Collaborative Topic Regression with Social Regularization for Tag Recommendation

Hao Wang, Binyi Chen, Wu-Jun Li
Shanghai Key Laboratory of Scalable Computing and Systems
Department of Computer Science and Engineering, Shanghai Jiao Tong University, China
{js05212, Charles_Chen}@sjtu.edu.cn, liwujun@cs.sjtu.edu.cn

Abstract

Recently, tag recommendation (TR) has become a very hot research topic in data mining and related areas. However, neither *co-occurrence based methods* which only use the item-tag matrix nor *content based methods* which only use the item content information can achieve satisfactory performance in real TR applications. Hence, how to effectively combine the item-tag matrix, item content information, and other auxiliary information into the same recommendation framework is the key challenge for TR. In this paper, we first adapt the *collaborative topic regression* (CTR) model, which has been successfully applied for article recommendation, to combine both item-tag matrix and item content information for TR. Furthermore, by extending CTR we propose a novel hierarchical Bayesian model, called *CTR with social regularization* (CTR-SR), to seamlessly integrate the item-tag matrix, item content information, and social networks between items into the same principled model. Experiments on real data demonstrate the effectiveness of our proposed models.

1 Introduction

Tagging systems have been playing very important role for us to better categorize and organize information. For example, Flickr[1] uses tags to label and organize photos, Last.fm[2] adopts tags to categorize artists and music, and CiteULike[3] allows users to tag articles. With the tagging systems, users are able to better organize their own content and find relevant resources (content) more easily.

However, finding the set of proper words (tags) to describe the resources often requires high mental focus. That is why tag recommendation (TR) [Gupta *et al.*, 2010; Wang *et al.*, 2012] has become more and more important on the Internet. With the tag recommendation system, users only need a few clicks to finish the tagging process. Moreover, tags created by various users can be inconsistent and idiosyncratic. Different users might use different words to express the same meaning, which makes it more difficult to utilize the tagging information. Tag recommendation can help to limit vocabulary of tags and thus alleviate the above problems. Furthermore, it can also help to prevent misspelt or meaningless words. Therefore, TR [Wang *et al.*, 2012] has become a very hot research topic in recent years, and many methods have been proposed by researchers.

Existing tag recommendation methods can be roughly categorized into three classes [Wang *et al.*, 2012]: content-based methods, co-occurrence based methods, and hybrid methods. Content-based methods [Chen *et al.*, 2008; Lipczak *et al.*, 2009; Shen and Fan, 2010; Lee *et al.*, 2010; Toderici *et al.*, 2010; Chen *et al.*, 2010], directly adopt the content of resources/items, such as abstract of articles, image content and description of images, to perform tag recommendation. Co-occurrence based methods [Benz *et al.*, 2006; Xu *et al.*, 2006; Hotho *et al.*, 2006; Marinho and Schmidt-Thieme, 2007; Sigurbjörnsson and van Zwol, 2008; Garg and Weber, 2008; Weinberger *et al.*, 2008; Wu *et al.*, 2009; Rendle and Schmidt-Thieme, 2010] mainly use the co-occurrence of tags among items (i.e., the item-tag matrix) for tagging. Actually, the underlying principle of co-occurrence based methods is similar to that of *collaborative filtering* (CF) methods [Adomavicius and Tuzhilin, 2005; Zhen *et al.*, 2009; Li and Yeung, 2011]. Because the TR problem is very complex and difficult, neither co-occurrence based methods nor content based methods can achieve satisfactory performance in real TR applications. Hence, the recent trend in TR research is to use hybrid methods [Wu *et al.*, 2009; Sevil *et al.*, 2010; Lops *et al.*, 2011; 2013] which try to combine both item-tag matrix and item content information together for recommendation.

However, it is still a challenge to find an effective way to combine both item-tag matrix and item content information for TR. Furthermore, in some applications there may exist social networks (relations) between items. For example, if we want to tag articles in CiteULike, there are citation relations or other social networks between articles [Li *et al.*, 2011; Wang and Li, 2013]. Typically, two articles with relation between them might be most likely to be about the same topic [Li *et al.*, 2009a; 2009b], and consequently they should have similar tags. Hence, how to effectively integrate social networks between items for tagging is another challenge.

[1] http://www.flickr.com
[2] http://www.lastfm.com
[3] http://www.citeulike.org

In this paper, we propose some novel methods to solve the above challenges. The main contributions of this paper can be outlined as follows:

- We adapt the *collaborative topic regression* (CTR) model [Wang and Blei, 2011], which has been successfully applied for article recommendation, to combine both item-tag matrix and item content information for tag recommendation in a principled way.
- By extending CTR, we propose a novel hierarchical Bayesian model, called *CTR with social regularization* (CTR-SR), to seamlessly integrate the item-tag matrix, item content information, and social networks between items into the same principled model.
- Extensive experiments on real-world data sets show that CTR can outperform the baselines which use only one kind of information, either item-tag matrix or item content information. Furthermore, CTR-SR can effectively utilize the social networks between items to further improve the performance.

