Forgetting the Words but Remembering the Meaning: Modeling Forgetting in a Verbal and Semantic Tag Recommender

Dominik Kowald[1,2], Paul Seitlinger[2], Christoph Trattner[1], and Tobias Ley[3]

[1] Know-Center, Austria
{dkowald,ctrattner}@know-center.at
[2] Graz University of Technology, Austria
paul.seitlinger@tugraz.at
[3] Tallinn University, Estonia
tley@tlu.ee

Abstract. We assume that recommender systems are more successful, when they are based on a thorough understanding of how people process information. In the current paper we test this assumption in the context of social tagging systems. Cognitive research on how people assign tags has shown that they draw on two interconnected levels of knowledge in their memory: on a conceptual level of semantic fields or topics, and on a lexical level that turns patterns on the semantic level into words. Another strand of tagging research reveals a strong impact of time dependent forgetting on users' tag choices, such that recently used tags have a higher probability being reused than "older" tags. In this paper, we align both strands by implementing a computational theory of human memory that integrates the two-level conception and the process of forgetting in form of a tag recommender and test it in two large-scale social tagging datasets (drawn from CiteULike and Flickr). As expected, our results reveal a selective effect of time: forgetting is much more pronounced on the lexical level of tags. Second, an extensive evaluation based on this observation shows that a tag recommender interconnecting both levels and integrating time dependent forgetting on the lexical level results in high accuracy predictions and outperforms other well-established algorithms, such as Collaborative Filtering, Pairwise Interaction Tensor Factorization, FolkRank and an alternative time dependent approach. We conclude that tag recommenders can benefit from going beyond the manifest level of word co-occurrences, and from including forgetting processes on the lexical level.

Keywords: personalized tag recommendations; time dependent recommender systems; Latent Dirichlet Allocation; LDA; human categorization; human memory model; CiteULike; Flickr

1 Introduction

Many interactive systems are designed in a way that they mimic human behavior and thinking. For example, intelligent tutoring systems make inferences similar to teachers when they draw on knowledge of the learning domain, knowledge about the learner and knowledge about effective teaching strategies. Similarly, recommender systems based on Collaborative Filtering use information about socially similar individuals to recommend items, much in the same way as humans are influenced by similar peers when they make choices. An implicit assumption behind this seems to be that interactive systems should be better the closer they correspond to human behavior. Such assumption seems to be sensible because it is humans that interact with these systems, and the systems often draw on data that humans have produced (such as in the case of the Collaborative Filtering approaches). It is therefore reasonable to assume that strategies that have evolved in humans over their individual or collective development are good models for interactive systems. However, the assumption that an interactive system should perform better the closer it mimics human behavior is not often tested directly.

In the current paper, we test this assumption in the context of a tag recommender algorithm. We draw on research that has explored how human memory is used in a dynamic and adaptive fashion to make sense of new information encountered in the environment. Sensemaking happens by dynamically forming ad-hoc categories that relate the new information with knowledge stored in their semantic memory (e.g.,[2]). For instance, when reading an article about personalized recommendations, a novice has to figure out meaningful connections between previously distinct topics such as cognition and information retrieval and hence, has to start developing an ad-hoc category about common features of both of them. When using a social tagging system in such a situation, people apply labels to their own resources which to some extent externalize this process of spontaneously generating ad-hoc categories [8]. Usually, a user describes a particular bookmark by a combination of about three to five tags verbalizing and associating aspects of different topics (e.g., "memory","retrieval", "recommendations", "collaborative filtering").

In previous work, we have shown that this behavior can be well described by differentiating between two separate forms of information processing in human memory, a semantic process that generates and retrieves topics or gist traces, and a verbal process that generates verbatim word forms to describe the topics [22]. In this paper, we put an emphasis on another fundamental principle of human cognition to improve this model. According to Polyn et al. [20], memory traces including recently activated features contribute more strongly to retrieval than traces including features that have not been activated for a longer period of time. This relationship provides a natural account of what is called the recency effect in memory psychology (e.g., [1]). Obviously, things that happened a longer time ago tend to be forgotten and influence our current behavior less than things that have happened recently.

