

A Collaborative Filtering Recommendation System Combining Semantics and Bayesian Reasoning

Jialing Li Li Li* Xiao Wen Jianwei Liao

Faculty of Computer and Information Science
Southwest University, Chongqing 400715, China
Email: {lj1333, lily, swnuwx}@swu.edu.cn

Abstract

Tag-based recommendation systems aim to improve the search experience of the end users. However, due to different backgrounds of the end users, descriptions of the same resources may be totally different in particle size and degree of specialization, which raises the question of how to tackle the growing discrepancy of public taxonomies (Folksonomy) in the social networks. In line with this, WordNet-based similarity is used to obtain semantic distance between tags and topic categories in order to reduce the divergence of tags. This in turn improves the search accuracy. The Bayesian reasoning is introduced to infer users' preferences through mining users' comments towards particular categories. Users' interaction behavior, which may facilitate preference estimation, is considered as well to enhance search efficiency. A series of experiments are conducted based on Flickr and Delicious datasets. The results show that the proposed recommendation algorithm can effectively improve search precision and provide a greater level of user satisfaction.

Keywords: Recommendation system, Collaborative filtering, Bayesian theory, Semantic similarity

1 Introduction

Overabundant growth of information on the Internet has raised many issues such as how to effectively manage vast amounts of heterogenous data, and how to help end users find resources according to their tastes with both accuracy and efficiency. Tags, as a kind of metadata, are used to address data heterogenous issues. Folksonomy formed by user tagging in social networking services such as Flickr¹, Delicious², etc., helps to classify and manage social data repositories in a uniform way. Tags also build a bridge between users and items, and can be used to improve search mechanisms (Karen et al. 2008). For taste matching issues, in addition to the familiar keyword searching technique, recommendation systems appear to complement potential interesting information with corresponding recipients (Resnick & Varian 1997).

The key point of recommendation is to capture users' preference as precisely as possible. However, in-

complete user preference modeling, preference learning and result presentation may hinder the system from achieving its full potential. In fact, there exists several discussions about what factors should be considered when modeling users' preference. Xu (Xu et al. 2011) model the user preference on various topics in a Topic Oriented Graph, and devise a topic-oriented tag-based recommendation system by preference propagation. Experiments show the approach outperforms several state-of-the-art collaborative filtering methods, yet it does not take users' contacts into account. Lemman's work (Lerman 2006) clearly shows that users' contact information can effectively improve search precision. In this paper, we build user's personalized file according to the content item liked by the user and topics liked by his/her friends, and then sort query results according to the file to improve user satisfaction.

Three research questions need to be answered in order to achieve a better recommendation result regarding the user's taste against a topic of an appointed contact, they are: 1) How to attain the topic-item relation? 2) How to learn the user's preference from his/her past rating behaviour? 3) How to cater the user by returning his/her most favourite items? This paper aims to address the above problems. The contributions of the paper include:

- Category identification problems tackled using semantics and category repositories;
- Bayesian probabilistic reasoning used to learn user's preference in social network tagging systems;
- Resulting items ranked accordingly to meet user's taste mostly.

The remainder of the paper is organized as follows. Section 2 is the problem definition. Section 3 details the proposed mining method, including Bayesian reasoning and item category identification with similarity calculation. Section 4 presents experimental results based on two social network datasets. Finally, Section 5 is the related works and Section 6 concludes the paper.

2 Terminology Definition

Firstly, suppose u_a is the centric user. Let $U = \{u_1, u_2, \dots, u_n\}$ denote the contact cluster of u_a . Then the items of u_i are presented as $I(u_i) = \{i_1, i_2, \dots, i_m\}$, the tags of $i_s, i_s \in I(u_i)$, are expressed as $T(u_i, i_s) = \{t_1, t_2, \dots, t_k\}$, topics of items are categorized into $C = \{c_1, c_2, \dots, c_r\}$. The organization relationship among them is shown in Fig. 1. Let us focus on the "interaction behaviour". Each item can be marked, commented or be liked by other users, which may include u_a . Thus we can mine u_a 's taste from its rating history, taking into account that the items u_a rated

*Corresponding author

Copyright ©2012, Australian Computer Society, Inc. This paper appeared at the 10th Australasian Data Mining Conference (AusDM 2012), Sydney, Australia, December 2012. Conferences in Research and Practice in Information Technology (CRPIT), Vol. 134, Yanchang Zhao, Jiuyong Li, Paul Kennedy, and Peter Christen, Ed. Reproduction for academic, not-for-profit purposes permitted provided this text is included.

