

Refining Frequency-Based Tag Reuse Predictions by Means of Time and Semantic Context

Dominik Kowald^{1,2}, Simone Kopeinik², Paul Seitlinger², Tobias Ley³, Dietrich Albert², and Christoph Trattner¹

¹ Know-Center, Graz University of Technology, Graz, Austria
`{dkowald,ctrattner}@know-center.at`

² Knowledge Technologies Institute, Graz University of Technology, Graz, Austria
`{simone.kopeinik,paul.seitlinger,dietrich.albert}@tugraz.at`

³ Institute of Informatics, Tallin University, Tallinn, Estonia
`tley@tlu.ee`

Abstract. In this paper, we introduce a tag recommendation algorithm that mimics the way humans draw on items in their long-term memory. Based on a theory of human memory, the approach estimates a tag’s probability being applied by a particular user as a function of usage frequency and recency of the tag in the user’s past. This probability is further refined by considering the influence of the current semantic context of the user’s tagging situation. Using three real-world folksonomies gathered from bookmarks in BibSonomy, CiteULike and Flickr, we show how refining frequency-based estimates by considering usage recency and contextual influence outperforms conventional “most popular tags” approaches and another existing and very effective but less theory-driven, time-dependent recommendation mechanism.

By combining our approach with a simple resource-specific frequency analysis, our algorithm outperforms other well-established algorithms, such as FolkRank, Pairwise Interaction Tensor Factorization and Collaborative Filtering. We conclude that our approach provides an accurate and computationally efficient model of a user’s temporal tagging behavior. We demonstrate how effective principles of recommender systems can be designed and implemented if human memory processes are taken into account.

Keywords: personalized tag recommendations; time-dependent recommender systems; base-level learning equation; ACT-R; human memory model; BibSonomy; CiteULike; Flickr

1 Introduction

In this paper, we suggest a tag recommendation mechanism that mimics how people use to access their memory to name things they have encountered in the past. In everyday communication, people are very effective and quick in retrieving relevant knowledge from the enormous amount of information units stored in their individual long-term memory (LTM). One example is tagging resources on the Web, a rudimentary variant of communication [1, 2]. Here, people name objects, such as images or music files, by means of social tags to create retrieval cues for personal and collective information organization [3]. The issue of how human memory ensures a fast and automatic information retrieval from its huge LTM has been extensively examined by memory psychology (e.g., [4]). Essentially, human memory is tuned to the statistical structure of an individual’s environment and keeps available those memory traces that have been used frequently and recently in the past and are relevant in the current context [5].

Social tagging provides an illustrative example of the strong interplay between external, environmental and internal memory structures and processes (e.g., [6]). For instance, the development of generative models of social tagging demonstrated that the probability of a tag being applied can be modeled through the preferential attachment principle (e.g., [7]): the higher the frequency of a tag’s past occurrence in the tagging environment is, the more likely it will be reused by an individual. Additionally, the same probability is also a function of the tag’s recency, which is the time elapsed since the tag last occurred in the environment [8]. In summary, the probability of applying a particular word reflects the individual’s probability of being exposed to the word in her environment [5].

The activation equation 1 of the cognitive architecture ACT-R (e.g., [4]) is an empirically well-established formula to estimate the activation A_i of a memory trace for an item i (e.g., the tag “recognition”), where the psychological construct of item activation is assumed to control for the item’s retrieval from the LTM. It is given by:

$$A_i = B_i + \underbrace{\sum_j W_j \times S_{j,i}}_{AssociativeComponent} \quad (1)$$

A wealth of empirical data (e.g., [4, 9]) demonstrates that A_i is a function of the item’s base-level activation B_i and associative activation caused by the associative component. B_i reflects the general usefulness of item i in the individual’s past and is given by the base-level learning equation:

$$B_i = \ln\left(\sum_{j=1}^n t_j^{-d}\right) \quad (2)$$

, where n is the frequency of item occurrences in the past and t_j is the recency, which is the time since the j^{th} occurrence. For example, if a user has applied the two tags “recognition” and “recommender” with equal frequency, i.e., for an equal number of bookmarks, but “recommender” has dominated the user’s

recent bookmarks, the equation predicts a higher activation and hence, larger probability of being reused, for “recommender” than for “recognition”. The exponent d accounts for the power-law of forgetting and models the phenomenon that each memory’s activation, caused by the j^{th} occurrence, decreases in time according to a power function. The exponent d is typically set to 0.5 [4].

The second component of A_i , the associative activation, is assumed to adjust B_i according to the individual’s current context that may consists of words included in a resource’s abstract or popular tags assigned to the resource. For example, even if the base-level activation for the tag “recognition” is smaller than for “recommender”, a particular set of contextual elements, such as the words “memory” and “recollection”, will spread associative activation to “recognition” and will substantially increase its probability being reused. In equation 1, W_j represents the weights of the items j , which are elements of the current context (e.g., “memory” and “recollection”); $S_{j,i}$ represents the strengths of association from the contextual elements to an item (e.g., “recognition”). Section 3 gives a detailed and formal description of how these two components are calculated.

In the present work, we test the assumption that the two components of equation 1 (time and semantic context) can be used to improve frequency-based tag reuse predictions. Specifically, we raise the following three research questions:

- *RQ1*: Does the base-level learning equation provide a valid model of a user’s tagging behavior in the past to predict future tag assignments?
- *RQ2*: Does the additional consideration of the associative component evoked by the current context further improve the accuracy of the base-level learning equation?
- *RQ3*: Can the whole activation equation, that considers base-level and associative activation, be applied and extended to create an effective and efficient tag recommendation mechanism compared to state-of-the-art baseline approaches?

The strategy we chose to address all research questions consists of three steps. In a first step, we implemented the “pure” BLL equation in form of a tag recommender and compared its performance with a MostPopular _{u} (MP _{u}) approach, which suggests the most frequent tags in a user’s tag assignments. As expected, the comparison with MP _{u} showed evidence of the incremented value that results when additionally processing the recency of tag use. Moreover, we compared our BLL recommender with the currently leading time-based tag recommender approach introduced by [10] and showed the advantages of our theory-driven approach.