2 Problem Statement

Assume we have a set of items $W = [\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_J]$ to be tagged, where $\mathbf{w}_j \in \mathbb{R}^d$ denotes the content (attributes) of item j. For example, if we want to tag articles (papers) in CiteULike, the items are papers, and the content information can be the bag-of-word representation of paper abstract. Assume there are I tags $\{\mathbf{t}_1, \mathbf{t}_2, \cdots, \mathbf{t}_I\}$ which are candidates to be recommended to tag each item. Then we can use a tag-item matrix[4] $R = [r_{ij}]_{I\times J}$ to represent the tagging information for all the items. r_{ij} is a binary variable, where $r_{ij} = 1$ means that the tag $\mathbf{t}_i$ is associated with item $\mathbf{w}_j$. Otherwise, $r_{ij} = 0$ means that tag $\mathbf{t}_i$ is not associated with item $\mathbf{w}_j$. The tag recommendation task is to predict the missing values in $r_j = [r_{1j}, r_{2j}, \cdots, r_{Ij}]^T$. Note that we focus on tag recommendation for articles (papers) in this paper. However, our models are flexible enough to be applied in other applications such as image and video tagging because we can also represent the image and video content as bag-of-words.

The content base methods use only the content information for recommendation. For example, if we want to recommend tags for item $\mathbf{w}_j$, we can use the tags from the nearest neighbor in W based on the content similarity. We can also treat each tag as a label and use multi-label methods to train classifiers based on content information.

Co-occurrence based methods use only the item-tag matrix R for recommendation. For example, if $\mathbf{t}_i$ and $\mathbf{t}_k$ occur simultaneously in many items' tags and $\mathbf{t}_i$ is associated with $\mathbf{w}_j$, we should also recommend $\mathbf{t}_k$ to $\mathbf{w}_j$. It is easy to see that the underlying principle of co-occurrence based methods is similar to that of collaborative filtering [Adomavicius and Tuzhilin, 2005].

Both content based methods and co-occurrence based methods discard some useful information. Hence, they can not achieve satisfactory performance in real applications.

[4]For ease of presentation, we use tag-item matrix and item-tag matrix interchangeably in this paper.

3 Collaborative Topic Regression

Collaborative topic regression (CTR) [Wang and Blei, 2011] combines CF and latent Dirichlet allocation (LDA) [Blei *et al.*, 2003] to perform recommendation. CTR is initially proposed to recommend articles (papers) to users by utilizing both user-article rating information and article content information. In this paper, we adapt CTR to our tag recommendation problem to seamlessly integrate both item-tag matrix information and item content information.

For ease of presentation, we use similar graphical models and notations as those in CTR [Wang and Blei, 2011] for our problem formulation. The graphical model of CTR is illustrated in Figure 1. Assume there are K topics $\boldsymbol{\beta} = \beta_{1:K}$. The generative process of CTR for tag recommendation is listed as follows:

1. Draw tag latent vector for each tag i:
$$u_i \sim \mathcal{N}(0, \lambda_u^{-1} I_K),$$
where $\mathcal{N}(\cdot)$ denotes the normal distribution, I_K is an identity matrix with K rows and columns.
2. For each item j,
   (a) Draw topic proportions $\theta_j \sim$ Dirichlet(α).
   (b) Draw item latent offset $\epsilon_j \sim \mathcal{N}(0, \lambda_v^{-1} I_K)$ and then set the item latent vector to be: $v_j = \epsilon_j + \theta_j$.
   (c) For each word w_{jn} of item (paper) $\mathbf{w}_j$,
      i. Draw topic assignment $z_{jn} \sim$ Mult(θ_j).
      ii. Draw word $w_{jn} \sim$ Mult$(\beta_{z_{jn}})$.
3. Draw the tagging information r_{ij} for each tag-item pair (i, j),
$$r_{ij} \sim \mathcal{N}(u_i^T v_j, c_{ij}^{-1}), \qquad (1)$$
where c_{ij} reflects the confidence of r_{ij}:
$$c_{ij} = \begin{cases} a, & \text{if } r_{ij} = 1, \\ b, & \text{if } r_{ij} = 0, \end{cases}$$
with a and b being tuning parameters and $a > b > 0$.

We can adopt the maximum a posteriori (MAP) estimation to learn the parameters of CTR. The details can be found in [Wang and Blei, 2011].

It is easy to see that the above process integrates matrix factorization (MF) [Koren *et al.*, 2009] based CF (Equation (1)) for tagging information and topic modeling for item content information into the same principled framework.

4 Collaborative Topic Regression with Social Regularization

By extending CTR, we propose a novel hierarchical Bayesian model, called *CTR with social regularization* (CTR-SR), to seamlessly integrate the item-tag matrix, item content information, and social networks between items into the same principled model. The graphical model of CTR-SR is shown in Figure 2.

The generative process of CTR-SR is listed as follows:

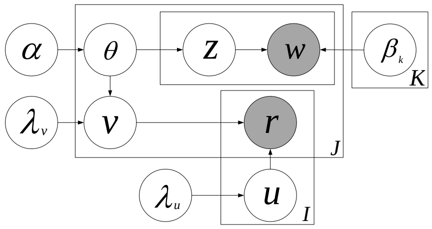


Figure 1: The graphical model of collaborative topic regression (CTR).

