A hybrid PLSA approach for warmer cold start in folksonomy recommendation

Alan Said Robert Wetzker Winfried Umbrath Leonhard Hennig
DAI Labor
Technische Universität Berlin
{alan.said, robert.wetzker, winfried.umbrath, leonhard.hennig}@dai-labor.de

ABSTRACT

We investigate the problem of item recommendation during the first months of the collaborative tagging community *CiteULike*. CiteULike is a so-called folksonomy where users have the possibility to organize publications through annotations - tags. Making reliable recommendations during the initial phase of a folksonomy is a difficult task, since information about user preferences is meager. In order to improve recommendation results during this *cold start* period, we present a probabilistic approach to item recommendation. Our model extends previously proposed models such as probabilistic latent semantic analysis (PLSA) by merging both user-item as well as item-tag observations into a unified representation. We find that bringing tags into play reduces the risk of overfitting and increases overall recommendation quality. Experiments show that our approach outperforms other types of recommenders.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Information Search and Retrieval

Keywords

recommendations, folksonomies, CiteULike, cold start, PLSA

1. INTRODUCTION

Recommender systems have become an integral part of almost every Web 2.0 site, allowing users to easily discover relevant content. Many social tagging communities, such as *CiteULike*[1], *Delicious*[2] and *Bibsonomy*[3], use recommendation techniques as part of their service. These communities, generally referred to as folksonomies, give users the possibility to annotate items with freely chosen keywords (tags) for easier content retrieval at later points in time [14]. Most of these folksonomy systems recommend suitable tags when a user tags a new item.

In this paper we consider a problem all new folksonomy websites and services offering recommendations encounter – the cold start phase, during which recommendations have to be made based on very little historical data. During this phase, the similarities between items are hard to calculate as the user-item graph is sparsely, if at all, connected. However, when turning to tags, we find a higher connectivity. Our objective is to increase the quality of item recommendations during this cold start phase by utilizing item-tag co-occurrences in conjunction with their user-item co-occurrence counterparts. By doing so, we improve standard collaborative filtering models by considering user-given annotations, i.e. tags.

Probabilistic latent semantic analysis (PLSA), as introduced by Hofmann in [8], is known for improving recommendation quality in different settings [1]. PLSA assumes a lower dimensional latent topic distribution of the observed co-occurrences. These latent distributions group similar items together. We use an extended model of the hybrid PLSA recommender described in [15]. Our recommender derives the latent topic distribution from user-item and item-tag co-occurrences in parallel. Furthermore we extend this model to cope with known issues related to overfitting.

We perform our evaluation on a subset of the CiteULike dataset. CiteULike is a service allowing users to share, organize and store scholarly papers by assigning tags. Daily snapshots of the CiteULike dataset are made available through the official website[4]. Since we are only interested in the startup period of recommendation systems we carry out our experiments on the first 12 months of the available data, starting on day one - when the first document was tagged on November 4, 2004. For comparison we also present how our model performs after 24 months, which corresponds to a point in time when CiteULike had become an established service. Our results clearly show that our Hybrid PLSA (HyPLSA) model produces higher quality recommendations during the cold start period compared to other models. Additionally we find that our approach performs well when the dataset has grown significantly in size, although these improvements are not as significant as the ones during the cold start period though.

[1]http://www.citeulike.org/
[2]http://delicious.com/
[3]http://www.bibsonomy.org/
[4]http://www.citeulike.org/faq/data.adp

Recommender Systems '09 New York, New York USA

1.1 Related Work

Recommender systems can be divided into three main categories; collaborative filtering-based, content-based, and so-called hybrid systems which combine both. Collaborative filtering approaches base their recommendations solely on co-occurrence observations between users and items. Content-based ones, as the name suggests, derive their similarities based on content, i.e. term distributions etc. Hybrid systems utilize data from both of these models. In folksonomies tags tend to reflect the content of the tagged item [2], thus even if we do not consider the actual item content itself, we group tag-based recommender system together with the content-based ones.