In a second step, we extended the “pure” BLL equation to the full activation equation proposed by Anderson et al. [4]. In this way, we also take into account relevant context information (i.e., tags applied to the target resource) and fine-tune base-level activation values. This has led to an improvement over the “pure” BLL equation, and showed particularly good results in settings where context

information is important (e.g., if there is a high probability that previously assigned resource tags get adapted by the users).

In a third step, we combined the activation equation with popular tags that have been applied to the target resource by other users. When also considering other users' tags, it allows us to introduce new tags to the target user, namely tags that have not been used by the target user before (e.g., [11, 12] or [13]). To this end we weighted the tags based on their frequency in the resource's tag assignments, hereinafter referred to as MostPopular_r (MP_r). We then compared the performance of the combination of the activation equation and MP_r with well-established approaches, such as Collaborative Filtering (CF), FolkRank (FR) and Pairwise Interaction Tensor Factorization (PITF), and showed that this approach outperforms the state-of-the-art algorithms in terms of recommender accuracy.

The remainder of this paper is organized as follows: we begin with discussing related work (Section 2) and describing our approach in Section 3. Sections 4 and 5 describe the experimental setup and the baseline algorithms we used for our evaluation. Section 6 addresses our three research questions and summarizes the settings and results of our extensive evaluation. Finally, in Section 7, we conclude the paper by discussing our findings in the light of the benefits of deriving tag recommender mechanisms from empirical, cognitive research.

2 Related Work

Recent years have shown that tagging is an important feature of the Social Web supporting the users with a simple mechanism to collaboratively organize and find content [14]. Although tagging has the ability to improve search (in particular tags provided by the individual), it is also known that users are typically lazy in providing tags for instance for their bookmarked resources. It is therefore not surprising that recent research has attempted to address this challenge to support the individual in her tag application process in the form of personalized tag recommenders. To date, the two following approaches have been established: graph-based and content-based tag recommender systems [12]. In our work we focus on graph-based approaches.

The probably most notable work in this context is the work of Hotho et al. [15] who introduced an algorithm termed FolkRank (FR) that has become the most prominent benchmarking tag recommender approach over the past few years. Subsequently, the work of Jäschke et al. [16] and Hamouda & Wanas [17] showed how the classic Collaborative Filtering (CF) approach could be adopted for the problem of predicting tags to the user in a personalized manner. More recent work in this context are studies of Rendle et al. [18], Wetzker et al. [19], Krestel et al. [20] or Rawashdeh et al. [21] who introduced a factorization model, a Latent Dirichlet Allocation (LDA) model or a Link-Prediction model, based on the Katz measure, to recommend tags to users.

Within the context of this paper, another relevant study addressing the potential of social tagging systems to model the user in a resource context is pre-

sented in [22]. In this work similarities between users are defined through firstly, their consensus in tagging behavior and secondly, their resource usage. The resulting network with actors, resources, tags and tag assignments as nodes, is modeled as a directed graph. The graph serves as a basis to spread activation from one actor to another, by going through multiple types of nodes that represent context information linking the actors. Stanley et al. [23] studied a tag recommendation model inspired by the declarative memory retrieval mechanism of ACT-R (e.g., [4]) on forum data. Recommendations are based on two aspects, the user’s tag history and co-occurrences between tag words and words extracted from the post’s content. Although the model initially aims to implement the entire activation equation of ACT-R, since it was tailored to the properties of a computer science forum, where assumingly the frequency of tags does not change significantly over time, the approach does not consider the time component as it is done in this paper. Sigurbjörnsson et al. [24] suggests a slightly different approach to calculate relatedness values between tags in order to recommend tags in the image-sharing portal Flickr. The proposed method is based on the Jaccard coefficient to normalize the co-occurrence of two tags

Although the mentioned approaches perform reasonably well, they are computational expensive compared to simple “most popular tags” approaches. Furthermore, they ignore recent observations with regard to social tagging systems, such as the variation of the individual tagging behavior over time [25]. To that end, recent research has made the first promising steps towards more accurate graph-based models that also account for the variable of time [26, 10]. The approaches have shown to outperform some of the current state-of-the-art tag recommender algorithms.

In line with the latter strand of research, in this paper we present a novel graph-based tag recommender mechanism that uses the activation equation that is part of the former mentioned ACT-R theory (e.g., [4]) to integrate item frequency and recency as well as current context information. We show that the approach is not only very simple and straightforward but also reveal that the algorithm outperforms current state-of-the-art graph-based (e.g., [19, 15, 16]) and the leading time-based [10] tag recommender approaches.

3 Approach

In Section 1 we formulated the assumption that the activation equation consisting of a base-level and associative component can be applied to explain a high variance in a tag’s probability of being applied. To address our first research question (as to whether the activation equation’s first component can be applied to improve a “most popular tags by user” approach) we have calculated the base-level activation $B(t, u)$ of a given tag t in a user u ’s set of tag assignments, $Y_{t,u}$. First, we determined a reference timestamp $timestamp_{ref}$ (in seconds) that is the timestamp of the most recent bookmark of user u . In our dataset samples, $timestamp_{ref}$ corresponded to the timestamp of the user’s bookmark that was selected for the test set (see Section 4.2).

If $i = 1 \dots n$ indexes all tag assignments in $Y_{t,u}$, the recency of a particular tag assignment is given by $timestamp_{ref} - timestamp_i$. Finally, $B(t, u)$ of tag t for a user u is given by the BLL equation:

$$B(t, u) = \ln\left(\sum_{i=1}^n (timestamp_{ref} - timestamp_i)^{-d}\right) \quad (3)$$

, where d is set to 0.5 based on [4]. In order to map the values onto a range of 0 to 1 we applied a normalization method as proposed in related work [27]:

$$\|B(t, u)\| = \frac{\exp(B(t, u))}{\sum_{t' \in Y_u} \exp(B(t', u))} \quad (4)$$

, where Y_u is the set of unique tags used by user u in the past.

To investigate our second research question (as to whether the BLL equation can be further improved by also considering the associative component evoked by the current context) we have implemented equation 1 in form of:

$$A(t, u) = \|B(t, u)\| + \underbrace{\sum_j W_j \times S_{j,i}}_{AssociativeComponent} \quad (5)$$

To calculate the variables of the associative component, i.e., to model a user's semantic context, we simply looked at the tags assigned by other users to the given resource r . A user's semantic context certainly consists of much more aspects, such as content words in the title or in the page text. However, since not all of our datasets contain title information or page text and other studies have convincingly demonstrated the impact of a resource's prominent tags on a user's tagging behavior (e.g., [11, 12]), we decided to approximate the context by means of other users' tags.

