Classifying Users and Identifying User Interests in Folksonomies

Elias Zavitsanos[1], George A. Vouros[1], and Georgios Paliouras[2]

[1]Department of Information and Communication Systems Engineering
University of the Aegean, Greece
izavits@iit.demokritos.gr, georgev@aegean.gr
[2]Inst. of Informatics and Telecommunications, NCSR Demokritos, Greece
paliourg@iit.demokritos.gr

Abstract. This paper presents a probabilistic method for classifying folksonomy users to folksonomy sub-domains and identifying their particular interests. In particular, we propose a method for mining topic hierarchies that may reveal either the collective or the user-specific conceptualization of those domains, as these are reflected by users' tags. We then propose two alternatives for identifying users' interests in the domains: The first exploits users' tags directly, and the second exploits users' specific conceptualizations of each domain. Both approaches use the collective domain conceptualizations as "reference", to which users' tags and conceptualizations are compared. The proposed statistical method is parametric-less and does not require any prior knowledge or external resources. We apply the proposed method on the Del.icio.us online bookmarking system and we provide experimental results.

1 Introduction

Collaborative tagging systems, also known as *Folksonomies*, comprise *content* (objects, resources), *annotations* (tags) and *users*. Popular examples of folksonomies include Del.icio.us, Flickr and CiteULike. Although the term "folksonomy" is based on the term "taxonomy", that implies a hierarchy, folksonomies constitute flat organizations of tags and resources: They do not include semantic relations between tags or any representation of tags' intended meaning. Folksonomy tags depend totally on the interests, preferences, conceptualization, nomenclature, whim and personal style of users. Therefore, there is a great potential for acquiring knowledge about folksonomy users by exploiting their tags ([5], [12], [11], [6], [18]), introducing a number of interesting challenges and opportunities in the context of Web 2.0 and its bridge to the Semantic Web.

In this context, the issues that this paper addresses are the following: (a) Automated identification of conceptualizations of domains, by exploiting the tags users introduce to resources related to those domains, and (b) exploitation of tags and induced conceptualizations for identifying individual users' interests to specific domains and topics. Numerous entities and organizations can make use of such capabilities: For advertisement, for recommendation and for educational purposes, identifying users' interests, preferences and needs is of high value. Moreover, the collaborative tagging systems themselves could be improved by these capabilities, guiding the users to specific topics of interest, and of course, influencing the future tagging activity.

To address these challenging issues, we present a probabilistic method for classifying folksonomy users to folksonomy sub-domains, mining users' interests and conceptualizations. In particular, the contributions made in this work are as follows: (a) The automated induction of topic hierarchies from tags, in a statistical and parametric-less way, without requiring any external resources or prior knowledge, using the method proposed in [19]. These hierarchies of latent topics represent (collective or user-specific) conceptualizations of domains. (b) The use of collective topic hierarchies for classifying and identifying particular interests of users to the specific domains, by means of two alternative methods. The first exploits users' tags directly, and the second builds and exploits a user-specific conceptualization of each domain. Both approaches use the collective conceptualizations as "references", to which users' tags and conceptualizations are compared. It must be emphasized that all the above methods are performed in an unsupervised and language agnostic way, without requiring training data for each user. In our experiments we use datasets gathered from Del.icio.us, a popular online bookmarking system that offers collaborative tagging of bookmarks.

2 Related Work

Regarding the induction of hierarchies from folksonomies, the work in [7] presents a method for converting a corpus of tags into a taxonomy. The corpus is represented as frequency vectors of tags. A similarity function is defined between vectors and then a threshold is established to prune irrelevant values. Finally, for a given dataset, a tag similarity graph is created exploiting the notion of graph centrality. Starting from this graph, a hierarchy is induced.

The work in [15] proposes the application of Sanderson's probabilistic subsumption model [13] to tag sets in order to induce a hierarchy of tags from Flickr. The method adjusts the statistical thresholds of the subsumption model and adds filters in order to control the highly idiosyncratic Flickr vocabulary. The aim is to eventually produce a hierarchy of tags.

Since folksonomies are actually triples, the authors in [14] present a formal model of folksonomies as a tripartite hyper-graph and explore possible projections of the folksonomy into two dimensions, in order to apply association rule mining methods and mine the relations between tags. Doing so, they manage to create a graph of tags connected with edges that represent mined rules.

