A Blending Method for Automated Social Tagging

Shenghua Liu, Yatao Zhu, Jiafeng Guo, Yuanzhuo Wang, Xueqi Cheng, Yue Liu
Institute of Computing Technology, Chinese Academy of Sciences
Beijing, China 100190
Email: {liushenghua, zhuyatao, guojiafeng, wangyuanzhuo, cxq, liuyue}@ict.ac.cn

***Abstract*—Social tagging has grown in popularity on the web due to its effectiveness in organizing and accessing webpages. This short paper addresses the problem of automated social tagging, which aims to predict tags for webpages automatically and help with future navigation, filtering or search. We explore and find three foundations of the collaborative tags in social tagging services, that are $consistency$, $sharability$ and $stability$. The complementary advantages are studied among three well-known methods, $i.e.$ TF-weighted keyword extraction, collaborative filtering approach, and Corr-LDA ($correspondence\ latent\ Dirichlet\ allocation$) topic model. We then propose a blending model for automated social tagging to emphasize all the foundations, which linearly combines those tags generated by the three methods, and a permutation probability model is built to learn the linear blending. With the experiments on 50,000 training and 10,000 testing webpages from Del.icio.us database, the results show that our blending method outperforms the four baselines. Furthermore, compared with both topic models, Corr-LDA and mixed membership LDA, our approach results in 14.2% and 25.6% of NDCG_{10} improvement separately.**

***Keywords*-social tagging, automatic annotation, topic model, collaborative filtering**

I. Introduction

Social tagging has grown in popularity on the web due to its effectiveness in organizing content for future navigation and filtering [1], [2], and improving the quality of web search [3], [4], [5], [6], [7] and query recommendation [8].

Despite the usefulness of social tags, most webpages have few or no annotations [9]. Thus automated social tagging is gaining more and more attention [10]. In this short paper, we develop an automated social tagging method for those webpages with text as their dominant content.

We explore the social tags in online bookmarking services, and find three important foundations, $i.e.$ $consistency$, $sharability$ and $stability$, that our method is built on. **Consistency** indicates that social tags always reflect the textual content of the annotated webpage in some degree. **Sharability** describes the phenomenon of collaborative annotation that users are more likely to assign similar tags for the similar web content. And **stability** shows that collaboratively annotated tags form a stable tag frequency distribution.

Based on $consistency$, keyword extraction mounts to an intuitive way to generate tags [11]. With the bag-of-words assumption, the term frequency (TF) is an important weight to identify keywords [12]. $Sharability$ gives the clue that a collaborative filtering (CF) method could find those tags other than the words in web content from its nearest neighbors, measured by web content similarity. $Stability$ guarantees that $correspondence\ latent\ Dirichlet\ allocation$ (Corr-LDA) model leverages a probabilistic topic model to capture the conditional distribution between lower dimensional representations (i.e. hidden topics) of words (in webpages) and tags.

In order to consider all the properties of social tags, our paper proposed a linear blending approach for automated social tagging via combining TF-weighted keywords, the inferred tags by CF approach and Corr-LDA topic model. A permutation probability model is built to learn the linear blending. With conducting experiments on the Del.icio.us database, our linear blending method achieves higher precision and recall than the baselines. Compared with both the well-known topic models, Corr-LDA and MM-LDA (mixed membership LDA), ours improves the NDCG_{10} by 14.2% and 25.6% separately.

The rest of the paper is organized as follows. Section II discusses the related work. In section III, we describe our blending model in details. The experimental results are presented in section IV, and we finally conclude the paper in section V.

II. Related Work

In the earlier research on social tags, Golder et al. [1] provided empirical study of the tagging behavior and the usage of tags in Del.icio.us. In [2], Quintarelli gave a general introduction of social annotation. Afterwards, the related research has given rise to various topics, such as tag recommendation [13], [14], [15], tag visualization [16], and information retrieval [3], [4], [5], [6], [7]. Different from above, the paper investigates the capability of the collaborative tags in tag prediction for unannotated webpages.

