

Recommending Items in Social Tagging Systems Using Tag and Time Information

Emanuel Lacic*
Knowledge Technology
Institute
Graz University of Technology
Graz, Austria
elacic@know-center.at

Dominik Kowald*
Know-Center
Graz University of Technology
Graz, Austria
dkowald@know-center.at

Paul Seitlinger
Knowledge Technology
Institute
Graz University of Technology
Graz, Austria
paul.seitlinger@tugraz.at

Christoph Trattner
Know-Center
Graz University of Technology
Graz, Austria
ctrattner@know-center.at

Denis Parra
CS Department
Pontificia Universidad Católica
de Chile
Santiago, Chile
dparra@ing.puc.cl

ABSTRACT

In this work we present a novel item recommendation approach that aims at improving Collaborative Filtering (CF) in social tagging systems using the information about tags and time. Our algorithm follows a two-step approach, where in the first step a potentially interesting candidate item-set is found using user-based CF and in the second step this candidate item-set is ranked using item-based CF. Within this ranking step we integrate the information of tag usage and time using the Base-Level Learning (BLL) equation coming from human memory theory that is used to determine the reuse-probability of words and tags using a power-law forgetting function.

As the results of our extensive evaluation conducted on datasets gathered from three social tagging systems (BibSonomy, CiteULike and MovieLens) show, the usage of tag-based and time information via the BLL equation also helps to improve the ranking and recommendation process of items and thus, can be used to realize an effective item recommender that outperforms two alternative algorithms which also exploit time and tag-based information.

*Both authors contributed equally to this work.

Categories and Subject Descriptors

H.2.8 [Database Management]: Database Applications—*Data mining*; H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval—*Information filtering*

Keywords

recommender systems; social tagging; collaborative filtering; item ranking; base-level learning equation

1. INTRODUCTION

Over the past few years social tagging gained tremendously in popularity, helping people for instance to categorize or describe resources on the Web for better information retrieval (e.g., BibSonomy or CiteULike) [13, 23]. Although the process of tagging has been well explored in the past and in particular the task of predicting the right tags to the user in a personalized manner [12, 20], studies on predictive models to recommend items to users based on social tags are still rare. To contribute to this sparse field of research, in this paper we present preliminary results of a study that aims at addressing this issue. In particular, we provide first results of a novel attempt to improve item recommendations by taking into account peoples' social tags and the information of the time the tags have been applied by the users. As shown in related work, recommending items to users in a collaborative manner relying on social tagging information is not an easy task in general (e.g., [24] or [17]). However, other related work has also proved that the information of time is an important factor to make the models more accurate in the end (e.g., [26] or [10]).

Contrary to the previous work mentioned above, we suggest a less data-driven approach that is inspired by principles of human memory theory about remembering things over time. As shown in our previous work on tag recommender systems [15], the base-level learning (BLL) equation introduced by Anderson and Schooler [16] (see also Anderson et al. [1]), which integrates tag frequency and recency (i.e., the time since the last tag usage), can be used to implement an effective tag recommendation and ranking algorithm. In partic-

ular, the BLL equation models the time-depended drift of forgetting of words and tags using a power-law distribution in order to determine a probability value that a specific tag will be reused by a target user.

In this work, we apply this equation for ranking and recommending items to users. To this end, we present a novel recommender approach called *Collaborative Item Ranking Using Tag and Time Information (CIRTT)* that firstly identifies a potentially interesting candidate item set and secondly, ranks this candidate set in a personalized manner (similar to [10]). In this second step of personalization, we integrate the BLL equation to include this information about tags and time. To investigate the question as to whether tag and time information can improve the ranking and recommendation process, we conducted an extensive evaluation using folksonomy datasets gathered from three social tagging systems (BibSonomy, CiteULike and MovieLens). Within this study we compared our approach to two alternative tag and time based recommender algorithms [26, 10] amongst others. The results show that integrating tag and time information using the BLL equation helps to improve item recommendations and to outperform state-of-the-art baselines in terms of recommender accuracy.