The work in [8] uses formal concept analysis to build tag hierarchies from tags of the blogosphere. The main assumption is that if a blog has relationships with other blogs, these blogs will use a similar sets of tags.

Regarding the clustering approaches reported in [18] and [21], our aim is not to cluster the tags per se, but to identify the latent topics that reveal the content of tag chains: Since tags are introduced by users, latent topics reflect users' conceptualizations. Additionally, tags may contribute to different topics with different proportions, and topics are represented as probability distributions over the tag space.

To a greater extent than existing approaches, in this paper we present a fully automated, parametric-less method for learning a hierarchy of latent topics from tag chains without the use of external resources or any prior knowledge. It must be pointed out that we consider the proposed method to be complementary to the approaches reported in [12] and [6], since it can be applied to different projections of the information concerning tags. However, further work is necessary to thoroughly compare the different approaches using commonly agreed datasets and evaluation criteria.

Regarding the construction of users' profiles, or the classification of users based on their interests in folksonomies, the work in [3] presents a framework that depends on external ontologies in order to build users' profiles, given their tagging activity and navigation in a folksonomy. A predefined ontology defines the concepts that are required to build a user profile. Profiles are exploited for recommendation purposes by a reasoner. The whole framework depends heavily on external ontologies and resources that are being used used for matching tags with elements of the domain ontology.

The work in [9] aims to cluster users based on their tagging activity. For a particular domain of interest, the main idea is to find the urls and the users that have labeled those urls with the tags in the domain. For each domain, a cluster of users is generated, comprising users with similar interests.

Moreover, the authors in [4] propose a method for generating and maintaining user profiles in a tag-based manner. The basic idea is to relate a user with a set of tagged objects and store them in an intermediate user profile. The representation of the user profile is based on the tags associated with the objects. Based on the user profile, recommendations can take place, since the tags define the interests of the users.

The work in [16] proposes an architecture for building user profiles by exploiting folksonomies, in four steps: (a) user account identification, (b) harvesting of user tags, (c) tag filtering to identify synonyms and deal with misspellings, and (d) profile building by matching tags to Wikipedia categories.

Finally, the work in [1] aims to exploit users' tags and additional knowledge inferred from the expertise profiles of other users to infer user's expertise. Our approach is rather generic: It induces and exploits collective and user-specific topic hierarchies, aiming to the classification of users to specific domains and to the identification of users' specific interests to these domains.

Therefore, to a greater extent than the existing approaches, the aim of this paper is to classify the users to specific domains, according to their interests, in an automated and unsupervised way, identifying also their specific interests to topics of these domains, given their tags. This is done with a probabilistic topic modeling approach, in order to avoid pitfalls related to surface appearance of tags.

3 The Del.icio.us Datasets

We provide experimental results using datasets compiled from the Del.icio.us collaborative tagging system: Crawling Del.icio.us we have gathered the tag chains (sets of tags related to a specific resource) of resources regarding a specific "domain". E.g. a tag chain for a resource in `http://www.Del.icio.us.com/popular/programming` includes the tags *python*, *c*, *compiler*, *performance*, *programming*. The crawler takes as input a single tag characterizing a domain (e.g. **programming**), and a number specifying the depth of the crawling process. For instance, for depth equal to 0, only the tag chains of the first page for the input tag are gathered. For depth equal to 1, the tag chains of the first page are gathered, and next, for each tag of each tag chain, the tag chains of the first page of that tag are also gathered. The above crawling process is done without considering the individual users tagging the resources. We are also interested in harvesting the tag chains of specific users. For this purpose we follow a similar process as that for harvesting the tag chains of a domain. Again, the depth parameter must be specified in order to retrieve the user-specific tag chains. These datasets provide all the necessary information in order to induce users'-specific conceptualizations and further classify the users to the domains. Obviously, deep crawling results in introducing tags that tend to be not closely related to the domain, thus introducing noise. For instance, starting with the tag **programming**, if the crawler reaches a depth of 4, then resources that are not closely related to programming appear, introducing also irrelevant tags. Having said that, we must point out that the specification of a domain with a single tag (as it is done here) is done without any loss of generality, since in case we were using a set of tags, the crawling process would consider the resources that have been tagged with all tags specified. However, starting the whole process from a single tag makes the whole task of inducing topic hierarchies more difficult, since a lot of heterogeneity is introduced in the set of tags gathered by the crawling process.