When applying the formula to a recommender system, related literature [24] [28] suggests to use a measure of normalized tag co-occurrence to represent the strength of an association. Accordingly, we define the co-occurrence between two tags as the number of bookmarks in which both tags are included. To add meaning to the co-occurrence value, the overall frequency of the two tags is also taken into consideration. This is done by normalizing the co-occurrence value according to the Jaccard coefficient (6) following the approach described in [24]:

$$S_{j,i} = \frac{|t_j \cap t_i|}{|t_j \cup t_i|} \quad (6)$$

In our implementation, $S_{j,i}$ is calculated as an association value between a tag previously given by the target user (t_i) and a tag that has been assigned to a resource of interest (t_j). Based on a tag co-occurrence matrix that depicts the tag relations of an entire data set, information about how many times two tags co-occur ($t_j \cap t_i$) is retrieved and set into relation with the number of bookmarks

in which at least one of the two tags appear ($t_j \cap t_i$). We set the attentional weight W_j to the number of times t_j occurred in the tag assignments of the target resource.

Finally, to examine our third research question (as to whether the activation equation can be implemented in form of an effective recommender mechanism) we extended equation 5 by also considering the most popular tags in the tag assignments of the resource Y_r (MP_r , i.e., $\arg \max t \in T(|Y_r|)$) [15]. This simple extension was necessary to allow for the prediction of new and plausible tags that a user has not assigned in her or his previous tagging history (e.g., (e.g., [11, 12]) or [13]). Finally, the list of recommended tags for a given user u and resource r is calculated by the following equation:

$$\tilde{T}(u, r) = \arg \max t \in T(\underbrace{\beta \underbrace{\|A(t, u)\|}_{BLL_{AC}} + (1 - \beta) \|Y_{t,r}\|}_{BLL_{AC} + MP_r}) \quad (7)$$

, where β is used to inversely weight the two components, i.e. the activation values $A(t, u)$ and the most popular tags of the target resource given by MP_r . The results presented in Section 6 were calculated using $\beta = 0.5$. However, we focused on the performance of $BLL_{AC} + MP_r$ in the experiments, i.e. on an approach estimating a tag's probability of being applied by means of user and corresponding resource information. The source-code of our approaches [29] is open-source and can be found online⁴.

4 Experimental Setup

In this section we describe in detail the datasets, the evaluation method and the metrics used in our experiments.

4.1 Datasets

For the purposes of our study and for reasons of reproducibility, we focused our investigations on three well-known and freely-available folksonomy datasets. To test our approach on both types of advocates, known as broad and narrow folksonomies [30] (in a broad folksonomy many users are allowed to annotate a particular resource while in a narrow folksonomy only the user who has uploaded the resource is permitted to apply tags), freely available datasets from the social bookmark and publication sharing system BibSonomy⁵, the reference management system CiteULike⁶ (broad folksonomies) and the image and video sharing platform Flickr⁷ (narrow folksonomy) were utilized.

⁴ <https://github.com/learning-layers/TagRec/>

⁵ <http://www.kde.cs.uni-kassel.de/bibsonomy/dumps>

⁶ <http://www.citeulike.org/faq/data.adp>

⁷ <http://www.tagora-project.eu/>

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.

Dataset	Core	$ B $	$ U $	$ R $	$ T $	$ TAS $
BibSonomy	-	400,983	5,488	346,444	103,503	1,479,970
	3	41,764	788	8,711	5,757	161,509
CiteULike	-	3,879,371	83,225	2,955,132	800,052	16,703,839
	3	735,292	17,983	149,220	67,072	2,242,849
Flickr	-	864,679	9,590	864,679	127,599	3,552,540
	3	860,135	8,332	860,135	58,831	3,465,346

Since automatically generated tags affect the performance of the tag recommender systems, we excluded all of those tags from the datasets (e.g., for BibSonomy and CiteULike we excluded the *no-tag*, *bibtex-import-tag*, etc.). Furthermore, we decapitalized all tags as suggested by related work in the field (e.g., [18]). In the case of Flickr we randomly selected 3% of the user profiles for reasons of computational effort (see also [31]). The overall dataset statistics can be found in Table 1. As can be seen in column “Core”, we applied both: a p -core pruning approach [32] (represented by “3”) to capture the issues of data sparseness, as well as no p -core pruning (represented by “-”) to capture the issue of cold-start users or items [33], respectively.

4.2 Evaluation Methodology

To evaluate our tag recommender approach we used a leave-one-out hold-out method as proposed by popular and related work in this area (e.g., [16]). Hence, we created two sets, one set for training and the other set for testing. To split up each dataset in two, we eliminated for each user her latest bookmark (in time) from the original dataset and added it to the test set. Each bookmark in the test set consists of a collection of one or more tags to which we further refer as relevant tags. The reduced original dataset was used for training, and the newly created one for testing. This procedure simulates a real-world environment well and is a recommended offline-evaluation procedure for time-based recommender systems [34]. To finally quantify the performance of our approaches, a set of well-known information retrieval performance standard metrics were utilized [16, 12]:

Recall ($R@k$) is calculated as the number of correctly recommended tags divided by the number of relevant tags, where t_u^k denotes the top k recommended tags and T_u the list of relevant tags of a bookmark of user u :

$$R@k = \frac{1}{|U|} \sum_{u \in U} \frac{|t_u^k \cap T_u|}{|T_u|} \quad (8)$$

Precision (P@k) is calculated as the number of correctly recommended tags divided by the number of recommended tags. As it can be seen in the following formula, in contrast to R@k, P@k usually decreases with a higher number of recommended tags k :

$$P@k = \frac{1}{|U|} \sum_{u \in U} \frac{|t_u^k \cap T_u|}{|t_u^k|} \quad (9)$$

F1-score (F1@k) combines recall and precision into one score. It is calculated as the product of P@k and R@k divided by the sum of R@k and P@k multiplied by 2 [12]:

$$F1@k = \frac{1}{|U|} \sum_{u \in U} (2 \times \frac{R@k \times P@k}{R@k + P@k}) \quad (10)$$