The remainder of this paper is organized as follows. We begin with explaining our tag and time based approach CIRTT in Section 2. Then we describe the experimental setup of our evaluation in Section 3 and summarize the results of this study in Section 4. Finally, in Section 5, we close the paper with a short conclusion and an outlook into the future.

2. APPROACH

In this section we provide a detailed description of our item recommendation approach called *Collaborative Item Ranking Using Tag and Time Information (CIRTT)*. In general, our CIRTT algorithm uses a similar strategy as the approach proposed by Huang et al. [10] and thus, consists of two steps relying on a combination of user- and item-based CF: in the first step, a potentially interesting candidate item set for the target user u is determined and in the second step, this candidate item set gets ranked using item similarities and tag and time information.

Step one (i.e., determining candidate items) is conducted using a simple user-based CF approach. Hence, we first find the most similar users for the target user u (i.e., the neighborhood) based on the binary user-item matrix $B_{u,i}$ (see also [26]) and then, use the bookmarked items of these neighbours as our candidate item set. We use a neighbourhood of $k = 20$ users and the Cosine similarity measure [7] (see also Section 3.3).

In the second step (i.e., ranking candidate items) we use an item-based CF approach in order to determine the relevance of each candidate item for the target user based on the items she has bookmarked in the past. Hence, for each candidate item i in the candidate item set we calculate this combined similarity value $sim(u, i)$ by the item-based CF formula:

$$sim(u, i) = \sum_{j \in items(u)} sim(i, j) \quad (1)$$

Dataset	$ B $	$ U $	$ R $	$ T $	$ TAS $
BibSonomy	82,539	2,437	28,000	30,919	339,337
CiteULike	36,471	3,202	15,400	20,937	99,635
MovieLens	53,607	3,983	5,724	14,883	92,387

Table 1: Properties of the datasets, where $|B|$ is the number of bookmarks, $|U|$ the number of users, $|R|$ the number of resources, $|T|$ the number of tags and $|TAS|$ the number of tag assignments.

, where $items(u)$ is the set of items the target user u has bookmarked in the past. This item-based CF step helps us to give a higher ranking to candidate items that are more similar to the items the target user has bookmarked in the past (see also [10]).

To finally realize CIRTT in order to integrate tag and time information we make use of the base-level learning (BLL) equation proposed by Anderson et al. [1]. As described in our previous work [15], the BLL equation can be used to determine a relevance value for a tag t in the tag assignments of a target user u based on tag frequency and recency:

$$BLL(u, t) = \ln\left(\sum_{i=1}^n t_i^{-d}\right) \quad (2)$$

, where n is the number of times t has been used by u and t_i is the recency, i.e., the time since the i^{th} occurrence of t in the tag assignments of u . The exponent d is used to model the power law of forgetting memory items and is usually set to .5 (see [1]). In order to map these BLL values on a range of 0 - 1, we used the same normalization method as used in our previous work [15].

We adopt this equation for the ranking of items in social tagging systems using a similar method as proposed in [26] and [10]. Thus, a user is assumed to prefer an item if it has been tagged with tags of high relevance for the user, that is, with tags exhibiting a high BLL value. Given this assumption, the BLL value of a given item i for the target user u is determined using the following formula:

$$BLL(u, i) = \sum_{t \in tags(u, i)} BLL(u, t) \quad (3)$$

, where $tags(u, i)$ is the set of tags u has used to tag i .

Taken together, the prediction value $pred(u, i)$ of a candidate item i using our CIRTT approach is given by:

$$pred(u, i) = \underbrace{\sum_{j \in items(u)} sim(i, j)}_{sim(u, i)} \times BLL(u, i) \quad (4)$$

This approach enables us to weight higher the items within the candidate set that are more important to the target user (i.e., items associated with tags exhibiting a high BLL value that integrates tag frequency and recency). CIRTT and the baseline algorithms presented in this work are implemented in the Java programming language, are open-source software and can be downloaded online from our Github Repository¹ [14].

¹<https://github.com/learning-layers/TagRec/>

3. EXPERIMENTAL SETUP

In this section we describe in detail the datasets, the evaluation methodology and metrics as well as the baseline algorithms used for our experiments.