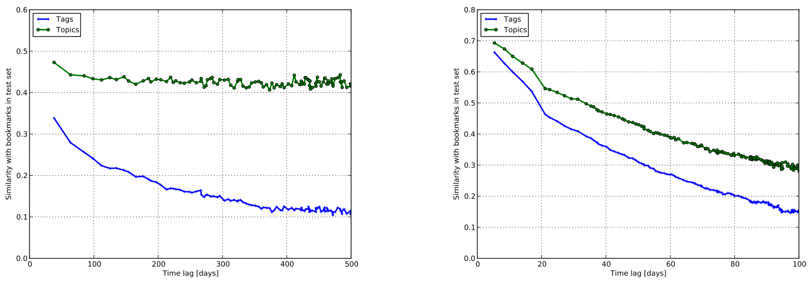
The purpose of this paper is twofold. First, we study the interaction between the effect of recency and the level of knowledge representation in human memory (semantic vs. verbal) in a social tagging system. In particular, we suspect that recency has a greater effect on the verbal level than on the semantic level, so that stronger forgetting can be observed on the word level than on the semantic level (*first research question*). The second purpose, then, is to improve our tag recommender by integrating the time-dependent forgetting process and to demonstrate that it is superior to other well-established tag recommender algorithms (e.g., Collaborative Filtering, Pairwise Interaction Tensor Factorization and FolkRank), as well as one alternative time-dependent approach called GIRPTM (*second research question*).

The remainder of this paper is organized as follows. We begin with reviewing some of the work concerning recency in memory research and its current use in social tagging in Section 2. Then we describe our approach and the experimental setup of our extensive evaluation in Sections 3 and 4. We then present the results of this evaluation in terms of recommender quality in Section 5 and discuss related work in the field in Section 6. Finally, we conclude the paper by discussing our findings und future work in Section 7.

2 Recency in Memory and in the Use of Social Tagging

Our previous work [22], a theoretical framework and corresponding model called *3Layers (3L)* differentiates between a semantic layer and an interconnected lexical layer. The former stores the topics of all bookmarks in the user's personomy, calculated with Latent Dirichlet Allocation (LDA) [15], while the latter stores the tags of those bookmarks. In a first step of calculation, 3L compares the activation of LDA topics that are true for the new bookmark to suggest tags for with the semantic layer to filter out other semantically relevant bookmarks. The resulting pattern of activation across the bookmarks can be regarded an ad-hoc category emerging through the interaction between the current context and the semantic layer. The pattern is then applied to the lexical layer to simply activate and recommend those tags that belong to relevant bookmarks. Thus, the result of the last step of calculation is a verbalization of the previously generated ad-hoc category (see also Section 3).

Concerning our first question about the impact of recency on the memory trace, research on Fuzzy Trace Theory (FTT; e.g., [3]) allows for deriving a hypothesis. FTT differentiates between two distinct memory traces, a gist trace and a verbatim trace, which represent general semantic information of e.g., a read sentence and the sentence's exact wording, respectively. These two types of memory traces share properties with our distinction between a semantic and a lexical layer (see also Section 3). While vectors of the semantic layer provide a formal account of each bookmark's gist (its general semantic content), vectors of the lexical layer correspond to a bookmark's verbatim trace (explicit verbal information in form of assigned tags). This assumption is also in line with Kintsch & Mangalath [13] who model gist traces of words by means of LDA topic vectors



(a) CiteULike.   (b) Flickr.

Fig. 1: Interaction between time dependent forgetting and level of knowledge representation for CiteULike and Flickr showing a more pronounced decline for tags than for topics (first research question).

and explicit traces of words by means of word co-occurrence vectors. An empirically well-established assumption of FTT is that verbatim traces are much more prone to time dependent forgetting than gist traces (e.g., [3]): while people tend to forget the exact wording, usually they can remember the gist of a sentence (or a bookmark). Taken together, we derived the hypothesis that a user's verbatim traces (vectors in the lexical layer that encode the user's tags) are more strongly affected by time dependent forgetting and therefore more variable over time than a user's gist traces (vectors in the semantic layer) that should be more similar to each other over time.