Mean reciprocal rank (MRR) is a ranking-dependent metric and is calculated as the sum of the reciprocal ranks of all relevant tags in the list of the recommended tags. This means that a higher MRR is achieved if the relevant tags occur at the beginning of the recommended tag list [21]:

$$MRR = \frac{1}{|U|} \sum_{u=1}^{|U|} (\frac{1}{|T_u|} \sum_{t \in T_u} \frac{1}{rank(t)}) \quad (11)$$

Mean average precision (MAP) is an extension of the precision metric that also looks on the ranking of the recommended tags. It is described in the formula below where B_n is 1 if the recommended tag at position n is relevant [21]:

$$MAP = \frac{1}{|U|} \sum_{u=1}^{|U|} (\frac{1}{|T_u|} \sum_{k=1}^{|T_u|} B_n \times P@k) \quad (12)$$

In particular, we report R@k, P@k, MRR and MAP for $k = 10$ and F1-Score ($F_1@k$) for $k = 5$ recommended tags⁸.

5 Baseline Algorithms

We compared our approaches to several baseline tag recommender algorithms. The algorithms range from simple frequency-based approaches to more complex mechanisms based on factorization or temporal models and have been selected based on their popularity in the community, performance and novelty [35, 36]. The used baselines are the following:

⁸ $F_1@5$ was also used as the main performance metric in the ECML PKDD Discovery Challenge 2009: <http://www.kde.cs.uni-kassel.de/ws/dc09/>.

MostPopular (MP): This approach is an unpersonalized tag-recommender algorithm that does not take into account information about the target user or resource. MP recommends for any user and any resource the same set of tags that is weighted by the frequency in all tag assignments [37].

MostPopular_u (MP_u): The *most popular tags by user* approach processes tagging information about the user but ignores the resource component, which means that a target user receives the same personalized tag suggestions, no matter which resource is going to be tagged. MP_u suggests the most frequent tags in the tag assignments of the user [37].

MostPopular_r (MP_r): The *most popular tags by resource* algorithm uses the previously assigned resource tags for the prediction process but ignores information about the target user. MP_r weights the tags based on their frequency in the tag assignments of the resource [37].

MostPopular_{u,r} (MP_{u,r}): This algorithm is a mixture of the most popular tags by user (MP_u) and most popular tags by resource (MP_r) approaches. MP_{u,r} joins both components using a simple linear combination [16].

Collaborative Filtering (CF): Marinho et al. [38] described how the classic Collaborative Filtering (CF) approach [39] can be adapted for tag recommendations. Since folksonomies have ternary relations (users, resources and tags), the classic CF approach can not be applied directly. Thus, the neighborhood of an user is formed based on the tag assignments in the user profile. Furthermore, in CF-based tag recommendations only the subset of users that have tagged the target resource is taken into account when calculating the user neighborhood (*Note*: if there are no users that have tagged the target resource in the past, all users are treated as candidates for the neighborhood). The set of recommended tags can then be determined based on this neighborhood [38, 16].

Adapted PageRank (APR): Hotho et al. [15] adapted the well-known PageRank algorithm in order to rank the nodes within the graph structure of a folksonomy. This is based on the idea that a resource is important if it is tagged with important tags by important users. Thus, the folksonomy is converted into an undirected graph, where the co-occurrences of users and resources, users and tags and resources and tags are treated as weighted edges.

FolkRank (FR): The FolkRank algorithm is an extension of the Adapted PageRank approach that was also proposed by Hotho et al. [15]. This extension gives a higher importance to the preference vector via a differential approach [16]. Our APR and FR implementations are based on the code and settings of

the open-source Java tag recommender framework provided by the University of Kassel⁹.

Factorization Machines (FM): Rendle [40] introduced Factorization Machines which combine the advantages of Support Vector Machines (SVM) with factorization models to build a general prediction model that is also capable of tag recommendations. In contrast to SVMs, FMs are able to estimate interactions between entities even in settings with huge sparsity (e.g., recommender systems).

Pairwise Interaction Tensor Factorization (PITF): This approach proposed by Rendle and Schmidt-Thieme [18] is an extension of Factorization Machines and is based on the Tucker Decomposition (TD) model. In contrast to TD, PITF explicitly models the pairwise interactions between users, resources and tags to provide personalized tag recommendations. The FM and PITF results presented in this paper were calculated using the open-source C++ tag recommender framework provided by the University of Konstanz¹⁰ with 256 factors, as suggested by [18].

Temporal Tag Usage Patterns (GIRP): This time-dependent tag recommender algorithm was presented by Zhang et al. [10] and is based on the frequency and the temporal usage of a user’s tag assignments. In contrast to BLL it models the temporal tag usage with an exponential distribution rather than a power-law distribution.

GIRP with Tag Relevance to Resource (GIRPTM): This is an extension of the GIRP algorithm that also takes the resource component into account [10]. This is done in the same manner as in $BLL_{AC} + MP_r$, thus adding the probability values of MP_r using a simple linear combination.

6 Results

The presentation of the evaluation results is organized in line with our three research questions, as introduced in Section 1. With respect to the recommender quality, we will turn our attention first to the BLL equation and its validity to model tagging behavior, second, to the impact of context information when added to the BLL equation (BLL_{AC}) and third, to a comparison of our context enriched BLL implementation ($BLL_{AC} + MP_r$) with state-of-the-art baseline approaches.

The BLL equation models the user’s tagging behavior with respect to frequency and recency. While the frequency of tag use is a fairly common parameter

⁹ <http://www.kde.cs.uni-kassel.de/code>

¹⁰ <http://www.informatik.uni-konstanz.de/rendle/software/tag-recommender/>

Table 2. $F_1@5$, MRR and MAP values for BibSonomy, CiteULike and Flickr (no core and core 3) showing that the BLL equation provides a valid model of a user’s tagging behavior to predict tags (first research question). Moreover, the results imply that using the activation equation (BLL_{AC}) to also take into account semantic cues (i.e, associations with resource tags) can further improve this model (second research questions).