3.1 Datasets

In order to evaluate our approach and for reasons of reproducibility we used freely-available folksonomies gathered from three well-known social-tagging systems. We used datasets of the social bookmark and publication sharing system BibSonomy², the reference management system CiteULike³ and the movie recommendation site MovieLens⁴. As suggested by related work in the field (e.g. [11, 9]), we excluded all automatically imported and generated tags (e.g., bibtex-import). In the case of CiteULike we randomly selected 10% of the user profiles for reasons of computational effort (see also [7]).

We did not use a full p -core pruning technique, since this would negatively influence the recommender evaluation results in social tagging system as shown by Doerfel and Jäschke [6], but excluded all unique resources (i.e., resources that have been bookmarked only by a single user). The final dataset statistics can be found in Table 1.

3.2 Evaluation Methodology

To evaluate our item recommender approach we used a training and test-set split method as proposed by popular and related work in this area [10, 26]. Hence, for each user we sorted her bookmarks in chronological order and used the 20% most recent bookmarks for testing and the rest for training. With the training set we examined then whether a recommender approach could predict the bookmarked resources of a target user in the test set. This procedure also simulates well a real environment where the bookmarking behavior of a user in the future is tried to be predicted based on the bookmarking behavior in the past [3].

To finally quantify the recommendation accuracy of our approaches, we used a set of well-known information retrieval metrics. In particular, we report Normalized Discounted Cumulative Gain (nDCG@20), Mean Average Precision (MAP@20), Recall (R@20), Diversity (D) and User Coverage (UC) [21, 8]. All performance metrics are calculated and reported based on the top-20 recommended items. Moreover we also show the performance of the algorithms in the plots of all three accuracy metrics (nDCG, MAP and Recall) for 1 - 20 recommended items (see also [4]).

3.3 Baseline Algorithms

In order to evaluate our tag and time based approach, we compared CIRTT to several baseline methods in terms of recommender accuracy. The algorithms have been selected with respect to their popularity, performance and novelty.

Most Popular (MP): The most basic approach we utilized is the simple *Most Popular (MP)* approach that recommends for any user the same set of items. These items are weighted

by their frequency in all bookmarks, meaning that the most frequently occurring items in the dataset are recommended.

User-based Collaborative Filtering (CF): Another approach we benchmarked against is the well-known *User-based Collaborative Filtering (CF)* recommendation algorithm [19]. The main idea of CF is that users that are more similar to each other (i.e., have similar taste), will probably also like the same items. Thus, the CF approach first finds the k most similar users for the target user and afterwards recommends their items that are new to her (i.e., have not been bookmarked before). We calculated the user-similarities based on both, the binary user-item matrix as proposed in [26] (hereinafter referred to as CF_B) and the tag-based user profiles as proposed in [10] (hereinafter referred to as CF_T). Although we also considered using *Item-based CF* [18], we dismissed this method based on the tag-based recommender experiments of Bogers et al. [2] showing that user-based CF always beats item-based CF. They explain the result given that the number of items in their dataset is larger than the number of users, and this is also the case in our three datasets (Table 1).

Collaborative Filtering Using Tag and Time Information (Z / H): We also compared our approach to two alternative algorithms that focus on improving Collaborative Filtering for social tagging systems using tag and time information. The first one has been proposed by Zheng et al. [26] (hereinafter referred to as Z) and improves the traditional CF approach based on the binary user-resource matrix using tag and time information. As in our CIRTT approach this is done using information about tag frequency and recency but in contrast to our solution the authors model the forgetting process using an exponential distribution rather than a power-law distribution. Moreover, this information is already used in the user similarity calculation step and not in the item ranking step as it is done in our approach.

The second tag and time-based mechanism we tried to benchmark against was proposed by Huang et al. [10] (hereinafter referred to as H). As in our approach, this algorithm uses a 2-step recommendation process, where in the first step a potentially interesting candidate item-set for the target user is determined using user-based CF and in the second step this candidate item-set is ranked using item-based CF. In contrast to CIRTT, the authors calculate the user and item similarities based on user tag-profiles rather than based on the binary user-item matrix. Furthermore, in this algorithm the forgetting process is modeled using a simple linear function rather than a power-law distribution.