In the framework of Semantic Web, metadata generation problem has been well studied. Dill et al. [17] presented

This paper is partially supported by National Grand Fundamental Research 973 Program of China (No. 2012CB316303, No. 2013CB329602), National Natural Science Foundation of China (No. 61202213, 61202215, 61173008, 61232010), Projects of Development Plan of the State High Technology Research (No. 2012AA011003) and Beijing nova program (No. Z121101002512063).

978-1-4799-2902-3/13 $31.00 © 2013 IEEE
DOI 10.1109/WI-IAT.2013.17

a platform for large-scale text analysis and automated semantic tagging. Handschuh et al. [18] considered the webpages that are generated from a database, and automatically produced semantic annotations from the database with an ontology. Our approach does not conform to any priori formal ontology.

With keyword extraction, Chirita et al. proposed P-TAG algorithm [11]. Nevertheless, social tags not only emphasize the keywords of a webpage, but also contain some additional information beyond the webpage text [19]. Collaborative filtering (CF) is firstly proposed by Goldberg et al. [20]. [10] proposed a CF approach to generate tags of a webpage from those tags of its nearest neighbors. [21] considered the context clues through tags and social connectivity among users in the CF approach. Corr-LDA was firstly proposed in [22] for modeling images and their captions. We apply such model to capture the hidden relationship between document words and tags.

III. Automated Social Tagging

With the bag-of-word assumption, the textual content of webpage i is denoted by a vector $\boldsymbol{w_i}$ in a word space $\boldsymbol{V}$, where each element $w_{i,j} \in \boldsymbol{w}_i$ indicates the word frequency in webpage i. The social tags of webpage i, likewise, are represented by a vector $\boldsymbol{t_i}$ in a tag space $\boldsymbol{U}$, where each element $t_{i,k} \in \boldsymbol{t_i}$ means the frequency of the tag that is used to annotate the webpage by web users. Thus webpage i with its social tags is represented as a 2-tuple ($\boldsymbol{w}_i$, $\boldsymbol{t}_i$), and the corpus consists of such tuples of the webpages. The corpus is divided into two parts that one is used for training the model, $i.e.$ training data set $\mathcal{R}$, and the other is for verifying the effectiveness of the tag prediction method, $i.e.$ testing data set $\mathcal{D}$. In order to learning our blending model, we randomly sample out a collection of webpages as $\mathcal{R}'$ from training data set $\mathcal{R}$. Since tags are annotated by a large number of users and contain a higher-level abstraction on the web content [23], they usually differ from the words in webpages literally, $i.e.$ $\boldsymbol{V} \neq \boldsymbol{U}$.

A. Foundations

consistency. With the "wisdom of the crowds", collaboratively annotated tags are more likely to be a less biased semantic description of web content.

sharability. Users are more likely to assign similar tags for the similar content. Thus most social tags are shared by a large variety of webpages. Fig. 1 illustrates the average document frequency (DF) all over the webpages for top frequently annotated tags. We see that each one of the top-10 tags of a webpage is shared by more than 300 webpages on average.

stability. The collaboratively annotated tags form a stable tag frequency distribution. Let a bookmark be a user's tagging activity to a webpage, and it contains one or more tags marked by the user. Fig. 2 shows that empirically after

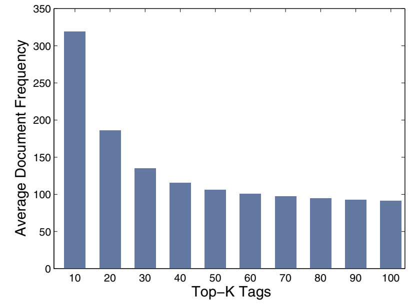

Figure 1. Average Frequency (DF) of top frequent tags.