The authors of [15] present a hybrid approach to item recommendation in collaborative tagging communities based on PLSA in which they exploit tags to improve recommendations on very large datasets.

Since the introduction of PLSA by Hoffman in [8] it has shown to perform very well in a wide area of topics, among others it continues to outperform multiple other recommendation and decomposition algorithms [1, 16]. A drawback of PLSA is that it does not necessary converge to the global optimum [5]. One way to overcome any effects that may arise from this is presented in [3, 6] where the authors show that multiple training cycles for the same test/train splits provide for more robust results.

Past research within the context of recommendation in folksonomies has, until recently, been focused on tag recommendation [7, 13]. We apply our extended HyPLSA approach on the task of item recommendation instead.

Another successful approach to recommendation within folksonomies is the FolkRank algorithm introduced by the authors of [10], we use this algorithm as a comparison to our approach.

The remainder of this paper is structured as follows. In Section 2 we present the algorithms utilized in our experiments followed by a description of our tests, dataset and experimental setup in Section 3. We present our results in Section 4 and draw final conclusions in Section 5.

2. ALGORITHMS

In the following section we describe our HyPLSA approach which is an extended version of the one presented by the authors of [15].

2.1 Model fusion using PLSA – HyPLSA

Hotho et al. [9] describe a folksonomy as tripartite graph in which the vertex set is partitioned into three disjoint sets of users $U = \{u_1, ..., u_l\}$, tags $T = \{t_1, ..., t_n\}$ and items $I = \{i_1, ..., i_m\}$. In [15] Wetzker et al. simplify this model into two bi-partite models; the collaborative filtering model IU built from the item user co-occurrence counts $f(i, u)$, and the annotation-based model IT analogously derived from the co-occurrence total between items and tags $f(i, t)$. In the case of social bookmarking IU becomes a binary matrix ($f(i, u) \in \{0, 1\}$) since each user can bookmark a given web resource one time only. Given this model, we want to recommend the most interesting new items from I to user u_l given his or her item history.

The PLSA aspect model associates the co-occurrence of observations with a hidden topic variable $Z = \{z_1, \ldots, z_k\}$. In the context of collaborative filtering, an observation corresponds to the bookmarking of an item by a user and all observations are given by the co-occurrence matrix IU. Users and items are assumed independent given the topic variable Z. When applying the aspect model, the probability of an item that has been bookmarked by a given user can be computed by summing over all latent variables Z:

$$P(i_m|u_l) = \sum_k P(i_m|z_k)P(z_k|u_l) \tag{1}$$

For the annotation-based scenario we assume the set of hidden topics to be the same as in the item tag co-occurrence observations given by IT. In compliance with (1), the conditional probability between tags and items can be written as:

$$P(i_m|t_n) = \sum_k P(i_m|z_k)P(z_k|t_n). \tag{2}$$

Following the procedure in [4], we can now merge both models based on the common factor $P(i_m|z_k)$ by maximizing the log-likelihood function:

$$\begin{aligned} L \quad = \quad & \sum_m \Bigg[\alpha \sum_l f(i_m, u_l) \log P(i_m|u_l) \\ & + (1 - \alpha) \sum_n f(i_m, t_n) \log P(i_m|t_n) \Bigg], \end{aligned} \tag{3}$$

where α is a predefined weight for the leverage of each two-mode model. Using the expectation-maximization (EM) algorithm [4] we subsequently perform maximum likelihood parameter estimation for the aspect model. During the expectation (E) step we begin with calculating the posterior probabilities:

$$P(z_k|u_l, i_m) = \frac{P(i_m|z_k)P(z_k|u_l)}{P(i_m|u_l)}$$

$$P(z_k|t_n, i_m) = \frac{P(i_m|z_k)P(z_k|t_n)}{P(i_m|t_n)},$$

and then re-estimate parameters in the maximization (M) step according to:

$$P(z_k|u_l) \quad \propto \quad \sum_m f(u_l, i_m)P(z_k|u_l, i_m) \tag{4}$$

$$P(z_k|t_n) \quad \propto \quad \sum_m f(t_n, i_m)P(z_k|t_n, i_m) \tag{5}$$

$$\begin{aligned} P(i_m|z_k) \quad \propto \quad & \alpha \sum_l f(u_l, i_m)P(z_k|u_l, i_m) \\ & + (1 - \alpha) \sum_n f(t_n, i_m)P(z_k|t_n, i_m) \end{aligned} \tag{6}$$