¹<http://www.flickr.com>

²<http://delicious.com>

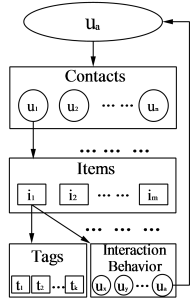


Figure 1: Tree-like structure in social tagging network

does not always come from his/her contacts' collections.

Our task is to track the centric user's indirect past ratings to learn his/her topic-specific preferences according to the appointed friend, which is represented as topic-contact matrix. The size of the matrix is $|C| \times |U|$, where $|C|$ is the number of topic categories. Let $Matrix(i, j) = f(c_i, u_j)$, so (c_i, u_j) are the items posted by u_j , and the concrete value represents the digitized extent that u_a shares the likes of u_b about topic category c_x , $c_x \in C$. The value is obtained by the method explained in Section 3.

Since the topic user cares about is different among websites, we initialize C in Flickr as {Animal, Art, City, Entertainment, Nature, News and Politics, People, Science and Technology, Sports, Travel and Places} on Flickr, and C in Delicious as {Animal, Art, City, Entertainment, Education, Fashion, Comedy, Lifestyle, Music, Nature, New, People, Science, Sport, Travel} on Delicious, these categories are obtained from the navigation classified on each website.

3 Development of Efficient Mining Method

Bayesian reasoning and item category identification method followed by similarity calculation are discussed in the following section. We also outline the mining algorithm.

3.1 Bayesian Reasoning

Since our aim is to analyze user's behaviours in different situations, we can figure out the distribution of user preferences for each friend on each topic category through calculating the conditional probability of collected statistics.

Let $likes(u_a, i)$ stand for items liked by the centric user u_a , and $posted(u_b)$ represent item collection posted by u_b . Let $topic(i)$ on behalf of the category that item i should belong to, and $topic(i) \in C$. The probability of item i posted by u_b , belonging to a certain category c_x and liked by u_a can be expressed as $p(likes(u_a, i) | topic(i) = c_x, i \in posted(u_b))$.

According to the Bayesian conditional theorem and Gursel et al.'work (Gursel & Sen 2009), the above probability can be rewritten as equation (1):

$$\frac{p(i \in posted(u_b) | topic(i) = c_x, likes(u_a, i)) \times p(likes(u_a, i) | topic(i) = c_x) \times p(topic(i) = c_x)}{p(i \in posted(u_b) | topic(i) = c_x) \times p(topic(i) = c_x)} \quad (1)$$

Factors in the above formula can be calculated by the following methods:

$$p(i \in posted(u_b) | topic(i) = c_x, likes(u_a, i)) = \text{(number of items proposed by } u_b \text{ and belonging to}$$

the category c_x and also liked by u_a)/(number of items belonging to the category c_x and also liked by u_a).

$$p(likes(u_a, i) | topic(i) = c_x) = \text{(number of items belonging to the category } c_x \text{ and also liked by } u_a) / \text{(number of items belonging to the category } c_x).$$

$$p(i \in posted(u_b) | topic(i) = c_x) = \text{(number of items proposed by } u_b \text{ and belonging to the category } c_x) / \text{(number of items belonging to the category } c_x).$$

3.2 Item Category Identification

In folksonomy, the key advantage of a tag is its straightforward way in describe the content of heterogeneous data like music data file, blog website, etc. Usually, through calculating the degree of its tags that also belongs to C , we can attain which topic the item should be indirectly be categorized into. Let the $topic(i)$ be the category of item i , t_k^i be the k -th element in its tag set, and $w(t_k^i)$ be the weight of the tag, the probability item i belonging to category c_x is computed as follows:

$$Pr(topic(i) = c_x) = \frac{\sum_{k=1}^{l_i} w(t_k^i) \times Level(c_x, t_k^i)}{\sum_{c_y \in C} \sum_{k=1}^{l_i} w(t_k^i) \times Level(c_y, t_k^i)} \quad (2)$$

