Semantic Contextualisation of Social Tag-based Profiles and Item Recommendations

Iván Cantador, Alejandro Bellogín, Ignacio Fernández-Tobías,
Sergio López-Hernández

Departamento de Ingeniería Informática, Universidad Autónoma de Madrid
28049 Madrid, Spain
{ivan.cantador, alejandro.bellogin}@uam.es, ign.fernandez01@estudiante.uam.es,
sergio.lopez@uam.es

Abstract. We present an approach that efficiently identifies the semantic meanings and contexts of social tags within a particular folksonomy, and exploits them to build contextualised tag-based user and item profiles. We apply our approach to a dataset obtained from Delicious social bookmarking system, and evaluate it through two experiments: a user study consisting of manual judgements of tag disambiguation and contextualisation cases, and an offline study measuring the performance of several tag-powered item recommendation algorithms by using contextualised profiles. The results obtained show that our approach is able to accurately determine the actual semantic meanings and contexts of tag annotations, and allow item recommenders to achieve better precision and recall on their predictions.

Keywords: social tagging, folksonomy, ambiguity, semantic contextualisation, clustering, user modelling, recommender systems.

1 Introduction

Among the formats of user generated content available in the so called Web 2.0, *social tagging* has become a popular practice as a lightweight mean to classify and exchange information. Users create or upload content (resources), annotate it with freely chosen words (tags), and share these annotations with others. In this context, the nature of tagged resources is manifold: photos (Flickr[1]), music tracks (Last.fm[2]), video clips (YouTube[3]), and Web pages (Delicious[4]), to name a few.

In a social tagging system, the whole set of tags constitutes an unstructured collaborative knowledge classification scheme that is commonly known as *folksonomy*. This implicit classification serves various purposes, such as for resource organisation, promotions, and sharing with friends or with the public. Studies have shown, however, that tags are generally chosen by users to reflect their interests. Golder and Huberman [9] analysed tags on Delicious, and found that (1) the overwhelming majority of tags identify the topics of the tagged resources, and (2)

[1] Flickr, Photo sharing, http://www.flickr.com
[2] Last.fm, Internet radio and music catalogue, http://www.last.fm
[3] YouTube, Online video-sharing, http://www.youtube.com
[4] Delicious, Social bookmarking, http://delicious.com

almost all tags are added for personal use, rather than for the benefit of the community. These findings lend support to the idea of using tags to derive precise user preferences and item descriptions, and bring with new research opportunities on personalised search and recommendation.

Despite the above advantages, social tags are free text, and thus suffer from various vocabulary problems [12]. Ambiguity (polysemy) of the tags arises as users apply the same tag in different domains (e.g., `bridge`, the architectonical structure vs. the card game). At the opposite end, the lack of synonym control can lead to different tags being used for the same concept, precluding collocation (e.g., `biscuit` and `cookie`). Synonym relations can also be found in the form of acronyms (e.g., `nyc` for `new york city`), and morphological deviations (e.g., `blog`, `blogs`, `blogging`). Multilinguality also obstructs the achievement of a consensus vocabulary, since several tags written in different languages can express the same concept (e.g., `spain`, `españa`, `spagna`). Moreover, there are tags that have single meanings, but are used in different semantic contexts that should be distinguished (e.g., `web` may be used to annotate items about distinct topics such as Web design, Web browsers, and Web 2.0).

To address such problems, in this paper, we present an approach that efficiently identifies semantic meanings and contexts of social tags within a particular folksonomy (Section 3), and exploits them to build contextualised tag-based user and item profiles (Section 4). These enhanced profiles are then used to improve a number of tag-powered item recommendation algorithms (Section 5). To evaluate our approach, we conduct two experiments on a dataset obtained from Delicious social bookmarking system (Section 6): a user study consisting of manual judgements of tag disambiguation and contextualisation cases, and an offline study that measures the performance of the above recommenders. The obtained results show that our approach is able to accurately determine the actual semantic contexts of tag annotations, and allows item recommenders to achieve better precision and recall on their predictions.

2 Related Work

Current social tagging systems facilitate the users with the organisation and sharing of content. The way users can access the resources, however, is limited to searching and browsing through the collections. User-centred approaches, such as personalised search and recommendation, are not yet supported by most of such systems, although these functionalities are proven to provide a better user experience, by facilitating access to huge amounts of content, which, in the case of social tagging systems, is created and annotated by the community of users.

Recent works in the research literature have investigated the adaptation of personalised search [10, 15, 21] and recommendation [5, 6, 14, 16, 22] techniques to social tagging systems, but they have a common limitation: they do not deal with **semantic ambiguities** of tags. For instance, given a tag such as `sf`, existing content retrieval strategies do not discern between the two main meanings of that tag: *San Francisco* (the Californian city) and *Science Fiction* (the literary genre). This phenomenon occurs too frequently to be ignored by a social tagging system. As an example, as for March 2011, Wikipedia contains[5] over 192K disambiguation entries.