Based on the iterative computation of the above E and M steps, the EM algorithm monotonically increases the likelihood of the combined model on the observed data. Using the α parameter, our new model can easily be reduced to a collaborative filtering, or annotation-based model, simply by setting α to 1.0 or 0.0 respectively.

Because of the random initialization of the EM algorithm utilized by PLSA, we employ an averaging approach to reduce any effects possibly caused by local maximum optimizations. Thus, following Equation (1), we repeat Equations (4) to (6) n times for every recommendation and average the probabilities obtained from Equation (1). Our contribution to the model is presented in Equation (7), where the final, averaged, probability is given.

$$\bar{P}(i_m|u_l) = \frac{\sum_n P_n(i_m|u_l)}{n} \quad (7)$$

We can now recommend items to a user u_l weighted by the probability $P(i_m|u_l)$ from Equation (7). For items already bookmarked by the user in the training data we set this weight to 0, thus they are appended to the end of the recommended item list.

3. EXPERIMENTS

We conduct our experiments on the CiteULike dataset, these experiments are described next.

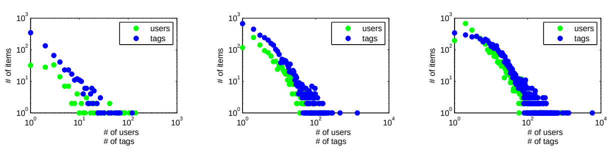
3.1 Dataset

The CiteULike bookmarking service provides daily snapshots of their data for research purposes. At the time of writing the overall dataset consisted of roughly 53 months of data. As noted earlier we are only interested in the initial phase of the service and therefore limit our analysis to the first 24 months, focusing on the first 12.

CiteULike was chosen as the experimental dataset because it is a well known real-world folksonomy and has been experimented on by numerous others [11, 12, 17].

We started by removing all users who had bookmarked less than 20 items as well as items and tags that occurred only once. We then created monthly snapshots where each of the snapshots accumulated all previous tagging events. By doing this we were able to simulate a growing folksonomy over time.

Figures 1(a), 1(c), 1(b) and 2 show some characteristics of our dataset.



(a) at 3 months. (b) at 12 months. (c) at 24 months.

Figure 1: Users and tags plotted against the number of items they are connected to.

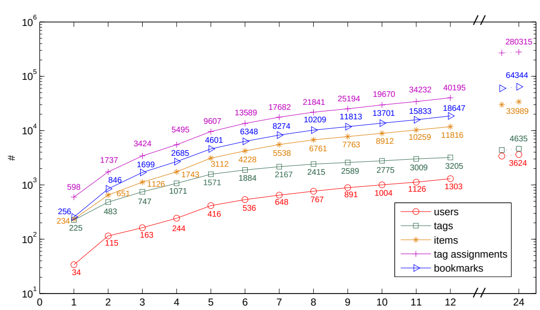


Figure 2: Accumulated number of items, tags, users, tag assignments and bookmarks of our data per month

3.2 Experimental setup

To create test and training sets for our algorithms, we split each monthly snapshot in two. For all users who had bookmarked at least 5 items in the current snapshot, we selected 80% of their items as the training set. The remaining items were consequently used for testing. We then trained all recommender types on the training sets and evaluated their performance on the test sets. The relatively small size of the dataset allowed us to optimize parameters through a brute force approach. Evaluation measures were averaged over all users in 10 independent test runs.

4. RESULTS

We evaluate the performance of each recommender with the well known and widely used precision at 10 measure (Prec@10). Other evaluation measures, such as *mean average precision* (MAP), *area under curve* (AUC)and *F1 score* (F1), showed similar results.