We have been running the crawler for a two-month period, for four domains, delineated by the following tags: **design**, **software**, **programming** and **web**. We have crawled Del.icio.us for each of these domains and for depth values 0 to 3. For each domain, we have created a corpus of documents. In particular, each tag chain is treated as a separate ("virtual") document. Regarding individual users, for the above time period, and for the aforementioned depths, we have gathered the tag chains of 300 users for each of the four domains. Table 1 summarizes the characteristics of the compiled dataset. The first column indicates the domain of interest. The next four columns indicate the different depths of crawling. Each cell shows the number of documents per data set. The last column indicates the number of users.

4 The Proposed Method

The proposed method is based on computing hierarchical topic models for specific domains. These models, constructed by exploiting the gathered tags, represent conceptualizations of those domains.

Table 1. Crawled data: Number of virtual documents per crawling depth and number of users.

Domain	d 0	d 1	d 2	d 3	Users
design	12	125	1511	4678	300
software	7	122	1417	2476	300
programming	7	84	931	1993	300
web	7	146	1073	5510	300

The dataset gathered for each domain (described in Section 3) constitutes the input for building the domain topic hierarchy. This process consists of two main steps: The first step creates a document - tag matrix of frequencies. The tags of the corpus constitute the features of the vector representation, whose values are the frequencies of the tags occurring in the documents. This matrix is the input to the second step, which induces the topic hierarchy. This is further described in detail below. We have to point out that the introduced method may skip this first step and instead use a domain ontology: In this case, the ontology concepts must be transformed to distributions over the common term space of the ontology and the crawled tags [20]. This is a subtlety that we plan to consider in our future work.

4.1 Hierarchy Learning

The proposed hierarchical learning algorithm is based on the Hierarchical Dirichlet Process (HDP) priors [17], as shown in figure 1b.

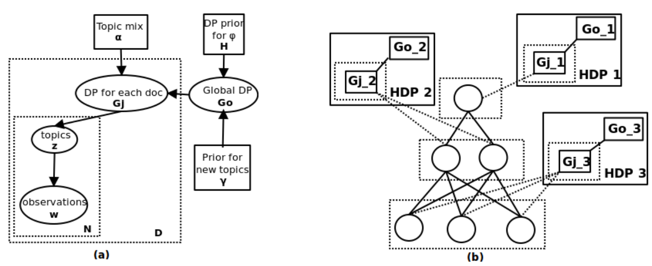

Fig. 1. (a): The HDP model. Assuming a corpus of D "virtual" documents, each of length N, there is a DP G_j for each document to draw tag distributions and a global, higher-level DP (G_0) that maintains the global distribution of tag distributions. (b): The hierarchy learning model. There is a HDP associated at each level, connected to all topics of that level.

A document (tag chain) consisting of N words (tags) is assumed to have been generated by a number of K latent topics. These topics have been drawn by a Dirichlet Process base measure G_j which in turn has been drawn from a Global Dirichlet Process G_0 that applies to the whole corpus of documents, assuring the sharing of topics among documents (Fig. 1a). The topics maintain a multinomial probability distribution over the words of the term space of the corpus (i.e. over the space of tags in the corpus). Thus, according to the generative process, to generate a document (i.e. a tag chain of a specific resource), topics are selected according to some probability, and then, for each selected topic, a word (tag) is selected from that topic, according again to some probability. In our case, where the corpus is given at hand, we perform the reverse process: The inference of the latent topics. We are interested therefore in the process where the model computes the topics and their hierarchical relations.

The proposed learning method, besides the fully automated estimation of the topic hierarchy per domain, supports the inference of the depth of the hierarchy, and by inheriting the characteristics of the HDP model, estimates the number of nodes at each level. All the above factors make the learning of the topic hierarchy completely parametric-less without relying on external resources or prior knowledge of the domain of interest.

In particular, according to the proposed method (Fig. 1b), at each level, there is a DP (G_j) for each document and a global DP (G_0) over all the DPs for that level. Therefore, each level of the topic hierarchy is associated with a HDP. An important characteristic of this approach is that the number of topics of each level is automatically inferred, due to the non-parametric Bayesian nature of the HDP, and it allows the topics at each level to be shared among the documents in the dataset.