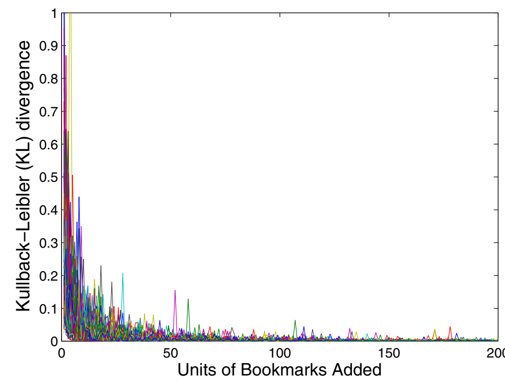

Figure 2. The KL divergence between the tag distribution at every step, and that of the last observation in data set $\mathcal{R}$.

the first 100 or so bookmarks, the KL divergence converges to zero, which means that the tag distribution gets stable with a plenty of bookmarks annotated. The similar observations have been obtained in [1], [24] as well.

B. Blending Model

TF-weighted keywords, CF approach and Corr-LDA inference model are primarily inspired with different aspects of foundations, which cannot actually overbeat each other. We define the ratio of tags that are only correctly predicted by TF-weighted keywords rather than Corr-LDA method as the $complementarity$ of TF to Corr-LDA, which is formally defined as follows.

$$N_{tf,cl}/\mathrm{K}$$

where $N_{tf,cl}$ is the number of tags that are correctly predicted by TF-weighted keywords, and not hit by Corr-LDA, and K is the size of predicted tags. The complementarity of other pair of methods is defined in the same way.

With learning and evaluating on training data set $\mathcal{R}$ and testing data set $\mathcal{D}$, the average complementarity among TF, CF, and Corr-LDA methods are illustrated in table I. The positive values in each row of the table represent the complementarity of the method to the other two. Considering

Table I
THE COMPLEMENTARITY AMONG TF, CF AND CORR-LDA METHODS.

Methods	TF	CF	Corr-LDA
TF	–	0.0966	0.1170
CF	0.3037	–	0.0998
Corr-LDA	0.3092	0.0849	–

* K equals to 10

the complementarity among those three approaches, we propose an blending method for automated tagging. We define the blending weights $\chi_j^{(i)}$ of tag t_j for webpage i based on the normalized weights from the three kinds of predictions as equation (1).

$$\chi_j^{(i)} = \chi(\tau_{i,j}, \zeta_{i,j}, \rho_{i,j}|\lambda) \tag{1}$$

where λ is the blending parameter, and its estimation is discussed later in section III-B3. $\tau_{i,j}$, $\zeta_{i,j}$ and $\rho_{i,j}$ are the normalized weights for tag $t_{i,j}$ annotated to webpage i by TF-weighted keywords, CF, and Corr-LDA methods separately as follows.

$$\tau_{i,j} = \begin{cases} w_{i,j}/\sum\limits_k w_{i,k} & \text{, word } w_{i,\cdot} \text{ is in the top-K tag list} \\ 0 & \text{, otherwise} \end{cases} \tag{2}$$

where k is the index of the word that appears in the top-K tag list.

$$\zeta_{i,j} = \begin{cases} r_{i,j}/\sum\limits_k r_{i,k} & \text{, tag } t_{i,\cdot} \text{ is in the top-K tag list} \\ 0 & \text{, otherwise} \end{cases} \tag{3}$$

where k is the index of the tag that appears in the top-K prediction list, and $r_{i,\cdot}$ (discussed later in section III-B1) is the weight of tag $t_{i,\cdot}$ predicted by CF method.

$$\rho_{i,j} = \begin{cases} p(t_{i,j}|\boldsymbol{w}_i)/\sum\limits_k p(t_{i,k}|\boldsymbol{w}_i) & \text{, tag } t_{i,\cdot} \text{ is in top-K list} \\ 0 & \text{, otherwise} \end{cases} \tag{4}$$

where $p(t_{i,\cdot}|\boldsymbol{w}_i)$ (discussed later in section III-B2) is the weight of tag $t_{i,\cdot}$ inferred by Corr-LDA model.

TF-weighted keywords for automated social tagging intuitively equals to the term frequency $w_{i,j}$. Thus, we only discussed how to evaluate the weights of predicted tags by CF and Corr-LDA methods in the following.