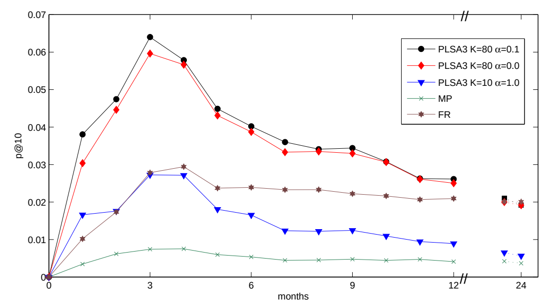


Figure 3: Prec@10 values for the item recommendation task on the *CiteULike* dataset plotted per month. The number of latent topics (k) is set to 80 for the purely annotation-based PLSA recommender ($\alpha = 0.0$) and to 10 for the purely collaborative version ($\alpha = 1.0$). The MP and FR lines represents the performances of a *most-popular* baseline classifier and the *FolkRank* recommender presented by Hotho et al. [10]. The results of the combined HyPLSA approach are seen in the topmost line, the parameters α and k were set to 0.1 and 80 respectively.

Figure 3 shows the Prec@10 values for the HyPLSA recommender obtained with an optimal parameter setting ($alpha = 0.1$, $k = 80$), the purely collaborative filtering-based PLSA recommender ($\alpha = 1$, $k = 10$), the purely annotation-based PLSA recommender ($alpha = 0$, $k = 80$), the baseline most-popular recommender and the FolkRank recommender. The figure shows a significant improvement when using the HyPLSA recommender, especially in the early months of the CiteULike. We also see that as the dataset grows, and the number of possible items to recommend increases, the precision values decrease. Nevertheless, the HyPLSA approach delivers significantly better results than the other evaluated approaches.

Figure 4 shows a figure similar to Figure 3, this time with Prec@10 values plotted against the number of items in the dataset, confirming the observation made earlier.

In Figure 5 we present the relative improvements in precision of the HyPLSA approach plotted against the other ones explored in this paper. The reason for these results

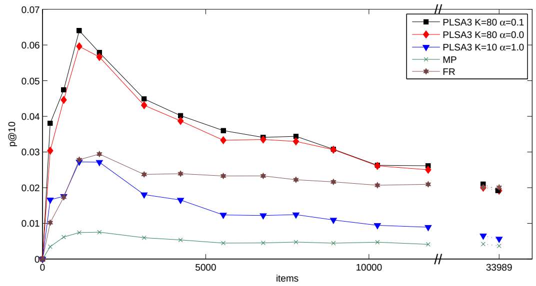


Figure 4: Prec@10 values for the same scenario and value as in Figure 3 plotted against the number of items in the dataset.

can be traced back to Figure 1(a) where we clearly see the differences in density in the user domain when comparing Figure 1(a) to Figures 1(b) and 1(c). At best our approach improves precision values roughly tenfold compared to the baseline MP recommender and twice as well as the FR recommender. As expected the improvement is highest in the first couple of months and slowly decreases (for MP and FR) or stabilizes (for CF) as the dataset grows. Comparing to the tag-based approach, the improvement is not as distinct as in the case with other recommenders. Relative improvements are in the order of $2-25\%$ during first seven months, decreasing in the long run.

These results confirm the notion that, for a small dataset, the number of user-item co-occurrences is too low to allow a collaborative filtering recommender to make good predictions. Tags and tag-item co-occurrences, on the other hand, provide higher item-item similarities as tags are more abundant and contain contextual information about the items. Therefore tags provide for recommendations on a finer level of granularity.

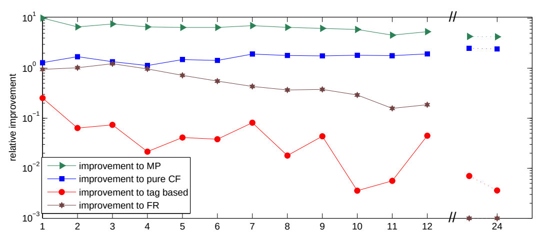


Figure 5: The relative improvement of the proposed HyPLSA recommender compared to the other explored ones. The higher the line, the bigger the improvement.