The dataset provides the observations, i.e. the occurrence of tags, for the inference of the latent hierarchy. The process starts by inferring the lowest level of the hierarchy, i.e. the leaves. During this process, tags are assigned to leaf topics. Having inferred the leaf topics, their mixture proportions for the documents is known. In other words we can infer which topics have contributed, and to what degree, to the "generation" of each document. This type of inference has been used for the classification of user-specific virtual documents (tag chains) described in the following section.

Furthermore, the assignment of a tag to a specific topic constitutes the observation for the inference of the next level up. At the next levels up, following the same procedure, each inferred topic maintains a distribution over the tags of the virtual documents and over the topics at the level below. Therefore, each internal node or topic maintains a distribution over tags and over subtopics. The procedure is repeated until it converges to a single topic, which serves as the root of the hierarchy. The sampling scheme that we propose for the taxonomy learning method is described in Algorithm 1. More details may be found in [19].

Regarding the induced hierarchies, these contain hierarchical relations among topics rather than tags. Induced topics may index documents (i.e. tag chains) even if their constituent tags do not actually appear in a document, if this is consistent with the major patterns of association in the data. Doing so, synonym terms may end up in the same topic, and a polysemous term may exist in several topics.

User Classification based on Maximum Likelihood: Having computed the collective topic hierarchy for each of the domains, the first users' classification

Algorithm 1 Estimation of latent topic hierarchy.

DATA: Document - Tag matrix of frequencies
RESULT: Estimated topic hierarchy
set M=number of documents
set V=vocabulary size
estimate leaf topics K
set $T = K$
while $\mid T \mid > 1$ **do**
  // **transform document space**
  set $M = K$
  set input=MxV matrix of frequencies
  estimate topics K of next level up
  set $T = K$
end while

alternative computes the log-likelihood of each hierarchical model, given the documents (i.e. the dataset) of each user. The user is classified to the model that has the maximum likelihood, since it is assessed that this is the model that is able to "generate" the dataset of that user. The log-likelihood of the models is measured by using the Left-to-Right Sequential sampler [2]. It must be pointed out that as a consequence of this computation, the log likelihood of the specific topics that may have generated users' document are also computed: Doing so, the interest of users to specific domain topics is revealed.

User Classification based on Hierarchy Comparison: The second alternative for user classification, in conjunction to the computation of the collective models, creates a topic hierarchy for each user, using as input the user's tag chains. The process is the same as the one used for computing the collective conceptualization of each domain, as explained in Subsection 4.1. Then, the classification process continues as follows: having the collective model of each domain and the domain model of a particular user, the topic hierarchies are compared and the corresponding user is classified to the domain whose model is "closest" to the user's model. Closeness is measured by the metrics described in the following paragraphs. In order to compare two hierarchical topic models, we use the DMA distributional alignment method proposed in [20]. This method is mainly used for evaluating learned ontologies with respect to a gold standard. The main idea is to align the two ontologies, and based on the matchings to derive some scores that are inspired by the notions of Precision, Recall and F-measure. In our case, we treat the collective hierarchy as the gold one and the user-specific hierarchy as the learned one. The extensive experimental tests in [20] show that this method succeeds to reflect the deviation between the two hierarchies, taking also into account the differences between the hierarchies' structures and the deviations of the induced topics. It is also shown that the effectiveness of the alignment computed by this method is comparable to that of state of the art methods. Therefore, this method constitutes a firm basis for classifying users by exploiting domain-specific hierarchical topic models. Again, as a consequence of this comparison between models, topics in the collective model are compared to user-specific topics: Doing so, the interest of users to specific domain topics is revealed.

Briefly, the DMA alignment method proceeds as follows. Given that all nodes in both hierarchical models are represented as multinomial probability distributions over the tags of the dataset, the method proceeds to compute a common term (tag) space. This contains the union of the tags that the two models comprise. The nodes of the two hierarchies are now transformed to distributions over the common term space. Then, the collective topic hierarchy is compared to the user-specific hierarchy, by comparing the topics from the two hierarchies. For the computation of the similarity SD between different topics we have used the Total Variational Distance Measure (TVD) specified in Equation (1) and ranging in $[0, 1]$.

$$TVD = \frac{1}{2}\sum_{i} \mid P(i) - Q(i) \mid \tag{1}$$

In Equation (1), $P(\cdot)$ and $Q(\cdot)$ are multinomial probability distributions over tags in the compared topics. Therefore, the matching scheme compares the distributional representations of topics and finds the best correspondences between topics. Finally, Matching Precision MP, Matching Recall MR and the Matching F-measure MF[1] provide an assessment of user's topic hierarchy "closeness" to the collective topic hierarchy. The formulae for these measures are given in Equations (2), (3) and (4).