1) Collaborative Filtering: We employ the cosine similarity $Sim_{i,j}$ defined in equation (5) to measure the content similarity of webpage i and j.

$$Sim_{i,j} = \frac{\boldsymbol{w}_i \cdot \boldsymbol{w}_j}{||\boldsymbol{w}_i||\,||\boldsymbol{w}_j||} \tag{5}$$

Furthermore, not all tags of webpage i carry the same information. Let $p_{i,j}$ be the normalized the frequency as equation (6) defines.

$$p_{i,j} = \frac{t_{i,j}}{\sum\limits_{l \in \boldsymbol{U}} t_{i,l}} \tag{6}$$

However, webpages with fewer users to bookmark are likely to contain inappropriate tags of heavier weight $p_{i,j}$. We include the total number of tags as another factor to adjust the tag significance between frequently and infrequently annotated webpages. Since the total number of tags given to a webpage follows a power law, we take its logarithm to avoid being over-weighted. Hence, finding the k-nearest neighbors (k-NN) $\boldsymbol{\kappa}_i$ of an unannotated webpage i, the CF method generates the ordered tags list according to the following weight.

$$r_{i,j} = \sum_{k \in \boldsymbol{\kappa}_i} (p_{k,j} \times s_{i,k} \times \log \sum_{l \in \boldsymbol{U}} t_{k,l}) \tag{7}$$

where $s_{i,k}$ is the normalized similarity in neighbors $\boldsymbol{\kappa}_i$ as defined in equation (8).

$$s_{i,k} = \frac{Sim_{i,k}}{\sum\limits_{k' \in \boldsymbol{\kappa}_i} Sim_{i,k'}} \tag{8}$$

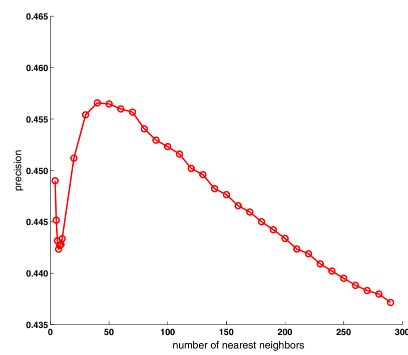

Figure 3. The precisions with variant number of nearest neighbors in CF method.

Finally, we evaluate the precision of top-10 predicted tags, with variant number of neighbors. As Fig. 3 shows, the precision reaches maximum with around 50 neighbors. Thus, we choose 50 nearest neighbors in the CF-based tag prediction method.

2) Corr-LDA Model: In the generative topic model, each pair $(\boldsymbol{w}_i, \boldsymbol{t}_i)$ in data set $\mathcal{R}$ serves as the observations for learning. Thus we view vector $\boldsymbol{w}_i$ in word space $\boldsymbol{V}$ as a sequence of samples generated from word topic model, and each $w_{i,j} \in \boldsymbol{w}_i$ indicates the times of the j-th word is observed in this sampling. In the same way, $t_{i,j} \in \boldsymbol{t}_i$ represents the frequency of the j-th tag while drawing samples from the tag topic model.

The Corr-LDA model in the paper is represented as the description in [22]. With the approximate posterior on the latent variables, the model calculates the conditional probability $p(t|\boldsymbol{w})$ and defined in equation (9).

$$p(t|\boldsymbol{w}) \approx \sum_{n} \sum_{z_n} q(z_n|\phi_n) p(t|z_n, \eta) \tag{9}$$

In order to determine the number of topics that are used to modeling the tagged webpages, we apply Corr-LDA on training data set $\mathcal{R}$ with various number of topics and iterations as Fig. 4 shows. It indicates that 300 hidden topics are basically enough to capture the major categories of the words and tags of the training webpages, while more topics are very likely to be redundant and cause over-fitting. Thus, the Corr-LDA model has 300 hidden topics in the paper.

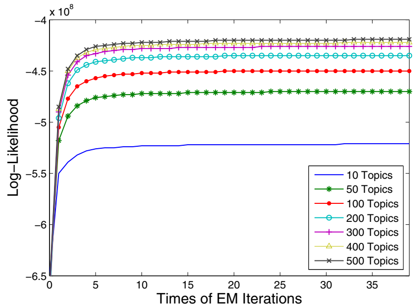

Figure 4. Log-Likelihood on the number of iterations between different topic sizes in Corr-LDA model.