[5] Wikipedia disambiguation pages, http://en.wikipedia.org/wiki/Category:All_disambiguation_pages

Semantic ambiguity of tags is being investigated in the literature. There are approaches that attempt to identify the actual meaning of a tag by linking it with **structured knowledge bases** [2, 7, 18]. These approaches, however, rely on the availability of external knowledge resources, and so far are preliminary and have not been applied to personalisation and recommendation.

Other works are based on the concept of tag co-occurrence, that is, on extracting the actual meaning of a tag by analysing the occurrence of the tag with others in describing different resources. These approaches usually involve the application of **clustering techniques** over the co-occurrence information gathered from the folksonomy [3, 4, 20], and have been exploited by recent personalisation and recommendation approaches [8, 17]. Their main advantage is that an external knowledge source is not required. Nonetheless, they present several problems:

- *Lack of scalability*. Current approaches are not incremental; small changes in the folksonomy imply re-computing clusters within the whole folksonomy. This lack of scalability is undesired for a social tagging system, as its community of users is constantly adding new resources and annotations, resulting in a highly dynamic folksonomy.
- *Need for a stop criterion*. Current approaches have to define a stop criterion for the clustering processes. For instance, a hierarchical clustering [17] needs to establish the proper level at which clusters are selected, whereas an approach using a partitional clustering technique such as K-means needs to define beforehand how many clusters to build [8]. These values are difficult to define without proper evaluation, and have a definite impact on the outcome of the clustering process, and ultimately, on the semantic disambiguation or contextualisation approach. Moreover, these approaches define and evaluate the above parameter values over static test collections, and thus may not be easily adjustable over real social tagging systems.
- *Lack of explicit contextualisation*. Current approaches do not use clustering information to explicitly build contextualised user and item models. This information is rather incorporated into the retrieval and filtering algorithms, and cannot be exploited by other systems. Thus, these approaches do not offer a real contextualisation of tags, since they do not extract the context in which tags are used. For instance, a desired outcome of a disambiguation approach would be to provide a new contextualised tag description of the user's interests rather than her original raw tag values. Following the previous example, `sf` tag would be properly contextualised if it is defined within one of its possible meanings, such as `sf|San_Francisco` and `sf|Science_Fiction`. Recent works have investigated the contextualisation of folksonomies [3], but lack proper user and item models, and usually require humans to manually label each context.

As explained in subsequent sections, the approach presented herein addresses the above limitations by exploiting a fast graph clustering technique proposed by Newman and Girvan [13], which automatically establishes an optimal number of clusters. Moreover, for a particular tag, the approach does not have to be executed in the whole folksonomy tag set but in a subset of it, and explicitly assigns semantic contexts to annotations with such tag.

3 Semantic Contexts of Social Tags

In the literature, there are approaches that attempt to determine the different semantic meanings and contexts of social tags within a particular folksonomy by clustering the tags according to their co-occurrences in item annotation profiles [3, 8, 17]. For example, for the tag `sf`, often co-occurring tags such as `sanfrancisco`, `california` and `bayarea` may be used to define the context "San Francisco, the Californian city", while co-occurring tags like `sciencefiction`, `scifi` and `fiction` may be used to define the context "Science Fiction, the literary genre".

In this paper, we follow a clustering strategy as well, but in contrast to previous approaches, ours provides the following benefits:

- Instead of using simple tag co-occurrences, we propose to use more sophisticated tag similarities, which were presented by Markines et al. in [11], and are derived from established information theoretic and statistical measures.
- Instead of using standard hierarchical or partitional clustering strategies, which require defining a stop criterion for the clustering processes, we propose to apply the graph clustering technique presented by Newman and Girvan [13], which automatically establishes an optimal number of clusters. Moreover, to obtain the contexts of a particular tag, we propose not to cluster the whole folksonomy tag set, but a subset of it.

In the following, we briefly describe the above tag similarities and clustering technique.

3.1 Tag Similarities

A folksonomy $\mathcal{F}$ can be defined as a tuple $\mathcal{F} = \{\mathcal{T}, \mathcal{U}, \mathcal{I}, \mathcal{A}\}$, where $\mathcal{T}$ is the set of tags that comprise the vocabulary expressed by the folksonomy, $\mathcal{U}$ and $\mathcal{I}$ are respectively the sets of users and items that annotate and are annotated with the tags of $\mathcal{T}$, and $\mathcal{A} = \{(u,t,i)\} \in \mathcal{U} \times \mathcal{T} \times \mathcal{I}$ is the set of assignments (annotations) of each tag t to an item i by a user u.

To compute semantic similarities between tags, we follow a two step process. First, we transform the tripartite space of a folksonomy, represented by the triples $\{(u,t,i)\} \in \mathcal{A}$, into a set of tag-item relations $\{(t,i,w_{t,i})\} \in \mathcal{T} \times \mathcal{I} \times \mathbb{R}$ (or tag-user relations $\{(t,u,w_{t,u})\} \in \mathcal{T} \times \mathcal{U} \times \mathbb{R}$), where $w_{t,i}$ (or $w_{t,u}$) is a real number that expresses the relevance (importance, strength) of tag t when describing item profile i (or user profile u). In [11], Markines et al. call this transformation as tag assignment "aggregation", and present and evaluate a number of different aggregation methods. In this paper, we focus on two of these methods, *projection* and *distributional* aggregation, which are described with a simple example in Figure 1. Projection aggregation is based on the Boolean use of a tag for annotating a particular item, while distributional aggregation is based on the popularity (within the community of users) of the tag for annotating such item.