1. Draw tag latent vector for each tag $\mathbf{t}_i$:
$$u_i \sim \mathcal{N}(0, \lambda_u^{-1} I_K).$$
2. For each item j,
   (a) Draw topic proportions $\theta_j \sim \text{Dirichlet}(\alpha)$.
   (b) For each word w_{jn} of item (paper) $\mathbf{w}_j$,
      i. Draw topic assignment $z_{jn} \sim \text{Mult}(\theta_j)$.
      ii. Draw word $w_{jn} \sim \text{Mult}(\beta_{z_{jn}})$.
3. Draw the *social latent matrix* $S = [s_1, s_2, \cdots, s_J]$ from a *matrix variate normal distribution* [Gupta and Nagar, 2000]:
$$S \sim \mathcal{N}_{K,J}(0, I_K \otimes (\lambda_l \mathscr{L}_a)^{-1}). \tag{2}$$
4. Draw the *item latent vector* for item j from the product of two Gaussians (PoG) [Gales and Airey, 2006]:
$$v_j \sim \text{PoG}(\theta_j, s_j, \lambda_v^{-1} I_K, \lambda_r^{-1} I_K). \tag{3}$$
5. Draw the tagging information r_{ij} for each tag-item pair (i, j),
$$r_{ij} \sim \mathcal{N}(u_i^T v_j, c_{ij}^{-1}).$$

In the above generative process, S denotes the *social latent matrix* of size $K \times J$, each column of which is the *social latent vector* s_j for item j, $\mathcal{N}_{K,J}(0, I_K \otimes (\lambda_l \mathscr{L}_a)^{-1})$ in (2) denotes a *matrix variate normal distribution* [Gupta and Nagar, 2000]:
$$\begin{aligned} p(S) &= \mathcal{N}_{K,J}(0, I_K \otimes (\lambda_l \mathscr{L}_a)^{-1}) \\ &= \frac{\exp\{\text{tr}[-\frac{\lambda_l}{2} S \mathscr{L}_a S^T]\}}{(2\pi)^{JK/2} |I_K|^{J/2} |\lambda_l \mathscr{L}_a|^{-K/2}}, \end{aligned} \tag{4}$$
where the operator $\otimes$ denotes the Kronecker product of two matrices [Gupta and Nagar, 2000], $\text{tr}(\cdot)$ denotes the trace of a matrix, $\mathscr{L}_a$ is the Laplacian matrix incorporating the social network information. $\mathscr{L}_a = D - A$ where D is a diagonal matrix whose diagonal elements $D_{ii} = \sum_j A_{ij}$. Here A is the adjacency matrix of the social networks with binary entries indicating the links (relations) between items. $A_{jj'} = 1$ indicates that there is a link between item j and item j'. Otherwise, $A_{jj'} = 0$. $\text{PoG}(\theta_j, s_j, \lambda_v^{-1} I_K, \lambda_r^{-1} I_K)$ in (3) denotes the product of the Gaussian $\mathcal{N}(\theta_j, \lambda_v^{-1} I_K)$ and the Gaussian $\mathcal{N}(s_j, \lambda_r^{-1} I_K)$, which is also a Gaussian [Gales and Airey, 2006]. The resulting Gaussian is $\mathcal{N}(\mu_{vr}, \lambda_{vr}^{-1} I_K)$ with
$$\mu_{vr} = \frac{\theta_j \lambda_v + s_j \lambda_r}{\lambda_v + \lambda_r},$$
$$\lambda_{vr} = \frac{\lambda_v \lambda_r}{\lambda_v + \lambda_r}.$$

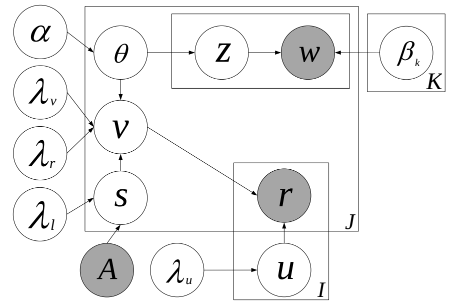


Figure 2: The graphical model of collaborative topic regression with social regularization (CTR-SR).

As shown in (2) and Figure 2, the social network information is seamlessly integrated into the CTR-SR by putting the Laplacian of the adjacency matrix into the prior distribution for S. The physical meaning is to make the latent factors (s_j and v_j) of linked items as close as possible, which will be discussed in detail in the following content.

Since it is obviously intractable to compute the full posterior of u_i, v_j, s_j, and θ_j, an EM-style algorithm is developed to learn the maximum a posteriori (MAP) estimation. We can maximize the posterior by maximizing the complete log-likelihood of $U = [u_1, u_2, \cdots, u_I]$, $V = [v_1, v_2, \cdots, v_J]$, S, $\theta_{1:J}$, and R given λ_u, λ_v, λ_r, λ_l and β,
$$\begin{aligned} \mathscr{L} = &-\frac{\lambda_l}{2} \text{tr}(S \mathscr{L}_a S^T) - \frac{\lambda_r}{2} \sum_j (s_j - v_j)^T (s_j - v_j) \\ &- \frac{\lambda_u}{2} \sum_i u_i^T u_i - \frac{\lambda_v}{2} \sum_j (v_j - \theta_j)^T (v_j - \theta_j) \\ &+ \sum_j \sum_n \log(\sum_k \theta_{jk} \beta_{k,w_{jn}}) - \sum_{i,j} \frac{c_{ij}}{2} (r_{ij} - u_i^T v_j)^2. \end{aligned} \tag{5}$$