3) Learning to Blend: Our goal of blending is to gain an ranking list of tags such that it coheres well with the social annotated tag list. We employ the permutation probability model[25], [26] to measure the ranking list. It assumes that given a list of tags with weights for each webpage, any permutation is possible, and the probability is defined as equation (10).

$$P(\pi|\boldsymbol{\chi}) = \prod_{j=1}^{k} \frac{\exp(\chi_{\pi(j)})}{\sum_{t=j}^{k} \exp(\chi_{\pi(t)})} \tag{10}$$

where π is a permutation on the tags, $\pi(j)$ denotes the tag $\pi(j)$ at position j in the permutation, and $\exp(\cdot)$ is the natural exponential function.

It is proved that $P(\pi|\boldsymbol{\chi})$ forms a probability distribution over the set of permutations, and it is monotonically increased with more pairs of tags in a descendent order. Thus $P(\pi|\boldsymbol{\chi})$ reaches its maxima with its permutation π is a ranking list ($i.e.$ in a descending order).

TF weights are directly obtained from the frequency of the words that occurs in the web content and belong to tag space $\boldsymbol{V}$. CF and Corr-LDA method are both trained on data set $\mathcal{R} - \mathcal{R}'$. Finally, with the supervised data set $\mathcal{R}'$, log-likelihood $\mathcal{L}$ is defined as follows.

$$\begin{aligned} \mathcal{L} = & \sum_{i=1}^{||\mathcal{R}'||} lnP^{(i)}(\pi|\boldsymbol{\chi}) \\ = & \sum_{i=1}^{||\mathcal{R}'||} \sum_{t=1}^{k} (\chi_{\pi(t)}^{(i)} - ln \textstyle\sum_{l=t}^{k} \exp(\chi_{\pi(l)}^{(i)})) \end{aligned} \tag{11}$$

where $||\mathcal{R}'||$ is the size of data set $\mathcal{R}'$. Let vector $x_j^{(i)}$ represents the normalized weights from the three different methods for tag j of webpage i, and define

$$x_j^{(i)} = [\tau_{i,j}, \zeta_{i,j}, \rho_{i,j}]^T.$$

In our blending model, we assume that the social tags are generated by the linear combination of the normalized TF, CF and Corr-LDA weights, $i.e.$

$$\begin{aligned} \chi_{\pi(j)}^{(i)} = & \lambda_1 \tau_{i,\pi(j)} + \lambda_2 \zeta_{i,\pi(j)} + \lambda_3 \rho_{i,\pi(j)} \\ = & \lambda^T \cdot x_{\pi(j)}^{(i)}. \end{aligned} \tag{12}$$

Thus the parameters λ are estimated by maximizing the log-likelihood, which equals to minimize its opposite $-\mathcal{L}$. The gradient of $-\mathcal{L}$ with respect to parameters λ is shown as equation (13).

$$\Delta\lambda = \frac{-\partial\mathcal{L}}{\partial\lambda}$$

where

$$\frac{-\partial\mathcal{L}}{\partial\lambda_m} = -\sum_i \sum_t (x_{\pi(t),m}^{(i)} - \frac{\sum_{l=t}^{k} x_{\pi(l),m}^{(i)} \cdot \exp(\chi_{\pi(l)}^{(i)})}{\sum_{l=t}^{k} \exp(\chi_{\pi(l)}^{(i)})}), \quad m = 1, 2, 3. \tag{13}$$

So the gradient descent algorithm is employed for learning the blending parameters, and backtracking line search algorithm is used for finding a descent step length, and $-\Delta\lambda$ decides the descent direction.

