

Understanding the user: Personomy translation for tag recommendation

Robert Wetzker¹, Alan Said¹, and Carsten Zimmermann²

¹ Technische Universität Berlin, Germany

² University of San Diego, USA

Abstract. This paper describes our approach to the challenge of graph-based tag recommendation in social bookmarking services. Along the ECML PKDD 2009 Discovery Challenge, we design a tag recommender that accurately predicts the tagging behavior of users within the Bibsonomy bookmarking service. We find that the tagging vocabularies among folksonomy users differ radically due to multilingual aspects as well as heterogeneous tagging habits. Our model overcomes the prediction problem resulting from these heterogeneities by translating user vocabularies, so called *personomies*, to the global folksonomy vocabulary and vice versa. Furthermore we combine our user-centric translation approach with item-centric methods to achieve more accurate solutions. Since our method is purely graph-based, it can also readily be applied to other folksonomies.

1 Introduction

Over the last years, social bookmarking services, such as Delicious³, Bibsonomy⁴ and CiteULike⁵, have grown rapidly in terms of usage and perceived value. One distinguishing feature provided by these services is the concept of tagging - the labeling of content with freely chosen keywords (tags). Tagging enables users to describe and categorize resources in order to organize their bookmark collections and ease later retrieval. Social bookmarking services are therefore the classic example of collaborative tagging communities, so called *folksonomies* [1][2]. The consumer-centric (collaborative) tagging aspect differentiates social bookmarking from other content sharing community services, such as Flickr⁶ or YouTube⁷, where tags are generally assigned by the content creator [3].

Most folksonomy solutions assist users during the bookmarking process by recommending tags. Thus, a user can select recommended tags from various sets in addition to entering tags manually. Despite their positive effect on usability, these recommenders are effective tools to limit tag divergence within

³ <http://delicious.com>

⁴ <http://bibsonomy.org>

⁵ <http://citeulike.org>

⁶ <http://flickr.com>

⁷ <http://www.youtube.com>

folksonomies as they are generally considered to lower the ratio of misspellings and to increase the likelihood of tag reassignments. The design of a folksonomy tag recommender was one of the tasks of the ECML PKDD 2009 Discovery Challenge⁸. In the following sections, we describe our solution to this task as submitted.

Our approach is based on the observation that the tag vocabularies of users, their personomies, differ within a folksonomy. This heterogeneity is mainly caused by differences in the tags users constantly assign to categorize content and the multilingualism of the user base, as apparent for Bibsonomy. To overcome the problems caused by this heterogeneity, we propose a tagging model that translates the personomy of each user to the folksonomy vocabulary and vice versa. We find that our model is highly accurate as it characterizes an item by its tag spectrum before translating this spectrum to a user’s personomy. We then combine the translational model with item-centric tag models to improve performance.

This paper is structured as follows: The introductory section presents a graph model for the underlying data structure and explains the actual goals of the challenge. This is followed by an analysis of different properties of the Bibsonomy dataset with respect to their impact on tag recommendation. Section 3 introduces and discusses the tag vocabularies found within folksonomies, before we present our recommendation algorithm and evaluation results.

1.1 Modeling folksonomies

According to [2], a folksonomy can be described as a tuple $F := (I, T, U, Y)$, where $I = \{i_1, \dots, i_k\}$, $T = \{t_1, \dots, t_l\}$ and $U = \{u_1, \dots, u_m\}$ are finite sets of items, tags and users, and Y is a ternary relation whose elements are called *tag assignments*. A *tag assignment* (TAS) is defined by the authors as relation $Y \subseteq U \times T \times I$, so that the tripartite folksonomy hypergraph is given by $G = (V, E)$, where $V = I \cup T \cup U$ and $E = \{(i, t, u) | (i, t, u) \in Y\}$. The set of all bookmarks is then given as $BM = \{(i, u) | \exists t : (i, t, u) \in Y\}$. This graph structure is characteristic for all folksonomies.

