The Journal of Web Science, 2016, 2: 45–61

Improving Collaborative Filtering Using a Cognitive Model of Human Category Learning

Simone Kopeinik[1], Dominik Kowald[1,2], Ilire Hasani-Mavriqi[1,2] and Elisabeth Lex[1]

[1]*Knowledge Technologies Institute*
Graz University of Technology
Graz, Austria

[2]*Know-Center*
Graz University of Technology
Graz, Austria

ABSTRACT

Classic resource recommenders like Collaborative Filtering treat users as just another entity, thereby neglecting non-linear user-resource dynamics that shape attention and interpretation. SUSTAIN, as an unsupervised human category learning model, captures these dynamics. It aims to mimic a learner's categorization behavior. In this paper, we use three social bookmarking datasets gathered from BibSonomy, CiteULike and Delicious to investigate SUSTAIN as a user modeling approach to re-rank and enrich Collaborative Filtering following a hybrid recommender strategy. Evaluations against baseline algorithms in terms of recommender accuracy and computational complexity reveal encouraging results. Our approach substantially improves Collaborative Filtering and, depending on the dataset, successfully competes with a computationally much more expensive Matrix Factorization variant. In a further step, we explore SUSTAIN's dynamics in our specific learning task and show that both memorization of a user's history and clustering, contribute to the algorithm's performance. Finally, we observe that the users' attentional foci determined by SUSTAIN correlate with the users' level of curiosity, identified by the SPEAR algorithm. Overall, the results of our study show that SUSTAIN can be used to efficiently model attention-interpretation dynamics of users and can help improve Collaborative Filtering for resource recommendations.

Keywords: Resource recommendations, Collaborative filtering, Hybrid recommendations, SUSTAIN, Attentional focus, Decision making, Social tagging, LDA

ISSN 2332-4031; DOI 10.1561/106.00000007

1 Introduction

The Web features a huge amount of data and resources that are potentially relevant and interesting for a user. However, users are often unable to evaluate all available alternatives due to the cognitive limitations of their minds. Thus, recommender systems have been proved as being a valid approach for Web users for coping with information overload Kantor *et al.*, 2011 – with Collaborative Filtering (CF) being one of the most successful methods Bar *et al.*, 2013. CF recommends resources to a user based on the digital traces she leaves behind on the Web, i.e., her interactions with resources and the interactions of other, similar users.

Recent advances in the interdisciplinary field of Web Science provide even more comprehensive digital traces of social actions and interactions that can be exploited in recommender systems' research. At least implicitly, research on recommender systems has implemented interesting assumptions about structures and dynamics in Social Information Systems (SIS), such as MovieLens, LastFM or BibSonomy. For instance, by computing matrices or high-dimensional arrays, approaches like CF represent and process SIS as networks or graphs, which relate entities of different quality (e.g., users, resources, time, ratings, tags, etc.) to each other. That way, a compositional view is taken that is reminiscent of a material-semiotic perspective (e.g., Law, 2009), assuming that we gain a deeper understanding of the intention or function of an entity, if we consider the associations it has established with other entities. In other words, "everything in the social and natural worlds [is regarded] as a continuously generated effect of the webs of relations within which they are located" (Law, 2009, p. 142).

Problem. If we look at the machinery underlying CF, it becomes clear that structurally the algorithm treats users as just another entity, such as a tag or a resource. We regard this indifference as a structuralist simplification abstracting from individuals' complexity. The structuralist stance also runs the risk of neglecting nonlinear, dynamic processes going on between different entities, such as a user's intentional state (e.g., attentional focus, interpretations, decision making) and resources (e.g., articles) consumed in the past.

Approach and methods. The main goal of this work, and also of our previous work Seitlinger *et al.*, 2015 is to take a closer look at these dynamics and to capture them by means of an appropriate model. Each user develops subjectivity, an idiosyncratic way of perceiving and interpreting things in the world, which manifests itself in particular preferences. Partially, this development evolves through a user's trajectory in the SIS (e.g., Fu and Dong, 2012). Every resource that we decide to collect corresponds to a learning episode: Depending on the resource's features, the episode causes a shift in attention, particularly in attentional tunings for certain features as well as a shift in mental categories (conceptual clusters), which influences our decision-making (e.g., Love *et al.*, 2004). The shape that mental patterns (e.g., attentional tunings and conceptual clusters) acquire, is governed by both the environment and the current mental state. The acquired pattern in turn orients the user towards particular resources and hence, closes the loop of the environment-user dynamics.

In order to capture these dynamics, we investigate the potential of SUSTAIN Love *et al.*, 2004, a particularly flexible cognitive model of human category learning. To this end, we slightly adapt the approach as described in Section 3.2 to train a model using a user's history (collected resources in a training set). The resulting user model is then applied to predict new resources from a preselected candidate set. For our empirical studies, we utilize three social bookmarking datasets from BibSonomy, CiteULike and Delicious. We chose social tagging systems for our study because their datasets are freely-available for scientific purposes and because tagging data can be utilized to derive semantic topics for resources Griffiths *et al.*, 2007 by means of LDA (see Section 3.3).

Research questions and findings. SUSTAIN, a learning model built upon theories of human category learning, can differentiate between users by means of attention and interpretation dynamics demonstrated towards observed aspects. We further talk about attentional and conceptual processes. Attentional processes describe the cognitive operation that decides which environmental aspects a user attends to (focuses on) and therefore determines what a user learns, while conceptual processes refer to the development and incremental refinement of a user's specific model of concepts and its interpretation. Our hypothesis is that these dynamics can be exploited to anticipate user-specific preferences and decisions on resource engagement. In this work, we therefore investigate a resource recommender that draws on SUSTAIN to model a user's traces (e.g., items a user has collected in the past) with an unsupervised clustering approach. The model incorporates individuals' attentional foci and their semantic clusters. Our main hypothesis is that given sufficient traces per user for training, a recommender equipped with SUSTAIN can be applied to simulate a user's decision making with respect to resource engagement, leading to improved recommender accuracy. This is based on the assumption that learning happens in categories and new resource items are likely to relate to previously visited categories. Thus, the first research question of our work is briefly stated as:

RQ1: *Do resource recommendations become more accurate if a set of resources identified by CF is processed by SUSTAIN to simulate user-specific attentional and conceptual processes?*

To tackle this research question, we first adapted and implemented the unsupervised learning paradigm of SUSTAIN to fit our learning task. In a second step, we combined our approach with user-based Collaborative Filtering (CF_U) to create our hybrid approach SUSTAIN+CF_U. Then, we compared this algorithm to SUSTAIN alone, CF_U as well as other state-of-the-art approaches like resource-based CF (CF_R) and an effective Matrix Factorization variant (WRMF) Hu *et al.*, 2008. Our results reveal that SUSTAIN+CF_U outperforms SUSTAIN, CF_U and CF_R in our setting. Furthermore, WRMF only reaches higher accuracy estimates in one of the datasets, which indicates that our approach can also compete with this much more computationally expensive method. This leads us to our next research question:

RQ2: *Which aspects of the SUSTAIN algorithm contribute to the improved performance?*

To address this question, we carried out a parameter study, in which a set of different parameters are simulated and observed. The resulting plots indicate the effect of recency that can be inferred from the optimal learning rate and the impact of the dynamic learning approach, i.e., how many semantic clusters work best for a specific dataset?

To validate the computational efficiency of SUSTAIN+CF_U compared to state-of-the-art methods such as WRMF, our third research question is:

RQ3: *To what extent can resource recommendations be calculated in a computationally efficient way using SUSTAIN+ CF_U in comparison to other state-of-the-art algorithms like matrix factorization?*

Addressing this research question, we analyzed the computational complexity of the approaches discussed when studying RQ1. We found that the most computationally expensive step of SUSTAIN+CF_U is the calculation of the resource-specific topics. Since our datasets do not contain topic information, Latent Dirichlet Allocation (LDA) was applied to extract 500 topics describing each resource. Because this step can be calculated offline, the complexity of our approach is much lower than that of WRMF.

With respect to evaluation, we tried to take a broader perspective on our hybrid approach by additionally investigating SUSTAIN-specific attentional entropy values. More specifically, we investigated the correlation between the attentional entropy values and a user's curiosity, since, as described by Loewenstein Loewenstein, 1994, *when attention becomes focused on a gap in one's knowledge*, curiosity arises[1].

The well known SPEAR algorithm Noll *et al.*, 2009; Yeung *et al.*, 2011 can be used to calculate expertise scores for users in a network based on their resource interaction patterns. Using these expertise scores, it is possible to determine discoverers among users, i.e., *curious users* who tend to be faster at finding resources of high quality. With this in mind, we raise the last research question of this work:

RQ4: *Do users' attentional foci, determined by SUSTAIN, correlate with users' expertise scores identified by the SPEAR algorithm?*

In order to address this research question, we correlated SUSTAIN attentional entropy values with SPEAR's expertise scores on our three datasets. We observed Spearman rank correlation values between .55 for Delicious and .83 for BibSon-

[1] http://ideas.time.com/2013/04/15/how-to-stimulate-curiosity

omy, which indicates that users with a high curiosity value determined by SUSTAIN also receive a high expertise score determined by SPEAR and thus, can be identified as discoverers.

Structure. The rest of this work is organized as follows: In Section 2, we discuss related work that has inspired our hybrid recommendation approach. A detailed description of the algorithm and its application can be found in Section 3. In Section 4, we first describe the methodology applied to compare the performance of our SUSTAIN+CF_U approach to several baseline algorithms. Second, the setup of a parameter investigation study is given and third, details on the algorithms' computational efficiencies are provided. Finally, we report how we used the SPEAR algorithm's curiosity values to compare with user-specific attentional preferences (tunings). Results addressing our four research questions are presented and discussed in Section 5. Conclusions and opportunities for future work are given in Section 6.

2 Related Work

At the moment, we identify three main research directions that are related to our work.

Collaborative filtering extensions. In Lacic *et al.*, 2014, the *Collaborative Item Ranking Using Tag and Time Information (CIRTT)* approach is introduced, which combines user-based and item-based CF with the information about tag frequency and recency through the base-level learning (BLL) equation from human memory theory. An extensive survey on CF was recently conducted by Shi *et al.*, 2014. In this survey, the authors classify CF approaches based on the type of information that is processed and the type of paradigm applied. Furthermore, CF extensions are defined as approaches that, enrich classic CF algorithms with valuable additional information on users and resources. Analogous categorization of CF studies is performed in Adomavicius and Tuzhilin, 2005 as well. Additionally, these studies have identified challenges that are crucial to future research on CF. In this context, authors state the fact that there is a lack of studies which address issues on recommender systems from the psychological perspective. To the best of our knowledge, there have been no remarkable endeavors which combine the implementation of a dynamic and connectionist model of human cognition, such as SUSTAIN, with existing CF algorithms. The work presented in Wang and Blei, 2011 is related to our study due to its focus on deriving semantic topics for resources. The approach presented in Wang and Blei, 2011 combines collaborative filtering and probabilistic topic modeling to recommend existing and newly published scientific articles to researchers in an online scientific community. Similarly, the author in Marlin, 2004 introduces the User Rating Profile Model for rating-based collaborative filtering, which combines a multinomial mixture model, the aspect model and LDA.