All CF-based approaches mentioned in this section use a neighborhood of 20 users and make use of the Cosine similarity measure as it is also done in CIRTT (see also [7]).

4. RESULTS

In this section, we present the results of the evaluation comparing our CIRTT approach to the baseline algorithms described in Section 3.3 with respect to recommender accuracy on three different folksonomy datasets (BibSonomy, CiteULike and MovieLens).

²<http://www.kde.cs.uni-kassel.de/bibsonomy/dumps>

³<http://www.citeulike.org/faq/data.adp>

⁴<http://grouplens.org/datasets/movielens/>

Dataset	Metric	MP	CF_T	CF_B	Z	H	$CIRTT$
BibSonomy	nDCG@20	.0143	.0448	.0610	.0621	.0564	.0638
	MAP@20	.0057	.0319	.0440	.0447	.0394	.0464
	R@20	.0204	.0618	.0820	.0834	.0816	.0907
	D	.8307	.8275	.8852	.8528	.6209	.8811
	UC	100%	99.76%	99.52%	99.52%	99.76%	99.76%
CiteULike	nDCG@20	.0062	.0407	.0717	.0762	.0706	.0912
	MAP@20	.0036	.0241	.0453	.0484	.0459	.0629
	R@20	.0077	.0630	.1033	.1077	.0928	.1225
	D	.8936	.7969	.8642	.8145	.6318	.8640
	UC	100%	98.38%	96.44%	97.32%	98.38%	97.61%
MovieLens	nDCG@20	.0198	.0361	.0602	.0614	.0484	.0650
	MAP@20	.0075	.0201	.0347	.0367	.0263	.0413
	R@20	.0366	.0561	.1031	.1013	.0763	.1058
	D	.9326	.8861	.9267	.9119	.7789	.9176
	UC	100%	97.82%	95.90%	98.43%	97.82%	95.90%

Table 2: nDCG@20, MAP@20, R@20, D and UC values for BibSonomy, CiteULike and MovieLens showing that CIRTT, that integrates tag and time information using the BLL-equation, outperforms state-of-the-art baseline algorithms (highest accuracy values are highlighted in bold).

In an extensive empirical study, Cremonesi et al. [5] have shown that standard Information Retrieval accuracy metrics (e.g., Recall or nDCG) are well suited to evaluate recommender systems, at least in case of top- N recommendation tasks. Therefore, Table 2 provides measures of accuracy (nDCG@20, MAP@20, R@20) and - additionally - measures of Diversity (D) and User Coverage (UC) for each approach and for each of the three datasets.

As expected, the MP baseline, which is not personalized at all, resulted in the lowest accuracy estimates. Regarding the two traditional CF algorithms, CF_B , which constructs a binary user-item matrix based on bookmarks, performs better than CF_T , which is based solely on the user tag-profiles. Regarding the two alternative tag- and time-based approaches, a same phenomenon can be observed as the algorithm of Zheng et al. (Z) [26], that is also based on the binary user-item matrix, performs better than the method of Huang et al. (H) [10], that is based on the user tag-profiles.

With respect to all accuracy metrics (nDCG@20, MAP@20, R@20), our CIRTT approach, that integrates tag and time information using the BLL-equation, performs best in all three datasets (BibSonomy, CiteULike and MovieLens). This may suggest that applying a power-law function as it is done via the BLL-equation is more appropriate to account for effects of recency than an exponential function (Zheng et al. [26]) or a linear function (Huang et al. [10]). A same pattern of results can be observed when looking at Figure 1 that reveals estimates of the nDCG, MAP and Recall measures for different sizes of the recommended item set. These plots show that only in the case of BibSonomy the approach of Zheng et al. reaches slightly higher accuracy estimates than our method for the first 7 recommended items. However, this changes when increasing the number of recommended items where our approach again produces the best recommender quality. Furthermore, we have also tried to integrate an exponential recency function [26] in our approach which resulted in lower accuracy estimates than the BLL power-law forgetting function.