1.2 The challenge: Graph-based tag recommendations (Task 2)

The ECML PKDD 2009 Discovery Challenge consists of different tasks related to the problem of tag recommendation. Testbeds for all tasks are different snapshots of the Bibsonomy bookmarking service. Our solution contributes to the task of ”Graph-Based Recommendations (Task 2)“ as it does not consider the content of the given resources. The recommendation task in this setting resembles the problem of link prediction within G given a user and an item node. The recommender thus needs to estimate $P(t|i, u)$, the probability of observing a given tag when a combination of user and item has been observed.

⁸ <http://www.kde.cs.uni-kassel.de/ws/dc09>

1.3 Related work

One of the first works on tag recommendation in folksonomies is [4], where the authors compare the performance of co-occurrence based tag recommenders and more complex recommenders based on the FolkRank algorithm [2]. Furthermore they report only minor improvements of the FolkRank method over the co-occurrence approaches. Parts of their analysis is performed on a snapshot of the Bibsonomy dataset.

Further related work was presented by the participants of last years challenge on tag recommendation. The authors of [5] enrich tag vocabularies by terms extracted from all bookmarks’ meta data, such as user given descriptions or the abstract, author, year etc. information in case of publications. More similar to our work is the approach in [6]. The team combines the keywords found within a resource’s title and the actual tags previously assigned to a resource with the tags from a user’s personomy. Each vocabulary is mapped to the global tag vocabulary using co-occurrence tables. This mapping is similar to our translation process. However, the fusion of different sources differs from our method and no user-centric optimization of parameters was reported.

2 The dataset

The Bibsonomy bookmarking service allows its users to bookmark URLs and publications in parallel. This hybrid approach makes Bibsonomy different from other bookmarking communities such as Delicious or CiteULike. For each web bookmark, participants are given the URL, the title and an optional description of the resource as provided by the user during the bookmarking process. Bookmarked publications generally come with information about the title, the authors, the abstract or other common bibliographic attributes. The completeness of this information is not guaranteed, and many attribute fields are left empty. Table 1 gives an overview of the node statistics found within the dataset for p-core levels one and two⁹.

Table 1. Different node set sizes of the Bibsonomy dataset for p-core levels 1 and 2.

p-core	$ E $	$ BM $	$ BM_{BIB} $	$ BM_{URL} $	$ I $	$ T $	$ U $
1	1,401,104	421,928	263,004	158,924	378,378	93,756	3,617
2	253,615	64,120	41,268	22,852	22,389	13,276	1,185

For the construction of our recommender we ignore the meta-data attached to the bookmarked content and only consider the graph G . Furthermore, we do not distinguish between URLs and publications, but merge both node sets to the item set I . Both decisions result in a loss of information which potentially

⁹ The p-core of a folksonomy graph has the characteristic that all contained nodes appear in at least p bookmarks. See [4] for details.

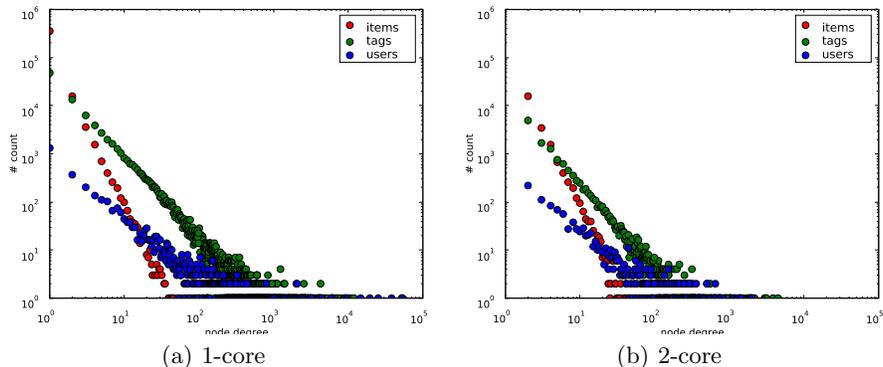


Fig. 1. Node degree distributions within the full Bibsonomy dataset (a) and the dataset at p-core level 2. Node degree distributions found within folksonomies generally exhibit power law characteristics with few highly connected nodes and many nodes with low occurrence. This characteristic is obtained when cutting the graph to its level 2-core.

reduces the overall accuracy of our recommender. However, we believe this loss to be compensated by the general applicability of our approach and the improved validity of the presented results.