A constant is omitted and the parameter of the topic model α is set to 1 as that in CTR. Note that the first term $-\frac{\lambda_l}{2} \text{tr}(S \mathscr{L}_a S^T)$ corresponds to $\log p(S)$ with a constant omitted and
$$\begin{aligned} \text{tr}(S \mathscr{L}_a S^T) &= \frac{1}{2} \sum_{j=1}^{J} \sum_{j'=1}^{J} A_{jj'} ||S_{*j} - S_{*j'}||^2 \\ &= \frac{1}{2} \sum_{j=1}^{J} \sum_{j'=1}^{J} [A_{jj'} \sum_{k=1}^{K} (S_{kj} - S_{kj'})^2] \\ &= \frac{1}{2} \sum_{k=1}^{K} [\sum_{j=1}^{J} \sum_{j'=1}^{J} A_{jj'} (S_{kj} - S_{kj'})^2] \\ &= \sum_{k=1}^{K} S_{k*}^T \mathscr{L}_a S_{k*}, \end{aligned} \tag{6}$$
where S_{r*} denotes the rth row of S and S_{*c} denotes the cth column of S. We can see that maximizing $-\frac{\lambda_l}{2} \text{tr}(S^T \mathscr{L}_a S)$ will make s_j close to $s_{j'}$ if item j and item j' are linked ($A_{jj'} = 1$).

The function $\mathscr{L}$ in (5) can be optimized using coordinate ascent. We first fix parameters β and optimize the collaborative filtering variables $\{u_i, v_j, s_j\}$ and the topic proportions θ_j iteratively. The parameter β is updated every time $\{u_i, v_j, s_j\}$ and θ_j are optimized.

The update rules for u_i and v_j are:

$$u_i \leftarrow (VC_iV^T + \lambda_u I_K)^{-1} VC_iR_i,$$

$$v_j \leftarrow (UC_iU^T + \lambda_v I_K + \lambda_r I_K)^{-1}(UC_jR_j + \lambda_v\theta_j + \lambda_r s_j),$$

where C_i is a diagonal matrix with $\{c_{ij}, j = 1, \dots, J\}$ as its diagonal entries and R_j is the jth row of R.

For *social latent matrix* S, we fix all rows of S except the kth one S_{k*} and then update S_{k*}. After taking the gradient of $\mathscr{L}$ with respect to S_{k*} and setting it to 0, we get the following linear system:

$$(\lambda_l \mathscr{L}_a + \lambda_r I)S_{k*} = \lambda_r V_{k*}. \quad (7)$$

One direct way to solve the linear system is to set $S_{k*} = \lambda_r(\lambda_l\mathscr{L}_a + \lambda_r I_J)^{-1}V_{k*}$. However, the complexity for one single update is $O(J^3)$ where J is the number of items. Inspired by [Li and Yeung, 2009], we use the *steepest descent* method [Shewchuk, 1994] to iteratively update S_{k*}:

$$\begin{aligned} S_{k*}(t+1) &\leftarrow S_{k*}(t) + \delta(t)r(t) \\ r(t) &\leftarrow \lambda_r V_{k*} - (\lambda_l\mathscr{L}_a + \lambda_r I_J)S_{k*}(t) \\ \delta(t) &\leftarrow \frac{r(t)^T r(t)}{r(t)^T(\lambda_l\mathscr{L}_a + \lambda_r I_J)r(t)} \end{aligned}$$

As discussed in [Li and Yeung, 2009], using the steepest descent method instead of solving the linear system directly can dramatically reduce the computation cost in each iteration from $O(J^3)$ to $O(J)$.

For θ_j, we first define $q(z_{jn=k}) = \psi_{jnk}$ as that in CTR and LDA [Blei *et al.*, 2003] and apply Jensen's inequality after items containing θ_j are separated,

$$\begin{aligned} \mathscr{L}(\theta_j) \geq & -\frac{\lambda_v}{2}(v_j - \theta_j)^T(v_j - \theta_j) \quad (8) \\ & + \sum_n \sum_k \phi_{jnk}(\log \theta_{jk}\beta_{k,w_{jn}} - \log \phi_{jnk}) \\ & = \mathscr{L}(\theta_j, \boldsymbol{\phi}_j). \end{aligned}$$

Here $\phi_j = (\phi_{jnk})_{n=1,k=1}^{N\times K}$. Obviously $\mathscr{L}(\theta_j, \boldsymbol{\phi}_j)$ is a tight lower bound of $\mathscr{L}(\theta_j)$ and we can use projection gradient to optimize θ_j. The optimal ϕ_{jnk} is

$$\phi_{jnk} \propto \theta_{jk}\beta_{k,w_{jn}}.$$

As for the parameter β, we follow the same M-step update as in LDA [Blei *et al.*, 2003],

$$\beta_{kw} \propto \sum_j \sum_n \phi_{jnk} 1[w_{jn} = w].$$

5 Experiments

We conduct experiments on two real-world data sets to demonstrate the effectiveness of our models. As stated in Section 2, although our focus is on tag recommendation for articles (papers) in this paper, our models are general enough to model other kinds of data like image tagging.