When item i has only one tag, the weight of the tag equals one. Otherwise the $w(t_k^i)$ can be obtained by formula (3), in which the variable v denotes the number of tags of item i . And $Level(c_x, t_k^i)$ used to calculate the degree tag t_k^i belonging to c_x is calculated as formula (4).

$$w(t_k^i) = \begin{cases} \frac{1}{2^k} + \frac{1}{2^{v(v-1)}} & k < v \\ \frac{1}{2^k} & k = j \end{cases} \quad (3)$$

$$Level(c_x, t_k^i) = (1 - \theta) \times IsCategory(c_x, t_i) + \theta \times similarity(c_x, t_i) \quad (4)$$

The extent of t_i belonging to c_x is calculated in formula (4) composed of 2 parts. As preprocessing, we build category repositories to establish "belong" relation between tags and topic categories, through clustering tags which co-occurrence with the topic, namely, if $t_j \in c_x$, $t_k \in T(u_i, i_s)$, and also $t_j \in T(u_i, i_s)$, then $t_k \in c_x$. On startup, category name used as the 1-st one belongs to the corresponding category. The obtained category files reserve irregular form of tags on Flickr and Delicious, and we also combine it with similarity calculation based on WordNet built by domain experts to keep semantic information between tags and category names.

In formula (4), $IsCategory(c_x, t_i)$ returns the 0-1 value which identify whether the tag exists in the repositories of c_x . And $similarity(c_x, t_i)$ returns the similarity value of t_i and c_x , where $0 < \theta < 1$ is used to coordinate the contribution of two parts.

3.3 WordNet-based similarity calculation

Many works devote to find out the semantic meaning wrapped by WordNet content of least common subsumer(LCS) (Jiang, & Conrath 1998), path length (Leacock, & Chodorow 1998) and relation structure (Wu, & Palmer 1994). Resnik (Resnik 1992) deems

that the common information the concept pairs share is the criterion to measure similarity value between them. Lin (Lin 1998) believes that similarity value between two concepts, like, concept 1 and concept 2, can be expressed as information content ratios of the common information between them. It is computed as follows:

$$\text{similarity}(\text{concept}_1, \text{concept}_2) = \frac{2 \times \log p(\text{lso}(\text{concept}_1, \text{concept}_2))}{\log p(\text{concept}_1) + \log p(\text{concept}_2)} \quad (5)$$

In which, $\text{lso}(\text{concept}_1, \text{concept}_2)$ represents the lowest common ancestor node in the classification tree shared by concept_1 and concept_2 . $p(c)$ denotes the probability of encountering an instance of concept c . In this paper, we use Lin's method to compute the similarity between tags and categories because of its good performance (Budanitsky & Hirst 2006).

3.4 Main Processes

Below are the main processes of the proposed recommendation system in the form of pseudo code. Symbols u_a , u_b , c_x , $T(u_b, i)$ are defined in Section 2, in which, u_a denoted the centric user, u_b denoted friends of u_a , c_x was the topic category and $T(u_b, i)$ represented the tag collection of item i which was posted by u_b . $\text{Contact}(u_a)$ represents the friends of u_a . Symbol a is used to count the number of items posted by u_b , as well as being liked by u_a , which belongs to category c_x . Symbol b denotes the number of items which are liked by u_a , and also belonging to c_x . Symbol c represents the number of items belonging to c_x . Symbol d denotes the number of items which are posted by u_b and also belonging to c_x .