Recommender systems and user modeling. The work by Cremonesi *et al.*, 2012b distinguishes between recommender systems that provide non-personalized and personalized recommendations. While non-personalized recommender systems are not based on user models, personalized ones choose resources by taking into account the user profile (e.g., previous user interactions or user preferences). Various techniques have been proposed to design user models for resource recommendations Jawaheer *et al.*, 2014; Coleho *et al.*, 2010. Some approaches aim to provide dynamically adapted personalized recommendations to users Dooms, 2013.

Another related field is human decision making in recommender systems Chen *et al.*, 2013. For example, the work presented in Cremonesi *et al.*, 2012a systematically analyzes recommender systems as decision support systems based on the nature of users' goals and the dynamic characteristics of the resource space such as e.g., availability of resources. Our recent work Kowald and Lex, 2016 shows that the type of folksonomy in a social tagging systems also determines the efficacy of a tag recommender approach. There is, however, still a lack of research focusing on investigating user decision processes in detail, considering insights from psychology. With this work, we contribute to this sparse area of research.

Long tail recommendations and user serendipity. In the recommender systems community, long tail recommendations have also gained in importance. Essentially, the long tail refers to resources of low popularity Shi *et al.*, 2014. However, enhancing recommendation results with long tail resources can impact user satisfaction. In this context, current research Shi *et al.*, 2014; Yin *et al.*, 2012; Shi, 2013 investigates whether additional revenue can be generated by the recommender systems from long tail resources. Various solutions have been proposed to overcome the problem of over-specialization and concentration-bias in recommender systems Adamopoulos and Tuzhilin, 2014; Lamprecht *et al.*, 2015. The problem of concentration-bias becomes evident since traditional CF algorithms recommend resources based on the users' previous history of activities. Hence, resources with the most occurrences in this history are typically repeatedly recommended to users, causing a narrowing of choices by excluding other resources which might be of interest. Additionally, recommending resources based on user's previous activities or preferences yields to over-specialization of recommendations. However, the balance between information overload and facilitating users to explore new horizons by recommending serendipitous choices is not tackled within the scope of this work.

3 Approach

In this section, we first introduce the main principles of the SUSTAIN model, followed by all steps of our approach and its implementation. This includes a delineation of how we designed a hybrid recommender based on SUSTAIN and how we derived semantic topics by means of LDA. Finally, we describe how we identified candidate resources using CF. Notations used throughout this paper are summarized in Table 1.

3.1 SUSTAIN

SUSTAIN (*Supervised and Unsupervised STratified Adaptive Incremental Network*) is a flexible model of human category learning that is introduced and thoroughly discussed in Love *et al.*, 2004. By means of a clustering approach, it represents

Symbol	Description
u	user
v	neighbor in the sense of CF
t	tag
r	resource
c	candidate resource
P	set of posts / bookmarks
U	set of users
$V_{u,r}$	neighbors of user u that bookmarked r
T	set of tags
R	set of resources
R_u	resources of user u
R_v	resources of neighbor v
S_u	similar resources of u based on topics
S_r	similar resources of resource r
C_u	resource candidate set of user u
Z	number of topics (i.e., n dimensions)
k	number of neighbors (CF)
k	number of Matrix Factorization factors
l	number of iterations
I	topic vector of a resource
I_{act}	activated topics of I (i.e., with value 1)
H_j	cluster j in a user's clusters
H_m	most activated (winning) cluster
H_j^{act}	activation value of cluster j
H_m^{act}	activation value of winning cluster m
μ_{ij}	distance to cluster j at dimension i
λ_i	attentional tuning (weight) of dimension i
r	attentional focus parameter
η	learning rate
τ	threshold for the creation of new clusters
$sim(u,v)$	similarity between users u and v
α	weighting parameter of SUSTAIN
$CF_U(u,r)$	Collaborative Filtering value for u and r
$RecRes(u)$	set of recommended resources for user u

Table 1: Overview of notations used in this paper.

the way humans build up and extend their category representations when learning by means of examples. The key points of the model are flexibility and simplicity, which are supported by the fact that the number of hidden units (i.e., clusters) is not chosen in advance, but is discovered incrementally through the learning trajectory. Initially, the model starts as very simple with one cluster representing the first example, and then grows with the complexity of the problem space. The model only recruits a new cluster if a new example cannot be accommodated in one of the already existing clusters.

SUSTAIN is described as a three layer model with (1) the input layer that encodes the input stimulus, (2) the intermediate layer, a cluster set representing learned categories and (3) the output layer that predicts which category an input stimulus belongs to. Depending on the requirements, the model can support either unsupervised or supervised learning processes, where the two approaches mainly differ through their means of cluster recruitment. Supervised learning requires an external feedback mechanism that verifies the correct categorization of new examples. A false categorization is interpreted as an error and leads to a new cluster recruitment. Unsupervised learning on the other hand, does not require an explicit feedback mechanism but instead uses the similarity of the input stimulus to the cluster set. In other words, if a given input stimulus' similarity to the existing clusters is below a threshold value τ, it is assumed that the input cannot be sufficiently represented in the existing cluster set. This leads to a new cluster representing the input stimulus. In order to explain the input stimulus, the existing clusters compete amongst each other. Therefore, for each cluster an activation value is calculated that reflects the similarity to the input stimulus. The highest activated cluster wins and will, if its activation is greater than τ, predict the input stimulus' category.

In line with the requirements of our learning task, this work focuses on the unsupervised learning process, clustering with interconnected input, hidden and output units.

To adjust to the peculiarities of different data sets, the approach additionally offers parameters such as the learning rate η and the attentional focus r (see also Table 2). The learning rate η determines the influence of an input stimuli on its accommodating cluster and consequently defines how fast the algorithm learns new patterns. The attentional focus r is a constant that represents a person's capability to focus on information aspects or features relevant to a given task, while suppressing minor features of that particular task. To capture a user's specific preferences for certain aspects, the attentional focus r is enhanced by attentional tunings (i.e., tunings of the attentional focus on input features that evolve with encounters with new exemplars).

In this work, we train a slightly adapted SUSTAIN model using a user's history (i.e., collected resources in a training set). The resulting user model is applied to predict new resources from a preselected candidate set. During training and testing, SUSTAIN maps the input features (e.g., topics identified by Latent Dirichlet Allocation) of a resource to a set of dimensions at the input layer. The activation of each dimension is controlled by the attentional tuning that is learned in the course of the training phase and reflects the importance of the corresponding feature dimension for a specific user. The hidden layer consists of a set of clusters each representing similar resources encountered in the past. Hence, one cluster corresponds to a user-specific field of interest. In our test phase, the set and the structure of recruited clusters are treated as fixed measurements that no longer change. The classification decision (i.e., the decision to choose or not choose a given resource) is a function of the activation of the most activated (winning) cluster.

3.2 A Hybrid Resource Recommender Based on SUSTAIN

First, to describe our Web resources using categories, we derive 500 LDA topics from tags assigned to resources of our datasets Griffiths *et al.*, 2007, as described in Section 3.3. The LDA topics of our resources represent the n input features of our model. Then, on the basis of the resources a user has bookmarked in the past (i.e., the training set of a user), each user's personal attentional tunings and cluster representations are created in the training phase and included in our user model. Subsequently, our user model based prediction algorithm is evaluated in the testing phase.

To better fit our learning task's specific needs, we slightly adapt SUSTAIN's unsupervised clustering approach; and our adaptions impact specifically the training and testing phase. More precisely, we make an adjustment to the very high number of 500 input dimensions by limiting the learning focus to the topics activated by the current learning resource (further referred to as I_{act}). This led to improved performance results, which explain the difference to results reported in our previous work Seitlinger *et al.*, 2015.

Training. Following an unsupervised learning procedure, we start simple, with one cluster and expand the number of clusters if necessary. Please note that all SUSTAIN-specific parameter settings are adopted from Love *et al.*, 2004 (see Table 2).

For each resource in the training set of a user u, we start by calculating the distance μ_{ij} to cluster j at dimension i as described in equation (1):

$$\mu_{ij} = \left| I^{pos_i} - H_j^{pos_i} \right| \tag{1}$$

where I is the n-dimensional input vector, which represents the topics of this resource, and vector H_j is cluster j's position in the n-dimensional feature space, which holds a value for each topic and is initially set to $\vec{0}$. In this setup, input and cluster vectors represent 500 topics of which only a few are activated by each resource. Adjusting to this setting, we set the distance μ_{ij} to 1 (maximal distance) for every topic i that is not activated in the input vector ($I^{pos_i} = 0$) and therefore $i \notin I_{act}$ for $I_{act} = \{i \in I \wedge i = 1\}$. In the next step, we consider only activated topics $i \in I_{act}$ to calculate the activation value H_j^{act} of the j^{th} cluster by equation (2):

$$H_j^{act} = \frac{\sum_{i \in I_{act}} (\lambda_i)^r e^{-\lambda_i \mu_{ij}}}{\sum_{i \in I_{act}} (\lambda_i)^r} \tag{2}$$

where λ_i represents the attentional tuning (weight) of dimension i and acts as a multiplier on i in calculating the activation. Initially, vector λ is set to $\vec{1}$ and evolves during the training phase according to equation (3) calculated at the end of every training iteration (i.e., after including a resource). r, which is set to 9.998, is an attentional focus parameter that accentuates the effect of λ_i: if $r = 0$. All dimensions are weighted equally.

If the activation value H_m^{act} of the most activated (i.e., winning) cluster is below a given threshold $\tau = .5$, a new cluster is created, representing the topics of the currently processed resource. At the end of an iteration, the tunings of vector λ are updated given by equation (3):

$$\Delta\lambda_i = \eta e^{-\lambda_i \mu_{im}} (1 - \lambda_i \mu_{im}) \tag{3}$$

where j indexes the winning cluster and the learning rate η is set to .096. In a final step, the position vector of the winning cluster, which holds a value for each of the n topics, is recalculated as described by equation (4):

$$\Delta H_m^{pos_i} = \eta (I^{pos_i} - H_m^{pos_i}) \tag{4}$$

The training phase is completed when steps (1) to (4) are subsequently processed for every resource in a user's training set. For each user, this results in a particular vector of attentional tunings λ and a set of j cluster vectors H_j. More formally, the training procedure of our approach is given by Algorithm 1.

Function	Symbol	Value
Attentional focus	r	9.998
Learning rate	η	.096
Threshold	τ	.5

Table 2: SUSTAIN's best fitting parameters for unsupervised learning as suggested in Love *et al.*, 2004.