IV. Experiments

A. Settings

In the experiment, a hybrid crawling strategy [9] was employed. 167,958,659 bookmarks made by 825,402 different users on 57,813,581 different URLs with 5,916,196 different tags were crawled from Del.icio.us website during October and November, 2008. We empirically filtered out those webpages annotated by less then 100 users. Finally, 50,000 tagged webpages are randomly selected as training data set $\mathcal{R}$, and another 10,000 ones are selected as testing data set $\mathcal{D}$. We also sample out 10,000 tagged webpages from training data set $\mathcal{R}$ as data set $\mathcal{R}'$ for learning the linear blending model. A different topic model, $i.e.$ mixed membership model(MM-LDA) [27] is implemented as one of baselines for comparison.

B. Training blending model

Our blending model is trained on data set $\mathcal{R}'$ with weights $\tau_{i,j}$, $\zeta_{i,j}$ and $\rho_{i,j}$ estimated by TF, CF and Corr-LDA methods. Since the ranking list keeps the same while λ is scaled by any positive value in the linear blending model, we finally normalized and get $\lambda = [0.305, 0.268, 0.427]$.

Table II
EXPERIMENTAL RESULTS.

(Top-k and Exact-k column is given in percentage (%))

	TF			**CF**			**Corr-LDA**			**MM-LDA**			**ours**		
k	Top-k	Exact-k	NDCG	Top-k	Exact-k	NDCG	Top-k	Exact-k	NDCG	Top-k	Exact-k	NDCG	top-k	Exact-k	NDCG
1	59.3	59.3	0.487	81.8	82.2	0.708	80.1	80.1	0.676	74.2	74.2	0.609	85.4	85.4	0.744
2	74.4	43.5	0.427	89.0	69.5	0.643	89.6	67.0	0.616	88.5	62.4	0.558	93.3	75.4	0.691
3	82.0	33.9	0.398	92.4	61.5	0.611	93.3	58.4	0.586	93.0	52.1	0.531	95.9	66.9	0.663
4	86.0	26.6	0.380	94.1	54.5	0.591	95.4	49.9	0.566	95.2	44.6	0.513	97.3	58.2	0.644
5	88.5	22.3	0.368	95.2	47.1	0.578	96.6	43.8	0.552	96.5	38.8	0.501	98.0	51.0	0.630
6	90.2	18.8	0.361	96.1	41.2	0.569	97.4	38.3	0.544	97.3	32.4	0.493	98.4	44.0	0.621
7	91.4	16.9	0.356	96.7	36.0	0.565	97.9	33.1	0.539	97.9	29.3	0.489	98.7	38.3	0.616
8	92.5	15.2	0.355	97.2	31.4	0.563	98.3	28.7	0.538	98.4	26.6	0.488	98.9	33.4	0.614
9	93.4	13.3	0.356	97.5	27.7	0.565	98.6	25.6	0.540	98.7	23.6	0.491	99.1	29.5	0.617
10	94.1	12.2	0.359	97.8	24.7	0.570	98.8	23.2	0.545	98.9	21.7	0.496	99.3	26.1	0.623
Imp	5.5%	114.5%	73.5%	1.5%	5.8%	9.2%	0.5%	12.6%	14.2%	0.4%	20.3%	25.6%	–	–	–

C. Performance Comparisons

In the experiment, we compare the baseline approaches with the blending method in the following metrics[15], $i.e$, Top-k accuracy, Exact-k accuracy, Recall, Precision and NDCG. Top-k accuracy is defined as the percentage of webpages correctly annotated by *at least* one of the top-k predicted tags. And Exact-k accuracy is defined as the percentage of webpages correctly annotated by the k-th predicted tag, which gives the indication that whether the tags ranked higher in prediction list are more likely to annotate webpages. MM-LDA model are also trained with 300 topics. In addition, NDCG_k of the top-k predicted tags gives an overall evaluation of the quality of predicted ranking list, and is calculated by the following equation (14).

$$\mathrm{NDCG}_k = \frac{\mathrm{DCG}_k}{\mathrm{IDCG}_k} \tag{14}$$

where DCG_k is the discounted Cumulative gain shown in equation (15), and $IDCG_k$ is the ideal DCG of top-k annotated tags.

$$\mathrm{DCG}_k = rel_1 + \sum_{i=2}^{k} \frac{rel_i}{\log_2 i} \tag{15}$$

where rel_i is the relevance value of the i-th tag in the prediction. A tag is usually annotated in a number of times for a webpage, so the ground truth of the ranks of individual tags are ordered by the number of annotating times. And the graded relevance values of the top 10 user-annotated tags are assigned from ten to one separately in the ground-truth order, while the relevance values of other tags are assigned to zeros.