To test this hypothesis, we performed an empirical analysis in CiteULike and Flickr. The topics for the resources of these datasets' bookmarks were calculated using Latent Dirichlet Allocation (LDA) [15] based on the Gibbs sampling implementation in the Java framework Mallet[4]. We chose 500 latent topics as suggested in [13] as a lower bound for topic models. For each user we selected the most recent bookmark (with the largest timestamp) and described the bookmark by means of two vectors: one encoding the bookmark's LDA topic pattern (gist vector) and one encoding the tags assigned by the user (verbatim vector). Then, we searched for all the remaining bookmarks of the same user, described each of them by means of the two vectors and arranged them in a chronologically descending order. Next, we compared the gist and the verbatim vector of the most recent bookmark with the two corresponding vectors of all bookmarks in the user's past by means of the cosine similarity measure. The obtained results are represented in the two diagrams of Figure 1, plotting the average cosine similarities of all users against the time lags in days. In the case of CiteULike we show these results for the last 500 days and in the case of Flickr for the last 100 days of tagging activity because in these time ranges there are enough users available for each bookmark to calculate mean values reliably. The

[4] http://mallet.cs.umass.edu/topics.php

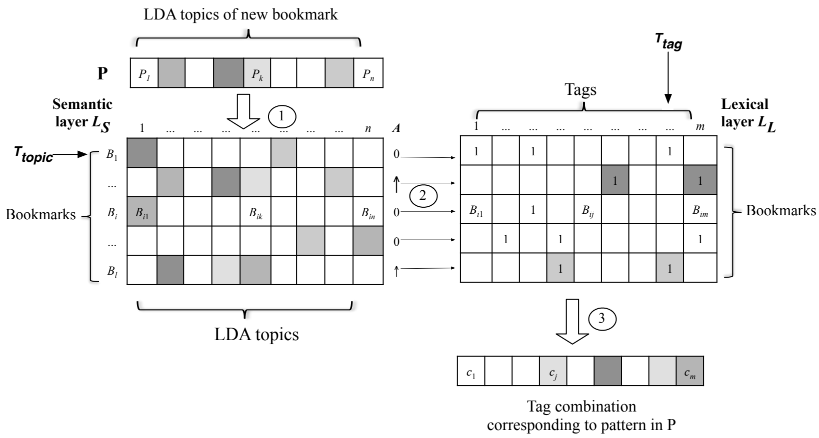


Fig. 2: Schematic illustration of the framework showing the connections between the semantic layer (L_S) encoding the LDA topics and the lexical layer (L_L) encoding the tags. *Note:* The upward-directed arrows symbolize the extent to which B_i is activated by P. Furthermore, T_{topic} and T_{tag} schematically demonstrate how the recency component is integrated in case of $3\mathrm{LT}_{topic}$ and $3\mathrm{LT}_{tag}$, respectively.

diagrams quite clearly reveal that – independent of the environment (CiteULike or Flickr) – the similarity between the most recent bookmark and all other bookmarks decreases monotonically as a function of time lag. More importantly and as expected, the time dependent decline is more strongly pronounced for the verbatim vectors (encoding tag assignments) in contrast to the gist vectors (encoding LDA topics).

3 Approach

As already mentioned in Section 1 and as Figure 2 schematically shows, we apply 3L and model a user's personomy as a network connecting a semantic layer (L_S; n LDA topics $\times$ l bookmarks matrix) and a lexical layer (L_L; l bookmarks $\times$ m tags matrix) and borrow a mechanism from MINERVA2, a computational theory of recall from memory [10], to process this network (similar to [16]). First, the LDA topics that are true for a new bookmark to be tagged are represented in a vector P with n features. Then, P is used as a cue to activate each bookmark (B_i) in L_S depending on the similarity (S_i) between both vectors. Similar to [16], S_i is given by the cosine between the two vectors:

$$S_i = \frac{\sum_{k=1}^{n} P_k \times B_{i,k}}{\sqrt{\sum_{k=1}^{n} P_k^2} \times \sqrt{\sum_{k=1}^{n} B_{i,k}^2}} \qquad (1)$$