Algorithm 1 Training procedure per user

```
Initialize a set of cluster H = ∅
Initialize a vector λ with λ_i = 1
for every resource topic vector I do
    for every cluster H_j ∈ H do
        Calculate μ_j
        Calculate H_j^act
    end for
    Identify H_m with max H_m^act
    if H_m^act <= τ then
        H_m ← I
        H ← H ∪ {H_m}
    end if
    λ ← λ + Δλ
    H_m ← H_m + ΔH_m
end for
return λ
return H
```

Testing. As described in Section 3.3, we determine the top 100 resources identified by CF_U as a candidate set C_u of potentially relevant resources for the target user u. Then, for each candidate c in C_u, we calculate H_m^{act} by applying equations (1) and (2). In order to compare the values resulting from SUSTAIN and CF_u, we normalize them such that $\sum_{c \in C_u} H_m^{act}(c) = 1$ and $\sum_{c \in C_u} CF_U(u,c) = 1$ holds. This leads to the normalized values $\overline{H_m^{act}(c)}$ and $\overline{CF_U(u,c)}$ that are finally put together as shown in equation (5) in order to determine the set of k recommended resources $RecRes(u)$ for user u:

$$RecRes(u) = \underset{c \in C_u}{\arg\overset{k}{\max}} (\underbrace{\alpha \underbrace{\overline{H_m^{act}(c)}}_{SUSTAIN} + (1-\alpha)\overline{CF_U(u,c)}}_{SUSTAIN+CF_U}) \tag{5}$$

where α can be used to inversely weigh the two components of our hybrid approach. For now, we set α to .5 in order to equally weight SUSTAIN and CF_U.

3.3 Technical Preliminaries

Our approach requires two steps of data preprocessing. First, the extraction of semantic topics to describe resources and second, the identification of candidate resources using CF. Candidate resources describe the user-specific set of Web resources that the algorithm considers recommending to a user.

Deriving semantic topics for resources. In order to derive semantic topics for the resources Griffiths *et al.*, 2007 of our social tagging datasets (see Section 4.1.1), we use Latent Dirichlet Allocation (LDA) Blei *et al.*, 2003. Categories or topics describing Web resources form the basis of our approach.

Since our datasets do not explicitly contain such properties for resources, we chose LDA to simulate an external categorization.

LDA is a probability model that helps find latent semantic topics for documents (i.e., resources). In the case of social tagging data, the model takes assigned tags of all resources as input and returns an identified topic distribution for each resource. We implemented LDA using the Java framework Mallet[2] with Gibbs sampling and $l = 2000$ iterations as suggested in the framework's documentation and related work (e.g., Krestel *et al.*, 2009). In order to reduce noise and to meaningfully limit the number of assigned topics, we set the number of latent topics Z to 500 (see also Kintsch and Mangalath, 2011) and only consider topics for a resource that show a minimum probability value of .01. The Latent Dirichlet Allocation can be formalized as follows:

$$P(t_i|d) = \sum_{j=1}^{Z} (P(t_i|z_i = j)P(z_i = j|d)) \tag{6}$$

Here $P(t_i|d)$ is the probability of the ith word for a document d and $P(t_i|z_i = j)$ is the probability of t_i within the topic z_i. $P(z_i = j|d)$ is the probability of using a word from topic z_i in the document.

Identifying candidate resources. Within the scope of this paper, the term *candidate resources* describes the set of resources that is considered when calculating most suitable items for a recommendation. To evaluate our approach, we use *User-based Collaborative Filtering (CF_U)* Schafer *et al.*, 2007 to identify 100 candidate resources per user. CF_U typically consists of two steps: first, the most similar users (the k nearest neighbors) for a target user are identified using a specific similarity measure. Second, resources of these neighbors are recommended that are new to the target user. This procedure is based on the idea that if two users had a similar taste in the past, they will probably share the same taste in the future and thus, will like the same resources Schafer *et al.*, 2007. We calculate the user similarities based on the binary user-resource matrix and the cosine-similarity measure (see Zheng and Li, 2011). In addition, we set the neighborhood size k to 20, as is suggested for CF_U in social tagging systems Gemmell *et al.*, 2009.

More formally, the prediction value $CF_U(u, i)$ for a target user u and a resource r is given by equation (7):

$$CF_U(u, r) = \sum_{v \in V_u} sim(u, v) \tag{7}$$

where $V_{u,r}$ is the set of most similar users of u that have bookmarked r. $sim(u, v)$ is the cosine similarity value between u and v.

Source code. Our approach as well as the baseline algorithms described in Section 4.1.4 (except for WRFM) and the evaluation method described in Section 4.1.2 are implemented in Java within our *TagRec* recommender benchmarking framework Kowald *et al.*, 2014, which is freely available via GitHub[3].

[2] http://mallet.cs.umass.edu/

[3] https://github.com/learning-layers/TagRec/

Dataset	Type	$\|P\|$	$\|U\|$	$\|R\|$	$\|T\|$	$\|P\|/\|U\|$
Bibsonomy	Full	400,983	5,488	346,444	103,503	73
	Sample	82,539	2,437	28,000	30,919	34
	Training	66,872	2,437	27,157	27171	27
	Test	15,667	839	11,762	12,034	19
CiteULike	Full	753,139	16,645	690,126	238,109	45
	Sample	105,333	7,182	42,320	46,060	15
	Training	86,698	7,182	40,005	41,119	12
	Test	18,635	2,466	14,272	16,332	8
Delicious	Full	104,799	1,867	69,223	40,897	56
	Sample	59,651	1,819	24,075	23,984	33
	Training	48,440	1,819	23,411	22,095	27
	Test	11,211	1,561	8,984	10,379	7

Table 3: Properties of the full datasets as well as the used dataset samples (including training and test set statistics) for BibSonomy, CiteULike and Delicious. Here, $|P|$ is the number of posts, $|U|$ is the number of users, $|R|$ is the number of resources and $|T|$ is the number of tags.

4 Experimental Setup

This section describes the methodology we selected to evaluate SUSTAIN based on recommender performance metrics and the SPEAR algorithm. It is structured in accordance with our four research questions.

4.1 Model Validation Based on Recommendation Accuracy (RQ1)

In this section, we describe datasets, method, metrics and baseline algorithms used in our recommender evaluation study.

4.1.1 Datasets

We used the social bookmark and publication sharing system BibSonomy[4] (2013-07-01), the citation sharing system CiteULike[5] (2013-03-10) and the social bookmarking system Delicious[6] (2011-05-01) to test our approach in three different settings that vary in their dataset sizes. To reduce computational effort, we randomly selected 20% of the CiteULike user profiles Gemmell *et al.*, 2009 (the other datasets were processed in full size). We did not use a p-core pruning approach to avoid a biased evaluation (see Kowald and Lex, 2015) but excluded all posts assigned to unique resources, i.e., resources that have only been bookmarked once (see Parra-Santander and Brusilovsky, 2010). The statistics of the full datasets, dataset samples we used (i.e., after the exclusion of posts assigned to unique resources), and training and test sets (see next section) are shown in Table 3.

[4] http://www.kde.cs.uni-kassel.de/bibsonomy/dumps/

[5] http://www.citeulike.org/faq/data.adp

[6] http://files.grouplens.org/datasets/hetrec2011/hetrec2011-delicious-2k.zip

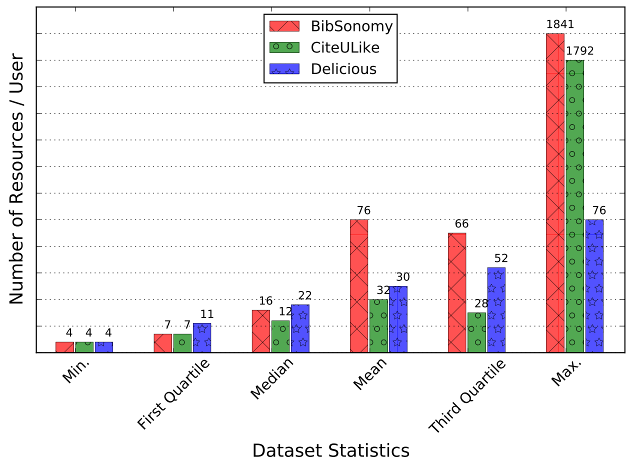


Figure 1: Resource statistics of the training datasets for BibSonomy, CiteULike and Delicious illustrating the number of resources users' have engaged with.

4.1.2 Evaluation Protocol

In order to evaluate our algorithm and to follow common practice in recommender systems research (e.g., Kowald and Lex, 2015; Huang *et al.*, 2014; Zheng and Li, 2011), we split our datasets into training and test sets. Therefore, we followed the method described in Lacic *et al.*, 2014 to retain the chronological order of the posts. Specifically, we used the 20% most recent posts of each user for testing and the rest for training the algorithms. The statistics of the training and test sets used can be found in Table 3. This evaluation protocol is in line with real-world scenarios, where user interactions in the past are used to try and predict future user interactions **campos2013time**

4.1.3 Evaluation Metrics

To finally determine the performance of our approach as well as of the baseline methods, we compared the top 20 recommended resources determined by each algorithm for a user with the relevant resources in the test set using a variety of well-known evaluation metrics Parra and Sahebi, 2013; Herlocker *et al.*, 2004 in recommender systems research. In particular, we took into account Normalized Discounted Cumulative Gain (nDCG@20), Mean Average Precision (MAP@20), Recall (R@20) and Precision (P@20). Moreover, we show the performance of the algorithms for different numbers of recommended resources ($k = 1 - 20$) by means of Precision/Recall plots.

4.1.4 Baseline Algorithms

We selected a set of well-known resource recommender baseline algorithms in order to determine the performance of our novel approach in relation to these approaches. Hence, we have not only chosen algorithms that are similar to our approach in terms of their processing steps (CF_U and CB_T) but also current state-of-the-art methods for personalized resource recommendations (CF_R and WRMF) along with a simple unpersonalized approach (MP).

Most Popular (MP). The simplest method we compare our algorithm to, is the *Most Popular (MP)* approach that ranks the resources by their total frequency in all posts Parra and Sahebi, 2013. In contrast to the other chosen baselines, the MP approach is non-personalized and thus recommends the same set of resources for any user.

User-based Collaborative Filtering ($\mathbf{CF}_U$). See Section 3.3 for a detailed description of the *User-based Collaborative Filtering (CF_U)* baseline.

Resource-based Collaborative Filtering ($\mathbf{CF}_R$). In contrast to CF_U, *Resource-based Collaborative Filtering (CF_R)* (also known as Item-based CF), identifies potentially interesting resources for a user by computing similarities between resources instead of similarities between users. Hence, this approach processes the resources a user has bookmarked in the past in order to find similar resources to recommend Sarwar *et al.*, 2001. As with CF_U, we calculated similarities based on the binary user-resource matrix using cosine similarity and focused on a resource-neighborhood size k of 20 Zheng and Li, 2011; Gemmell *et al.*, 2009.

Content-based Filtering using Topics ($\mathbf{CB}_T$). Content-based filtering (CB) methods recommend resources to users by comparing the resource content and the user profile Basilico and Hofmann, 2004. Hence, this approach does not need to calculate similarities between users or resources (as done in CF methods) but directly tries to map resources and users. We implemented this method in the form of *Content-based Filtering using Topics (CB_T)* since topics are the only content-based features available in our social tagging datasets (see Section 4.1.1). The similarity between the topic vector of a user and a resource has been calculated using the cosine similarity measure.