Figure 1 shows the node degree distributions of the full dataset and its 2-core. As previously reported for other folksonomies (e.g. [2][7]), we find the node degree distributions of Bibsonomy to exhibit power law characteristics, with very few and very frequent nodes on one end and many infrequent nodes on the other. This characteristic is basically obtained when reducing G to its 2-core. However, we find that this reduction drastically reduces the impact of some previously very influential users. These users tend to have a rather "organizational" background and bookmark large collections of similar resources which are of no interest to other users. Most of these resources therefore fall victim to the p-core reduction, which also explains the drastic cut on items in Table 1.

3 Tag vocabularies

As users bookmark items only once¹⁰, we cannot estimate $P(t|i, u)$ directly from previous observations. Instead, we base our estimates on other distributions. The most basic sources are the overall tag distribution and the tags previously assigned by the user, or to the item.

Global tag distributions. This is the distribution of all observed tag assignments within the training data. By definition, each tag of the test data has

¹⁰ Note that a small percentage of user item combinations found in the given dataset occur in more than one bookmark.

to occur within this distribution. A recommender that only considers the global tag distribution would assume $P(t|u, i) \approx P(t)$. We will refer to such a recommender as *MostPopular* recommender.

Item tag distributions. This is the distribution of previously assigned tags for a given item. It was shown that the tag distributions of items converge to a characteristic tag spectrum over time. Furthermore, as reported by [8] and [9], the resulting tag distributions often follow a power law with few tags being assigned very frequently and most tags occurring in the long tail. If we neglect the personalization aspect of tagging, we can recommend tags by assuming $P(t|i, u) \approx P(t|i)$. However, our observation of $P(t|i)$ within the training data may be limited, i.e. information about most tags will be missing.

User tag distributions (Personomies). Each user develops his own vocabulary of tags over time called his *personomy*. Users will generally be interested in reassigning previously used tags as this will simplify content search later on. The interest in tag convergence often results in the frequent assignment of a limited number of category tags, and it was shown that user vocabularies develop power law characteristics over time [10]. A personomy based recommender would assume $P(t|i, u) \approx P(t|u)$. Once again, the distribution $P(t|u)$ estimated over the training data is likely to miss a variety of tags especially for users with few bookmarks.

The authors of [4] report that a tag model which combines user and item tag distributions into a unified distribution achieves sufficient recommendation accuracy. We will consider a hybrid recommender with $P(t|i, u) \approx \alpha P(t|i) + (1 - \alpha)P(t|u)$ as an additional baseline approach during our evaluations (*MostPopular2d*).

4 Our approach

The design of our tag recommender is based on two intuitive assumptions:

1. **Tags are personalized.** Different users will assign different tags to the same item. This effect cannot only be explained by statistical variance. Instead, we find users developing their own category tags over time. One of the implications for the tag recommendation task is the problem of recommending the right version of a tag, especially in cases where synonymous tags exist. This includes different spellings, such as "web20" versus "web2.0". Even though semantically equal, these will be different for a user who assigns tags for content categorization. Furthermore, especially in multilingual folksonomies, we find that users often assign keywords from their mother tongue. This is of particular importance for Bibsonomy, where many users seem to come from Germany, with the effect that the tag distribution is a mixture of German and English words. Whereas some users tagged a site as "searchengine" related we also find the German translation ("suchmaschine") among the frequent tags.

2. **Tags describe items.** The authors of [11] report that the vast majority of assigned tags on Delicious identify the topic, the type or the owner of a URL. We can therefore assume that users will assign personomy tags depending on the item. This assumption is a simplification as it excludes tags, such as "toread" or "self", which actually refer to the user item relationship instead of the item itself. However, we believe that these tags cannot be easily predicted based on the given training data but require deeper knowledge about the user. Luckily, as reported by [12] for Delicious, *usage context* and *self reference* tags are relatively scarce compared to descriptive tags.

These assumptions directly influenced the basic design decisions for our recommender which suggests tags from a user's personomy with respect to the community opinion about the underlying item.