5.1 Dataset

Two real-world datasets are used in our experiments. Both of them are from CiteULike[5], but they are collected in different ways. The first dataset, called *citeulike-a*, is from [Wang and Blei, 2011]. Note that there is not tag information in the original dataset of [Wang and Blei, 2011]. We collect the tag information from CiteULike. We collect the second dataset, called *citeulike-t*, by ourselves. Specifically, we manually select 273 seed tags and collect all the articles with at least one of these tags. Note that the final number of tags (19107 and 52946 respectively for two datasets) corresponding to all the collected articles is far more than the number of seed tags (273). We remove tags used less than 5 times and get 7386 and 8311 tags for *citeulike-a* and *citeulike-t*, respectively. There are 16980 items (articles) and 25975 items in the datasets *citeulike-a* and *citeulike-t*, respectively. The ratios of non-empty entries (equal to 1-*sparsity*) in the item-tag matrices of *citeulike-a* and *citeulike-t* are 0.00145 and 0.00104 respectively, which means that the second dataset is sparser than the first one.

We preprocess the text information (content of items) following the same procedure as that in [Wang and Blei, 2011]. As in [Wang and Blei, 2011], we also use the titles and abstracts of articles as content information of *citeulike-t*. We choose the top 20000 distinct words according to the tf-idf values as our vocabulary after removing the stop-words.

Because citation information is not provided in CiteULike, we use the user-article information which is available in CiteULike to construct the social networks between items. For each dataset, we construct the *social network* with a threshold of 4 using the user-article matrices. More specifically, if two items have 4 or more users in common, they are linked in the *social network*. This is meaningful because two papers with similar users (readers) typically have similar topics. We then merge this *social network* and the *citation network* between papers to get the *final network*. After network constructing, the numbers of links in the *final networks* are 294072 and 180103 for *citeulike-a* and *citeulike-t*, respectively.

5.2 Evaluation Scheme

In each dataset, we randomly select P items associated with each tag to construct the training set and use all the rest of the dataset as test set. We vary P from 1 to 10 in our experiments and the smaller P is, the sparser the training set is. Note that when $P = 1$, only 4.1% of the tagging entries are put in training set for dataset *citeulike-a* and the number is 3.7% for dataset *citeulike-t*. For each P we repeat the evaluation 5 times with randomly selected training set, and the average performance will be reported.

As in [Wang and Blei, 2011] and [Marinho and Schmidt-Thieme, 2007], we use recall as our evaluation metric since zero entries may be caused either by irrelevance between the tag and the item or by users who do not know the existence

[5]CiteULike allows users to create their own collections of articles. There are abstracts, titles, and tags for each article. Other information like authors, groups, posting time, and keywords is not used in this paper. The detailed information can be found at http://www.citeulike.ort/faq/data.adp

of the tags when tagging items, which means precision is not a proper metric here. Like most recommender systems, we sort the predicted ratings of candidate tags and recommend the top M tags to the target item. For each item, recall@M is defined as

$$\text{recall@M} = \frac{\text{number of tags the item is associated with in top M}}{\text{total number of tags the item is associated with}}.$$

The final reported result is the average of all the items' recall.

Besides, as in [Sigurbjörnsson and van Zwol, 2008], we use success@M to be another evaluation metric. success@M is defined as the probability of finding a true tag among the top M recommended tags.

5.3 Baselines and Experimental Settings

We use the following baselines for comparison:

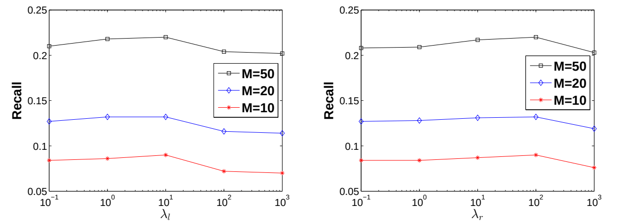
- TAGCO: This method belongs to the category of co-occurrence based methods, which is described in [Sigurbjörnsson and van Zwol, 2008].
- SCF: This is a similarity-based collaborative filtering [Marinho and Schmidt-Thieme, 2007]. It finds k nearest neighbors of the paper's existing tags and recommends other tags according to its neighbors' tags. It only uses the item-tag matrix information.
- CF: This is a matrix factorization based collaborative filtering [Koren *et al.*, 2009] method. It factorizes the training matrix into two low-rank matrices U, V, and recovers the original matrix by UV^T. It only uses the item-tag matrix information.
- SCF+LDA: This method integrates similarity-based collaborative filtering with LDA. It falls into the category of *hybrid methods* and is adapted from [Sevil *et al.*, 2010].
- CTR: The method introduced in Section 3.

We use a validation set to find the optimal parameters. More specifically, we find that CTR achieves good prediction performance when $\lambda_v = 10$, $\lambda_u = 0.1$, $a = 1$, $b = 0.01$, and $K = 200$. For CF, the parameters are $\lambda_v = 1$, $\lambda_u = 1$, $a = 1$, $b = 0.01$, and $K = 200$. For CTR-SR, the parameters are $\lambda_v = 10$, $\lambda_u = 0.1$, $a = 1$, $b = 0.01$, $K = 200$, $\lambda_r = 100$ and $\lambda_l = 10$.