Require: the topic-contact matrix of u_a and the *keywords* to be queried

Ensure: recommendation items

```

1: for each  $u_b \in \text{Contact}(u_a)$  do
2:   for each category  $c_x$  do
3:     for each item  $i, i \in I(u_b)$  do
4:        $\text{WeightList} \leftarrow \text{WeightDistribution}(T(u_b, i));$ 
5:        $\text{CategoryList} \leftarrow$ 
6:          $\text{SimilarityCalculation}(T(u_b, i));$ 
7:        $\text{Topic}(i) \leftarrow$ 
8:          $\text{TopicIdentify}(\text{WeightList}, \text{CategoryList});$ 
9:       if  $i \in \text{Posted}(u_b)$  AND  $i \in \text{Interacted}(u_a)$  AND
10:         $\text{Topic}(i) = c_x$  then
11:          $a + 1 \leftarrow a;$ 
12:       end if
13:       if  $\text{Topic}(i) = c_x$  AND  $i \in \text{Interacted}(u_a)$  then
14:          $b + 1 \leftarrow b;$ 
15:       end if
16:       if  $i \in \text{Posted}(u_b)$  AND  $\text{Topic}(i) = c_x$  then
17:          $c + 1 \leftarrow c;$ 
18:       end if
19:       if  $\text{Topic}(i) = c_x$  then
20:          $d + 1 \leftarrow d;$ 
21:       end if
22:        $\text{Topic-ContactMatrix}(a, b, c, d);$ 
23:     end for
24:   end for
25: end for
26:  $\text{KeywordsItems} \leftarrow \text{keywords};$ 
27:  $\text{Sort}(\text{KeywordsItems}, \text{Topic-ContactMatrix})$ 
    
```

4 Experiments and Evaluations

4.1 Implementation

Using Visual Studio 2010 as IDE, the system is coded in C# language, and run on Microsoft Windows XP, .NET Framework 3.0 and above. The framework is

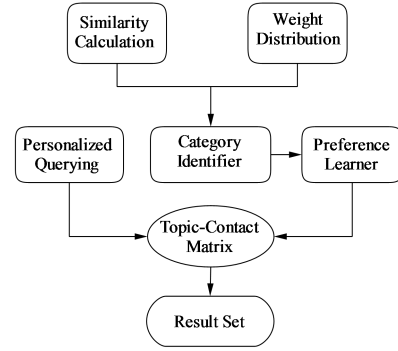


Figure 2: The framework of recommendation system

shown in Fig. 2. It includes modules such as Similarity Calculation, Weight Distribution, Category Identifier, Preference Learner and Personalized Querying.

Similarity Calculation: returning the semantic similarity between the lexical form of category and a certain tag.

Weight Distribution: returning weights of items tags sequentially computed by formula (3).

Category Identifier: returning the decreasing list of degrees of the item belonging to each category by formula (4).

Preference Learner: learning preference values from statistical data yield how much the centric user likes the items posted by one of his/her contacts. From statistical data in collections, we can see although u_a has a lot of contacts, the amount of items which u_a interacted is very little. Therefore, the topic-contact matrix is faced with a data sparsity problem. To improve efficiency of the preference matrix querying, we record the personalized information in triple formalized as $\langle \text{contact ID}, \text{topic category and preference value} \rangle$ stored in XML Schema after the first time preference has been learned to be reused later.

Personalized Querying: receiving a list of keywords as input, we obtain item collection with keyword searching. Then we recommend the result collection ranked by the topic-contact matrix in descending order according to the category sequence the keywords belong to.

4.2 Data Sets and Evaluations

The system was evaluated with two real-life datasets.

DataSet I: the dataset was crawled from Flickr during 2009.12-2010.1. We record the contact ID for the centric user and the photo information posted by each contact in two files. The photo information includes photo ID, contact ID (who posted the photo), sets of user IDs who had commented on it, and also its tag list. Before crawling, we check that the centric user is active enough to mine its preference.

DataSet II: It is the bookmark repository in Delicious, provided by the 2-nd International Workshop on Information Heterogeneity and Fusion in Recommendation system³ in 2011. The information was organized into 7 files. The statistics of the above two datasets are listed in Table 1.

In the experiment, for a given user, we select the top 20 items as recommendations ranked by the user's topic-contact matrix after firstly selected through keyword matching. The candidate items collected are assumed occur in the database. Suppose that I_1 are the top 20 items ranked by the matrix, or the top 20

³<http://ir.ii.uam.es/hetrec2011>

Table 1: Statistics of Experimental Datasets

Property	Flickr	Delicious
The number of users	10	1867
The number of contacts	3026	7668
The number of items	22969	69227
The number of tag assignments	124845	437593
Average tag number per user	45	48
Average tag number per resource	5	6

of the item collection returned by keywords searching as baseline. I_2 are the items in database. For each user, I_1 is different according to its topic-contact matrix and searching keywords each time. I_2 is the same on each website. It is meaningful to return the top 20 items to cater to user's taste preferences. Evaluation criterions are precision, recall and F-measure in this paper. Our criteria for the user "like" this photo is that the user's ID is contained in the photo's comment userID list.