Weighted Regularized Matrix Factorization (WRMF). *WRMF* is a model-based recommender method for implicit data (e.g., posts) based on the state-of-the-art Matrix Factorization (MF) technique. MF factorizes the binary user-resource matrix into latent user- and resource-factors, which represent these entities, in a common space. This representation is used to map resources and users and thus, to find resources to be recommended for a specific user. WRMF defines this task as a regularized least-squares problem based on a weighting matrix, which differentiates between observed and unobserved activities in the data Hu *et al.*, 2008. The results for WRFM presented in Section 5 have been calculated using the MyMediaLite 3.10 framework[7] (2013-09-23) with $k = 500$ latent factors, $l = 100$ iterations and a regularization value $\lambda = .001$.

4.2 Parameter Investigation to Understand the Dynamics of SUSTAIN (RQ2)

This section describes the setup and rationale of a parameter investigation that we conducted to tackle our second research question: Which aspects of the SUSTAIN algorithm contribute to the improved performance? In an initial study that has been reported in Seitlinger *et al.*, 2015 and in the comparative studies that will be presented in Section 5.1, we used the best fitting

[7]http://www.mymedialite.net/

parameters for unsupervised learning as suggested in Love *et al.*, 2004. This parameter set results from extensive parameter studies, applying a genetic algorithm to fine tune SUSTAIN for a variety of learning data and learning problems. The paper concluded that SUSTAIN does not show great sensitivity to single parameter values but rather succeeds due to its principles.

However, our learning task differs from the presented studies in multiple aspects, for instance in the amount of training data, in the application domain and most significantly in the format of the input stimuli. In Love *et al.*, 2004 the input stimuli are characterized by multiple dimensions of input units. For instance a dimension (e.g., color) with 3 input units (e.g., green, yellow, blue) could have an input vector of [0,0,1]. In our case an input stimulus consists of 500 dimensions (i.e., LDA topics) of binary input units. Furthermore, data that is typically available in non-commercial learning environments, and equally, the social bookmarking datasets we use in our study, are sparse and premature. With this in mind, we conducted a short parameter study to better understand the underlying dynamics of our adapted approach and to investigate possible inconsistencies. The priority was to look into SUSTAIN's parameters r, η in a first step, but secondly, also to find the best fitting α value to optimally weight the impact of CF_u.

The results in Section 5.1 were generated using the default SUSTAIN parameters stated in Love *et al.*, 2004, to avoid tuning our approach and thus favoring it over the baseline algorithms. Additionally, the parameter study was performed on separate holdout sets extracted from the training data (using the same method as described in Section 4.1.2) in order to prevent a biased study conducted on the test data.

SUSTAIN. First, we determined plausible ranges for r and η, and defined sequential steps within these ranges. Additionally, the simulation includes the originally suggested values as presented in Table 2.

For r, which strengthens the impact of input dimensions by potentiating λ_i (see equation (2)), we start with $r = 1$ as a lower bound. This leads to a simulation with plain λ values. From there, we continue linearly with $r = r + 2$ for $r <= 21$. As λ shows rather small values, with a great percentage varying from 1.0 to 1.3, a relatively high value of r seems to be reasonable.

For the learning rate η, we set the simulation span such that $\eta_{min} > \frac{1}{N_{\max}}$ where $N_{\max}$ is the maximal amount of training resources per user. Thus, the learning rate η is set to 7.5 E-4 on the lower bound, while 1 was chosen as an upper bound. In between those bounds, three learning rates per decimal power were tested. As the median values for resources per user in our training sets are 12, 16 and 22 (see Figure 1), we expect the optimal learning rate to be fairly high.

As described in the original study setup, we initially simplify the parameter study by treating $\tau = 0.5$ as a fixed value. τ is the threshold responsible for whether a new cluster is formed or not and may range from 0 to 1.

When interpreting the first set of plots, additional questions appeared, such as, to what extent the training datasets and the topic distribution of their users may shift the optimal amount of clusters. To this end, we looked into the distribution of clusters and resources per user and dataset that were calculated with the recommended parameter setting outlined in Table 2. Finally, we investigated the performance development of SUSTAIN with different learning rates when varying τ within its range of 0 and 1, monitoring steps of .1. Considering insights from the first parameter setting, we fixed r to 9, and the learning rate to a range from .01 to 1.

Weighting CF$_U$. For α, which is the only parameter that is not part of SUSTAIN, but inversely weights the impact of the SUSTAIN and CF_u components (see equation 5), we examine α values between .1 and .9.

4.3 Comparing the Computational Efficiency of Discussed Algorithms (RQ3)

In order to answer *RQ3*, we determined the computational complexity of our discussed recommender algorithms using $\mathcal{O}$-notations. We distinguished between offline components of the algorithms, which can be calculated without any knowledge of the target user, and online components, which need to be calculated for the target user on-the-fly. In order to validate our complexity analysis, we also measured the complete runtime (i.e., training + testing time) of the algorithms. We conducted the runtime measurement on an IBM System x3550 server with two 2.0 GHz six-core Intel Xeon E5-2620 processors and 128 GB of RAM using Ubuntu 12.04.2 and Java 1.8.

4.4 Relation between SUSTAIN attentional entropy values and SPEAR scores (RQ4)

One of the important factors when considering user behavior in social bookmarking systems is the level of the user's expertise. Expert users tend to provide high quality tags that describe a resource in a more useful way Lorince *et al.*, 2014; Lorince *et al.*, 2015, and they also tend to discover and tag high quality resources earlier, bringing them to the attention of other users in the community Noll *et al.*, 2009.

To calculate user's expertise levels, literature provides a very well established algorithm known as SPEAR - *SPamming-resistant Expertise Analysis and Ranking* Noll *et al.*, 2009; Yeung *et al.*, 2011, which is based on the HITS (Hypertext Induced Topic Search) algorithm. The authors determine the level of the user's expertise based on two principles: (1) mutual reinforcement between user expertise and resource quality and (2) experts are discoverers, curious users who tend to identify high quality resources before other users (followers). This indicates that expert users are the first to collect many high quality resources and, in turn, high quality resources are tagged by users showing high expertise levels.

Expertise scores. Based on the work of Noll *et al.*, 2009, we calculated SPEAR expertise scores for users and resources in our datasets described in Table 3.

For M users and N resources we define a set of activities: $activity = (user, resource, tag, timestamp)$, which describes at which timestamp a user has tagged a resource. *User expertise scores* and *resource quality scores* vectors are defined as $\vec{E} = (e_1, e_2, ..., e_M)$ and $\vec{Q} = (q_1, q_2, ..., q_N)$, respectively. Initially, the values of these two vectors are set to 1.0. As has already been mentioned, SPEAR implements the mutual reinforcement principle, which indicates that the expertise score of a user depends on the quality scores of the tagged resources and the quality score of a resource depends on the expertise score of the users who tagged that resource.

Thus, an adjacency matrix A of size $M \times N$ is constructed next, containing one of the following values: (1) $l+1$ if user i has tagged resource j before l other users or (2) 0 if user i has not tagged resource j. Assigning adjacency matrix values this way also enables the implementation of the discoverer/follower principle, i.e., if user i was the first that tagged resource j, then the corresponding value A_{ij} would be the total number of users that tagged j, and if user i tagged the resource j most recently, $A_{ij} = 1$. We applied the *credit score function* suggested by Noll *et al.*, 2009 to A, so that $A_{ij} = \sqrt{A_{ij}}$. Finally, user expert scores and resource quality scores are calculated through an iterative process based on equations 8 and 9:

$$\vec{E} = \vec{Q} \times A^T \tag{8}$$

$$\vec{Q} = \vec{E} \times A \tag{9}$$

To relate SUSTAIN attentional focus values to the SPEAR scores, we only considered the expertise score vector. The calculated expertise scores for the highest ranked users in our datasets vary between .01 in Delicious and CiteULike, and .03 in BibSonomy. The low values are due to data sparsity, i.e., many resources were only tagged by a single user.

Attentional entropy values. The expertise scores were correlated with the entropy of the users' attentional tunings derived from SUSTAIN. Thus, SUSTAIN gives us for each of the Z topics a user-specific attentional tuning, which can be combined using the Shannon entropy. We calculated the entropy of the distribution of users' attentional tunings applying the following equation:

$$S = -\sum_{i=1}^{Z} p(x_i) \cdot log(p(x_i)) \tag{10}$$

where $p(x_i)$ is the probability that the attentional tuning value x_i occurs. In this respect, a user with a high attentional entropy is interested in a rich set of topics and thus, can be seen as curious user (discoverer), which should also correlate with a high SPEAR score if our hypothesis is correct. The results of this correlation are presented in Section 5.4.

5 Results and Discussion

In this section, we present and discuss the results of our evaluation aligned to our four research questions presented in Section 1.

5.1 *Model Validation Based on Recommendation Accuracy (RQ1)*

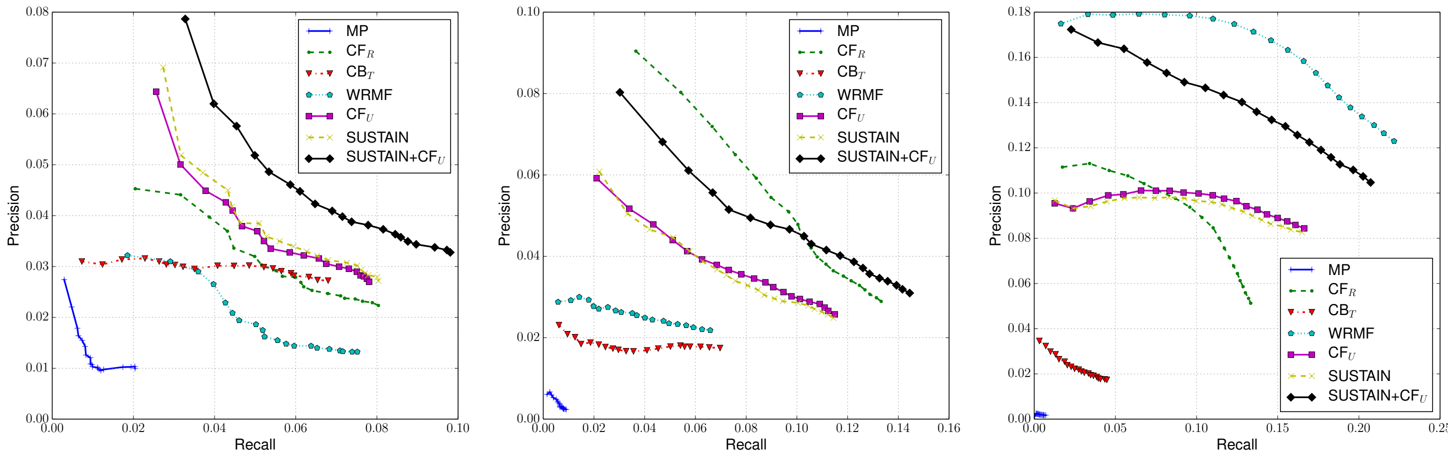
In order to tackle our first research question, we compared our approach to a wide set of state-of-the-art resource recommender algorithms. The results in Figure 2 and Table 4 reveal that the simplest baseline algorithm, i.e., the unpersonalized MP approach, achieves very low estimates of accuracy. Across all datasets, the other baseline algorithms reach larger estimates and therefore seem to be successful in explaining a substantial amount of variance in user behavior. Figure 2 reveals the evolution of accuracy values with a growing number of recommendations (i.e., one to 20). Note that recall (per definition) increases with the number of recommended items. Finally, Table 2 presents the results achieved with 20 recommended items.