5. CONCLUSIONS

We have shown that tags improve the quality of item recommendation during the cold start period of a folksonomy. This is due to the fact that tags offer denser, more detailed item information than usage patterns do.

Furthermore we have presented a hybrid probabilistic approach that combines user and tag similarities in order to boost recommendation quality. The recommendation quality improvement created by using this approach peaks during the cold start period, although the approach continues to provide for better recommendations as the size of dataset increases. We believe that the reason for the relative improvement being higher in the beginning can be traced back to the pattern seen in Figures 1(a) to 1(c) where we initially see a very much higher density in tag usage compared to usage patterns. As the tag usage pattern density becomes more and more similar to the tag density the recommendation results of all PLSA (tag, CF and HyPLSA) approaches become similar.

6. REFERENCES

[1] J. Arenas-García, A. Meng, K. B. Petersen, T. L. Schiøler, L. K. Hansen, and J. Larsen, ‘Unveiling music structure via PLSA similarity fusion’, in *Proc. of MLSP’07*, (2007).

[2] K. Bischoff, C. S. Firan, W. Nejdl, and R. Paiu, ‘Can all tags be used for search?’, in *CIKM ’08*, (2008).

[3] T. Brants, F. Chen, and I. Tsochantaridis, ‘Topic-based document segmentation with probabilistic latent semantic analysis’, in *Proc. of CIKM ’02*, (2002).

[4] D. A. Cohn and T. Hofmann, ‘The missing link - a probabilistic model of document content and hypertext connectivity’, in *NIPS*, pp. 430–436, (2000).

[5] A. P. Dempster, N. M. Laird, and D. B. Rubin, ‘Maximum likelihood from incomplete data via the em algorithm’, *Journal of the RSS. Series B*, **39**, (1977).

[6] L. Hennig, ‘Topic-based multi-document summarization with probabilistic latent semantic analysis’, in *RANLP ’09*, (Sep. 2009).

[7] P. Heymann, D. Ramage, and H. Garcia-Molina, ‘Social tag prediction’, in *Proc. of SIGIR ’08*, (2008).

[8] T. Hofmann, ‘Probabilistic latent semantic analysis’, in *Proc. of UAI ’99*, (1999).

[9] A. Hotho, R. Jäschke, C. Schmitz, and G. Stumme, ‘Information retrieval in folksonomies: Search and ranking’, in *ESWC*, pp. 411–426, (2006).

[10] A. Hotho, R. Jäschke, C. Schmitz, and G. Stumme, ‘FolkRank: a ranking algorithm for folksonomies’, in *Proc. FGIR 2006*, (2006).

[11] S. Noël and R. Beale, ‘Sharing vocabularies: tag usage in citeulike’, in *Proc. of BCS-HCI ’08*, (2008).

[12] S. Oldenburg, M. Garbe, and C. Cap, ‘Similarity cross-analysis of tag / co-tag spaces in social classification systems’, in *Proc. of SSM ’08*, (2008).

[13] P. Symeonidis, A. Nanopoulos, and Y. Manolopoulos, ‘Tag recommendations based on tensor dimensionality reduction’, in *Proc. of RecSys ’08*, pp. 43–50, (2008).

[14] T. Vander Wal. Folksonomy coinage and definition. `http://www.vanderwal.net/folksonomy.html` (retrieved on June 10, 2009), February 2007.

[15] R. Wetzker, W. Umbrath, and A. Said, ‘A hybrid approach to item recommendation in folksonomies’, in *Proc. of WSDM - ESAIR ’09*, pp. 25–29, (2009).

[16] G.-R. Xue, W. Dai, Q. Yang, and Y. Yu, ‘Topic-bridged plsa for cross-domain text classification’, in *Proc. of SIGIR ’08*, (2008).

[17] V. Zanardi and L. Capra, ‘Social ranking: uncovering relevant content using tag-based recommender systems’, in *Proc. of RecSys ’08*, pp. 51–58, (2008).