4.1 Translating personomies

We assume that each user has a distinctive vocabulary of tags. These tags can be translated to the community tag vocabulary by looking at co-occurrences within the shared item space. We are thus interested in the probability $P(t^u|t, u)$ that a user will assign a tag t^u from his personomy as next tag given that the item was tagged as t by another user. Based on previous knowledge about the users tagging behavior we can estimate $P(t^u|t, u)$ as

$$P(t^u|t, u) = \sum_{i \in I_u} P(t^u|i, u)P(i|t), \quad (1)$$

where I_u is the set of items previously bookmarked by the user. Based on $P(t^u|t, u)$ we can now translate the global folksonomy language to the personomy of a user.

For the recommendation task we are interested in the probability $P(t^u|i, u)$ for previously unseen user item combinations. For a new item with an observed tag vocabulary of $P(t|i)$, the probability that a user will assign one of his tags next is given as

$$P(t^u|i, u) = \sum_{t \in T_i} P(t^u|t, u)P(t|i). \quad (2)$$

Note that $\sum_{t^u \in T_u} P(t^u|i, u) = 1$ is only true if all tags in T_i have a mapping in T_u which is rather unlikely. Instead we expect $\sum_{t^u \in T_u} P(t^u|i, u)$ to decrease the more the given item deviates from the items previously bookmarked by the user.

4.2 The tag recommender

The tag recommender we propose, selects tags coming from three sources: the personomy of a user where tags are weighted by $P(t^u|i, u)$, the item vocabulary ($P(t|i)$) and the global vocabulary ($P(t)$). Including the item vocabulary is

important, as many tags may be item specific and are thus unlikely contained within the personomy. We also include the global tag distribution $P(t)$ for cases where little is known about user and item alike. We assume a weighted linear combination of sources and estimate $P(t|i, u)$ as

$$P(t|i, u) \approx \alpha_u P(t^u|i, u) + \alpha_i P(t|i) + (1 - \alpha_u - \alpha_i) P(t), \quad (3)$$

with $0 \leq \alpha_u, \alpha_i, \alpha_u + \alpha_i \leq 1$. We then recommend the N tags with highest probability $P(t|i, u)$, where N is a parameter that needs to be optimized together with α_u and α_i . Optimization can take place on a global or a user-centric scale.

Global optimization. For the globally optimized model, we assume equal parameter settings for all users. As the Bibsonomy dataset is rather small, we can use a brute-force approach to find the combination of N , α_u and α_i that maximizes the F-measure. We do so by performing a 10-fold cross-validation on the training data. We call this user-centric tag recommender with global optimization UC_G .

User-centric optimization. As users are heterogeneous, it is not intuitive to assume shared parameter preferences. Instead, it seems straightforward to optimize parameters for each user separately. Once again, we do so by performing a cross-validation on the training data. We then use a brute-force approach to find the combination of N , α_u and α_i that maximizes the F-measure of each user averaged over all folds. We will refer to the user-centric tag recommender with local optimization as UC_L .

5 Evaluation

We trained the models of all recommender types on the 2-core version of the dataset. The parameters of the *MostPopular2d*, UC_G and the UC_L models were fine-tuned in a 10-fold cross-validation as described above. For the *MostPopular2d* recommender we found an α value of 0.5 to perform best. For the user-centric tag recommender with global optimization the maximal F_1 measure was achieved when setting $\alpha_u = 0.6$ and $\alpha_i = 0.4$. The weight of the global tag distribution thus resulted 0 which means that including the global vocabulary did not yield performance gains. For the *MostPopular2d* as well as the UC_G recommender the best number of tags to recommend was 5. Evaluating the accuracy of all recommender types during the cross validation, we found the user-centric tag recommender with local optimization (UC_L) to constantly outperform all other versions. We therefore submitted the predicted tags of the UC_L approach as our solution to the challenge.

The released test dataset consists of 778 bookmarks from 136 users linking to 667 items. Table 2 presents the achieved F_1 measures on the first five ranks for the various recommender types. The UC_L recommender we submitted achieved a

Table 2. Performance of various recommender types on the test data. Underlined values represent the configuration that performed best during training. The submitted recommender (UC_L) achieved an F_1 measure of 0.314. The best F_1 measure could have been achieved with a UC_G recommender always suggesting 3 tags (bold).