Dataset	Core	Measure	MP_u	GIRP	BLL	BLL_{AC}
BibSonomy	-	$F_1@5$.152	.157	.162	.169
		MRR	.114	.119	.125	.133
		MAP	.148	.155	.162	.172
	3	$F_1@5$.215	.221	.228	.292
		MRR	.202	.210	.230	.286
		MAP	.238	.247	.272	.345
CiteULike	-	$F_1@5$.185	.194	.201	.211
		MRR	.165	.182	.193	.205
		MAP	.194	.213	.227	.242
	3	$F_1@5$.272	.291	.300	.336
		MRR	.268	.294	.319	.365
		MAP	.305	.337	.366	.424
Flickr	-	$F_1@5$.435	.509	.523	.523
		MRR	.360	.445	.466	.466
		MAP	.468	.590	.619	.619
	3	$F_1@5$.488	.577	.592	.592
		MRR	.407	.511	.533	.533
		MAP	.527	.676	.707	.707

for tag recommendations, the factor of time, that models the effects of a user’s long term memory (as described through recency), is expected to bring additional value to tag recommendation approaches. That is why we investigate our first research question by determining the effect of the recency component on tag assignments.

When comparing BLL with MP_u and GIRP, the results reported in Table 2 and Figure 1 clearly show that the time-dependent algorithms BLL and GIRP both outperform the frequency-based MP_u approach. Looking further at the two time-dependent algorithms, BLL reaches higher levels of accuracy than the less theory-driven GIRP algorithm in both settings (with and without p -core pruning). Even more apparent is the impact of the recency component in the narrow folksonomy (Flickr). Unlike the broad folksonomies (BibSonomy, CiteULike), the Flickr dataset has no tags of other users available for the target resource. Therefore, a user needs to assign tags without having the inspiration of previously given tags. We assume that the user, to this end, needs to draw on her long term memory that the BLL equation aims to mimic. In summary, these results prove that the BLL equation provides a valid model of a user’s tagging behavior to predict tags (first research question).

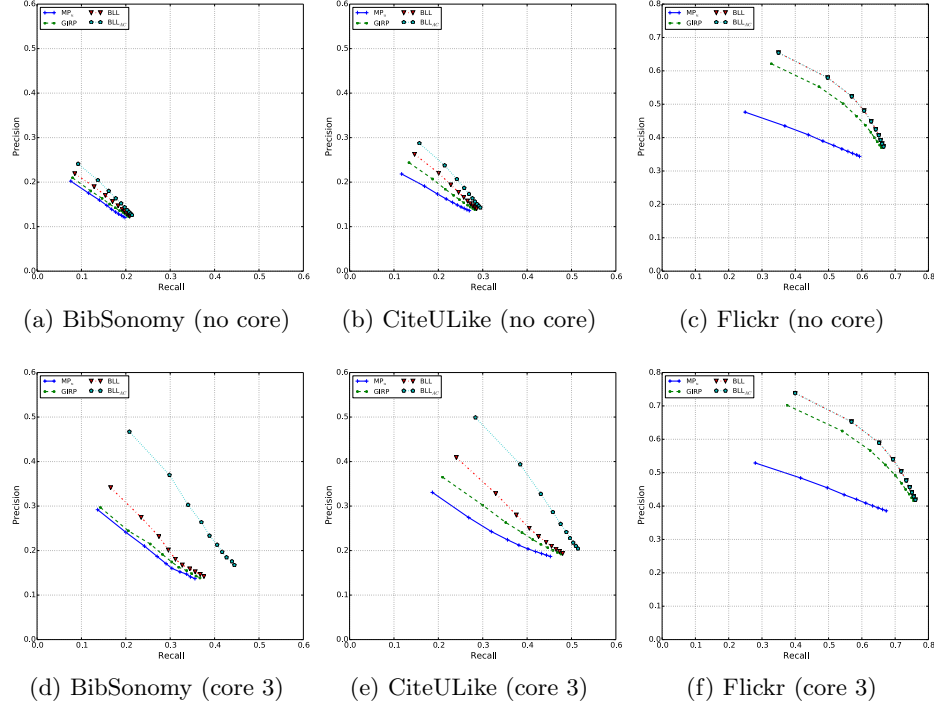


Fig. 1. Recall/Precision plots for BibSonomy, CiteULike and Flickr (no core and core 3) showing the performance of BLL and BLL_{AC} along with MP_u and GRP for 1 - 10 recommended tags (k).

By expanding BLL to BLL_{AC} , we implement the activation equation as explained in Section 3. The activation equation enriches the base-level activation (i.e., frequency and recency of tag use) by adding contextual activation through tags previously assigned to the target resource. Looking at the results of this experiment, as illustrated in Table 2 and Figure 1, a number of interesting aspects appear. For one thing, it demonstrates that BLL_{AC} reveals only a small improvement over BLL, when applied on the unfiltered datasets (no p -core) of the broad folksonomies (BibSonomy and CiteULike). However, this changes when looking at the results for the p -core pruned datasets (core 3). Caused by the higher number of tags assigned to each resource, the contextual activation gains impact. This leads to considerably increased values for all of the used metrics ($F1@5$, MRR, MAP). One might wonder why the results of BLL and BLL_{AC} are the same in the case of the narrow folksonomy (Flickr). This is, in fact, an expected outcome. As resources in the Flickr dataset are tagged by only one user (i.e, the one that has uploaded it), the model of the resource component does not generate additional value. According to these results we can answer the second research question positively, since the fine-tuning or re-ranking of

the user tags based on context cues increases the recommender accuracy in the broad folksonomies BibSonomy and CiteULike.

To address our third and last research question we combine our BLL_{AC} approach with MP_r , which leads to $BLL_{AC}+MP_r$. Hereby, BLL_{AC} models the context-aware user component while MP_r further models the resource component in order to also take into account new tags that have not been used by the target user in the past. The results presented in Table 3 show that this approach outperforms a set of state-of-the-art baseline algorithms as well as $BLL+MP_r$ (without contextual activation of the user tags). Further looking into the results, it becomes apparent that the three time-dependent algorithms (GIRPTM, $BLL+MP_r$ and $BLL_{AC}+MP_r$) produce higher estimates ($F_1@5$, MRR and MAP) across all datasets as well as in both settings (with and without p -core pruning). Moreover, an important observation is that our $BLL_{AC}+MP_r$ approach also outperforms GIRPTM, the currently leading, graph-based time-dependent tag recommendation algorithm. Particularly good results are shown here for the ranking-dependent metrics such as MRR and MAP. This observation clearly shows the advantages of our approach, that is build upon long-standing models of human memory theory, over the less-theory driven GIRPTM algorithm.