Second, in the obtained bipartite tag-item (or tag-user) space, we compute similarities between tags based on co-occurrences of the tags in item (or user) profiles. In [11], the authors compile a number of similarity metrics derived from established information theoretic and statistical measures. In this paper, we study some of these metrics, whose definitions are given in Table 1.

Tag assignments [user, tag, item]

Alice	conference	recommender	research
dexa.org/ecweb2011	1	1	
delicious.com		1	
ir.ii.uam.es		1	1

Bob	conference	recommender	research
dexa.org/ecweb2011	1	1	1
delicious.com		1	
ir.ii.uam.es			

⇩

Tag assignment aggregation [tag, item]

Projection	conference	recommender	research
dexa.org/ecweb2011	1	1	1
delicious.com		1	
ir.ii.uam.es		1	1

Distributional	conference	recommender	research
dexa.org/ecweb2011	2	2	1
delicious.com		2	
ir.ii.uam.es		1	1

Figure 1. An example of projection and distributional tag assignment aggregations. Two users, Alice and Bob, annotate three Web pages with three tags: *conference*, *recommender* and *research*.

Table 1. Tested tag similarity metrics. $I_1, I_2 \subseteq I$ are the sets of items annotated with $t_1, t_2 \in \mathcal{T}$.

Similarity	***Projection aggregation***	***Distributional aggregation***
Matching	$sim(t_1,t_2) = \lvert I_1 \cap I_2 \rvert$	$sim(t_1,t_2) = -\sum_{t \in I_1 \cap I_2} \log \mathrm{p}(t)$
Overlap	$sim(t_1,t_2) = \frac{\lvert I_1 \cap I_2 \rvert}{\min(I_1, I_2)}$	$sim(t_1,t_2) = \frac{\sum_{t \in I_1 \cap I_2} \log \mathrm{p}(t)}{\max\left(\sum_{t \in I_1} \log \mathrm{p}(t), \sum_{t \in I_2} \log \mathrm{p}(t)\right)}$
Jaccard	$sim(t_1,t_2) = \frac{\lvert I_1 \cap I_2 \rvert}{\lvert I_1 \cup I_2 \rvert}$	$sim(t_1,t_2) = \frac{\sum_{t \in I_1 \cap I_2} \log \mathrm{p}(t)}{\sum_{t \in I_1 \cup I_2} \log \mathrm{p}(t)}$
Dice	$sim(t_1,t_2) = \frac{2\lvert I_1 \cap I_2 \rvert}{\lvert I_1 \rvert + \lvert I_2 \rvert}$	$sim(t_1,t_2) = \frac{2\sum_{t \in I_1 \cap I_2} \log \mathrm{p}(t)}{\sum_{t \in I_1} \log \mathrm{p}(t) + \sum_{t \in I_2} \log \mathrm{p}(t)}$
Cosine	$sim(t_1,t_2) = \frac{I_1}{\sqrt{\lvert I_1 \rvert}} \cdot \frac{I_2}{\sqrt{\lvert I_2 \rvert}} = \frac{\lvert I_1 \cap I_2 \rvert}{\sqrt{\lvert I_1 \rvert \cdot \lvert I_2 \rvert}}$	$sim(t_1,t_2) = \frac{I_1}{\lVert I_1 \rVert} \cdot \frac{I_2}{\lVert I_2 \rVert}$

3.2 Tag Clustering

We create a graph G, in which nodes represent the social tags of a folksonomy, and edges have weights that correspond to semantic similarities between tags. By using the similarity metrics presented in Section 3.1, G captures global co-occurrences of tags within item annotations, which in general, are related to *synonym* and *polysemy* relations between tags. Note that G is undirected. Using asymmetric metrics (e.g. those of [11] based on collaborative filtering), we may obtain directed graphs that would provide different semantic relations between tags, e.g. *hypernym* and *hyponym*.

Once G is built, we apply the graph clustering technique presented by Newman and Girvan [13], which automatically establishes an optimal number of clusters. However, we do not cluster G, but subgraphs of it. Specifically, for each tag $t \in \mathcal{T}$, we select its T_1 most similar tags and then, for each of these new tags, we select its T_2 most similar tags[6] to allow better disinguising semantic meanings and contexts ot t within the set of T_1 tags. With all the obtained tags (at most $1 + T_1 T_2$), we create a new graph G_t, whose edges are extracted from G. We have implemented an online demo[7] that obtains the contexts of tags in stored folksonomies. Table 2 shows examples of contexts retrieved by our system for Delicious tags. Centroids are representative tags of the contexts, and are automatically identified by our approach, as explained in Section 4.