Our evaluation results indicate that our SUSTAIN+CF$_U$ approach outperforms CF$_U$ and SUSTAIN in all settings. For instance, in the Precision/Recall plots in Figure 2, we can see that there is no overlap between corresponding curves, with SUSTAIN+CF$_U$ always reaching higher values than SUSTAIN and CF$_U$ separately. Moreover, results of the ranking-dependent metric nDCG@20 in Table 4 particularly show a remarkably better value for SUSTAIN+CF$_U$ than CF$_U$, demonstrating that our approach, through its improved personalization, can be used to successfully re-rank candidate resources identified by CF$_U$. We attribute this to the fact that the user-based CF cannot rank the resources of a neighbor. This possibly leads to a list of recommendations that contains only the resources of a user's nearest neighbor with no ranking. With our hybrid approach, we tackle this issue. Thus, we can answer our first research question positively. Interestingly, the performance of the algorithms varies greatly across BibSonomy, CiteULike and Delicious. Regarding nDCG@20 a different algorithm wins in each of the three datasets. For instance, in the case of CiteULike, the best results are achieved with CF$_R$. We can explain this by studying the average topic similarity per user. In CiteULike (18.9%), it is much higher than in BibSonomy (7.7%) and Delicious (4.5%), indicating a more thematically consistent resource search behavior. Note that we define the average topic similarity per user as the average pairwise cosine similarity between the topic vectors of all resources a user has bookmarked. This is averaged over all users. The higher consistency positively impacts predictions that are based on resources collected in the past, such as CF$_R$-based predictions.

In the case of Delicious, the users in the dataset are chosen using a mutual-fan crawling strategy (see Cantador *et al.*, 2011) and thus, are not independent from each other. This is conducive to methods that capture relations between users with common resources by means of high-dimensional arrays, such as WRMF. However, compared to the other algorithms, especially to CF$_R$ and WRMF, SUSTAIN+CF$_U$ demonstrates relatively robust estimates (especially in terms of Precision and Recall) as our approach provides fairly good results in all three datasets. SUSTAIN+CF$_U$ shows particularly good results on BibSonomy, where it outperforms all baseline algorithms.

5.2 *Parameter Investigation to Understand the Dynamics of SUSTAIN (RQ2)*

This section presents and discusses insights from a parameter study that we conducted to address our second research question. Specifically, we aim to identify the core aspects of the SUSTAIN model that have the greatest effects on the performance of our model on our datasets. We were also able to verify the impact of user traces and detect and explain particularities



(a) BibSonomy (b) CiteULike (c) Delicious

Figure 2: Precision/Recall plots for BibSonomy, CiteULike and Delicious showing the recommender accuracy of our approach SUSTAIN+CF$_U$ in comparison to the baseline methods for k = 1 - 20 recommended resources. The results indicate that SUSTAIN+CF$_U$ provides higher Precision and Recall estimates than CF$_U$ (RQ1) and SUSTAIN for each k and in all three datasets. In the case of BibSonomy, SUSTAIN+CF$_U$ even outperforms all baseline methods, including WRMF.

Dataset	Metric	MP	CF$_R$	CB$_T$	WRMF	CF$_U$	SUSTAIN	SUSTAIN+CF$_U$
BibSonomy	nDCG@20	.0142	.0569	.0401	.0491	.0594	.0628	**.0739**
	MAP@20	.0057	.0425	.0211	.0357	.0429	.0436	**.0543**
	R@20	.0204	.0803	.0679	.0751	.0780	.0902	**.0981**
	P@20	.0099	.0223	.0272	.0132	.0269	.0295	**.0328**
CiteULike	nDCG@20	.0064	**.1006**	.0376	.0411	.0753	.0828	.0977
	MAP@20	.0031	**.0699**	.0170	.0210	.0468	.0503	.0634
	R@20	.0090	.1332	.0697	.0658	.1149	.1344	**.1445**
	P@20	.0023	.0289	.0174	.0218	.0257	.0279	**.0310**
Delicious	nDCG@20	.0038	.1148	.0335	**.1951**	.13	.131	.1799
	MAP@20	.0011	.0907	.0134	**.1576**	.0743	.0936	.1275
	R@20	.0071	.1333	.0447	**.2216**	.1599	.1649	.2072
	P@20	.0017	.0512	.0173	**.1229**	.0785	.0826	.1047

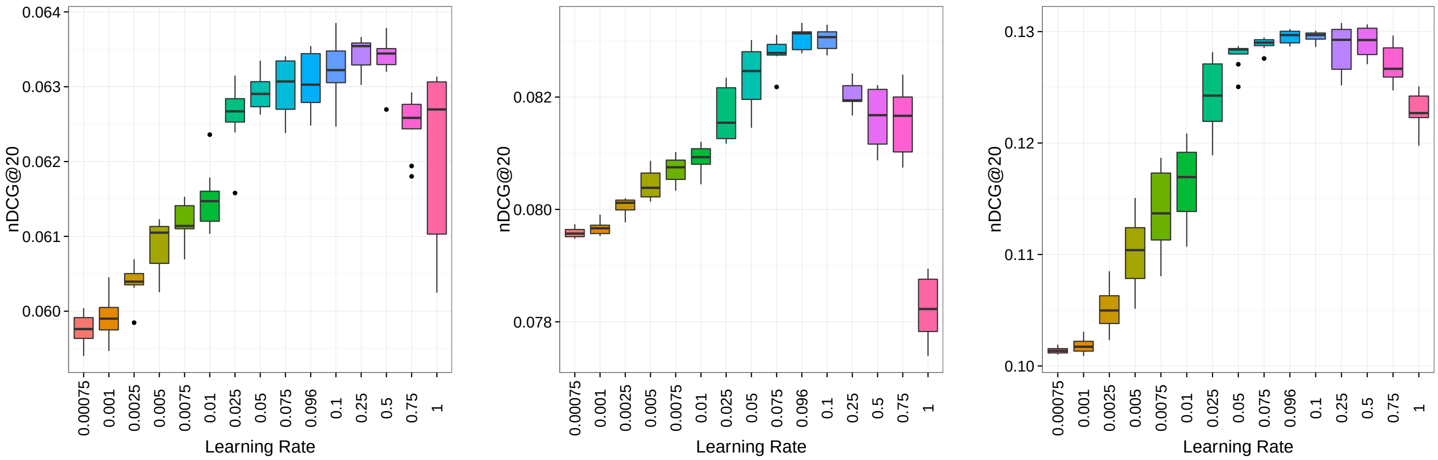
Table 4: nDCG@20, MAP@20, R@20 and P@20 estimates for BibSonomy, CiteULike and Delicious in relation to *RQ1*. The results indicate that our proposed approach SUSTAIN+CF$_U$ outperform CF$_U$ (RQ1) and SUSTAIN in all settings. Furthermore, SUSTAIN+CF$_U$ is able to compete with the computationally more expensive WRMF approach. *Note*: highest accuracy values per dataset over all algorithms are highlighted in bold.

of our three datasets.

SUSTAIN. In Figure 3, results of the first simulation are illustrated. In this setup, we treated $\tau = .5$ as a fixed variable, similar to the original parameter study (see Love *et al.*, 2004), and solely varied learning rate η and attentional focus parameter r within a parameter range, as explained in 4.2. The plots show SUSTAIN's performance on the y-axis given as nDCG@20 values and the learning rates on the x-axis. The shape of the box plot indicates the distribution of the performance values caused by a set of different r's, which means, the higher the box plot, the greater the influence of r. Even though some variation can be observed, for the best performing η, the influence of r seems to be marginal in this setting.

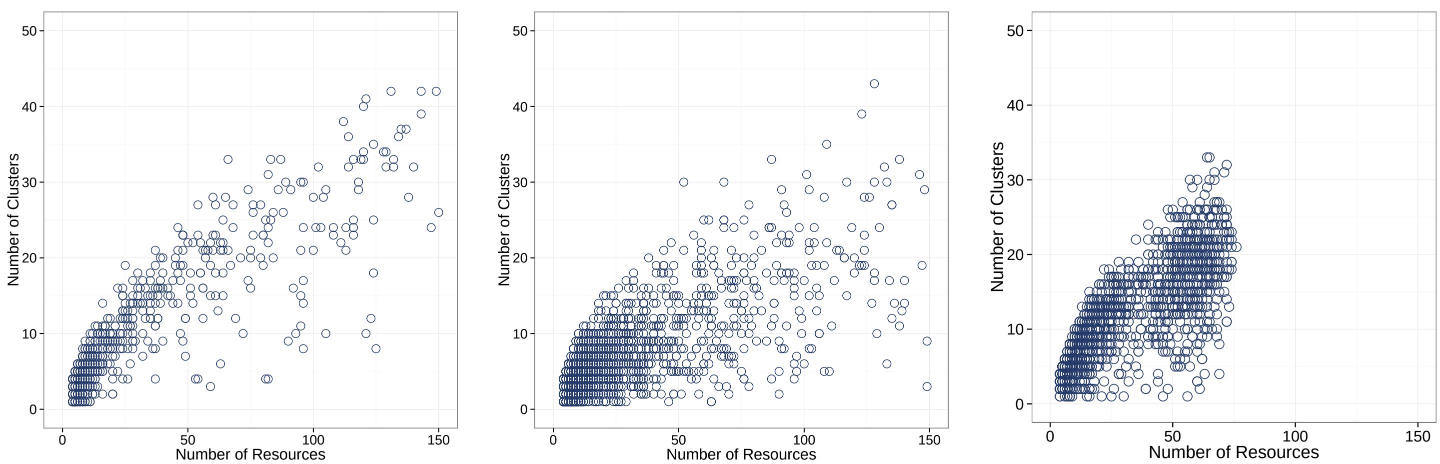
In our case, the learning rate tends to be the most important factor to consider. We identify two scenarios: (i) if the learning rate is too small, a user's behavior cannot be tracked fast enough and (ii) if the learning rate is too high, the algorithm forgets previous resources too quickly. The first scenario is likely to apply to users with few resources, whereas, the second scenario is potentially problematic for users with many resources. As illustrated in Figure 1, our training datasets show a large variation in the distribution of training resources per user, within and between datasets. However, the common trend shows that about 50 percent of users have less than 25 resources available for training the algorithm. In line with these observations, SUSTAIN's performance peaks at an intermediate value around .1. In our case, this particularly proves that the browsing history of a user needs to be taken into account for optimal predictions, and not just the most recent item.

Among the three datasets, the learning rate has the greatest impact on Delicious (note the ranges of nDCG@20). An ex-



(a) BibSonomy (b) CiteULike (c) Delicious

Figure 3: Recommendation effectiveness influenced by learning rate and attentional focus parameter.