When looking at the other two not accuracy-based metrics, interestingly, the approach of Huang et al. (H) always results in the lowest Diversity (D) of recommended items. This result might appear because this approach is based on the user tag-profiles and the Diversity metric is calculated based on tags. Finally, as all personalized approaches utilize a user-based CF approach for finding similar users, the measure of User Coverage (UC) does not appear to deviate between the different algorithms. We observed the maximum deviation of 2.53% within the MovieLens dataset.

5. CONCLUSIONS & FUTURE WORK

In this work we have presented preliminary results of a novel recommendation approach called *Collaborative Item Ranking Using Tag and Time Information (CIRTT)* that aims at improving Collaborative Filtering in social tagging systems. Our algorithm follows a two-step approach as also done in [10], where in the first step a potentially interesting candidate item set is found performing user-based CF and in the second step this candidate item set is ranked performing item-based CF. Within this ranking step we integrate the information of frequency and recency of tag use applying the Base-Level Learning (BLL) equation [1]. Thus, in contrast to existing approaches that also consider information about tags and time (e.g., [26, 10]), CIRTT draws on an empirically well established formalism modeling the reuse probability of memory items (tags) in form of a power-law forgetting function. In recent work, the same formalism has turned out to substantially improve the ranking and recommendation of tags [15].

The current evaluation conducted on datasets gathered from three social tagging systems (BibSonomy, CiteULike and MovieLens) reveals that applying the BLL equation also helps to improve the ranking and recommendation process of items. Most important, the results speak in favor of an integrative research endeavor that places a data-driven approach on a theoretical foundation provided by research on human cognition and semiotics.

Our future work will aim at improving the approach presented in this paper. For example, we will examine as to