To transform the resulting similarity values into activation values (A_i) and to further reduce the influence of bookmarks with low similarities, MINERVA2 raises S_i to the power of 3, i.e. $A_i = S_i^3$. Next, these activation values are propagated to L_L to activate tags that are associated with highly activated bookmarks on the semantic layer L_S (circled numbers 2 and 3 in Figure 2). This is realized by the following equation that yields an activation value c_j for each of the m tags:

$$\underbrace{c_j = \sum_{i=1}^{l} (B_{i,j} \times A_i)}_{3L} \tag{2}$$

Finally, we use the following method to map c_j onto the range of 0 - 1:

$$\|c_j\| = \frac{exp(c_j)}{\sum_{i=1}^{m} exp(c_i)} \tag{3}$$

The tags can then be ordered according to these normalized values and recommended to the users. These simple calculation steps constitute 3L. To realize 3LT$_{topic}$ and 3LT$_{tag}$, we integrate a recency component on the level of topics (hereinafter called T$_{topic}$) and on the level of tags (T$_{tag}$), respectively. Both recency components are calculated by the following equation that is based on the base-level learning (BLL) equation [1]:

$$BLL(t) = ln((tmstp_{ref} - tmstp_t)^{-0.5}) \tag{4}$$

, where $tmstp_{ref}$ is the timestamp of the most recent bookmark of the user and $tmstp_t$ is the timestamp of the last occurrence of t, encoded as the topic in the case of T$_{topic}$ or as the tag in the case of T$_{tag}$, in the user's bookmarks. While 3LT$_{topic}$ can be realized by using equation (5), 3LT$_{tag}$ can be realized by using equation (6):

$$\underbrace{c_j = \sum_{i=1}^{l} (B_{i,j} \times \sum_{k=1}^{n} (BLL(k)) \times A_i)}_{3LT_{topic}} \tag{5}$$

$$\underbrace{c_j = \sum_{i=1}^{l} (B_{i,j} \times BLL(j) \times A_i)}_{3LT_{tag}} \tag{6}$$

As described in [14], it is also important to take into account semantic cues in a user's current context (i.e., the resource to be tagged). We used the same approach here and modeled these semantic cues by simply taking into account the most popular tags of the resource in addition. The code we used to implement these approaches is open-source and can be found online[5].

[5] https://github.com/domkowald/tagrecommender

4 Experimental Setup

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

4.1 Datasets

We used two well-known folksonomy datasets that are freely available for scientific purposes in order to conduct our study and for reasons of reproducibility. Thus, we utilized datasets from the reference management system CiteULike[6] and the image and video sharing platform Flickr[7] to evaluate our approach on both types of folksonomies, broad (CiteULike; all users are allowed to annotate a particular resource) and narrow (Flickr; only the user who has uploaded a resource is allowed to tag it) ones. We excluded all automatically generated tags from the datasets (e.g., *no-tag*, *bibtex-import*, etc.) and decapitalized all tags as suggested by related work in the field (e.g., [21]). In the case of CiteULike we randomly selected 10% and in the case of Flickr 2% of the user profiles for reasons of computational effort (see also [7]). A p-core pruning approach has not been applied in order to capture also the issue of cold-start users or items and to prevent a biased evaluation [5]. This resulted in 379,068 bookmarks, 8,322 users, 352,343 resources, 138,091 tags and 1,751,347 tag assignments in the case of CiteULike and 542,672 bookmarks/resources, 6,393 users, 86,948 tags and 2,104,006 tag assignments in the case of Flickr.

4.2 Evaluation Methodology

To evaluate our tag recommender approaches, we split the two datasets into training and test sets based on a leave-one-out hold-out method as proposed by related work in this field (e.g., [12]). Hence, for each user we selected her most recent bookmark (in time) and put it into the test set. The remaining bookmarks were then used for the training of the algorithms. This procedure simulates well a real-world environment because the tagging behavior of a user in the future is tried to be predicted based on the tagging behavior in the past. Furthermore, it is a standard procedure for the evaluation of time-based recommender systems [4]. To finally quantify the recommender quality and to benchmark our recommender against other tag recommendation approaches, a set of well-known metrics for information retrieval and recommender systems were used. In particular, we report Recall (R@k), Precision (P@k), F1-Score (F_1@k), Mean Reciprocal Rank (MRR) and Mean Average Precision (MAP), where k is between 1 and 10 and MRR and MAP are calculated for 10 recommended tags ($k = 10$) [17].

[6] http://www.citeulike.org/faq/data.adp
[7] http://www.tagora-project.eu/

4.3 Baseline Algorithms

We compared the results of our approach to several "baseline" tag recommender algorithms. The algorithms were selected in respect to their popularity in the community, performance and novelty [18]. The most basic approach we utilized is the unpersonalized *MostPopular (MP)* algorithm that recommends for any user and any resource the same set of tags that is weighted by the frequency in all tag assignments [12]. A personalized extension of MP is the *MostPopular*$_{u,r}$ *(MP*$_{u,r}$*)* algorithm the suggests the most frequent tags in the tag assignments of the user and the resource [12]. Another simple and classic recommender approach is *Collaborative Filtering (CF)* which was adapted for tag recommendations by Marinho et al. [19]. Here the neighborhood of an user is formed based on the tag assignments in the user profile and the only variable parameter is the number of users k in this neighborhood. k has been set to 20 based on the work of Gemmell et al. [7].