Another aspect worth discussing is the contrast of the results illustrated in Table 2, where BLL_{AC} reaches substantially higher levels of accuracy than BLL , to the results outlined in Table 3, where $BLL_{AC}+MP_r$ only indicate marginal improvements over $BLL+MP_r$. In our opinion, this effect appears because the resource tag information depicted in MP_r is congruent with data used for the contextual activation in BLL_{AC} . This finding suggests that the use of other resource metadata, such as title or body-text, may be valuable when specifying the context in BLL_{AC} (see also Section 7). Similar observations can be made when looking at the Recall/Precision curves in Figure 2.

In summary, our results clearly imply that the activation equation by Anderson et al. [4] can be used to implement a highly effective recommender approach. Overall, the simulations demonstrate that it exceeds the performance of well-established and effective recommenders, such as $MP_{u,r}$, CF, APR, FM and even the currently leading time-dependent approach GIRPTM [10] (third research question). Finally, it is indispensable to highlight that $BLL_{AC}+MP_r$, despite its simplicity, appears to be even more successful than the sophisticated FR and PITF algorithms.

7 Discussion and Conclusion

In this study we followed a three-step strategy. We started by comparing the performance of BLL with MP_u to determine the effect of considering the recency of each tag use. Results of an additional comparison may differentiate our cognitive-psychological model from the less theory-driven GIRP approach introduced by Zhang et al. [10]. Our findings, tackling the first research question, clearly demonstrate that regardless of the evaluation metric and across

Table 3. $F_1@5$, MRR and MAP values for BibSonomy, CiteULike and Flickr (no core and core 3) showing that our $BLL_{AC}+MP_r$ approach outperforms state-of-the-art baseline algorithms (third research question).

Dataset	Core	Measure	MP	MP_r	$MP_{u,r}$	CF	APR	FR	FM	PITF	GIRPTM	$BLL+MP_r$	$BLL_{AC}+MP_r$
BibSonomy	-	$F_1@5$.013	.074	.192	.166	.175	.171	.122	.139	.197	.201	.202
		MRR	.008	.054	.148	.133	.149	.148	.097	.120	.152	.158	.159
		MAP	.009	.070	.194	.173	.193	.194	.120	.150	.200	.207	.209
	3	$F_1@5$.047	.313	.335	.325	.260	.337	.345	.356	.350	.353	.358
		MRR	.035	.283	.327	.289	.279	.333	.329	.341	.334	.349	.350
		MAP	.038	.345	.403	.356	.329	.414	.408	.421	.416	.435	.439
CiteULike	-	$F_1@5$.002	.131	.253	.218	.195	.194	.111	.122	.263	.270	.271
		MRR	.001	.104	.229	.201	.233	.233	.110	.141	.246	.258	.259
		MAP	.001	.134	.280	.247	.284	.284	.125	.158	.301	.315	.317
	3	$F_1@5$.013	.270	.316	.332	.313	.318	.254	.258	.336	.346	.351
		MRR	.012	.243	.353	.295	.361	.366	.282	.290	.380	.409	.415
		MAP	.012	.294	.420	.363	.429	.436	.326	.334	.455	.489	.497
Flickr	-	$F_1@5$.023	-	.435	.417	.328	.334	.297	.316	.509	.523	.523
		MRR	.023	-	.360	.436	.352	.355	.300	.333	.445	.466	.466
		MAP	.023	-	.468	.581	.453	.459	.384	.426	.590	.619	.619
	3	$F_1@5$.026	-	.488	.493	.368	.378	.361	.369	.577	.592	.592
		MRR	.026	-	.407	.498	.398	.404	.375	.390	.511	.533	.533
		MAP	.026	-	.527	.663	.513	.523	.481	.502	.676	.707	.707

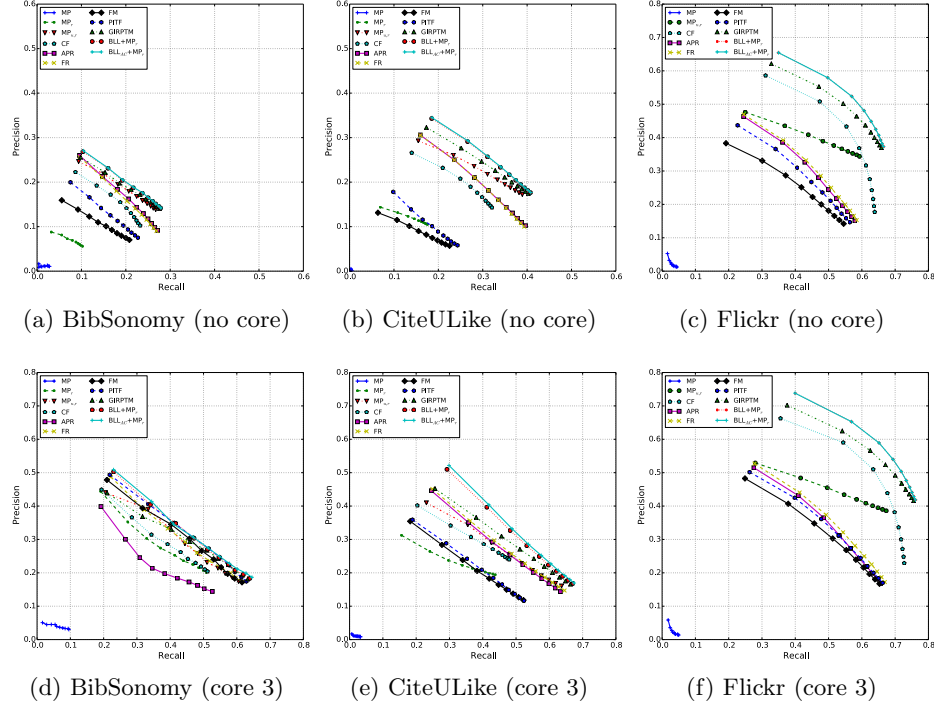


Fig. 2. Recall/Precision plots for BibSonomy, CiteULike and Flickr (no core and core 3) showing the performance of $BLL+MP_r$ and $BLL_{AC}+MP_r$ along with state-of-the-art baseline mechanisms for 1 - 10 recommended tags (k).

all datasets, BLL reaches higher levels of accuracy than MP_u and even outperforms GIRP. Thus, processing the recency of tag use is effective to account for additional variance of users' tagging behavior and therefore, a reasonable extension of simple "most popular tags" approaches. Furthermore, the advantage over GIRP indicates that drawing on memory psychology guides the application of a reliable and valid model built upon long-standing, empirical research. The equations that Zhang et al. [10] used to implement their approach were developed from scratch rather than derived from existing research described above. As a consequence, [10] models the recency of tag use by means of an exponential function, which is clearly at odds with the power law of forgetting described in related work (e.g., [5]).