$$MP = \frac{1}{M}\sum_{i=1}^{M}(1 - SD_i)PCP_i \tag{2}$$

$$MR = \frac{1}{M}\sum_{i=1}^{M}(1 - SD_i)PCR_i \tag{3}$$

$$MF = \frac{(\beta^2 + 1)MP * MR}{(\beta^2 MR) + MP} \tag{4}$$

In Equations (2) - (4), M is the number of matchings between topics in both induced hierarchies. The PCP and PCR (*Probabilistic Cotopy Precision and Recall*) factors in Equations (2) and (3) respectively, are influenced by the notion of Semantic Cotopy [10]. The cotopy set of a topic C is the set of all its direct and indirect super and subtopics, including also the topic C itself. Thus, for a matching i, of a topic T in the user-specific hierarchy and a topic C in the collective hierarchy, PCP_i is defined as the number of topics in the cotopy set of T matched to topics in the cotopy set of C, divided by the number of topics participating in the cotopy set of T. For the same matching i, PCR_i is defined as the number of topics in the cotopy set of T matched to topics in the

[1] Originally, these measures are called P, R and F values, but since we use these standard measures for the evaluation of the proposed methods, we have renamed them .

cotopy set of C, divided by the number of topics participating in the cotopy set of C. Values of the MP, MR and MF measures close to 1 indicate that the user-specific topic hierarchy is close to the collective one, while values close to 0 indicate the opposite.

5 Empirical Evaluation

The empirical evaluation of the proposed methods concerns the classification of different users into four main domains: **design**, **programming**, **software** and **web**. This process constitutes a multi-class classification problem (in the sense that we have more than two classes) that we address in an unsupervised way. We provide quantitative results in terms of Precision, Recall and F-measure per domain, for both alternatives of user classification described in Section 4: Precision is defined as the ratio of the number of users correctly classified to a domain to the total number of users that are classified to that domain. Recall is the ratio of the number of users correctly classified to that domain to the number of users that should have been classified to that domain. The F-measure is the harmonic mean of Precision and Recall. The experiments have been performed using a 3 GHz PC with one core. In the worst case, the learning of the hierarchy of the collective conceptualizations requires approximately 32 minutes, while the classification task requires less than 5 minutes. The CPU intensive task corresponds to the learning of the hierarchies, which depends on the size of the dataset (i.e. the number of the tags).

We provide experimental results using the datasets for crawling depth (i.e. the depth for gathering the user-specific and user-independent tag chains per domain) equal to 1. We do consider this crawling depth for two reasons: (a) Tags gathered from greater depths result to hierarchies that contain certain portions maybe from different domains. (b) Given the tags of users and the computed collective conceptualizations, we have asked three external evaluators to classify the users into the four categories (**design**, **programming**, **software** and **web**) in order to use this classification as the ground truth. Gathering tags from depths greater than 1 would make the set of virtual documents per user much larger and the topics in the hierarchies would be more as well: This would make the job of evaluators much more harder and thus, error-prone.

To show an example of the induced hierarchies, Figure 2 illustrates the hierarchy for the domain **web**, using the dataset compiled for crawling depth equal to 1, as well the induced hierarchies of two users: One belonging to that domain and one that is not. The figure shows the estimated latent topics with the four most probable words from their multinomial probability distributions.

The evaluators have agreed for the classification of 285 different users per domain. Each user was classified to only one domain (multi-label classification, in the sense that a user may belong to more than one domain at the same time, is left for future work). For evaluation purposes, all 1140 users were put in a single directory. The aim is to classify each of these users in one of the four categories in an unsupervised way. Having said that, we must point out that the evaluation method is rather strict, since the classification problem is handled as a multi-class classification problem.