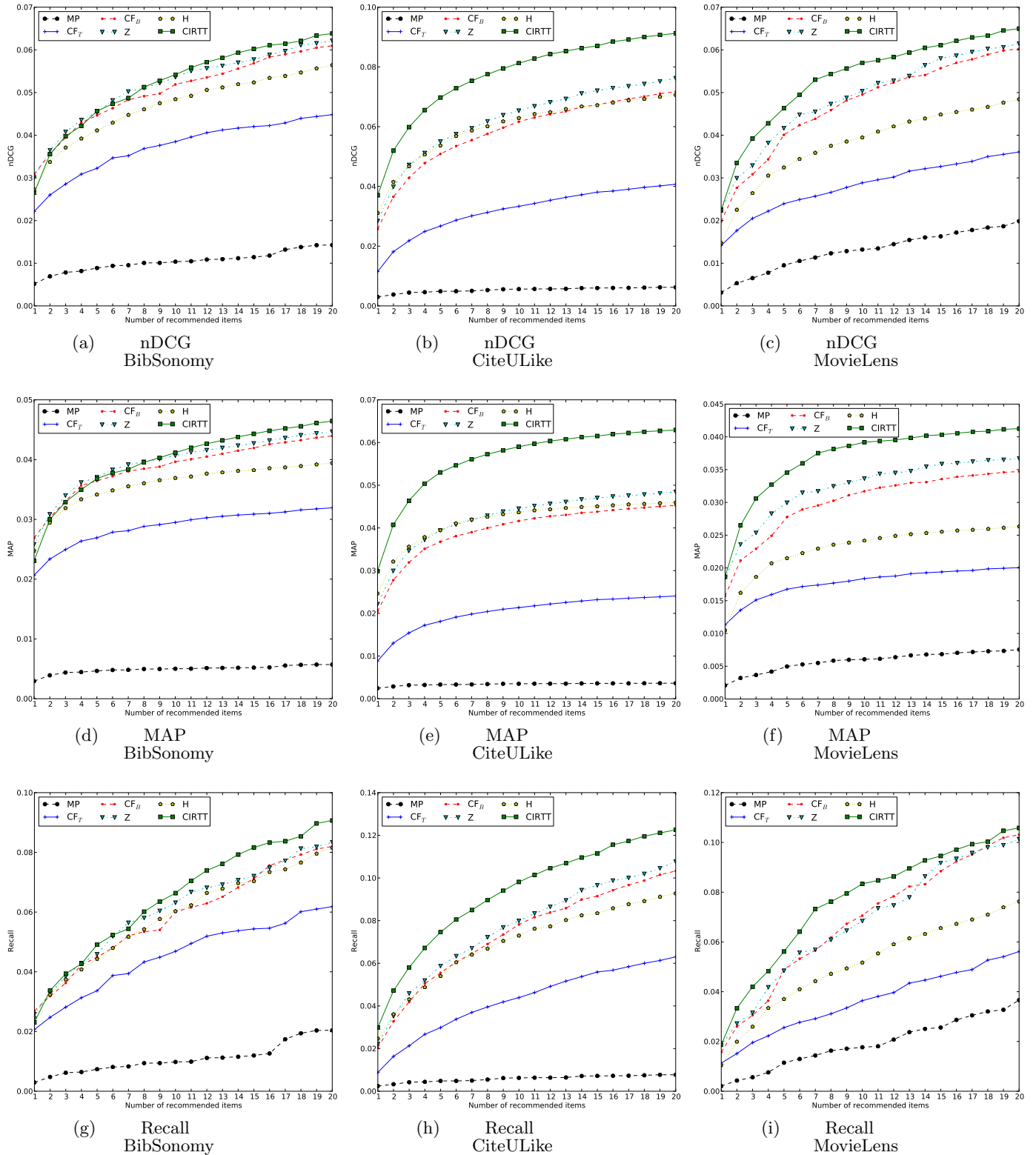


Figure 1: nDCG, MAP and Recall plots for BibSonomy, CiteULike and MovieLens showing the recommendation accuracy of our tag and time based CIRT approach along with state-of-the-art baseline algorithms for 1 - 20 recommended items (k). We can see that CIRT reaches the highest levels of recommender accuracy over all three metrics and on all datasets.

whether the BLL equation can also help to improve the calculation of user similarities and thus, to find more suitable user neighborhoods and candidate items. Additionally, we will put more emphasis on dynamics that have been found to play out in tagging systems (e.g., [22]) and how individual learning and forgetting processes are influenced by other individuals' behavior in the system. Moreover, we also plan to further improve the item ranking process using insights of relevant research dealing with recommender novelty and diversity (e.g., [25]) in order to increase the user acceptance. Finally, it would also be interesting to evaluate our proposed approach against state-of-the-art matrix factorization item recommender methods (e.g., SLIM or CLiMF).

Acknowledgments: This work is supported by the Know-Center, the EU funded project Learning Layers (Grant Nr. 318209) and the Austrian Science Fund (FWF): P 25593-G22. The Know-Center is funded within the Austrian COMET Program - Competence Centers for Excellent Technologies - under the auspices of the Austrian Ministry of Transport, Innovation and Technology, the Austrian Ministry of Economics and Labor and by the State of Styria. COMET is managed by the Austrian Research Promotion Agency (FFG).