Recommender	$F_1@1$	$F_1@2$	$F_1@3$	$F_1@4$	$F_1@5$
<i>MostPopular</i>	0.021	0.038	0.051	0.051	<u>0.059</u>
<i>MostPopular2d</i> $_{\alpha=0.5}$	0.229	0.286	0.306	0.313	<u>0.310</u>
$UC_G, \alpha_u=0.6, \alpha_i=0.4$	0.246	0.326	0.335	0.334	<u>0.330</u>
UC_L	0.230	0.294	0.306	0.311	<u>0.314</u>

performance of 0.314. This result is somewhat disappointing as it is only slightly above the result of the simpler *MostPopular2d* recommender. However, we find that the approach of vocabulary translation is generally superior as the results of the UC_G recommender are significantly better. We observe similar performance patterns when looking at the precision/recall curves plotted in Figure 2.

Investigating the reasons for the weak performance of the UC_L recommender, we find that the user distribution of the test set deviates from the trained one as shown in Figure 3. This deviation is likely to have a negative impact on the prediction quality as parameters have been tuned in expectation of a user distribution similar to the one of the training set. However, this problem is not

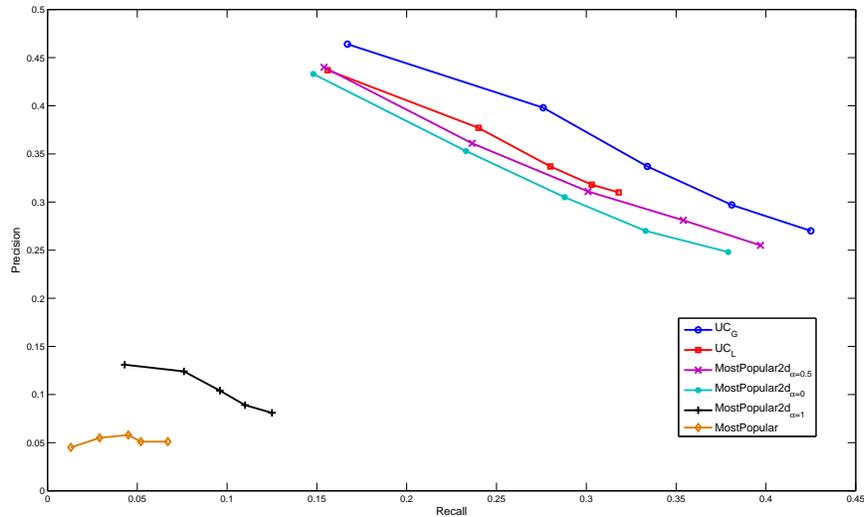


Fig. 2. Precision/Recall curves for various recommenders on the provided test data. The curve of the UC_L recommender appears "shorter" as this recommender suggests a variable number of tags.

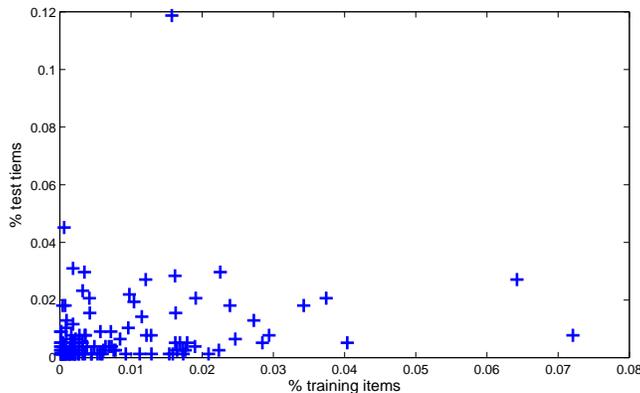


Fig. 3. Relative number of bookmarks per user in test and training data. The correlation between the user participation in the test set and the trained distribution is rather low ($\rho = 0.24$).

unique to the UC_L recommender but is expected to have a negative impact on the performance of all recommender types. Instead, we believe that the weak performance of the UC_L recommender is caused by an inadequate parameter tuning for users less present in the training data but frequent in the test set. Tuning α_u , α_i and N for these users often results in bad estimates due to missing data. Whereas the implications of these shortcomings are rather minor when users are distributed as in the training data, they seem to become major for the test distribution.