[6] In the conducted experiments, $T_1 = 25$ and $T_2 = 3$ gave the best results

[7] CTag Context Viewer, http://ir.ii.uam.es/reshet/results.html

Table 2. Examples of semantic contexts identified for different Delicious tags.

tag	context centroid	context popularity	context tags
sf	fiction	0.498	fiction, scifi, sciencefiction, schi-fi, stores, fantasy, literature
	sanfrancisco	0.325	sanfrancisco, california, bayarea, losangeles, la
	restaurants	0.082	restaurants, restaurant, dining, food, eating
	events	0.016	events, event, conferences, conference, calendar
web	webdesign	0.434	webdesign, webdev, web_design, web-design, css, html
	web2.0	0.116	web2.0, socialnetworks, social, socialmedia
	javascript	0.077	javascript, js, ajax, jquery
	browser	0.038	browser, browsers, webbrowser, ie, firefox
london	england	0.263	england, uk, britain, british, english
	transport	0.183	transport, tube, underground, transportation, train, bus, map
	theatre	0.030	theatre, theater, tickets, entertainment, arts
	travel	0.030	travel, vacation, flights, airlines
holiday	christmas	0.336	christmas, xmas
	travel	0.274	travel, trip, vacation, tourism, turismo, planner
	airlines	0.104	airlines, arline, flights, flight, cheap
	rental	0.019	rental, apartment, housing, realestate

4 Tag-based Profiles

We define the profile of user u as a vector $\mathbf{u} = (u_1, \dots, u_T)$, where u_t is a weight (real number) that measures the "informativeness" of tag t to characterise contents annotated by u. Similarly, we define the profile of item i as a vector $\mathbf{i} = (i_1, \dots, i_T)$, where i_t is a weight that measures the relevance of tag t to describe i. There exist different schemes to weight the components of tag-based user and item profiles. Some of them are based on the information available in individual profiles, while others draw information from the whole folksonomy.

TF Profiles

The simplest approach for assigning a weight to a particular tag in a user or item profile is by counting the number of times such tag has been used by the user or the number of times the tag has been used by the community to annotate the item. Thus, our first profile model for user u consists of a vector $\mathbf{u} = (u_1, \dots, u_T)$, where

$$u_t = tf_u(t),$$

$tf_u(t)$ being the tag frequency, i.e., the number of times user u has annotated items with tag t. Similarly, the profile of item i is defined as a vector $\mathbf{i} = (i_1, \dots, i_T)$, where

$$i_t = tf_i(t),$$

$tf_i(t)$ being the number of times item i has been annotated with tag t.

TF-IDF Profiles

In an information retrieval environment, common keywords that appear in many documents of a collection are not informative, and are generally not helpful to distinguish relevant documents for a given query. To take this into account, the TF-IDF weighting scheme is usually applied to the document profiles. We adopt that principle, and adapt it to social tagging systems, proposing a second profile model, defined as follows:

$$u_t = tfiuf_u(t) = tf_u(t) \cdot iuf(t),$$
$$i_t = tfiif_i(t) = tf_i(t) \cdot iif(t)$$

where $iuf(t)$ and $iif(t)$ are inverse frequency factors that penalise tags that frequently appear (and thus are not informative) in tag-based user and item profiles respectively. Specifically, $iuf(t) = \log(M/m_t), m_t = |\{u \in \mathcal{U}|u_t > 0\}|$, and $iif(t) = \log(N/n_t), n_t = |\{i \in \mathcal{I}|i_t > 0\}|$. Note that we incorporate both user and item tag distribution global importance factors, iuf and iif, following the vector space model principle that as more rare a tag is, the more important it is for describing either a user's interests or an item's content.

BM25 Profiles

As an alternative to TF-IDF, the Okapi BM25 weighting scheme follows a probabilistic approach to assign a document with a ranking score given a query. We propose an adaptation of such model by assigning each tag with a score (weight) given a certain user or item. Our third profile model has the following expressions:

$$u_t = bm25_u(t) = \frac{tf_u(t)\cdot(k_1+1)}{tf_u(t) + k_1\left(1-b+b\cdot {|u|}/{avg(|u|)}\right)} \cdot iuf(t),$$

$$i_t = bm25_i(t_l) = \frac{tf_i(t)\cdot(k_1+1)}{tf_i(t) + k_1\left(1-b+b\cdot {|i|}/{avg(|i|)}\right)} \cdot iif(t)$$

where b and k_1 are set to the standard values 0.75 and 2, respectively.

Profiles with Semantically Contextualised Tags

We propose to apply our semantic contextualisation approach to each of the profile models defined before – TF, TF-IDF and BM25. A tag t is transformed into a semantically contextualised tag t^u (or t^i), which is formed by the union of t and the semantic context $c_{t,u}$ (or $c_{t,i}$) of t within the corresponding user profile u (or item profile i). For instance, tag `sf` in a user profile with tags like `city`, `california` and `bayarea` may be transformed into a new tag `sf|sanfrancisco`, since in that profile, "sf" clearly refers to San Francisco, the Californian city. With this new tag, matchings with item profiles containing contextualised tags such as `sf|fiction`, `sf|restaurants` or `sf|events` would be discarded by a personalised search or recommendation algorithm because they may annotate items related to Science Fiction, or more specific topics of San Francisco like restaurants and events in the city.