The last row in Table II illustrates the improvement of our approach in Top-10 accuracy, Exact-10 accuracy, and NDCG_{10}, compared with other four baselines.

In addition, the precision-recall curves are drawn as well in Fig. 5 to show the performance of our blending method.

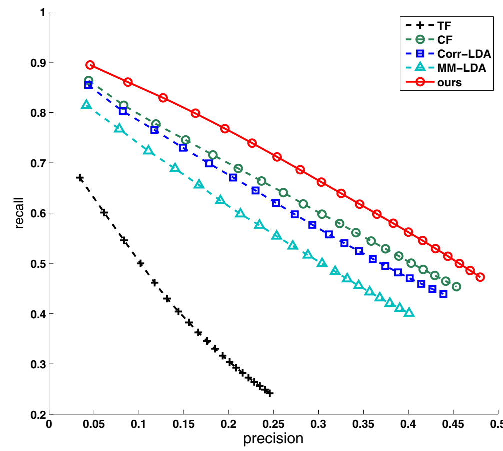

Figure 5. Precision-Recall graph.

V. CONCLUSIONS

In the short paper, we investigate the problem of automated social tagging, aiming at generating tags automatically for webpages. We firstly explore and find three important foundations of the collaboratively annotated tags, that are $consistency$, $sharability$ and $stability$. Furthermore, with the study of the complementary advantages among the three methods, TF-weighted keywords, CF and Corr-LDA, on the corpus, we finally propose a blending method for automated social tagging by linearly combining the three methods. With a permutation probability model to learn the linear blending, experimental results show that our approach consistently outperforms all the baselines.

A further study on modeling the relationship of such three types of methods for automated social tagging is our future work. Applying the auto-generated social tags for recommendation, filtering and search is another future research topic.

References

[1] S. A. Golder and B. A. Huberman, "Usage patterns of collaborative tagging systems," *J. Inf. Sci.*, vol. 32, no. 2, pp. 198–208, 2006.

[2] E. Quintarelli, "Folksonomies: power to the people," ISKO Italy-UniMIB Meeting, June 2005. [Online]. Available: http://www.iskoi.org/doc/folksonomies.htm

[3] S. Bao, G. Xue, X. Wu, Y. Yu, B. Fei, and Z. Su, "Optimizing web search using social annotations," in *WWW '07: Proceedings of the 16th international conference on World Wide Web*, 2007, pp. 501–510.

[4] D. Zhou, J. Bian, S. Zheng, H. Zha, and C. L. Giles, "Exploring social annotations for information retrieval," in *WWW '08: Proceeding of the 17th international conference on World Wide Web*, 2008, pp. 715–724.

[5] S. Xu, S. Bao, B. Fei, Z. Su, and Y. Yu, "Exploring folksonomy for personalized search," in *Porc. of international ACM SIGIR conference on Research and development in information retrieval*, 2008, pp. 155–162.

[6] J. M. Conde, D. Vallet, and P. Castells, "Inferring user intent in web search by exploiting social annotations," in *Porc. of international ACM SIGIR conference on Research and development in information retrieval*, 2010, pp. 827–828.

[7] D. Vallet, I. Cantador, and J. M. Jose, "Personalizing web search with folksonomy-based user and document profiles," *In ECIR*, vol. 5993, pp. 420–431, 2010.

[8] J. Guo, X. Cheng, G. Xu, and H. Shen, "A structured approach to query recommendation with social annotation data," in *Proc. of the ACM conference on Information and knowledge management*, 2010, pp. 619–628.

[9] P. Heymann, G. Koutrika, and H. Garcia-Molina, "Can social bookmarking improve web search?" in *In WSDM'08: Proc. of international conference on Web search and web data mining*, 2008, pp. 195–206.

