

Combining Tag Recommendations Based on User History

Ilari T. Nieminen

Helsinki University of Technology
ilari.nieminen@tkk.fi

Abstract. This paper describes our attempt at Task 2 of ECML PKDD Discovery Challenge 2009. The task was to predict which tags a given user would use on a given resource using methods that only utilize the graph structure of the training dataset, which was a snapshot of BibSonomy. The approach combines simple recommendation methods by weighting recommendations based on the tagging history of the user.

1 Introduction

Collaborative tagging systems or folksonomies have steadily gained popularity in the recent years. Users are free to choose the tags they want to use, and while this may be a main reason behind the popularity of these systems, it is also one of the biggest problems these systems face. As users come up with new tags they forget the tags they used to use, making it difficult to find the previously tagged content. Tag recommendation can help both in search and in keeping the users' tagging practices consistent. Tag recommendation can be defined as the problem of finding suitable tags or labels to a given resource for a given user.

Tag recommendation can be an important element in a folksonomy as it can help users employ the tags consistently as well as help users to use same tags for similar resources. This can improve searching within the users' own resources as well as the folksonomy.

We present a method for tag recommendation that combines several baseline methods and collaborative filtering. Combining the results makes use of the past performance of the recommenders.

2 Tag Recommendation

2.1 Collaborative Filtering for Folksonomies

Collaborative filtering (CF), a popular method used in recommender systems can be adapted for tag recommendation. The description here is based on [1].

Folksonomy can be understood as a tuple $\mathcal{F} = (U, R, T, Y)$, where U is the set of users, T is the set of tags and R is the set of resources (bookmarks and BibTeX entries in the case of BibSonomy [2]) and $Y \subseteq U \times R \times T$ is the tag assignment relation.

Projections $\pi_{UR}Y \in \{0, 1\}^{|U| \times |R|}$, $(\pi_{UR}Y)_{u,r} := 1$ iff $\exists t \in T$ s.t. $(u, r, t) \in Y$ and $\pi_{UT}Y \in \{0, 1\}^{|U| \times |T|}$, $(\pi_{UT}Y)_{u,t} := 1$ iff $\exists r \in R$ s.t. $(u, r, t) \in Y$ let us define the “tag neighbourhood” and “resource neighbourhood” of the users. The set of k nearest neighbours for a user u using the neighbourhood matrix X is

$$N_u^k := \operatorname{argmax}_{v \in U}^k \operatorname{sim}(x_u, x_v) \quad (1)$$

where sim is the cosine similarity

$$\operatorname{sim}(x, y) := \frac{x \cdot y}{\|x\| \|y\|} \quad (2)$$

The set of recommendations for a given user-resource pair (u, r) is

$$T'(u, r) := \operatorname{argmax}_{t \in T}^n \sum_{v \in N_u^k} \operatorname{sim}(x_u, x_v) \delta(v, r, t) \quad (3)$$

where $\delta(v, r, t) := \text{iff}(v, r, t) \in Y$.

2.2 Baseline Methods

The following are a collection of simple recommendation methods, which do not produce very good recommendations and have few redeeming qualities except that they are computationally inexpensive.

Popular tags for a resource. If the users of the folksonomy are homogenous, this method can be expected to perform almost as well as CF methods. However, if the users have very different tagging habits or if people use different tags from different languages, performance for the minorities can be expected to suffer.

Popular tags for a user. Some users use relatively few but obscure tags, which means that the popular tags for resource -recommender will not work. Collaborative recommendations also will not work well, as the user will probably have very few applicable “tag neighbours” and the “resource neighbours” will most likely not use the same tags. For example, user 483 used the tag “allgemein” a total of 2237 times in the 9003 posts. In other words, given a post by this user at random, there is almost a 25% chance it is tagged “allgemein”.

Globally popular tags. Recommending the most used tags is perhaps the simplest possible method.

We used several variants of the aforementioned recommenders. These and the method used to combine the recommendations are described in chapter 4.1.

3 Data Description and Preprocessing

The provided training data contains three files: *bibtex*, *bookmark* and *tas*. The *bibtex* and *bookmark* files describe the content of the links and BibTeX entries, respectively. The *tas* file contains the tag assignments. Also provided was the post-core at level 2 [3], which contained a reduced set, which contained only

those users, resources and tags that appear at least in two posts. The test set for this task was known to have the users, resources and tags from this set.

We processed bookmarks and BibTeX entries identically. The only information extracted from the “bookmark” and “bibtex” tables were the hash values which identified the resources. We used the `url_hash` and `simhash1` columns and did not attempt to combine duplicate resources. The `url_hash` considers two resources different if there are any differences in the url, such as a trailing slash.

To retain a slightly better neighbourhoods for the collaborative filtering approach we used full training set to calculate the neighbourhoods, but removed the tags that could not appear in the results. The difference between this and the post-core at level 2 was that this left several partial posts to the training data.

No effort was made to separate functional tags (such as “myown” and “toread”) from descriptive tags, which are considerably more interesting in tag recommendation.

Some of the most used tags in BibSonomy are used by a small minority, such as “juergen” (3101 posts, 2 users). In total, in the subset of tags that are contained in the post-core 2 there are 273 tags that have been used at least 100 times by at most 5 people. A measure for the popularity of the tag, which takes into account the number of users of a tag can be defined as

$$\text{popularity}(t) = \log(N_t) * \log(N_t^*), \quad (4)$$

where N_t is the number of times the tag t has been used and N_t^* is number of users for the tag t .

This measure can be used to improve tag recommendation methods which would not otherwise give weights to different tags.