Another approach we utilized is the well-known *FolkRank (FR)* algorithm which is an improvement of the Adapted PageRank approach [12]. FR adapts the PageRank algorithm in order to rank the nodes within the graph structure of a folksonomy[12] based on their importance in the network. Our implementation of it is based on the code and the settings of the open-source Java tag recommender framework provided by the University of Kassel[8]. A different popular and recent tag recommender mechanism is *Pairwise Interaction Tensor Factorization (PITF)* proposed by Rendle & Schmidt-Thieme [21]. It is an extension of factorization models that explicitly models the pairwise interactions between users, resources and tags. The PITF results presented in this paper were calculated using the open-source C++ tag recommender framework provided by the University of Konstanz[9] with 256 factors as suggested in [21].

The final approach we try to benchmark against is the time dependent *GIRPTM* algorithm presented by Zhang et al. [24] which is based on the frequency and the temporal usage of a user's tag assignments. It models the temporal tag usage with an exponential distribution based on the first- and last-time usage of the tags. The resource component is modeled by a simple most popular tags by resource approach as it is also done in 3Layers.

5 Results

In this section we present the evaluation of our three approaches in two steps that correspond to our two research questions. In step 1, we compared the three 3Layer approaches (3L, 3LT$_{topic}$ and 3LT$_{tag}$) with one another to examine the first question of whether recency has a differential effect on topics and tags. Referring to our empirical analysis in Section 2, 3LT$_{tag}$ should yield more accurate predictions than 3LT$_{topic}$ and 3L.

[8] http://www.kde.cs.uni-kassel.de/code

[9] http://www.informatik.uni-konstanz.de/rendle/software/tag-recommender/

	Measure	MP	LDA	$MP_{u,r}$	CF	FR	PITF	GIRPTM	3L	$3LT_{topic}$	$3LT_{tag}$
CUL	F_1@5	.007	.068	.199	.157	.160	.130	.207	.211	.214	**.218**
	MRR	.005	.065	.179	.168	.181	.149	.196	.194	.196	**.208**
	MAP	.005	.073	.210	.196	.212	.169	.229	.229	.232	**.246**
Flickr	F_1@5	.020	.169	.430	.413	.332	.321	.507	.516	.521	**.531**
	MRR	.020	.171	.355	.439	.351	.336	.445	.438	.445	**.476**
	MAP	.020	.205	.463	.582	.455	.433	.588	.587	.597	**.633**

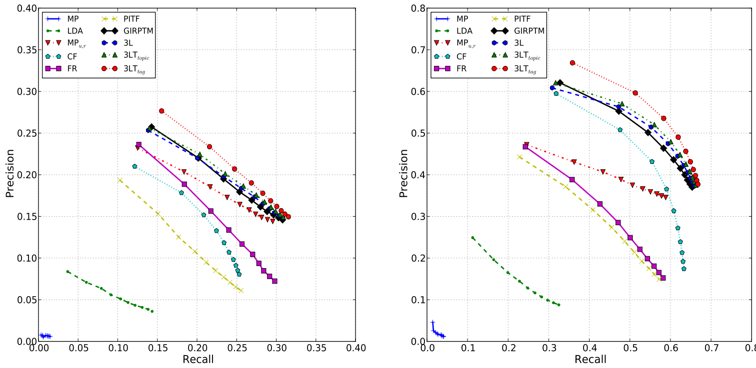
Table 1: F_1@5, MRR and MAP values for CiteULike and Flickr showing that $3LT_{topic}$ and $3LT_{tag}$ outperforms current state-of-the art algorithms (second research question).

The results in Table 1 are well in accordance with this assumption since - independent of the measure applied (F_1@5, MRR and MAP) - the difference between $3LT_{tag}$ and 3L appears to be larger than the difference between $3LT_{topic}$ and 3L. Hence, a user's gist traces (LDA topics associated with the user's bookmarks) are less prone to "forgetting" than a user's verbatim traces (tags associated with the bookmarks). Interestingly, this effect seems to be more strongly pronounced under the narrow folksonomy condition (Flickr) than under the broad folksonomy condition (CiteULike).