5.4 Performance

Figure 3 (a) and Figure 4 (a) show the recall@50 when P is set to be 1, 2, 5, 8, 10, on *citeulike-a* and *citeulike-t*, respectively. The random baselines are 0.68% and 0.60% respectively. As we can see, the hybrid method SCF+LDA outperforms those methods use only one kind of information, and CTR outperforms SCF+LDA. Furthermore, our CTR-SR model achieves the best performance for most cases by effectively integrating the social networks between items.

Figure 3 (b) and Figure 4 (b) show the recall of all the methods when P is fixed to be 5 by setting M=2, 5, 10, 20, 50 in dataset *citeulike-a* and *citeulike-t*. Figure 3 (c) and Figure 4 (c) show the success@M of all the methods when P is fixed to be 5 by setting M=2, 5, 10, 20, 50 in two datasets. Once again, we can see that CTR outperforms other baselines and CTR-SR is significantly better than other methods for most cases. Similar phenomena are observed for other values of P, which are omitted due to space constraint.



(a) (b)

Figure 5: Sensitivity to parameters. (a) The effect of λ_l in CTR-SR. (b) The effect of λ_r in CTR-SR.

5.5 Sensitivity to Parameters

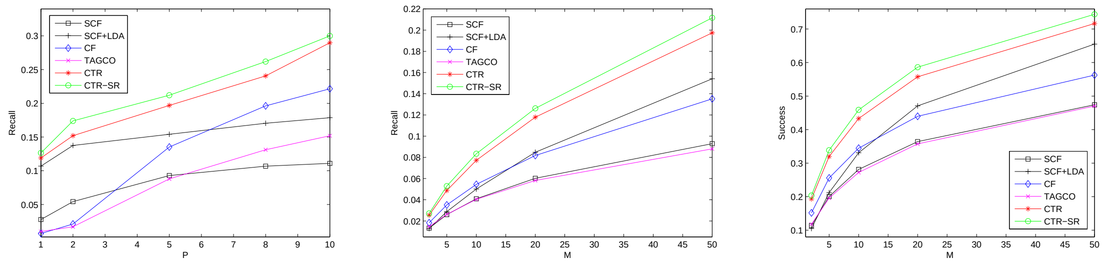
Figure 5 (a) shows how the prediction performance of CTR-SR is affected by the parameter λ_l. P is set to 5, $\lambda_v = 10$, $\lambda_u = 0.1$, and $\lambda_r = 100$. As we can see, the prediction performance first increases with λ_l and starts to slightly decrease at some point after $\lambda_l = 10$ for all values of M. It is not too sensitive in a large range of values.

Figure 5 (b) demonstrates the sensitivity of CTR-SR to parameter λ_r. In this experiment, P is also set to 5 and $\lambda_v = 10$, $\lambda_u = 0.1$, and $\lambda_l = 10$. As the figure shows, the performance first increases with λ_r and begins to decrease at some point after $\lambda_r = 100$ for all values of M. It is also not too sensitive in a large range of values.

5.6 Interpretability

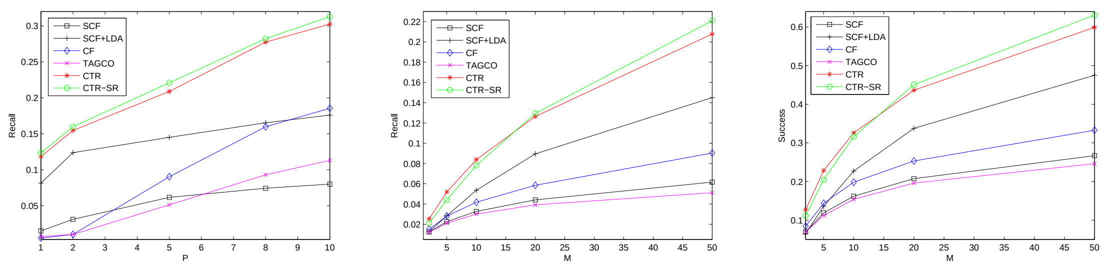
Besides promising prediction performance, our proposed model can also provide a very good interpretation. Two example articles (items) with their top 2 topics are presented in Table 1. Note that although the learned topic proportions of CTR are different from those of CTR-SR, the ranking of top 2 topics are the same. In this case study, CTR-SR and CTR are trained using the extremely sparse training data ($P = 1$) and recommend tags to articles. Note that in the training data, each tag is associated with only one single article, which makes tag recommendation very challenging. As we can see from Table 1, for Article I, precisions of the top 10 tags for CTR-SR and CTR are 50% and 10%, respectively. For Article II, the precisions are 60% and 10%, respectively. We can find that the social network information among items are very informative and our CTR-SR model can effectively exploit it.

When examining more closely, we can find that Article I '*How much can behavioral targeting help online advertising?*' is about *online advertising*, which can also be verified by the true tags shown in the table. As we can see, the recommended tags by CTR focuses more on the technical details while those returned by CTR-SR are closer to the essence of the articles. Similarly, Article II '*Lowcost multitouch sensing through frustrated total internal reflection*' focuses on *multitouch sensing*. Tags recommended by CTR like 'nanoparticles', 'dna-nanotechnology', and 'gamma' seem a lot more technical and achieve a low precision of 10%. On the contrary, tags recommended by CTR-SR like 'multi-touch' and 'screen' can better describe the main points of the article and achieve a high precision of 60%.