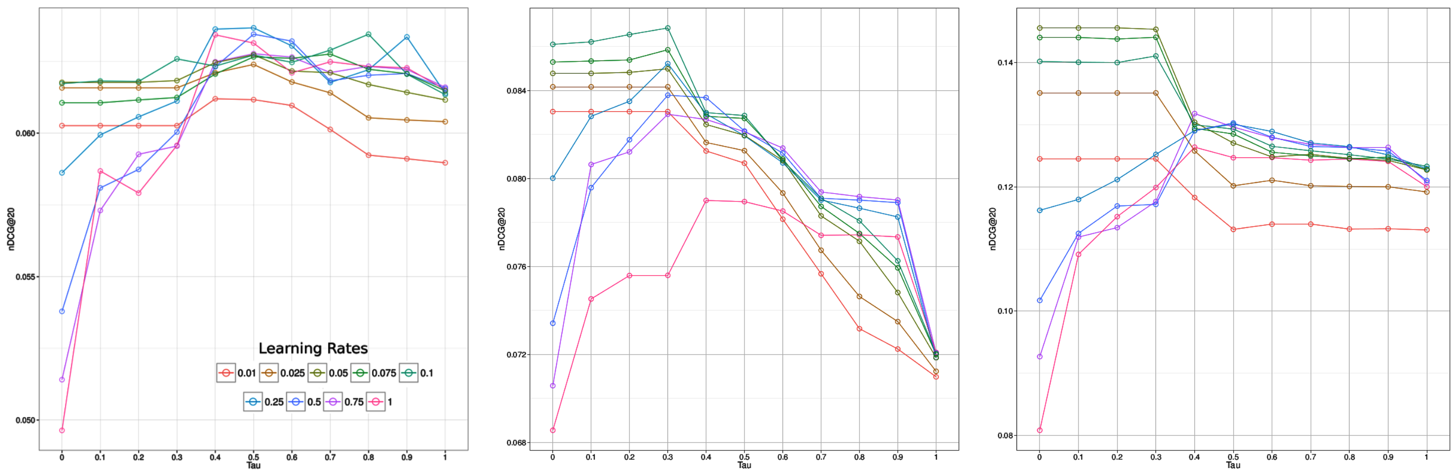
(a) BibSonomy (b) CiteULike (c) Delicious

Figure 4: Snapshot of the distribution of the clusters and resources appearing with parameters recommended in the literature. Please note that the range of the plots is restricted in order to improve readability. BibSonomy and CiteULike have both about 100 users with more than 150 resources, which are not depicted in this plot.

planation of this behavior can be derived from Figure 4, which presents a snapshot of the cluster resource distribution per user and dataset. In the case of Delicious, the overall trend shows that a new cluster is created for each second or third resource. Since only the cluster with the highest activation learns in our approach, the strong influence of the learning rate, or in other words the need for faster learning per cluster, seems reasonable.

Given that a new cluster is created whenever a new resource is added that cannot be integrated into any of the existing clusters due to a lack of similarities, the cluster distribution also presents the level of topic overlap among the resources of a typical user. For instance, when calculating basic statistics for the resource to cluster ratio of Delicious, we find that the average value is 2.8 resources per cluster in comparison to 4.2 resources per cluster for CiteULike, for instance. This indicates a large overlap between resources of users in CiteULike. Furthermore, we can observe a decreasing trend of the resource-to-cluster ratio as the number of resources grows. Furthermore, the plot for CiteULike highlights the rather weak relationship between clusters and resources, which signifies a great variety among users.

These results made us question how the number of clusters impacts the performance, and whether a dynamic clustering approach is even necessary for our task. In particular, we wanted to investigate if a different τ could lead to a better performance with the training sets. Thus, in a second simulation, we observed the performance development when varying τ and η. This time $r = 9$ was treated as a fixed variable, due to the marginal difference it caused in our first study. Line charts in Figure 5 present our findings. Regarding the optimal number of clusters, we can see that the three datasets vary greatly in their behavior. Delicious performs best with only one cluster



(a) BibSonomy (b) CiteULike (c) Delicious

Figure 5: Recommendation effectiveness influenced by learning rate and the number of clusters.

Algorithm	Component	Type	Complexity	Reference
MP	Complete	Offline	$\mathcal{O}(\|P\|)$	Parra & Sabhebi Parra and Sahebi, 2013
CB_T	Similarity	Offline	$\mathcal{O}(\|U\| \cdot \|R\| \cdot Z)$	
	Recommendation	Online	$\mathcal{O}(\|U\| \cdot \|S_u\|)$	
	Complete	Online	$\mathcal{O}(\|U\| \cdot \|S_u\|)$	Basilico & Hofmann Basilico and Hofmann, 2004
CF_U	Similarity	Offline	$\mathcal{O}(\|U\|^2)$	
	Recommendation	Online	$\mathcal{O}(\|U\| \cdot \|V_u\| \cdot \|R_v\|)$	
	Complete	Online	$\mathcal{O}(\|U\| \cdot \|V_u\| \cdot \|R_v\|)$	Schafer et al. Schafer *et al.*, 2007
CF_R	Similarity	Offline	$\mathcal{O}(\|R\|^2)$	
	Recommendation	Online	$\mathcal{O}(\|U\| \cdot \|R_u\| \cdot \|S_r\|)$	
	Complete	Online	$\mathcal{O}(\|U\| \cdot \|R_u\| \cdot \|S_r\|)$	Sarwar et al. Sarwar *et al.*, 2001
SUSTAIN /	Topic Extraction	Offline	$\mathcal{O}(\|R\| \cdot \|T\| \cdot Z)$	Blei et al. Blei *et al.*, 2003
SUSTAIN+CF_U	Candidates	Online	$\mathcal{O}(\|U\| \cdot \|V_u\| \cdot \|R_v\|)$	Schafer et al. Schafer *et al.*, 2007
	SUSTAIN Training	Online	$\mathcal{O}(\|U\| \cdot \|R_u\| \cdot Z)$	
	SUSTAIN Testing	Online	$\mathcal{O}(\|U\| \cdot \|C_u\| \cdot Z)$	
	Complete	Online	$\mathcal{O}(\|U\| \cdot (\|R_u\| + \|C_u\|) \cdot Z)$	Love et al. Love *et al.*, 2004
WRMF	Complete	Online	$\mathcal{O}(\|U\| \cdot \|R\| \cdot k^2 \cdot l)$	Ning et al. Ning and Karypis, 2011

Table 5: Computational complexity of the algorithms showing that our SUSTAIN+CF_U approach provides a lower complexity than WRMF. We distinguish between *offline* (i.e., can be calculated without any knowledge of the target user) and *online* complexity (i.e., can only be calculated at runtime) components.

(i.e., $\tau = 0$), CiteULike and BibSonomy show better results with $\tau = .3$ and $\tau = .5$, respectively.

Delicious is the dataset most sensitive to τ (note the ranges of nDCG@20). Again, we think this is due to the high variation of topics, which leads to overfitting when too many clusters are formed. BibSonomy exhibits a larger topic overlap than Delicious. At the same time, in the case of Bibsonomy, we are provided with a much larger amount of training data per user than is the case with Delicious and CiteULike. Figure 1 for instance shows that 25 percent of users have between 66 and 1841 resources available for training. CiteULike differs due to its small amount of training data per user. Note the comparably low values for median and third quartile. This results in an optimal number of clusters between one and seven with the mean = 1.05. Thus, results clearly suggest that the optimal number of clusters varies with the properties of the training data. We conclude that this value relates to the available number of training samples and the topic density.

Weighting CF_U. We completed a simulation varying α from 0 to 1 to find the best fit for the weighting of CF_U to SUSTAIN (see 5). Results identified $\alpha = .65$ as the best fitting value for all datasets. Moreover, all values in the range of .3 to .8 perform close to optimal.

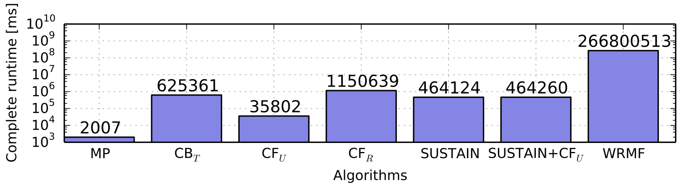


Figure 6: Complete runtime (i.e., training + testing time) of the algorithms in milliseconds (log scale) for the Delicious dataset. The plot verifies our findings regarding the computational complexities presented in Table 5 since our SUSTAIN-based approach provides a much lower complete runtime than WRMF. *Please note* that the other datasets provided similar results.

5.3 Comparing the Computational Efficiency of Discussed Algorithms (RQ3)

In this section, we investigate our third research question, considering the extent to which recommendations can be calculated in a computationally efficient way using SUSTAIN+CF$_u$ in comparison to other state-of-the-art algorithms like WRMF. The computational complexity of the approaches is shown in Table 5. In order to validate our complexity analysis, we also present the complete runtime (i.e., training + testing time) of the algorithms for the Delicious dataset in Figure 6 (the other datasets provided similar results). We discuss our findings for each algorithm as follows:

MP. The unpersonalized MostPopular approach has the lowest complexity. It has to analyze all posts in the dataset only once in order to calculate the overall frequencies.

CF$_U$. User-based Collaborative Filtering consists of an offline and an online component. The offline component calculates similarities between all users, whereas the online component analyzes the resources R_v of the most similar users (i.e., the neighbors V_u) of user u to calculate recommendations. Thus, the complete computational complexity only depends on the online component.

CF$_R$. Resource-based Collaborative Filtering works much like CF$_U$. It needs to first calculate similarities between all resources offline and then calculate recommendations online. In the online step, CF$_R$ analyzes the most similar resources S_r for each resource r in the set of the resources R_u of user u. Since our datasets' $|R|$ and $|R_u|$ are larger than $|U|$ and $|V_u|$ (20 in our case) respectively, CF$_R$ also has a higher complexity than CF$_U$.

CB$_T$. The Content-based Filtering using Topics approach mainly consists of the offline similarity calculation between users and resources, which is highly dependent on the number of topics Z (i.e., 500 in our case). For the online recommendation step, only the most similar resources S_u for a user u have to be analyzed, which is computationally efficient.

SUSTAIN+CF$_U$. Our hybrid SUSTAIN+CF$_U$ approach consists of a computationally expensive topic extraction step that is based on LDA. The complexity of LDA depends on the number of tags $|T|$, the number of resources $|R|$ and the number of topics Z. Furthermore, SUSTAIN+CF$_U$ requires an online recommendation calculation step, where candidate resources are identified and the SUSTAIN model is trained and tested. The identification of candidate resources is performed by CF$_U$ and the training of the SUSTAIN model is completed for all resources R_u of user u based on the topic space of size Z. The testing (or prediction) step is carried out for each candidate resource in the set of candidates C_u for a user u. Taken together, the computational complexity of our approach is given by $\mathcal{O}(|U| \cdot (|R_u| + |C_u|) \cdot Z)$ which is asymptotically comparable to CF$_R$. The same holds for the pure SUSTAIN approach as the candidate set needs to be calculated as well.