In a second step, we contrasted the performance of our approaches, especially $3LT_{tag}$, with several state-of-the-art algorithms to address our second research question of whether 3L and its two extensions can be implemented in form of effective and efficient tag recommendation mechanisms. First, Table 1 reveals that all personalized recommendation mechanisms clearly outperform our baseline mechanism, i.e. MP, which simply takes into account the tag's usage frequency independent of information about a particular user or resource. Second and more important, 3L and its two extensions ($3LT_{topic}$ and $3LT_{tag}$) appear to reach substantially higher accuracy estimates than the well-established mechanisms LDA, $MP_{u,r}$, CF, FR and PITF. From this we conclude that predicting tags in form of psychologically plausible steps of calculation that turn a user's gist traces into words (see Section 3) yields tag recommendations that correspond well to the user's tagging behavior. Third, with respect to MRR, the time dependent GIRPTM algorithm reaches higher estimates of accuracy than 3L. However, this relationship between the two mechanisms dramatically changes if we enhance 3L by the recency component at the two levels. Actually, $3LT_{topic}$ and especially $3LT_{tag}$ appear both to substantially outperform GIRPTM in terms of all three measures and in both datasets. Finally, as Figure 3 shows, a very similar pattern of results becomes apparent if the different approaches are evaluated by plotting recall against precision.

6 Related Work

In contrast to our own work, previous research on tag recommender systems has taken a more pragmatist stance, typically not considering models of cognitive



(a) CiteULike. (b) Flickr.

Fig. 3: Recall/Precision plots for CiteULike and Flickr showing the performance of the algorithms for 1 - 10 recommended tags (k).

processes that lead to the data that is being predicted. The probably most prominent work in this context is the work of Hotho et al. [11] who introduced an algorithm called FolkRank (FR) that has established itself as the most prominent benchmarking tag recommender approach over the past few years. Subsequent and other popular works in this context are the studies of Jäschke et al. [12] or Hamouda & Wanas [9] who introduced a set of Collaborative Filtering (CF) approaches for the problem of recommending tags to the user in a personalized manner. More recent and to some extent also well-know works in this context are e.g., the studies of Rendle et al. [21], Krestel et al. [15], Yin et al. [23] or Zhang et al. [24] who introduce a factorization model, a semantic model (based on LDA) or a time-based model to recommend tags to users (see Section 4.3).

Although the latter mentioned approaches perform reasonable well in accuratly predicting the users tags, all of them ignore well-established and long standing research from cognitive psychology on how humans process information. By applying formal models of human semantic memory, we presented in this work not only a cognitive plausable model for tag recommender systems but also revealed that the proposed approach outperforms the current state-of-the-art.

7 Discussion and Conclusion

In this study we have provided empirical evidence for an interaction between the level of knowledge representation (semantic vs. lexical) and time-based forgetting in the context of social tagging. Based on the analysis of two large-scale tagging datasets we conclude that - as expected - the gist traces of a user's personomy (the combination of LDA topics associated with the bookmarks) are more stable over time than the verbatim traces (the combination of associated tags). This pattern of results is well in accordance with research on human memory (e.g., [3]) suggesting that while people tend to forget surface details they keep quite robust memory traces of the general meaning underlying the experiences of the past (e.g., the meaning of read words). The interaction effect suggests that it is worthwhile to differentiate both time-based forgetting as well as level of knowledge representation in social tagging research.

Furthermore, the differential affect of forgetting on the two levels of processing has further substantiated the differences between tagging behavior on a semantic level of gist traces and a lexical level of verbatim traces. This in turn is in line with cognitive research on social tagging (e.g., [6]) that suggests to consider a latent, semantic level (e.g., modeled in form of LDA topics) when trying to understand the variance in the statistical patterns on the manifest level of users' tagging behavior.

Finally, we have gathered further evidence for our assumption that interactive systems can be improved by basing them on a thorough understanding of how humans process information. We note in particular that integrating two fundamental principles of human information processing, time-based forgetting and differentiating into semantic and lexical processing, significantly enhances the accuracy of tag predictions as compared to a situation when only one of the principles is considered. 3L, that is enhanced by forgetting on the lexical level ($3LT_{tag}$), outperforms both the traditional 3L, as well as other well-established algorithms, such as CF, PITF, FR and the time-based GIRPTM.