[10] Y.-T. Lu, S.-I. Yu, T.-C. Chang, and J. Y. Hsu, "A content-based method to enhance tag recommendation," in *In Proc. of IJCAI'09*, 2009, pp. 2064–2069.

[11] P. A. Chirita, S. Costache, W. Nejdl, and S. Handschuh, "P-tag: large scale automatic generation of personalized annotation tags for the web," in *WWW '07: Proceedings of the 16th international conference on World Wide Web*, 2007, pp. 845–854.

[12] G. Salton and M. McGill, *Introduction to Modern Information Retrieval*. McGraw-Hill, 1983.

[13] B. Sigurbjörnsson and R. van Zwol, "Flickr tag recommendation based on collective knowledge," in *WWW '08: Proceeding of the 17th international conference on World Wide Web*, 2008, pp. 327–336.

[14] P. Heymann, D. Ramage, and H. Garcia-Molina, "Social tag prediction," in *SIGIR '08: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, 2008, pp. 531–538.

[15] Y. Song, Z. Zhuang, H. Li, Q. Zhao, J. Li, W.-C. Lee, and C. L. Giles, "Real-time automatic tag recommendation," in *SIGIR '08: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, 2008, pp. 515–522.

[16] M. Dubinko, R. Kumar, J. Magnani, J. Novak, P. Raghavan, and A. Tomkins, "Visualizing tags over time," *ACM Trans. Web*, vol. 1, no. 2, p. 7, 2007.

[17] S. Dill, N. Eiron, D. Gibson, D. Gruhl, R. V. Guha, A. Jhingran, T. Kanungo, S. Rajagopalan, A. Tomkins, J. A. Tomlin, and J. Y. Zien, "SemTag and Seeker: Bootstrapping the Semantic Web via Automated Semantic Annotation," in *WWW '03: Proc. of the Twelfth International World Wide Web Conference*, May 20-24 - Budapest, Hungary 2003, pp. 178–186.

[18] S. Handschuh, S. Staab, and R. Volz, "On deep annotation," in *WWW '03: Proceedings of the 12th international conference on World Wide Web*, 2003, pp. 431–438.

[19] K. Bischoff, C. S. Firan, W. Nejdl, and R. Paiu, "Can all tags be used for search?" in *Proc. of the ACM conference on Information and knowledge management*, 2008, pp. 193–202.

[20] D. Goldberg, D. Nichols, B. Oki, and D. Terry, "Using collaborative filtering to weave an information tapestry," in *Communications of the ACM*, 1992.

[21] R. Nakamoto, S. Nakajima, J. Miyazaki, and S. Uemura, "Tag-based contextual collaborative filtering," *IAENG International Journal of Computer Science*, vol. 34, no. 2, pp. 214–219, 2007.

[22] D. M. Blei and M. I. Jordan, "Modeling annotated data," in *SIGIR '03: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, 2003, pp. 127–134.

[23] X. Li, L. Guo, and Y. E. Zhao, "Tag-based social interest discovery," in *WWW '08: Proceeding of the 17th international conference on World Wide Web*, 2008, pp. 675–684.

[24] H. Halpin, V. Robu, and H. Shepherd, "The complex dynamics of collaborative tagging," in *WWW '07: Proceedings of the 16th international conference on World Wide Web*, 2007, pp. 211–220.

[25] Z. Cao, T. Qin, T.-Y. Liu, M.-F. Tsai, and H. Li, "Learning to rank: From pairwise approach to listwise approach," in *Proceedings of the 24th International Conference on Machine Learning*, 2007, p. 129136.

[26] F. Xia, T.-Y. Liu, J. Wang, W. Zhang, and H. Li, "Listwise approach to learning to rank - theory and algorithm," in *Proceedings of the 25 International Conference on Machine Learning*, 2008, pp. 1192–1199.

[27] E. Erosheva, S. Fienberg, and J. Lafferty, "Mixed membership models of scientific publications," in *Proceedings of the National Academy of Sciences*, vol. 101. press, 2004, pp. 5220–5227.