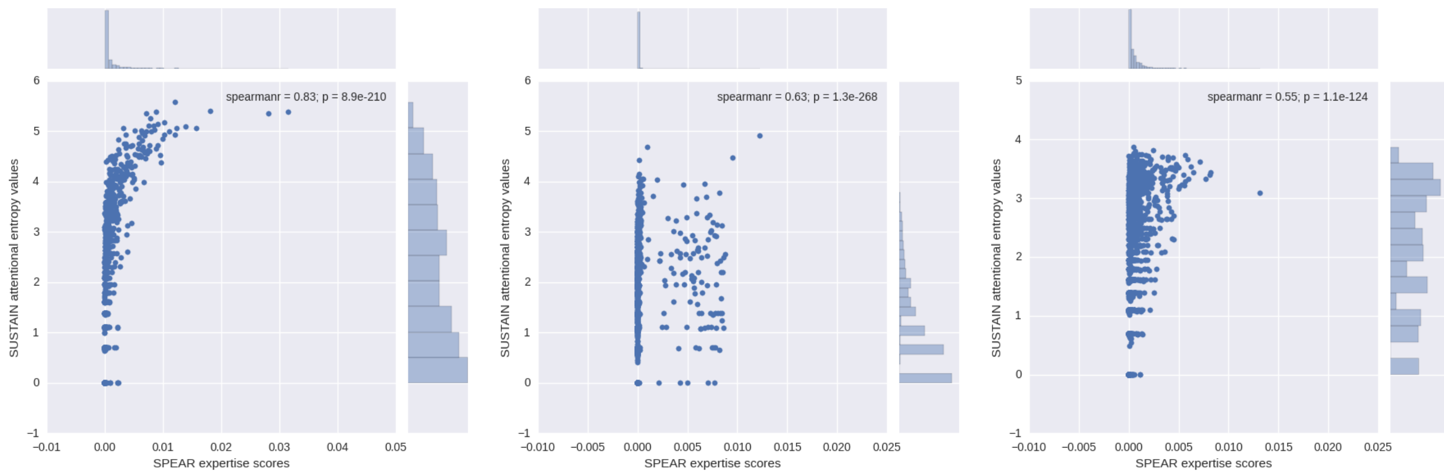
WRMF. The computationally most complex algorithm used in our study is the matrix factorization based WRMF approach. For each user u in U, WRMF needs to analyze all resources R depending on the squared factor dimension k (i.e., 500 in our case) and the number of iterations l (i.e., 100 in this paper). Since $|R|$ is far larger than $|R_u| + |C_u|$ and k^2 is the squared value of Z, it is obvious that our SUSTAIN+CF$_U$ approach is computationally much more efficient than WRMF. Additionally, WRMF is an iterative approach, which further increases its complexity by this factor.

Overall, our analysis shows the computationally efficiency of our approach compared to other state-of-the-art algorithms. This is further validated by the overall runtime results for the Delicious dataset shown in Figure 6. Hence, we can also answer our third research question positively.

5.4 Relation between SUSTAIN attentional entropy values and SPEAR scores (RQ4)

This section addresses our fourth research question (see Section 1) that inquires whether users' attentional entropies, determined by SUSTAIN, correlate with users' expertise scores identified by the SPEAR algorithm. To this end, we followed the procedure described in Section 4.4 to compare SUSTAIN's attentional entropy values with SPEAR's expertise scores four our three datasets. Results of this correlation study are presented in Figure 7.

Again, the plots show clear differences between the three datasets. Although we reach high Spearman rank correlation values in all three settings there is a considerable variation between Delicious (.55), CiteULike (.62) and BibSonomy (.83). This is in line with results presented in Sections 5.1 and 5.2, where we discuss recommender accuracy and SUSTAIN's model dynamics. In all experiments, we find that SUSTAIN+CF$_U$ performs best on BibSonomy and worst on Delicious when compared to baseline algorithms. In Figure 7, we can observe power-law like distributions for the SPEAR expertise scores in all three datasets, whereas, the distributions of SUSTAIN attentional entropy values vary strongly. The Delicious dataset shows an almost random distribution. Therefore, we presume that these findings are closely related to how well SUSTAIN and its parameter settings suit the properties of a specific dataset. However, the overall high correlation suggests that users, who reach high SPEAR expertise scores and can thus be identified as discoverers, also reach a high SUSTAIN attentional entropy value. This corroborates our hypothesis that attentional entropy values, and thus a user's attentional focus, correlate with a user's curiosity. This also provides a positive answer to the last research question in this work.



(a) BibSonomy (b) CiteULike (c) Delicious

Figure 7: Relation between SUSTAIN attentional entropy values and SPEAR's expertise scores for BibSonomy, CiteULike and Delicious (RQ4). Each plot illustrates the correlation between these values in the main panel and the data distributions in the upper and right plots. We observe Spearman Rank Correlation values between .55 for Delicious and .83 for BibSonomy, which indicates that users with a high attentional entropy value also receive a high expertise score.

6 Conclusions and Future Work

In this work, we investigated a model of human category learning, SUSTAIN Love *et al.*, 2004, which is applied to mimic a user's attentional focus and interpretation and its applicability to the recommender domain. Using offline studies on three social bookmarking datasets (BibSonomy, CiteULike and Delicious), we demonstrated its potential to personalize and improve user-based CF predictions. We attribute this improvement to the cognitive plausibility of SUSTAIN. The dynamically created user model allows for a more flexible and thorough representation of a user's decision making on a given set of resources: Reconstructing the user history in the form of an iteratively trained model with history-specific patterns of attentional tunings and clusters does more justice to a user's individuality than a CF-based representation of user-resource relations. Deepening our investigations, we show that both aspects, i.e., memorization of a user's history as well as clustering, contribute to the algorithm's performance. Our parameter study revealed that restricting cluster growth can prevent overfitting in sparse data environments. Additionally, we observed that our hybrid SUSTAIN+CF_U model is more robust in terms of accuracy estimates and less complex in terms of computational complexity than the Matrix Factorization-based approach WRMF.

Finally, we utilized the SPEAR algorithm to identify curious users. In SPEAR, curiosity is defined as a discoverer behavior (i.e., curious users tend to be faster at finding high quality resources). We connected the Spear score for the users in our dataset with their SUSTAIN-specific attentional entropy values and we found that a user's attentional focus indeed correlates with their curiosity. The highest correlation is achieved with the BibSonomy dataset, for which the SUSTAIN approach is also most effective.

We conclude that our attempt to keep the translation from theory into technology as direct as possible holds advantages for both technical and conceptual aspects of recommender systems' research. By applying computational models of human cognition, we can improve the performance of existing recommender mechanisms and at the same time gain a deeper understanding of fine-grained level dynamics in Social Information Systems.

Limitations and future work. We aim to improve and further evaluate our model in various ways. First, we are working on a variant that is independent of a resource candidate set obtained by CF_U and searches for user-specific recommendations only by means of the correspondingly trained SUSTAIN network. Second, we will make use of the network's sensitivity towards a user's mental state to realize a more dynamic recommendation logic. In particular, based on creative cognition research (e.g., Finke *et al.*, 1992) and in line with the findings of our evaluation studies, we assume a broader attentional focus (i.e., higher curiosity) to be associated with a stronger orientation toward novel or more diverse resources. If the algorithm integrates this association, depending on the user model, recommendations should become either more accurate or diverse.

With respect to recommender evaluation, the question arises whether SUSTAIN can realize its potential of providing additional benefits in cold-start and sparse data environments to improve real-life learning experiences. Online evaluations are less prone to error and misinterpretation, since they provide a direct user feedback in comparison to offline studies, where wrong predictions could be the result of a user's poor searching abilities.

A Appendix: Sustain Results for Different Numbers of LDA Topics

This section is an extension to *RQ1*, with Table 6 presenting simulation results for SUSTAIN in the Delicious dataset when applied to LDA topic sizes of 100, 500 and 1000. We see that the best results are reached when using 500 LDA topics, which verifies our choice to use this number of topics for our experiments. We observed the same results for BibSonomy and CiteULike. Furthermore, this table also provides SUSTAIN results for different numbers of recommended resources k.

Metric	Z	k=1	k=3	k=5	k=10	k=20
nDCG	100	.0036	.0089	.0128	.0202	.0374
	500	**.0232**	**.0471**	**.0649**	**.0958**	**.1310**
	1000	.0066	.0142	.0188	.0295	.0481
MAP	100	.0021	.0043	.0056	.0078	.0120
	500	**.0127**	**.0287**	**.0419**	**.0684**	**.0936**
	1000	.0043	.0082	.0099	.0138	.0189
Recall	100	.0021	.0071	.0119	.0234	.0589
	500	**.0127**	**.0347**	**.0556**	**.0999**	**.1658**
	1000	.0043	.0127	.0183	.0351	.0708
Precision	100	.0147	.0182	.0195	.0201	.0256
	500	**.0967**	**.0942**	**.0977**	**.0965**	**.0826**
	1000	.0224	.0231	.0239	.0275	.0317

Table 6: nDCG, MAP, R and P estimates for SUSTAIN in the Delicious dataset based on different numbers of LDA topics. The results show that 500 LDA topics lead to the best results. *Note*: highest accuracy values are highlighted in bold.

Acknowledgments

The authors are very grateful to Paul Seitlinger for providing the theoretical underpinning for this work, and for giving great advice on structure, study setups and interpretation. We would also like to thank Tobias Ley, Emanuel Lacic and Sebastian Dennerlein for their many valuable comments on this work. This work is supported by the Know-Center, the EU funded projects Learning Layers (Grant Agreement 318209) and weSPOT (Grant Agreement 318499) 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).

References

Adamopoulos, P. and A. Tuzhilin. 2014. "On Over-specialization and Concentration Bias of Recommendations". In: *Proc. of RecSys'14*. Foster City, Silicon Valley, California, USA: ACM. 153–160. ISBN: 978-1-4503-2668-1. DOI: 10.1145/2645710.2645752.

Adomavicius, G. and A. Tuzhilin. 2005. "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions". *IEEE Transactions on Knowledge and Data Engineering*. 17(6): 734–749.

Bar, A., L. Rokach, G. Shani, B. Shapira, and A. Schclar. 2013. "Improving simple collaborative filtering models using ensemble methods". In: *Multiple Classifier Systems*. Springer. 1–12.

Basilico, J. and T. Hofmann. 2004. "Unifying collaborative and content-based filtering". In: *Proc. of ICML'04*. ACM. 9.

Blei, D. M., A. Y. Ng, and M. I. Jordan. 2003. "Latent dirichlet allocation". *The Journal of machine Learning research*. 3: 993–1022.

Cantador, I., P. Brusilovsky, and T. Kuflik. 2011. "2nd Workshop on Information Heterogeneity and Fusion in Recommender Systems (HetRec 2011)". In: *Proc. of RecSys'11*. Chicago, IL, USA: ACM.

Chen, L., M. de Gemmis, A. Felfernig, P. Lops, F. Ricci, and G. Semeraro. 2013. "Human Decision Making and Recommender Systems". *ACM Trans. Interact. Intell. Syst.* 3(3): 17:1–17:7. ISSN: 2160-6455. DOI: 10.1145/2533670.2533675.

Coleho, B., C. Martins, and A. Almeida. 2010. "Web Intelligence in Tourism: User Modeling and Recommender System". In: *Proc. of WI-IAT '10*. Washington, DC, USA: IEEE Computer Society. 619–622. ISBN: 978-0-7695-4191-4. DOI: 10.1109/WI-IAT.2010.236.