$$precision = \frac{|number\ of\ items\ liked\ by\ u_a\ in\ I_1|}{|number\ of\ items\ in\ I_1|}$$

$$recall = \frac{|number\ of\ items\ liked\ by\ u_a\ in\ I_1|}{|number\ of\ items\ liked\ by\ u_a\ in\ I_2|}$$

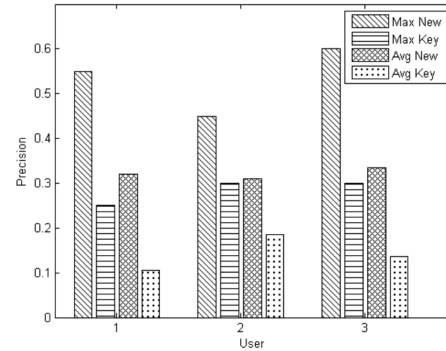
$$F - measure = \frac{2 \times precision \times recall}{precision + recall}$$

4.3 Results and Discussion

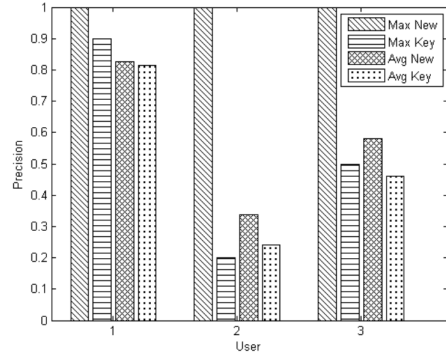
Firstly, we evaluate the proposed approach against the method of keyword searching only. There are 100 keyword tags for test randomly selected from tags of items in I_2 . Based on WordNet 2.0, we respectively learn 3 user preference in each dataset, we choose them for the reason that they are the most active users in each website. The precisions are compared in Fig. 3(a) and Fig. 3(b).

In Fig. 3(a) and Fig. 3(b), axis x presents users, axis y presents precision value corresponding to each user. Since we had tested 100 different tags for each user, we record their max precision and average precision value. Since user's purposes vary on different websites, the 100 testing tags are not the same between Flickr and Delicious, only the same in each. Legends sequentially denote max precision value of the proposed method (*New*), max precision value of keyword method (*Key*), average precision value of *New*, average precision value of *Key*. Method *New* utilize similarity calculation, where $\theta = 0.4$ in formula (4), the usage of θ will be discussed in Fig.7.

The results show that the proposed method could influence ranking result items to make them better to the centric user's tastes. Let us focus on Flickr in Fig. 3(a). Our proposed method outperforms the baseline method in some cases. The results vary among users in Flickr and Delicious for difference of keywords for searching each time and each user's specialized topic-contact matrix. With respect to Delicious user in Fig. 3(b), the contact number of each Delicious user is smaller than the Flickr user in dataset we collected, and the biggest number is 90. Taking the 1-st Delicious user for example, it only contains one record in its preference matrix. We studied it as a limiting case for regarding proposed method. The number of contacts increases from user 1 to user 3 in Delicious in Fig. 3(b). The more the system learns, the higher the precision it achieves. However we can also see that the



(a) Flickr



(b) Delicious

Figure 3: Precision comparisons on two methods

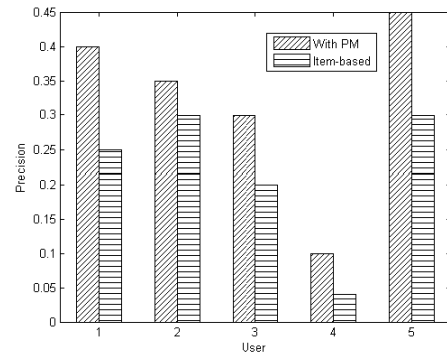
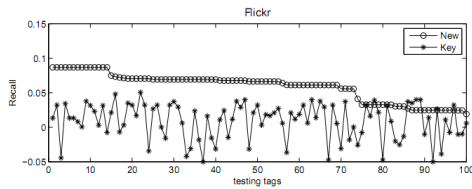


Figure 4: Precision comparisons on two methods

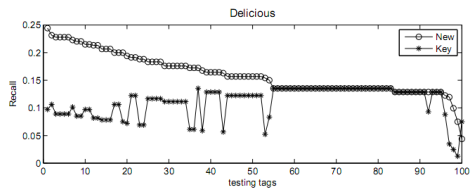
gap between the proposed method and the baseline method augments little. According with that average contact number of each user in Flickr can reach 300, it is evident that users in Flickr more frequently contact with their friends since their friends contribute more to their searching needs. The data also show that the proposed system improves precision by nearly 56% in Flickr and 21% in Delicious datasets.