More formally, the context (centroid) $c_{t,u}$ (or $c_{t,i}$) of tag t within the user profile u (or item profile i), and the corresponding contextualised tag t^u (or t^i) are defined as follows:

$$\forall(u,t,i)\in\mathcal{A},\quad \begin{array}{l} c_{t,u} = \mathrm{c}(t,u) = \arg\max_{c_t} cos(\mathbf{c}_t, \mathbf{u}) \Rightarrow t^u = t \cup c_{t,u} \\ c_{t,i} = \mathrm{c}(t,i) = \arg\max_{c_t} cos(\mathbf{c}_t, \mathbf{i}) \Rightarrow t^i = t \cup c_{t,i} \end{array}$$

where $\mathbf{c}_t = (c_1, \dots, c_T)$ is the weighted list of tags that define each of the contexts c_t of tag t within the folksonomy (see Table 2).

Table 3 shows some examples of contextualised tag-based profiles generated by our approach. We have implemented another online demo[8] that allows contextualising profiles manually defined by the user or automatically extracted from Delicious.

[8] CTag Profile Builder, http://ir.ii.uam.es/reshet/results.html

Table 3. Examples of 4 semantically contextualised tag-based item profiles. Each original *tag* is transformed into a *tag|context* pair.

culture|philosophy	essay|interesting	fiction|sf	future|scifi	futurism|philosophy
god|science	interesting|science	literature|scifi	mind|philosophy	read|philosophy
religion|philosophy	research|science	sci-fi|sf	sciencefiction|sf	scifi|writing
sf|fiction	storytelling|fiction	toread|philosophy	universe|philosophy	writing|fiction
bayarea|sf	california|sf	city|sustainability	conservation|green	eco|green
environment|recycle	government|activism	green|environment	home|green	local|sanfrancisco
recycle|environment	recycling|environment	sanfrancisco|sf	**sf|sanfrancisco**	solar|environment
sustainability|recycling	sustainable|green	trash|green	urban|sustainability	volunteer|environmental
ajax|javascript	css|javascript	design|web	embed|webdesign	framework|javascript
gallery|jquery	html|javascript	icons|web	javascript|ajax	jquery|webdev
js|javascript	library|javascript	plugin|webdev	programming|javascript	site|webdev
toolkit|webdev	tutorials|webdev	**web|javascript**	web2.0|web	webdev|javascript
articles|web	blogs|web2.0	idea|community	internet|tools	library|opensource
network|tools	podcasts|education	rdf|web	reading|education	school|educational
semantic|semanticweb	semanticweb|web	semweb|semanticweb	software|utilities	technology|web2.0
tim|web	trends|technology	**web|web2.0**	web2.0|social	wiki|web2.0

5 Tag-powered Item Recommenders

Adomavicius and Tuzhilin [1] formulate the recommendation problem as follows. Let $\mathcal{U}$ be a set of users, and let $\mathcal{I}$ be a set of items. Let $g{:}\,\mathcal{U}\times\mathcal{I}\to\mathcal{R}$, where $\mathcal{R}$ is a totally ordered set, be a utility function such that $g(u,i)$ measures the gain of usefulness of item i to user u. Then, for each user $u \in \mathcal{U}$, we want to choose items $i^{\max,u} \in \mathcal{I}$, unknown to the user, which maximise the utility function g:

$$\forall u \in \mathcal{U}, \quad i^{\max,u} = \arg\max_{i \in \mathcal{I}} g(u,i)$$

In content-based recommendation approaches, g is formulated as:

$$g(u,i) = sim(ContentBasedUserProfile(u), Content(i)) \in \mathcal{R}$$

where $ContentBasedUserProfile(u) = \mathbf{u} = (u_1,\ldots,u_K) \in \mathbb{R}^K$ is the content-based preferences of user u, i.e., the item content features that describe the interests, tastes and needs of the user, and $Content(i) = \mathbf{i} = (i_1,\ldots,i_K) \in \mathbb{R}^K$ is the set of content features characterising item i. These descriptions are usually represented as vectors of real numbers (weights) in which each component measures the "importance" of the corresponding feature in the user and item representations. The function *sim* computes the similarity between a user profile and an item profile in the content feature space. From the previous formulations, in this paper, we consider social tags as the content features that describe both user and item profiles (as explained in Section 4), and present a number of recommenders that we presented and evaluated in [6].