(a) (b) (c)

Figure 3: Experimental results on dataset *citeulike-a*. (a) shows the recall@50 of all the methods. (b) shows the recall of all the methods when $P = 5$ and M ranges from 2 to 50. (c) shows the success@M of all the methods when $P = 5$ and M ranges from 2 to 50.



(a) (b) (c)

Figure 4: Experimental results on dataset *citeulike-t*. (a) shows the recall@50 of all the methods. (b) shows the recall of all the methods when $P = 5$ and M ranges from 2 to 50. (c) shows the success@M of all the methods when $P = 5$ and M ranges from 2 to 50.

Table 1: Example Articles with Recommended Tags

Article I	Title: How much can behavioral targeting help online advertising?			
	Top topic 1: web, search, engine, pages, keyword, click, hypertext, html, searchers, crawler			
	Top topic 2: mobile, phones, attitudes, advertising, consumer, marketing, commerce, sms, m-learning			
	True tags: behavioral_targeting, advertising, ads, computational_advertising, recommend, user-behavior, user_profile			
Top 10 recommended tags	CTR	True tag?	CTR-SR	True tag?
	1. random-walks	no	1. behavioral_targeting	yes
	2. page-rank	no	2. ads	yes
	3. computational_advertising	yes	3. computational_advertising	yes
	4. citizen-science	no	4. random-walks	no
	5. natural_history	no	5. page-rank	no
	6. search_engine	no	6. developing	no
	7. engine	no	7. recommend	yes
	8. searchengine	no	8. advertising	yes
	9. what	no	9. what	no
	10. re-ranking	no	10. need	no
Article II	Title: Lowcost multitouch sensing through frustrated total internal reflection			
	Top topic 1: molecular, molecules, surface, chemical, formation, forces, reaction, shapes, sensing, kinetics			
	Top topic 2: design, interface, principles, interfaces, interactive, devices, usability, application			
	True tags: tech, screen, gestures, touch, interface, multitouch, multi-touch, table, visualization, computer_vision			
Top 10 recommended tags	CTR	True tag?	CTR-SR	True tag?
	1. guide	no	1. touch	yes
	2. gamma	no	2. field	no
	3. optical	no	3. gestures	yes
	4. nanoparticles	no	4. table	yes
	5. nano	no	5. multi-touch	yes
	6. dna-nanotecnology	no	6. screen	yes
	7. tirf	no	7. multitouch	yes
	8. sms	no	8. dna-nanotecnology	no
	9. touch	yes	9. nano	no
	10. field	no	10. superlist	no

6 Conclusion

In this paper, we first adapt CTR to combine both item-tag matrix and item content information for tag recommendation. Furthermore, by extending CTR we propose a novel hierarchical Bayesian model, called CTR with social regularization (CTR-SR), to seamlessly integrate the item-tag matrix, item content information, and social networks between items into the same principled model. Experiments on real-world datasets successfully demonstrate the effectiveness of our proposed models.

7 Acknowledgements

This work is supported by the NSFC (No. 61100125), the 863 Program of China (No. 2012AA011003), and the Program for Changjiang Scholars and Innovative Research Team in University of China (IRT1158, PCSIRT).

References

[Adomavicius and Tuzhilin, 2005] G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6):734–749, 2005.

[Benz *et al.*, 2006] Dominik Benz, Karen H. L. Tso, and Lars Schmidt-Thieme. Automatic bookmark classification - a collaborative approach. In *Proceedings of the 2nd Workshop in Innovations in Web Infrastructure (IWI2) at WWW2006*, Edinburgh, Scotland, May 2006.

[Blei *et al.*, 2003] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

[Chen *et al.*, 2008] Hong-Ming Chen, Ming-Hsiu Chang, Ping-Chieh Chang, Min-Chun Tien, Winston H. Hsu, and Ja-Ling Wu. Sheepdog: group and tag recommendation for flickr photos by automatic search-based learning. In *ACM Multimedia*, pages 737–740, 2008.

[Chen *et al.*, 2010] Lin Chen, Dong Xu, Ivor Wai-Hung Tsang, and Jiebo Luo. Tag-based web photo retrieval improved by batch mode re-tagging. In *CVPR*, pages 3440–3446, 2010.

[Gales and Airey, 2006] M. J. F. Gales and S. S. Airey. Product of gaussians for speech recognition. *CSL*, 20(1):22–40, 2006.

[Garg and Weber, 2008] Nikhil Garg and Ingmar Weber. Personalized, interactive tag recommendation for flickr. In *RecSys*, pages 67–74, 2008.

[Gupta and Nagar, 2000] A.K. Gupta and D.K. Nagar. *Matrix Variate Distributions*. Chapman & Hall/CRC Monographs and Surveys in Pure and Applied Mathematics. Chapman & Hall, 2000.

[Gupta *et al.*, 2010] Manish Gupta, Rui Li, Zhijun Yin, and Jiawei Han. Survey on social tagging techniques. *SIGKDD Explorations*, 12(1):58–72, 2010.

[Hotho *et al.*, 2006] Andreas Hotho, Robert Jäschke, Christoph Schmitz, and Gerd Stumme. Information retrieval in folksonomies: Search and ranking. In *ESWC*, pages 411–426, 2006.

[Koren *et al.*, 2009] Yehuda Koren, Robert M. Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *IEEE Computer*, 42(8):30–37, 2009.