Secondly, we evaluate the proposed approach against the traditional item-based method. The similarity between items is calculated according to the similarity among each item's tag vector. Using Latent Dirichlet Allocation(LDA) we easily dig the weights of which each item belongs to the latent topics in a certain number, and on the basis of the common among weight vector, we calculate the cosine similarity between weight vector.

Fig.4 is the comparison of the item-based method and the proposed method. It shows that the later one is better in our discussion. The more we learn from user's rating behaviour the more precise we can recommend items for him/her. As shown in Fig.4,

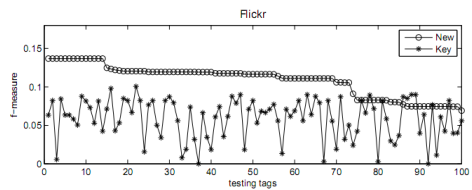


(a) Recall on Flickr

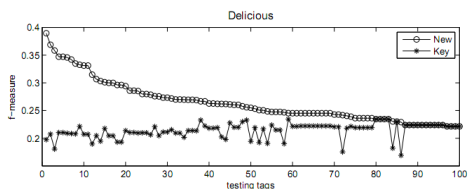


(b) Recall on Delicious

Figure 5: Recall comparisons on two methods



(a) F-measure on Flickr



(b) F-measure on Delicious

Figure 6: F-measure comparisons on two methods

the proposed method could gain more better result for the more active ones like user 1 and user 5 are. So the proposed method has certain learning ability.

Thirdly, two other criteria, the Recalls and F-measures are measured and the results are shown in Fig. 5 and Fig. 6, respectively. In Fig. 5 and Fig. 6, the axis x presents testing tags, the axis y presents recall measurement according to each tag in Fig. 5 and the corresponding f-measure. The legends “New” and “Key” respectively denotes the method with PM and the item-based method. The two methods are compared with the obtained data from Flickr (Fig. 5(a) and Fig. 6(a)) and Delicious (Fig. 5(b) and Fig. 6(b)) datasets, respectively. The results are sorted by the value of “New” in descending order. It shows that the proposed recommendation method outperforms those in which only keyword searching is used.

Finally, we study the impact of WordNet similarity calculation on recommendation precision as shown in Fig. 7. In Fig. 7 the axis x presents testing tags, the axis y presents precision value corresponding to each tag, legends represent proposed method with WordNet-based similarity calculation (*Sim*) and proposed method without similarity calculation (*NoSim*). In *Sim*, let $\theta = 1$ in formula (4) and in *NoSim*, let $\theta = 0$ in the same formula.

We find that contribution of similarity to the precision rate is not large, since when testing, the tags we choose is mostly already “known” in the tag repository,

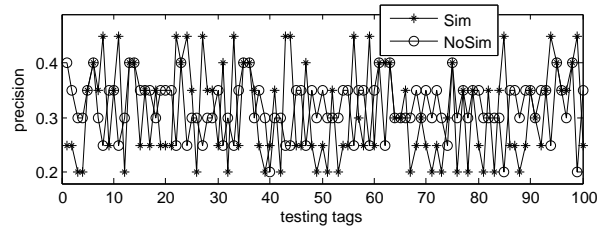


Figure 7: Precision comparisons on two methods

ry, this ensures that result can be found by keyword searching. Take searching keyword “cat” for example, no similarity-based method can identify the word since it already exists in category repositories, while using its synonym “kitty” as a replacement in the query, the proposed method without similarity failed to identify the word while the proposed similarity-based method can identify the word not in the tag repository. According to the results in Fig. 7, the proposed method with similarity calculation is preferable to the non-similarity method.