TF-based Recommender

To compute the preference of a user for an item, Noll and Meinel [15] propose a personalised similarity measure based on the user's tag frequencies. In their model, we introduce a normalisation factor that scales the utility function to values in the range [0,1], without altering the user's item ranking:

$$g(u,i) = tf(u,i) = \frac{\sum_{t:i_t>0} tf_u(t)}{\max_{v\in\mathcal{U},t\in\mathcal{T}}\left(tf_v(t)\right)}$$

TF-IDF Cosine-based Recommender

Xu et al. [21] use the cosine measure to compute the similarity between user and item profiles. As profile component weighting scheme, they use TF-IDF. We adapt their approach with the proposed tag-based profile models as follows:

$$g(u,i) = cos_{tf\text{-}idf}(u,i) = \frac{\sum_t tf_u(t) \cdot iuf(t) \cdot tf_i(t) \cdot iif(t)}{\sqrt{\sum_t \big(tf_u(t) \cdot iuf(t)\big)^2} \cdot \sqrt{\sum_t \big(tf_i(t) \cdot iif(t)\big)^2}}$$

BM25 Cosine-based Recommender

Xu et al. [21] also investigate the cosine measure with a BM25 weighting scheme. They use this model on personalised Web Search. We adapt and define it for social tagging as follows:

$$g(u,i) = cos_{bm25}(u,i) = \frac{\sum_t \big(bm25_u(t) \cdot bm25_i(t)\big)}{\sqrt{\sum_t \big(bm25_u(t)\big)^2} \cdot \sqrt{\sum_t \big(bm25_i(t)\big)^2}}$$

Recommenders with Semantically Contextualised Tag-based Profiles

We propose to evaluate the previous recommenders (1) by using tag-based user and item profiles existing in a real dataset, and (2) by contextualising these profiles with the approach presented in Section 4.

6 Experiments

To evaluate our tag-based profile contextualisation approach and its impact on the presented tag-powered recommendation models, we used a dataset obtained from Delicious system. Delicious is a social bookmarking site for Web pages. By the end of 2008, the service claimed more than 5.3 million users and 180 million unique bookmarked URLs. As a collaborative social tagging platform, Delicious contains tagged items (Web pages) belonging to practically any domain.

Our dataset was formed by 2,203 Delicious users, randomly selected from the set of users who tagged top Delicious bookmarks of 14th May 2009, and had at least 20 bookmarks in their profiles. By extracting the latest 100 bookmarks of each user, and filtering out those bookmarks with less than 20 tags, the final dataset contained 146,382 different bookmarks and 54,618 distinct tags. On average, each user profile had 77 bookmarks and 195 tags, and each item profile had 19 tags.

Once the dataset was built, we ran our clustering technique to obtain the semantic contexts of 2,893 tags: those belonging to at least 200 bookmarks. Although these tags are only 5.3% of the total set of tags in our dataset, they appear in 80.6% of the gathered tag assignments, and as we shall show in Section 6.2, they were enough to improve significantly the performance of the recommenders. Before that, in Section 6.1, we present an experiment to evaluate the accuracy of the contextualisation approach.

6.1 Evaluating Tag Contextualisation

We performed a preliminary user study to manually evaluate context assignments to tag annotations of user and item profiles. 30 PhD students and academic staff of our department participated in the experiment. They were requested to select the proper semantic context of 360 annotations (50% of them in user profiles and the remaining

50% in item profiles) of 78 distinct tags. Each annotation was evaluated by 3 different subjects, providing a total of 1,080 evaluation tests. An evaluation test consisted of presenting a subject with a particular tag, the profile the tag belonged to, and the set of possible semantic contexts of the tag. These semantic contexts were shown as coloured clusters in a tag co-occurrence based graph to ease the evaluation task. In each test, a subject could select one, two or three options for the proper semantic context of the tag. These options had to be selected sorted by decreasing preference. Moreover, in case a subject did not feel confident with the evaluation of a certain test, she could state that test was "unknown" for her. There was a substantial agreement among subjects. Fleiss' Kappa statistic measuring subjects' agreement was $\kappa = 0.636$ (a value $\kappa = 1$ means complete agreement) for the first context choice in known tests.

The contexts provided by the subjects were then used as ground truth to measure the accuracy of our contextualisation approach. For each test, we made a ranked list with the contexts selected by the subjects, ordered according to their positions in the subjects' choices lists (the more preferred choice, the higher the ranking score), and the number of such lists in which they appeared (the higher the number of lists, the higher the ranking score). Figure 2 shows the percentages of correct context assignments corresponding to the 1^{st} to 5^{th} positions in the rankings. Position 0 means the contexts assigned by our approach was not selected by any subject in the tests. For known tests, our approach assigned the correct context in 63.8% of the cases in the 1^{st} positions of the ranked lists. The accuracy was 60.6% for annotations in user profiles, and 66.7% for annotations in item profiles, which was expected since user profiles contain more diverse tags (user preferences) than item profiles (content descriptions). Summing the correct context assignments for the 2 and 3 top choices of each subject, we respectively obtained accuracy values of 81.1% and 88.4% (being 86.3% for user profiles, and 90.5% for item profiles). Only 8.2% of the context assignments were wrong.