[Lee *et al.*, 2010] Sihyoung Lee, Wesley De Neve, Konstantinos N. Plataniotis, and Yong Man Ro. Map-based image tag recommendation using a visual folksonomy. *Pattern Recognition Letters*, 31(9):976–982, 2010.

[Li and Yeung, 2009] Wu-Jun Li and Dit-Yan Yeung. Relation regularized matrix factorization. In *IJCAI*, pages 1126–1131, 2009.

[Li and Yeung, 2011] Wu-Jun Li and Dit-Yan Yeung. Social relations model for collaborative filtering. In *AAAI*, 2011.

[Li *et al.*, 2009a] Wu-Jun Li, Dit-Yan Yeung, and Zhihua Zhang. Probabilistic relational PCA. In *NIPS*, pages 1123–1131, 2009.

[Li *et al.*, 2009b] Wu-Jun Li, Zhihua Zhang, and Dit-Yan Yeung. Latent wishart processes for relational kernel learning. *Journal of Machine Learning Research - Proceedings Track*, 5:336–343, 2009.

[Li *et al.*, 2011] Wu-Jun Li, Dit-Yan Yeung, and Zhihua Zhang. Generalized latent factor models for social network analysis. In *IJCAI*, pages 1705–1710, 2011.

[Lipczak *et al.*, 2009] Marek Lipczak, Yeming Hu, Yael Kollet, and Evangelos Milios. Tag sources for recommendation in collaborative tagging systems. In *ECML PKDD Discovery Challenge 2009 (DC09)*, volume 497 of *CEUR-WS.org*, pages 157–172, September 2009.

[Lops *et al.*, 2011] Pasquale Lops, Marco de Gemmis, and Giovanni Semeraro. Content-based recommender systems: State of the art and trends. In Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul B. Kantor, editors, *Recommender Systems Handbook*, pages 73–105. Springer, 2011.

[Lops *et al.*, 2013] Pasquale Lops, Marco de Gemmis, Giovanni Semeraro, Cataldo Musto, and Fedelucio Narducci. Content-based and collaborative techniques for tag recommendation: an empirical evaluation. *J. Intell. Inf. Syst.*, 40(1):41–61, 2013.

[Marinho and Schmidt-Thieme, 2007] Leandro Balby Marinho and Lars Schmidt-Thieme. Collaborative tag recommendations. In *GFKL*, pages 533–540, 2007.

[Rendle and Schmidt-Thieme, 2010] Steffen Rendle and Lars Schmidt-Thieme. Pairwise interaction tensor factorization for personalized tag recommendation. In *WSDM*, pages 81–90, 2010.

[Sevil *et al.*, 2010] Sare Gul Sevil, Onur Kucuktunc, Pinar Duygulu, and Fazli Can. Automatic tag expansion using visual similarity for photo sharing websites. *Multimedia Tools Appl.*, 49(1):81–99, 2010.

[Shen and Fan, 2010] Yi Shen and Jianping Fan. Leveraging loosely-tagged images and inter-object correlations for tag recommendation. In *ACM Multimedia*, pages 5–14, 2010.

[Shewchuk, 1994] Jonathan R Shewchuk. An introduction to the conjugate gradient method without the agonizing pain. Technical report, Carnegie Mellon University, Pittsburgh, PA, USA, 1994.

[Sigurbjörnsson and van Zwol, 2008] Börkur Sigurbjörnsson and Roelof van Zwol. Flickr tag recommendation based on collective knowledge. In *WWW*, pages 327–336, 2008.

[Toderici *et al.*, 2010] George Toderici, Hrishikesh Aradhye, Marius Pasca, Luciano Sbaiz, and Jay Yagnik. Finding meaning on youtube: Tag recommendation and category discovery. In *CVPR*, pages 3447–3454, 2010.

[Wang and Blei, 2011] Chong Wang and David M. Blei. Collaborative topic modeling for recommending scientific articles. In *KDD*, pages 448–456, 2011.

[Wang and Li, 2013] Hao Wang and Wu-Jun Li. Online egocentric models for citation networks. In *IJCAI*, 2013.

[Wang *et al.*, 2012] Meng Wang, Bingbing Ni, Xian-Sheng Hua, and Tat-Seng Chua. Assistive tagging: A survey of multimedia tagging with human-computer joint exploration. *ACM Comput. Surv.*, 44(4):25, 2012.

[Weinberger *et al.*, 2008] Kilian Q. Weinberger, Malcolm Slaney, and Roelof van Zwol. Resolving tag ambiguity. In *ACM Multimedia*, pages 111–120, 2008.

[Wu *et al.*, 2009] Lei Wu, Linjun Yang, Nenghai Yu, and Xian-Sheng Hua. Learning to tag. In *WWW*, pages 361–370, 2009.

[Xu *et al.*, 2006] Z. Xu, Y. Fu, J. Mao, and D. Su. Towards the semantic web: Collaborative tag suggestions. In *Proceedings of the Collaborative Web Tagging Workshop at the WWW 2006*, Edinburgh, Scotland, 2006.

[Zhen *et al.*, 2009] Yi Zhen, Wu-Jun Li, and Dit-Yan Yeung. TagiCoFi: tag informed collaborative filtering. In *RecSys*, pages 69–76, 2009.