5 Related Work

Topic-based, tag-based and user-based methods are the dominant techniques applied in recommendation system (Shepitsen et al. 2010)(Rashid et al. 2002). Jin et al. (Jin et al. 2011) use Latent Dirichlet Allocation(LDA) to model topics probability distribution over tags and resources of multiple topics. It is uncommon that each tag has the same topic vector. Thus we consider different topics for different website users. Ziegler et al. (Ziegler et al. 2005) present topic diversification method and introduce a intra-list similarity metric to assess the topical diversity of recommendation lists. Although the proposed method can improve user satisfaction with recommendation lists, it is also detrimental to the average accuracy. Abel et al. (Abel et al. 2011) find that temporal profile build by hashtag-based, entity-based, and topic-based benefit from semantic enrichment improve recommendation quality. Results demonstrate that topic-based methods perform better than the hashtag-based strategy and requires less run-time and memory.

Cantador et al. (Cantador et al. 2011) argue that in some cases, tags are used to depict subjective qualities of a item or be related to organisational aspects. They map the concepts of filtered tags and classified by their mechanism to semantic entities like WordNet and Wikipedia. The obtained concepts are then transformed into semantic classes that can be uniquely assigned to context-based categories. They should dig into the semantic relations between the obtained tag clusters. Krestel et al. (Krestel & Fankhauser 2012) explore an approach to personalized tag recommendation that combining a probabilistic model of tags from the resource with tags from the user. They use LDA to estimate the topic-tag distribution and the resource topic distribution from an unlabeled corpus of documents using Dirichlet priors for the distributions and a fixed number of topics. Shepitsen et al. (Shepitsen et al. 2010) present a personalization algorithm for recommendation in folksonomies which rely on hierarchical tag clusters. Experiments show that clusters of tags can be effectively used as a means to ascertain the user’s interest to determine the topic of a resource. Durao et al. (Durao & Dolog 2010) present a tag-based recommendation system which suggests similar resources based on cosine similarity

calculus with additional factors such as tag popularity, tag representativeness and affinity between user and tag. Semantic similarity can be used to overcome ambiguity problems and further improve recommendation accuracy.

Nocera et al. (Nocera & Ursino 2011) cluster users based on the tags they share, and provide a user with recommendations of similar users and resources. While Nakatsuji et al. (Nakatsuji et al. 2009) allow user profiles to be constructed as a hierarchy of classes, user's interest weight is assigned to each class and instance, then generates user with the highest similarity into group. Yin et al. (Yin et al. 2012) deem that it is worthwhile to emphasize the significance of trust information providing reliable personal friend relationships, and provide a simple framework for the design of the prediction model making use of both listening and trust information. Firan et al. (Firan et al. 2007) focus on how tags characterized the user and enable personalized recommendations. Through analysing tag usage in contrast to conventional user profiles, they specify recommendation algorithms based on tag user profiles.

6 Conclusions

In order to improve users' searching experience by returning items most catering to his/her tastes on social networking websites, we propose a recommendation system by utilizing the Bayesian rule and WordNet-based similarity calculation to address problems such as user profile modeling, item category identification, preference learning and result presentation. We have achieved an improved average recommending precision by nearly 56% and 21% with the obtained data from Flickr and Delicious, respectively. It also helps to diversify users' needs and may foster new interests around the original search tags with the proposed semantic extension. The methods proposed in this paper can be applied to other social tagging networks such as Blog systems. Finally, our method shows potential in providing a personalized spin on the social network experience. In the future, we will focus on the improvement of preference calculations.

Acknowledgements

This work was supported by National Natural Science Foundations of China(61170192), and Natural Science Foundations of CQ.