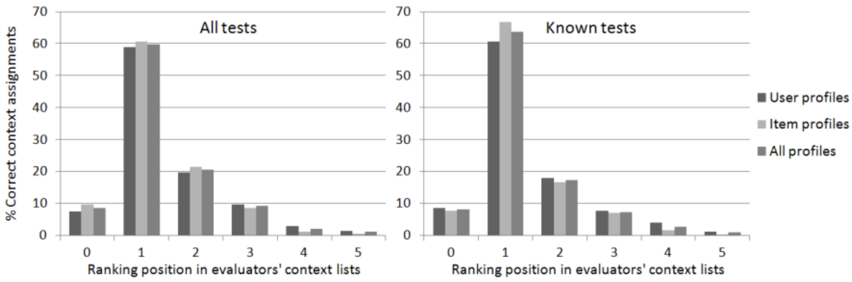

Figure 2. Accuracy of the proposed semantic contextualisation approach.

6.2 Evaluating Contextualised Tag-powered Item Recommendations

To evaluate the performance of each recommender, we assume a content retrieval scenario where a system provides the user a list of *N* recommended items based on her tag-based profile. We take into account the percentage and ranking of relevant items appearing in the provided lists, computing four metrics often used to evaluate information retrieval systems: Precision and Recall at the top *N* ranked results (P@*N*, R@*N*), Mean Average Precision (MAP), and Discounted Cumulative Gain (DCG). *Precision* is defined as the number of retrieved relevant items divided by the total number of retrieved items. *MAP* is a precision metric that emphasises ranking relevant items higher. *Recall* is the fraction of relevant items that are successfully retrieved by the system. Finally, *DCG* measures the usefulness of an item based on its position in a result list. In our evaluation framework, retrieved items were all the items belonging to each test set (see below). Thus, a test set may contain (1) items belonging to the active user's profile, considered thus as "relevant", and (2) items from other users' profiles, assumed as "non relevant" for the active user.

We randomly split the set of items in the database into two subsets. The first subset contained 80% of the items for each user, and was used to build the recommendation models (training). The second subset contained the remaining 20% of the items, and was used to evaluate the recommenders (test). We built the recommendation models with the whole tag-based profiles of the training items, and with those parts of the users' tag-based profiles formed by tags annotating the training items. We evaluated the recommenders with the tag-based profiles of the test items. In the evaluation, we performed a 5-fold cross validation procedure.

The results are shown in Table 4. As found in previous studies [6], BM25 recommender achieved the best precision and recall values. But more importantly, all the recommenders were improved by using contextualised tag-based profiles. The table also shows the performance improvement percentages, which range from 24% for the TF recommender to 13% for the BM25 recommender, in all the computed metrics. It is important to note that these improvements were obtained by using a simple contextualisation approach (Section 4) that achieved 63.8% of accuracy according to our user study (Section 6.1), and which was applied to only 5.3% of the tags.

Table 4. Improvements on the performance of the recommenders, by using contextualised profiles (those marked with *). The results were achieved with the *cosine similarity* and *distributional aggregation*. No significant differences were obtained with the other similarities.

	P@5	**P@10**	**P@20**	**MAP**	**R@5**	**R@10**	**R@20**	**NDCG**
tf	0.073	0.056	0.041	0.023	0.024	0.036	0.054	0.061
tfidf	0.135	0.103	0.074	0.044	0.044	0.067	0.096	0.113
bm25	0.149	0.109	0.077	0.048	0.048	0.071	0.100	0.121
tf*	0.093	0.069	0.049	0.029	0.030	0.045	0.064	0.077
tfidf*	0.162	0.117	0.083	0.052	0.053	0.076	0.107	0.131
bm25*	**0.171**	**0.123**	**0.085**	**0.069**	**0.055**	**0.080**	**0.109**	**0.136**
tf*	27.20%	23.18%	18.54%	23.77%	28.40%	23.98%	19.25%	24.81%
tfidf*	19.68%	14.49%	12.15%	18.07%	19.37%	14.18%	11.62%	18.07%
bm25*	15.25%	13.09%	9.85%	16.97%	15.09%	12.57%	9.13%	12.64%

7 Conclusions

In this paper, we have presented an approach to semantically contextualise social tag-based profiles within a particular folksonomy. Our approach utilises a clustering technique that exploits sophisticated co-occurrence based similarities between tags, and is very efficient since it is not executed on the whole tag set of the folksonomy, and provides an automatic stop criterion to establish the optimal number of clusters.

We have applied the approach on tag-based user and item profiles extracted from Delicious bookmarking system, and evaluated it with a number of state of the art tag-powered item recommenders. The obtained results are encouraging. By contextualising 5.3% of the tags available in the dataset, we achieved an accuracy on context assignments of 63.8% (according to manual judgements of a conducted user study), and 13% to 24% precision/recall improvements on the tested recommenders.

For future work, we plan to extend our study by investigating alternative contextualisation strategies, evaluating them on additional (collaborative filtering and hybrid) recommenders, and using larger datasets from different social tagging systems. An empirical comparison with other clustering approaches, and a deep analysis to determine which folksonomy characteristics have more impact on the effectiveness of contextualised tag-based profiles in recommendation will be done as well.