Cremonesi, P., A. Donatacci, F. Garzotto, and R. Turrin. 2012a. "Decision-Making in Recommender Systems: The Role of User's Goals and Bounded Resources." In: *Proc. of Decisions@RecSys'12 Workshop*. Vol. 893. *CEUR Workshop Proceedings*. CEUR-WS.org. 1–7.

Cremonesi, P., F. Garzotto, and R. Turrin. 2012b. "Investigating the Persuasion Potential of Recommender Systems from a Quality Perspective: An Empirical Study". *ACM Trans. Interact. Intell. Syst.* 2(2): 11:1–11:41. ISSN: 2160-6455. DOI: 10.1145/2209310.2209314.

Dooms, S. 2013. "Dynamic Generation of Personalized Hybrid Recommender Systems". In: *Proc. of RecSys'13*. Hong Kong, China: ACM. 443–446. ISBN: 978-1-4503-2409-0. DOI: 10.1145/2507157.2508069.

Finke, R. A., T. B. Ward, and S. M. Smith. 1992. *Creative cognition: Theory, research, and applications*. MIT press Cambridge, MA.

Fu, W.-T. and W. Dong. 2012. "Collaborative Indexing and Knowledge Exploration: A Social Learning Model". *IEEE Intelligent Systems*. 27(1): 39–46.

Gemmell, J., T. Schimoler, M. Ramezani, L. Christiansen, and B. Mobasher. 2009. "Improving folkrank with item-based collaborative filtering". *Recommender Systems & the Social Web*.

Griffiths, T. L., M. Steyvers, J. B. Tenenbaum, *et al.* 2007. "Topics in semantic representation". *Psychological review*. 114(2): 211.

Herlocker, J. L., J. A. Konstan, L. G. Terveen, and J. T. Riedl. 2004. "Evaluating collaborative filtering recommender systems". *ACM Trans. Inf. Syst.* 22(1): 5–53. ISSN: 1046-8188. DOI: 10.1145/963770.963772.

Hu, Y., Y. Koren, and C. Volinsky. 2008. "Collaborative filtering for implicit feedback datasets". In: *Proc. of ICDM '08*. IEEE. 263–272.

Huang, C.-L., P.-H. Yeh, C.-W. Lin, and D.-C. Wu. 2014. "Utilizing user tag-based interests in recommender systems for social resource sharing websites". *Knowledge-Based Systems*. 56: 86–96.

Jawaheer, G., P. Weller, and P. Kostkova. 2014. "Modeling User Preferences in Recommender Systems: A Classification Framework for Explicit and Implicit User Feedback". *ACM Trans. Interact. Intell. Syst.* 4(2): 8:1–8:26. ISSN: 2160-6455. DOI: 10.1145/2512208.

Kantor, P. B., L. Rokach, F. Ricci, and B. Shapira. 2011. *Recommender systems handbook*. Springer.

Kintsch, W. and P. Mangalath. 2011. "The construction of meaning". *Topics in Cognitive Science*. 3(2): 346–370.

Kowald, D., E. Lacic, and C. Trattner. 2014. "TagRec: Towards A Standardized Tag Recommender Benchmarking Framework". In: *Proc. of HT'14*. New York, NY, USA: ACM.

Kowald, D. and E. Lex. 2015. "Evaluating Tag Recommender Algorithms in Real-World Folksonomies: A Comparative Study". In: *In Proc of RecSys '15*. Vienna, Austria: ACM. 265–268. ISBN: 978-1-4503-3692-5. DOI: 10.1145/2792838.2799664.

Kowald, D. and E. Lex. 2016. "The Influence of Frequency, Recency and Semantic Context on the Reuse of Tags in Social Tagging Systems". In: *Proc. of HT '16*.

Krestel, R., P. Fankhauser, and W. Nejdl. 2009. "Latent dirichlet allocation for tag recommendation". In: *Proc. of Recsys'09*. ACM. 61–68.

Lacic, E., D. Kowald, P. Seitlinger, C. Trattner, and D. Parra. 2014. "Recommending Items in Social Tagging Systems Using Tag and Time Information". In: *Proc. of the Social Personalization Workshop colocated with HT'14*.

Lamprecht, D., F. Geigl, T. Karas, S. Walk, D. Helic, and M. Strohmaier. 2015. "Improving Recommender System Navigability Through Diversification: A Case Study of IMDb". In: *Proc. of i-KNOW '15*. Graz, Austria: ACM. 21:1–21:8. ISBN: 978-1-4503-3721-2. DOI: 10.1145/2809563.2809603.

Law, J. 2009. "Actor network theory and material semiotics". *The new Blackwell companion to social theory*: 141–158.

Loewenstein, G. 1994. "The psychology of curiosity: A review and reinterpretation". *Psychological Bulletin*: 75–98.

Lorince, J., S. Zorowitz, J. Murdock, and P. M. Todd. 2014. ""Supertagger" Behavior in Building Folksonomies". In: *Proc. of WebSci '14*. Bloomington, Indiana, USA: ACM. 129–138. ISBN: 978-1-4503-2622-3. DOI: 10.1145/2615569.2615686.

Lorince, J., S. Zorowitz, J. Murdock, and P. M. Todd. 2015. "The Wisdom of the Few? "Supertaggers" in Collaborative Tagging Systems". *The Journal of Web Science*. 1(1): 16–32. DOI: 10.1561/106.00000002.

Love, B. C., D. L. Medin, and T. M. Gureckis. 2004. "SUSTAIN: a network model of category learning." *Psychological review*. 111(2): 309.

Marlin, B. M. 2004. "Modeling User Rating Profiles For Collaborative Filtering". In: *Advances in Neural Information Processing Systems 16*. Ed. by S. Thrun, L. K. Saul, and B. Schölkopf. MIT Press. 627–634.

Ning, X. and G. Karypis. 2011. "Slim: Sparse linear methods for top-n recommender systems". In: *Data Mining (ICDM), 2011 IEEE 11th International Conference on*. IEEE. 497–506.

Noll, M. G., C.-m. Au Yeung, N. Gibbins, C. Meinel, and N. Shadbolt. 2009. "Telling Experts from Spammers: Expertise Ranking in Folksonomies". In: *Proc. of SIGIR'09*. Boston, MA, USA: ACM. 612–619. ISBN: 978-1-60558-483-6. DOI: 10.1145/1571941.1572046.

Parra, D. and S. Sahebi. 2013. "Recommender Systems : Sources of Knowledge and Evaluation Metrics". In: *Advanced Techniques in Web Intelligence-2: Web User Browsing Behaviour and Preference Analysis*. Springer-Verlag. 149–175.

Parra-Santander, D. and P. Brusilovsky. 2010. "Improving collaborative filtering in social tagging systems for the recommendation of scientific articles". In: *Proc. of WIIAT'10*. Vol. 1. IEEE. 136–142.

Sarwar, B., G. Karypis, J. Konstan, and J. Riedl. 2001. "Item-based collaborative filtering recommendation algorithms". In: *Proc. of WWW'01*. ACM. 285–295.

Schafer, J. B., D. Frankowski, J. Herlocker, and S. Sen. 2007. "Collaborative filtering recommender systems". In: *The adaptive web*. Springer. 291–324.

Seitlinger, P., D. Kowald, S. Kopeinik, I. Hasani-Mavriqi, T. Ley, and E. Lex. 2015. "Attention Please! A Hybrid Resource Recommender Mimicking Attention-Interpretation Dynamics". In: *Proc. of WWW'15*. International World Wide Web Conferences Steering Committee. 339–345.

Shi, L. 2013. "Trading-off Among Accuracy, Similarity, Diversity, and Long-tail: A Graph-based Recommendation Approach". In: *Proc. of RecSys'13*. Hong Kong, China: ACM. 57–64. ISBN: 978-1-4503-2409-0. DOI: 10.1145/2507157.2507165.

Shi, Y., M. Larson, and A. Hanjalic. 2014. "Collaborative Filtering Beyond the User-Item Matrix: A Survey of the State of the Art and Future Challenges". *ACM Comput. Surv.* 47(1): 3:1–3:45. ISSN: 0360-0300. DOI: 10.1145/2556270.

Wang, C. and D. M. Blei. 2011. "Collaborative Topic Modeling for Recommending Scientific Articles". In: *Proc. of KDD'11*. San Diego, California, USA: ACM. 448–456. ISBN: 978-1-4503-0813-7. DOI: 10.1145/2020408.2020480.

Yeung, C.-m. A., M. G. Noll, N. Gibbins, C. Meinel, and N. Shadbolt. 2011. "SPEAR: SPAMMING-RESISTANT EXPERTISE ANALYSIS AND RANKING IN COLLABORATIVE TAGGING SYSTEMS". *Computational Intelligence*. 27(3): 458–488.

Yin, H., B. Cui, J. Li, J. Yao, and C. Chen. 2012. "Challenging the Long Tail Recommendation". *Proc. of VLDB Endow.* 5(9): 896–907. ISSN: 2150-8097. DOI: 10.14778/2311906.2311916.

Zheng, N. and Q. Li. 2011. "A recommender system based on tag and time information for social tagging systems". *Expert Systems with Applications*. 38(4): 4575–4587.

Simone Kopeinik graduated from Graz University of Technology with a master degree in Computer Science. During her studies she specialized in Software Design and Development, User Adaptive Systems, Recommender Systems and Virtual

Learning Environments. She was involved in various national and international EU-funded research projects in the area of adaptive learning systems. Currently, Simone is working on her PhD-thesis focusing on tag and resource recommendation strategies for virtual learning environments. Her research interest comprises cognitive user modelling, user adaptive and personalized (learning) systems and recommender systems.

Dominik Kowald is a university assistant and PHD student at the Knowledge Technologies Institute of Graz University of Technology. Additionally, he is a software developer in the Social Computing division at the Know-Center. He has a MSc. (with hons) and BSc. in Software Engineering and Economy from Graz University of Technology and is doing his PhD in the course of the EU-IP Learning Layers on cognitive-inspired recommender systems for social tagging environments. His research interests are in the fields of recommender systems, social tagging and microblogging systems, information retrieval and social network analysis.

Ilire Hasani-Mavriqi LIlire Hasani-Mavriqi is a university project assistant and a PhD student at the Knowledge Technologies Institute (Graz University of Technology, Austria). She is also part of the Social Computing Team (led by Elisabeth Lex) at the Know-Center (Austria's Research Center for Data-Driven Business and Big Data Analytics). She received her MSc (with distinction) and BSc in Computer Science and Telematics from Graz University of Technology. Her research interests include social network analysis, dynamics in online social and collaboration networks, opinion diffusion and dynamics of social influence in complex networks.

Elisabeth Lex Elisabeth Lex is assistant professor at Graz University of Technology and she heads the Social Computing research area at Know-Center. Her research interests include Social Computing, Dynamics in Complex Networks, Recommender Systems, Web Science, Open Science, Data Mining and Machine Learning. Currently, Elisabeth is work package leader in the FP7 IP Learning Layers and Task Leader in the H2020 AFEL project. She is also member of the Expert Group on Altmetrics, which advises the European Commission, DG Research and Innovation.