6. REFERENCES

- [1] J. R. Anderson, M. D. Byrne, S. Douglass, C. Lebiere, and Y. Qin. An integrated theory of the mind. *Psychological Review*, 111(4):1036–1050, 2004.
- [2] T. Bogers and A. van den Bosch. Recommending scientific articles using citeulike. In *Proc., RecSys '08*, pages 287–290, New York, NY, USA, 2008. ACM.
- [3] P. G. Campos, F. Díez, and I. Cantador. Time-aware recommender systems: a comprehensive survey and analysis of existing evaluation protocols. *User Modeling and User-Adapted Interaction*, pages 1–53, 2013.
- [4] P. Cremonesi, P. Garza, E. Quintarelli, and R. Turrin. Top-n recommendations on unpopular items with contextual knowledge. In *Workshop on Context-aware Recommender Systems '11*.
- [5] P. Cremonesi, Y. Koren, and R. Turrin. Performance of recommender algorithms on top-n recommendation tasks. In *Proc., RecSys '10*, New York, NY, USA. ACM.
- [6] S. Doerfel and R. Jäschke. An analysis of tag-recommender evaluation procedures. In *Proceedings of the 7th ACM conference on Recommender systems*, pages 343–346. ACM, 2013.
- [7] J. Gemmell, T. Schimoler, M. Ramezani, L. Christiansen, and B. Mobasher. Improving folkrank with item-based collaborative filtering. *Recommender Systems & the Social Web*, 2009.
- [8] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)*, 22(1):5–53, 2004.
- [9] A. Hotho, R. Jäschke, C. Schmitz, and G. Stumme. Information retrieval in folksonomies: Search and ranking. In *The semantic web: research and applications*. Springer, 2006.
- [10] C.-L. Huang, P.-H. Yeh, C.-W. Lin, and D.-C. Wu. Utilizing user tag-based interests in recommender systems for social resource sharing websites. *Knowledge-Based Systems*, 2014.
- [11] R. Jäschke, L. Marinho, A. Hotho, L. Schmidt-Thieme, and G. Stumme. Tag recommendations in folksonomies. In *Knowledge Discovery in Databases: PKDD 2007*, pages 506–514. Springer, 2007.
- [12] R. Jäschke, L. Marinho, A. Hotho, L. Schmidt-Thieme, and G. Stumme. Tag recommendations in social bookmarking systems. *Ai Communications*, 21(4):231–247, 2008.
- [13] C. Körner, D. Benz, A. Hotho, M. Strohmaier, and G. Stumme. Stop thinking, start tagging: tag semantics emerge from collaborative verbosity. In *Proceedings of the 19th international conference on World wide web*, pages 521–530. ACM, 2010.
- [14] D. Kowald, E. Lacic, and C. Trattner. Tagrec: Towards a standardized tag recommender benchmarking framework. In *Proc., HT '14*, New York, NY, USA, 2014. ACM.
- [15] D. Kowald, P. Seitlinger, C. Trattner, and T. Ley. Long time no see: The probability of reusing tags as a function of frequency and recency. In *Proc. WWW '14*. ACM.
- [16] J. R. A. Lael J. Schooler. Reflections of the environment in memory. *Psychological Science*, 1991.
- [17] D. Parra-Santander and P. Brusilovsky. Improving collaborative filtering in social tagging systems for the recommendation of scientific articles. In *WI-IAT, 2010 IEEE/WIC/ACM*.
- [18] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. In *Proc., WWW '01*, pages 285–295, New York, NY, USA, 2001. ACM.
- [19] J. B. Schafer, D. Frankowski, J. Herlocker, and S. Sen. Collaborative filtering recommender systems. In *The adaptive web*, pages 291–324. Springer, 2007.
- [20] P. Seitlinger, D. Kowald, C. Trattner, and T. Ley. Recommending tags with a model of human categorization. In *Proc., CIKM '13*, pages 2381–2386, New York, NY, USA, 2013. ACM.
- [21] B. Smyth and P. McClave. Similarity vs. diversity. In D. Aha and I. Watson, editors, *Case-Based Reasoning Research and Development*, LNCS. Springer, 2001.
- [22] L. Steels. Semiotic dynamics for embodied agents. *Intelligent Systems, IEEE*, 21(3):32–38, 2006.
- [23] C. Trattner, Y.-I. Lin, D. Parra, Z. Yue, W. Real, and P. Brusilovsky. Evaluating tag-based information access in image collections. In *Proc., HT '12*, pages 113–122, New York, NY, USA, 2012. ACM.
- [24] K. H. Tso-Sutter, L. B. Marinho, and L. Schmidt-Thieme. Tag-aware recommender systems by fusion of collaborative filtering algorithms. In *Proc. of SAC '08*. ACM.
- [25] S. Vargas and P. Castells. Rank and relevance in novelty and diversity metrics for recommender systems. In *Proc., RecSys '11*. ACM.
- [26] N. Zheng and Q. Li. A recommender system based on tag and time information for social tagging systems. *Expert Syst. Appl.*, 2011.