Acknowledgements. This work was supported by the Spanish Ministry of Science and Innovation (TIN2008-06566-C04-02), and the Community of Madrid (CCG10-UAM/TIC-5877).

References

1. Adomavicius, G., Tuzhilin, A.: Toward the Next Generation of Recommender Systems: A Survey and Possible Extensions. IEEE Transactions on Knowledge & Data Engineering, 17(6), 734-749 (2005)
2. Angeletou, S., Sabou, M., Motta, E.: Improving Folksonomies Using Formal Knowledge: A Case Study on Search. In: 4th Asian Semantic Web Conference, 276-290. Springer-Verlag (2009)
3. Au Yeung, C. M., Gibbins, N., Shadbolt, N.: Contextualising Tags in Collaborative Tagging Systems. In: 20th Conference on Hypertext and Hypermedia, pp. 251-260. ACM Press (2009)
4. Benz, D., Hotho, A., Stützer, S., Stumme, G.: Semantics Made by You and Me: Self-emerging Ontologies Can Capture the Diversity of Shared Knowledge. In: 2nd Web Science Conference. (2010)
5. Bogers, T., Van Den Bosch, A.: Recommending Scientific Articles Using Citeulike. In: 2nd ACM Conference on Recommender Systems, 287-290. ACM Press (2008)
6. Cantador, I., Bellogín, A., Vallet, D.: Content-based Recommendation in Social Tagging Systems. In: 4th ACM Conference on Recommender Systems, pp. 237-240. ACM Press (2010)
7. García-Silva, A., Szomszor, M., Alani, H., Corcho, O.: Preliminary Results in Tag Disambiguation using DBpedia. In: 1st International Workshop on Collective Knowledge Capturing and Representation (2009)
8. Gemmell, J., Ramezani, M., Schimoler, T., Christiansen, L., Mobasher, B.: The Impact of Ambiguity and Redundancy on Tag Recommendation in Folksonomies. In: 3rd ACM Conference on Recommender Systems, pp. 45-52. ACM Press (2009)
9. Golder, S. A., Huberman, B. A.: Usage Patterns of Collaborative Tagging Systems. Journal of Information Science, 32(2), 198-208 (2006)
10. Hotho, A., Jäschke, R., Schmitz, C., Stumme, G.: Information Retrieval in Folksonomies: Search and Ranking. In: 5th International Semantic Web Conference, pp. 411-426. Springer-Verlag (2006)
11. Markines, B., Cattuto, C., Menczer, F., Benz, D., Hotho, A., Stumme, G.: Evaluating Similarity Measures for Emergent Semantics of Social Tagging. In: 18th Intl. Conference on WWW, pp. 641-650. ACM Press (2009)
12. Mathes, A.: Folksonomies - Cooperative Classification and Communication through Shared Metadata. Computer Mediated Communication, University of Illinois Urbana-Champaign, IL, USA (2004)
13. Newman, M. E. J., Girvan, M.: Finding and Evaluating Community Structure in Networks. Physical Review, E 69, 026113 (2004)
14. Niwa, S., Doi, T., Honiden, S.: Web Page Recommender System based on Folksonomy Mining for ITNG'06 Submissions. In: 3rd International Conference on Information Technology: New Generations, pp.388-393. IEEE Press (2006)
15. Noll, M. G., Meinel, C.: Web Search Personalization via Social Bookmarking and Tagging. In: 6th International Semantic Web Conference, pp. 367-380. Springer-Verlag (2007)
16. Sen, S., Vig, J., Riedl, J.: Tagommenders: Connecting Users to Items through Tags. In: 18th International Conference on WWW, pp. 671-680. ACM Press (2009)
17. Shepitsen, A., Gemmell, J., Mobasher, B., Burke, R. 2008. Personalized Recommendation in Social Tagging Systems using Hierarchical Clustering. In: 2nd ACM Conference on Recommender Systems, pp. 259-266. ACM Press (2008)
18. Specia, L., Motta, E.: Integrating Folksonomies with the Semantic Web. In: 4th European Semantic Web Conference, pp. 624-639. Springer-Verlag (2007)
19. Vallet, D., Cantador, I., Jose, J. M.: Personalizing Web Search with Folksonomy-Based User and Item Profiles. In: 32nd European Conference on Information Retrieval, pp. 420-431. Springer-Verlag (2010)
20. Weinberger, K. Q., Slaney, M., Van Zwol, R.: Resolving Tag Ambiguity. In: 16th International ACM Conference on Multimedia, pp. 111-120. ACM Press (2008)
21. Xu, S., Bao, S., Fei, B., Su, Z., Yu, Y. Exploring Folksonomy for Personalized Search. In: 31st Annual Intl. Conf. on Research and Development in Information Retrieval, pp. 155-162. ACM Press (2008)
22. Zanardi, V., Capra, L.: Social Ranking: Uncovering Relevant Content using Tag-based Recommender Systems. In: 2nd Conference on Recommender Systems, pp. 51-58. ACM Press (2008)