In a second step, we have extended BLL to BLL_{AC} using current context information based on the activation equation of Anderson et al. [4]. Where BLL gives the prior probability of tag reuse that is learned over time, the associative component tunes this prior probability to the current context by exploiting the current semantic cues from the environment (i.e., the previously assigned tags of the target resource). This is in line with how ACT-R models the retrieval from

long-term memory. Our results show that this step clearly improves the “pure” BLL equation, especially in case of the p -core pruned datasets, where more context information (i.e, tag assignments of the target resource) are available to calculate the associative component.

In a third step, we combined BLL_{AC} with the most popular tags that have been applied by other users to the target resource in the past (i.e., MP_r) in order to be able to also recommend new tags, i.e. tags that have not been used by the target user before. Despite their simplicity, our results show that this combination ($BLL_{AC}+MP_r$) has potential to outperform well-established mechanisms, such as CF, FR and PITF. We assume this is the case because, in following some fundamental principles of human memory, our approaches are better adapted to the statistical structure of the environment. Moreover, the results of this experiment also show that there is only a small difference between $BLL_{AC}+MP_r$ and $BLL+MP_r$ (without contextual activation of the user tags), which suggests the use of additional context information, such as content-based features (e.g., the resource’s title or body-text). This would also be in line with the study presented in [41], where the authors show that the resource title has a big impact on tags in collaborative tagging systems and so could be a better alternative to represent context cues than the popular tags of the resource used in the current work.

Finally, a glance on the results shows an interdependency between the examined dataset and the performance of our approaches. While the distance to other strongly performing mechanisms does not appear to be large in case of broad folksonomies (BibSonomy and CiteULike), this distance gets substantially larger in a narrow folksonomoy (Flickr), where no tags of other users are available for the target user’s resources. From this interdependency we conclude that applying a model of human memory is primarily effective if tag assignments are mainly driven by individual habits unaffected by the behavior of other users, such as it is done in Flickr.

In future work, we will continue examining memory processes that are involved in categorizing and tagging Web resources. For instance, in a recent study [42], we introduced a mechanism by which memory processes involved in tagging can be modeled on two levels of knowledge representation: on a semantic level (representing categories or LDA topics) and on a verbal level (representing tags). Next, we will aim at combining this integrative mechanism with the activation equation to examine a potential correlation between the impact of recency (time-based forgetting) and the level of knowledge representation. We believe that conclusions drawn from cognitive science will help to develop an effective and psychologically plausible tag recommendation mechanism.

Acknowledgments: This work is supported by the Know-Center, the Austrian Science Fund (FWF): P 25593-G22 and the European Community’s Seventh Framework Program (FP7/2007-2013) funded projects, Learning Layers (Grant Agreement 318209) and weSPOT (Grant Agreement 318499).

References

1. Halpin, H., Robu, V., Shepherd, H.: The complex dynamics of collaborative tagging. In: Proceedings of the 16th international conference on World Wide Web. WWW '07, New York, NY, USA, ACM (2007) 211–220
2. Steels, L.: Semiotic dynamics for embodied agents. *IEEE Intelligent Systems* **21** (2006) 32–38
3. Marlow, C., Naaman, M., Boyd, D., Davis, M.: Ht06, tagging paper, taxonomy, flickr, academic article, to read. In: Proceedings of the seventeenth conference on Hypertext and hypermedia. HYPERTEXT '06, New York, NY, USA, ACM (2006) 31–40
4. Anderson, J.R., Byrne, M.D., Douglass, S., Lebiere, C., Qin, Y.: An integrated theory of the mind. *Psychological Review* **111** (2004) 1036–1050
5. Anderson, J.R., Schooler, L.J.: Reflections of the environment in memory. *Psychological Science* **2** (1991) 396–408
6. Held, C., Kimmerle, J., Cress, U.: Learning by foraging: The impact of individual knowledge and social tags on web navigation processes. *Computers in Human Behavior* **28** (2012) 34 – 40
7. Dellschaft, K., Staab, S.: An epistemic dynamic model for tagging systems. In: Proceedings of the nineteenth ACM conference on Hypertext and hypermedia. HT '08, New York, NY, USA, ACM (2008) 71–80
8. Cattuto, C., Loreto, V., Pietronero, L.: Semiotic dynamics and collaborative tagging. *Proceedings of the National Academy of Sciences* **104** (2007) 1461–1464
9. Pirolli, P.L., Anderson, J.R.: The role of practice in fact retrieval. *Journal of experimental psychology: Learning, memory, and cognition* **11** (1985) 136
10. Zhang, L., Tang, J., Zhang, M.: Integrating temporal usage pattern into personalized tag prediction. In: *Web Technologies and Applications*. Springer (2012) 354–365
11. Lorince, J., Todd, P.M.: Can simple social copying heuristics explain tag popularity in a collaborative tagging system? In: Proceedings of the 5th Annual ACM Web Science Conference. WebSci '13, New York, NY, USA, ACM (2013) 215–224
12. Lipczak, M.: Hybrid Tag Recommendation in Collaborative Tagging Systems. PhD thesis, Dalhousie University (2012)
13. Kowald, D., Seitlinger, P., Trattner, C., Ley, T.: Long time no see: The probability of reusing tags as a function of frequency and recency. In: *Proc. WWW '14*, New York, NY, USA, ACM (2014)
14. Körner, C., Benz, D., Hotho, A., Strohmaier, M., Stumme, G.: Stop thinking, start tagging: tag semantics emerge from collaborative verbosity. In: Proceedings of the 19th international conference on World wide web. WWW '10, New York, NY, USA, ACM (2010) 521–530
15. Hotho, A., Jäschke, R., Schmitz, C., Stumme, G.: Information retrieval in folksonomies: Search and ranking. In: *The semantic web: research and applications*. Springer (2006) 411–426
16. Jäschke, R., Marinho, L., Hotho, A., Schmidt-Thieme, L., Stumme, G.: Tag recommendations in folksonomies. In: *Knowledge Discovery in Databases: PKDD 2007*. Springer (2007) 506–514
17. Hamouda, S., Wanas, N.: Put-tag: personalized user-centric tag recommendation for social bookmarking systems. *Social network analysis and mining* **1** (2011) 377–385

