp.43–52.

[21] Huang Z, Zeng D D, Chen H. Analyzing consumer-product graphs: Empirical findings and applications in recommender systems. *Management Science*, 53(7): 1146–1164.

[22] Zeng D, Li H. How useful are tags? – An empirical analysis of collaborative tagging for Web page recommendation. In *Proc. IEEE Intelligence and Security Informatics 2008 International Workshops*, 2008, pp.320–330.

Hui-Qian Li is a research scientist in the Key Lab of Complex Systems and Intelligent Science at Institute of Automation, Chinese Academy of Sciences. Her research interests include complex networks, recommendation systems, statistical machine learning.

Fen Xia is an assistant professor in the Key Lab of Complex Systems and Intelligent Science at Institute of Automation, Chinese Academy of Sciences. His research interests include statistical machine learning, ranking, regularization methods, efficient algorithms, and information retrieval.

Daniel Zeng received his Ph.D. degree in industrial administration from Carnegie Mellon University in 1998. He is a research professor at the Institute of Automation, Chinese Academy of Sciences. He is also affiliated with the University of Arizona as an associate professor and the director of the Intelligent Systems and Decisions Laboratory. Prof. Zeng is a senior member of IEEE. His research interests include software agents and multi-agent systems, intelligence and security informatics, social computing and recommender systems.

Fei-Yue Wang received his Ph.D. degree in computer and systems engineering from Rensselaer Polytechnic Institute in 1990. He is the director of the Key Laboratory of Complex Systems and Intelligence Science at the Chinese Academy of Sciences. He is also a professor in the University of Arizona's Systems & Industrial Engineering Department and the director of the university's Program in Advanced Research of Complex Systems. Prof. Wang is a fellow of IEEE, INCOSE, IFAC, ASME, and AAAS. His current research interests include social computing, Web and services science, modeling, analysis and control of complex systems, especially social and cyber-physical systems.

Wen-Ji Mao received her Ph.D. degree in computer science from the University of Southern California in 2006. She is an associate professor at the Institute of Automation, Chinese Academy of Sciences. Prof. Mao is a member of ACM and AAAI, and a senior member of the China Computer Federation. Her research interests include artificial intelligence, multi-agent systems and social modeling.