Table 1. Tags ordered by number of uses and “popularity”

Number of uses	Popularity
bookmarks	software
zzztosort	web
video	web20
software	video
programming	blog
web20	bookmarks
books	programming
media	internet
tools	tools
web	social

As can be seen from Table 1, sorting the tags by their “popularity” removes the unlikely tag “zzztosort” while preserving a sensible selection of popular tags.

4 Results

4.1 Combining Recommendations

The baseline methods can yield good results on certain users, but they are generally worse than the alternatives. However, combining the baseline results with results from collaborative filtering or other methods can be used to improve the general results. The problem of combining results is in evaluating the trustworthiness of the recommender results.

In tag recommendation, there are multiple “items” that are recommended, and besides the similarity between the user and the neighbours of the user there are few evident factors that could be used to weight the tags when combining different methods. In our method, we used the training data to predict the recent posts of the users (1-100 posts, but at most 20% of the user’s all posts)

In our approach, we took the arbitrary set of methods shown in Table 2 and assigned weights to different tags by calculating the weighted sum over all recommenders using the per-user per-post weighted sum

$$w_t := \sum_p [t \in T'] * 0.9^k f_p \quad (5)$$

where f_p is the F-measure of the method $p \in 1, \dots, 7$ on the validation set, and k is the position of the tag in the recommendation. This reduces the weight of the tag slightly so that the methods with smaller F-measure have a better possibility of getting a likely tag in the final results. The final recommendation are the five $t \in T$ with the highest w_t .

Table 2. Recommendation methods

Method
Collaborative filtering (UR neighbourhood)
Collaborative filtering (UT neighbourhood)
Most frequent tags by resource
Most frequent tags by resource (popularity > 3)
Most frequent user tags
Most frequent user tags (popularity > 3)
Most popular global tags

Prior to the competition, we performed a test with the training data. The posts were divided into three sets based on the post date. The first 80% was selected to work as a training set, the following 10% as the validation set and the last 10% were used for testing. The method weights were computed from the validation set. The resulting weights were tested on the test set, showing a modest 5% improvement in the F-measure over the best baseline method in the test.

4.2 Experiment on the Competition Set

The weights for the methods were assigned to the users in the competition set by generating recommendations for recent posts with all the methods listed in the previous section. The amount of posts was chosen was up to 100 posts, but at most 20% of the user’s all posts. After this, the F-measure for each method was used to generate a mixing profile for each user. Then the recommendations were made for the competition set and these were combined using the equation 5. The results are summarized in Table 3.

Table 3. Results on the competition set

Method	F-measure with 5 tags
CF-UR	0.2084
CF-UT	0.2317
resource tags	0.3067
resource tags (popularity > 3)	0.2940
user tags	0.0935
user tags (popularity > 3)	0.0050
popular tags	0.0354
combined	0.2952

One of the baselines (resource tags) outperforms the combined result slightly on the competition set. Some of the recommendations, such as “resource tags”, can contain very unlikely tags when the resource itself is tagged only a few times and contains unpopular tags; this was not taken into account when combining the recommendations. A possible solution for this problem is to not recommend unlikely (unpopular) tags if the user hasn’t used them in the past.

5 Conclusion

In these experiments, the weights of the recommenders are based on their past performance, but it is likely that there are several features that can be used to estimate these weights from statistical features of the user, such as the average “popularity” of the user’s tags and the number of distinct tags. We would like to study these numbers for correlations. Recommendations by other methods, such as FolkRank [1] could be added to improve the performance on the dense parts of the data.

The obtained results were less than stellar; in retrospect, more attention should have been paid to the combining of the results and especially the fact that the results of the recommendations were far from independent. Some method for filtering the results should have been applied, perhaps by modifying the weights for the individual tags by using the information whether the target user has used

a certain tag before and how popular the tag is. Simple methods should not be completely neglected, as they can provide useful results for users who do not conform to the tagging practices of the mainline users of the folksonomy.

6 Discussion

F-measure works as a performance measure for tag recommendation to a certain extent, but the utility of tag recommendation methods for usability and search within a folksonomy should be confirmed with user tests. Combining different tag recommendation results with different weights at different times may cause the recommendation to feel inconsistent.

Searching within a folksonomy is sometimes unnecessarily difficult. A part of the problem is that users tend to use only a few tags per post. One improvement for these tagging systems would be to ask for applicability of a set of tags that are similar to the ones user has already chosen. It might make sense to distinguish between the problems of tag prediction, that is, predicting the tags user will choose, and tag recommendation, the problem of finding descriptive tags for a resource.

7 Acknowledgements

The author acknowledges Heikki Kallasjoki's technical assistance and Mari-Sanna Paukkeri's comments. This work was supported by the Academy of Finland through the Adaptive Informatics Research Centre that is a part of the Finnish Centre of Excellence Programme.

References

1. Jäschke, R., Marinho, L.B., Hotho, A., Schmidt-Thieme, L., Stumme, G.: Tag recommendations in folksonomies. In Kok, J.N., Koronacki, J., de Mántaras, R.L., Matwin, S., Mladenic, D., Skowron, A., eds.: PKDD. Volume 4702 of Lecture Notes in Computer Science., Springer (2007) 506–514
2. Hotho, A., Jäschke, R., Schmitz, C., Stumme, G.: BibSonomy: A social bookmark and publication sharing system. In: Proc. Workshop on Conceptual Structure Tool Interoperability at the Int. Conf. on Conceptual Structures. (2006) 87–102
3. Batagelj, V., Zaversnik, M.: Generalized cores (2002) cs.DS/0202039, <http://arxiv.org/abs/cs/0202039>.