References

- Karen,H.L., Tso-Sutter, Leandro,B.M.,(2008), Tag-aware Recommender Systems by Fusion of Collaborative Filtering Algorithms. *SAC'08: Proceedings 23rd Annual ACM Symposium on Applied Computing*, Fortaleza, Brazil.
- Resnick,P., Varian,H.R. (1997), Recommender Systems, *Communications of the ACM*, Vol. 40, pp. 56-58.
- Xu,G.D., Gu,Y.H., Zhang,Y.C. (2011), TOAST: A Topic-oriented Tag-based Recommender System, *WISE'11: Proceedings of the 12th international conference on Web information system engineering*, Berlin, Heidelberg.
- Lerman,K. (2006), Social Networks and Social Information Filtering on Digg, *ICWSM2007 Boulder*, Colorado, USA.
- Gursel,A., Sen,S. (2009), Improving Search in Social Networks by Agent based Mining, *IJCAI'09: Proceedings of the 21st international joint conference on Artificial intelligence*, San Francisco, USA.
- Jiang,J.J., Conrath,D.W. (1998), Semantic Similarity based on Corpus Statistics and Lexical Taxonomy, In *Proceedings of the International Conference on Research in Computational Linguistics*, TaiWan, China.
- Leacock,C., Chodorow,M. (1998), Combining Local Context and WordNet Similarity for Word Sense Identification. In *WordNet: An Electronic Lexical Database*, Christiane Fellbaum, MIT Press, Vol. 49,pp. 265-283.
- Wu,Z.B., Palmer,M. (1994), Verb Semantics and Lexical Selection, *ACL-94: Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics*, Las Cruces, NM.
- Resnik,P. (1992), WordNet and Distributional Analysis: A Class-based Approach to Lexical Discover. in *Proceedings of the AAAI Symposium on Probabilistic Approaches to Natural Language*, San Joe, CA.
- Lin,D.K. (1998), An Information-theoretic Definition of Similarity. In *Proceedings of the 15th International Conf. on Machine Learning*, San Francisco, CA.
- Budanitsky,A., Hirst,G. (2006), Evaluating WordNet-based Measures of Lexical Semantic Relatedness, *Computational Linguistics*, Vol. 32, pp. 13-48.
- Shepitsen,A., Gemmel,J., Mobasher,B. (2010), Personalized Recommendation in Social Tagging Systems Using Hierarchical Clustering. In *RecSys'10: Proceedings of the fourth ACM conference on Recommender systems*, New York, USA.
- Rashid,A.M., Albert,I., Cosley,D. (2002), Getting to Know You: Learning New User Preferences in Recommender Systems, In *Proceedings of the 7th International conference on Intelligent user interfaces*, San Francisco, USA.
- Jin,Y.A., Li,R., Wen,K. (2011), Topic-based Ranking in Folksonomy via Probabilistic Model, *Artificial Intelligence Review*, Vol. 36, pp. 139-151.
- Ziegler,C.N., McNee,S.M., Konstan,J.A. (2005), Improving Recommendation Lists through Topic Diversification, In *Proceedings of the 14-th international conference on World Wide Web*, New York, USA.
- Abel,F., Gao,Q., Houben,G.J., et al. (2011), Analyzing User Modeling on Twitter for Personalized News Recommendations, *Lecture Notes in Computer Science*, Vol. 6787,pp. 1-12.
- Cantador,I., Konstas,I., Jose,J.M. (2011), Categorising Social Tags to Improve Folksonomy-based Recommendations, *Web Semantics: Science, Services and Agents on the World Wide Web*, Vol. 9,pp. 1-15.
- Krestel,R., Fankhauser,P. (2012), Personalized Topic-based Tag Recommendation, *Neurocomputing*, Vol. 76, pp. 61-70.
- Durao,F., Dolog,P. (2010), A Personalized Tag-Based Recommendation in Social Web Systems. In *Proceedings of the 2010 ACM Symposium on Applied Computing*, New York, USA.

- Nocera,A., Ursino,D. (2011), An Approach to Providing a User of a social folksonomy with Recommendations of Similar Users and Potentially Interesting Resources, *Knowledge-Based Systems*, Vol. 24, pp. 1277-1296.
- Nakatsuji,M., Yoshida,M., Ishida,T. (2009), Detecting Innovative Topics based on User-interest Ontology, *Web Semantics:Science, Services and Agents on the World Wide Web*, Vol. 7, pp. 107-120.
- Yin,C.X., Peng,Q.K., Chu,T. (2012), Personal Artist Recommendation via a Listening and Trust Preference Network, *Physica A:Statistical Mechanics and its Applications*, Vol. 391, pp. 1991-1999.
- Firan,C.S., NejdI,W., Paiu,R. (2007), The Benefit of Using Tag-Based Profiles, *In Proceeding of the 2007 Latin American Web Conference IEEE Computer Society*, Washington, USA.