The fact that the test set is dominated by users with rather small training vocabularies is also reflected by the performance of the *MostPopular2d* recommenders with α set to 0 and 1 as shown in Figure 2. Here, we find that a recommender which only suggest tags from a user’s personomy ($\alpha = 1$) performs very bad, whereas an item based recommender ($\alpha = 0$) achieves nearly as good results as the mixture model $\alpha = 0.5$. This implies, that most tags of the test data are not present within a user’s personomy at training time or, less likely, that the tagging behavior of users drastically changed in the test phase.

The inadequate modeling of infrequent users (and items) is an expected shortcoming of a purely graph-based recommendation approach. This is especially true for our personomy translation approach which requires the tags of the given item to have a mapping within the conditional distribution $P(t^u|t, u)$ (see equation 2). Incorporating the provided item meta-data may be a promising alternative to improve accuracy in scenarios where little is known about users and items from a graph perspective.

6 Conclusions

In this paper, we presented a novel approach to the challenge of graph-based tag recommendation in folksonomies. Building on the assumption that all users of a folksonomy have their own tag vocabulary, our approach translates the personomies of users to the global folksonomy vocabulary. Evaluation results show that this translation helps to significantly improve tag prediction performance. Furthermore, we fine-tuned our model by estimating parameters on a per user basis. Even though this user-centric approach performed rather disappointing during the challenge, we believe that user-level optimization will be essential for the success of future (tag) recommenders.

References

1. Wal, T.V.: Folksonomy coinage and definition, February 2, 2007, <http://www.vanderwal.net/folksonomy.html>.
2. Hotho, A., Jäschke, R., Schmitz, C., Stumme, G.: Information retrieval in folksonomies: Search and ranking. In: Proc. of the 3rd European Semantic Web Conference (ESWC), Springer (2006)
3. Marlow, C., Naaman, M., Boyd, D., Davis, M.: Ht06, tagging paper, taxonomy, flickr, academic article, to read. In: HYPERTEXT '06, New York, NY, USA, ACM (2006)
4. Jäschke, R., Marinho, L., Hotho, A., Schmidt-Thieme, L., Stumme, G.: Tag recommendations in folksonomies. In: Lernen - Wissensentdeckung - Adaptivität (LWA) Workshop Proc. (2007)
5. Tatu, M., Srikanth, M., D'Silva, T.: Rsd08: Tag recommendations using bookmark content. In: Proc. of the ECML PKDD Discovery Challenge (RSDC08). (2008)
6. Lipczak, M.: Tag recommendation for folksonomies oriented towards individual users. In: Proc. of the ECML PKDD Discovery Challenge (RSDC08). (2008)
7. Heymann, Paul Garcia-Molina, H.: Collaborative creation of communal hierarchical taxonomies in social tagging systems. Technical report, Computer Science Department, Stanford University (2006)
8. Cattuto, C., Loreto, V., Pietronero, L.: Collaborative tagging and semiotic dynamics. PNAS **104** (2007)
9. Halpin, H., Robu, V., Shepherd, H.: The complex dynamics of collaborative tagging. In: WWW '07: Proc. of the 16th int. conf. on World Wide Web, New York, NY, USA, ACM (2007)
10. Zhang, D., Mao, R., Li, W.: The recurrence dynamics of social tagging. In: WWW '09: Proc. of the 18th int. conf. on World wide web, New York, NY, USA, ACM (2009)
11. Golder, S.A., Huberman, B.A.: Usage patterns of collaborative tagging systems. J. of Information Science **32**(2) (2006)
12. Bischoff, K., Firan, C.S., Nejdl, W., Paiu, R.: Can all tags be used for search? In: CIKM '08: Proc. of the 17th ACM conf. on Information and knowledge management, New York, NY, USA, ACM (2008)