18. Rendle, S., Schmidt-Thieme, L.: Pairwise interaction tensor factorization for personalized tag recommendation. In: Proceedings of the third ACM international conference on Web search and data mining. WSDM '10, New York, NY, USA, ACM (2010) 81–90
19. Wetzker, R., Zimmermann, C., Bauckhage, C., Albayrak, S.: I tag, you tag: translating tags for advanced user models. In: Proceedings of the third ACM international conference on Web search and data mining, ACM (2010) 71–80
20. Krestel, R., Fankhauser, P.: Language models and topic models for personalizing tag recommendation. In: Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on. Volume 1., IEEE (2010) 82–89
21. Rawashdeh, M., Kim, H.N., Alja'am, J.M., El Saddik, A.: Folksonomy link prediction based on a tripartite graph for tag recommendation. *Journal of Intelligent Information Systems* (2012) 1–19
22. Troussov, A., Parra, D., Brusilovsky, P.: Spreading activation approach to tag-aware recommenders: modeling similarity on multidimensional networks. In: Proceedings of Workshop on Recommender Systems and the Social Web at the 2009 ACM conference on Recommender systems, RecSys. Volume 9. (2009)
23. Stanley, C., Byrne, M.D.: Predicting tags for stackoverflow posts. In: Proceedings of ICCM. (2013)
24. Sigurbjörnsson, B., Van Zwol, R.: Flickr tag recommendation based on collective knowledge. In: Proceedings of the 17th international conference on World Wide Web, ACM (2008) 327–336
25. Yin, D., Hong, L., Xue, Z., Davison, B.D.: Temporal dynamics of user interests in tagging systems. In: Twenty-Fifth AAAI conference on artificial intelligence. (2011)
26. Yin, D., Hong, L., Davison, B.D.: Exploiting session-like behaviors in tag prediction. In: Proceedings of the 20th international conference companion on World wide web, ACM (2011) 167–168
27. McAuley, J., Leskovec, J.: Hidden factors and hidden topics: Understanding rating dimensions with review text. In: Proceedings of the ACM Conference Series on Recommender Systems, New York, NY, USA, ACM (2013)
28. Van Maanen, L., Marewski, J.N.: Recommender systems for literature selection: A competition between decision making and memory models. In: Proceedings of the 31st Annual Conference of the Cognitive Science Society. (2009) 2914–2919
29. Kowald, D., Lacic, E., Trattner, C.: Tagrec: Towards a standardized tag recommender benchmarking framework. In: Proc. HT'14, New York, NY, USA, ACM (2014)
30. Helic, D., Körner, C., Granitzer, M., Strohmaier, M., Trattner, C.: Navigational efficiency of broad vs. narrow folksonomies. In: Proceedings of the 23rd ACM conference on Hypertext and social media. HT '12, New York, NY, USA, ACM (2012) 63–72
31. Gemmell, J., Schimoler, T., Ramezani, M., Christiansen, L., Mobasher, B.: Improving folkrank with item-based collaborative filtering. *Recommender Systems & the Social Web* (2009)
32. Batagelj, V., Zaveršnik, M.: Generalized cores. *arXiv preprint cs/0202039* (2002)
33. Doerfel, S., Jäschke, R.: An analysis of tag-recommender evaluation procedures. In: Proceedings of the 7th ACM Conference on Recommender Systems. RecSys '13, New York, NY, USA, ACM (2013) 343–346

34. Campos, P.G., Díez, F., Cantador, I.: Time-aware recommender systems: a comprehensive survey and analysis of existing evaluation protocols. *User Modeling and User-Adapted Interaction* (2013) 1–53
35. Marinho, L., Nanopoulos, A., Schmidt-Thieme, L., Jäschke, R., Hotho, A., Stumme, G., Symeonidis, P.: Social tagging recommender systems. In Ricci, F., Rokach, L., Shapira, B., Kantor, P.B., eds.: *Recommender Systems Handbook*. Springer US (2011) 615–644
36. Balby Marinho, L., Hotho, A., Jschke, R., Nanopoulos, A., Rendle, S., Schmidt-Thieme, L., Stumme, G., Symeonidis, P.: *Recommender Systems for Social Tagging Systems*. SpringerBriefs in Electrical and Computer Engineering. Springer (2012)
37. Jäschke, R., Marinho, L., Hotho, A., Schmidt-Thieme, L., Stumme, G.: Tag recommendations in social bookmarking systems. *Ai Communications* **21** (2008) 231–247
38. Marinho, L.B., Schmidt-Thieme, L.: Collaborative tag recommendations. In: *Data Analysis, Machine Learning and Applications*. Springer (2008) 533–540
39. Schafer, J.B., Frankowski, D., Herlocker, J., Sen, S.: Collaborative filtering recommender systems. In: *The adaptive web*. Springer (2007) 291–324
40. Rendle, S.: Factorization machines. In: *Data Mining (ICDM)*, 2010 IEEE 10th International Conference on, IEEE (2010) 995–1000
41. Lipczak, M., Milios, E.: The impact of resource title on tags in collaborative tagging systems. In: *Proceedings of the 21st ACM Conference on Hypertext and Hypermedia*. HT '10, New York, NY, USA, ACM (2010) 179–188
42. Seitlinger, P., Kowald, D., Trattner, C., Ley, T.: Recommending tags with a model of human categorization. In: *The ACM International Conference on Information and Knowledge Management*, New York, NY, USA, ACM (2013)