In future work, we plan to include our algorithms in an online recommender. Only in such setting is it possible to test the recommendation performance by looking at user acceptance. It will also be interesting to check whether lexical/semantic processes and forgetting are also observable in other interactive systems and Web paradigms such as in Web curation, and improve recommendations there.

References

1. J. R. Anderson and L. J. Schooler. Reflections of the environment in memory. *Psychological Science*, 2(6):396–408, 1991.
2. L. Barsalou. Situated simulation in the human conceptual system. *Language and cognitive processes*, 18(5-6):513–562, 2003.
3. C. Brainerd and V. Reyna. Recollective and nonrecollective recall. *Journal of memory and language*, 63(3):425–445, 2010.
4. 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.
5. S. Doerfel and R. Jäschke. An analysis of tag-recommender evaluation procedures. In *Proceedings of the 7th ACM Conference on Recommender Systems*, RecSys '13, pages 343–346, New York, NY, USA, 2013. ACM.
6. W.-T. Fu and W. Dong. Collaborative indexing and knowledge exploration: A social learning model. *IEEE Intelligent Systems*, 27(1):39–46, 2012.
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. R. J. Glushko, P. P. Maglio, T. Matlock, and L. W. Barsalou. Categorization in the wild. *Trends in cognitive sciences*, 12(4):129–135, 2008.
9. S. Hamouda and N. Wanas. Put-tag: personalized user-centric tag recommendation for social bookmarking systems. *Social network analysis and mining*, 1(4):377–385, 2011.
10. D. L. Hintzman. Minerva 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers*, 16(2):96–101, 1984.
11. A. Hotho, R. Jäschke, C. Schmitz, and G. Stumme. Information retrieval in folksonomies: Search and ranking. In *The semantic web: research and applications*, pages 411–426. Springer, 2006.
12. 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.
13. W. Kintsch and P. Mangalath. The construction of meaning. *Topics in Cognitive Science*, 3(2):346–370, 2011.
14. 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 *Proceedings of the 23rd international conference on World Wide Web*, WWW '14, New York, NY, USA, 2014. ACM.
15. R. Krestel, P. Fankhauser, and W. Nejdl. Latent dirichlet allocation for tag recommendation. In *Proceedings of the third ACM conference on Recommender systems*, pages 61–68. ACM, 2009.
16. P. J. Kwantes. Using context to build semantics. *Psychonomic Bulletin & Review*, 12(4):703–710, 2005.
17. M. Lipczak. *Hybrid Tag Recommendation in Collaborative Tagging Systems*. PhD thesis, Dalhousie University, 2012.
18. L. Marinho, A. Nanopoulos, L. Schmidt-Thieme, R. Jäschke, A. Hotho, G. Stumme, and P. Symeonidis. Social tagging recommender systems. In F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, editors, *Recommender Systems Handbook*, pages 615–644. Springer US, 2011.
19. L. B. Marinho and L. Schmidt-Thieme. Collaborative tag recommendations. In *Data Analysis, Machine Learning and Applications*, pages 533–540. Springer, 2008.
20. S. M. Polyn, K. A. Norman, and M. J. Kahana. A context maintenance and retrieval model of organizational processes in free recall. *Psychological review*, 116(1):129, 2009.
21. S. Rendle and L. Schmidt-Thieme. Pairwise interaction tensor factorization for personalized tag recommendation. In *Proceedings of the third ACM international conference on Web search and data mining*, WSDM '10, pages 81–90, New York, NY, USA, 2010. ACM.
22. P. Seitlinger, D. Kowald, C. Trattner, and T. Ley. Recommending tags with a model of human categorization. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, CIKM '13, pages 2381–2386, New York, NY, USA, 2013. ACM.
23. D. Yin, L. Hong, and B. D. Davison. Exploiting session-like behaviors in tag prediction. In *Proceedings of the 20th international conference companion on World wide web*, pages 167–168. ACM, 2011.
24. L. Zhang, J. Tang, and M. Zhang. Integrating temporal usage pattern into personalized tag prediction. In *Web Technologies and Applications*, pages 354–365. Springer, 2012.