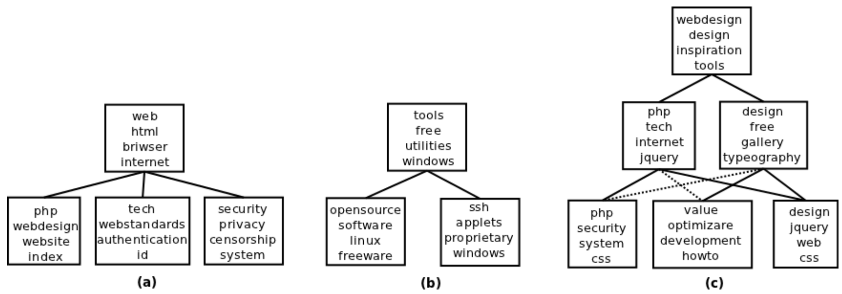

Fig. 2. The induced hierarchies (a) for a user belonging in the domain "web", (b) for a user that does not belong to that domain, and (c) for the domain "web" for crawling depth=1.

Table 2 provides experimental results for both classification alternatives. Regarding the classification process based on the log-likelihood, we observe that the F-measure ranges between 0.80 and 0.90, while the *Accuracy* of this method is equal to 0.865. We observe that the effectiveness of this classification method for the domains **programming** and **software** is lower than that reported for the other domains. This is so, since these two domains share many tags and it is rather difficult to classify users to one of them: This is something also experienced by the evaluators.

Table 2. Evaluation results for the two classification approaches.

	LogLikelihood Approach			DMA		
Domain	Precision	Recall	F-measure	Precision	Recall	F-measure
Design	0.98	0.82	0.89	0.91	0.80	0.85
Programming	0.99	0.67	0.80	0.96	0.68	0.80
Software	0.75	0.97	0.85	0.75	0.96	0.84
Web	0.83	1.0	0.90	0.82	0.92	0.87

One could choose to treat this task as a binary classification process, since we have a distinct model for each class (domain). In other words, one could choose to assess the method by measuring the classification accuracy per domain if as no other classes existed. That is, to assess for instance whether a particular user is classified to a specific domain (i.e. binary classification) without penalizing mis-classification to other domains. In that case, for all domains the precision would be equal to 1.0, boosting accordingly the corresponding F-measures to 0.90 for the domain **design**, 0.80 for the domain **programming**, 0.98 for the domain **software** and 1.0 for the domain **web**.

Regarding the second classification method (DMA) we observe that the F-measure ranges between 0.80 and 0.87, with *Accuracy* equal to 0.841. Again, if we address classification as a binary problem, then the precision of each domain would be equal to 1.0, and the F-measures would become 0.89 for the domain **design**, 0.81 for the domain **programming**, 0.98 for the domain **software** and 0.96 for the domain **web**. In order to increase the distinctive power of this classification alternative we experimented with different values of the parameter β of the MF measure so as to give more emphasis to the MR measure: This is motivated by the fact that by comparing the model of each user with the collective one, we prefer having a large number of correspondences between topics, rather than having few precise correspondences. By increasing β, we have observed a significant improvement of the evaluation results. Specifically, when setting $\beta = 2$, the F-measure of the domain **design** increases to 0.92, while the F-measure of the domain **web** increases to 0.93. The *Accuracy* of the method for this setting is equal to 0.876. Finally, when setting $\beta = 3$, the evaluation results are further improved. In particular, the F-measure of the domain **design** increases to 1.0, the domain **programming** increases to 0.82, the domain **software** increases to 0.861 and **web** is close to 1.0. The *Accuracy* of the method for this setting is equal to 0.92.

In order to gain a better insight on how the two proposed classification alternatives are related, we performed the following experiment: For each of the users that were classified correctly by both classification methods, we measured the log-likelihood of each topic in the collective model of each domain. This task computes the likelihood of each domain topic to index the tag chains of a user. The soft clustering that is performed during the learning of the hierarchy imposes that a tag chain may have been generated by more than one topic, with different proportions. Experimental results showed that in case a user U is classified under a specific domain D (i.e. the user's documents are indexed by some of the topics in the hierarchy for D), then all these topics in the domain hierarchy correspond to topics in the user's topic hierarchy. These topics show the particular interests of users to the specific domains.

6 Conclusions

Folksonomies are rapidly gaining momentum in the context of the Social Web. In this paper we presented methods for classifying the users of a folksonomy into hierarchical models that are induced by folksonomy data corresponding to a specific domain of interest. Specifically, given a set of tags, the proposed method is able to create a hierarchical topic model for a particular domain. From this point, two alternatives were proposed for user classification. One, based on the log-likelihood of the collective models to generate/index users' tag chains, and another, based on computing correspondences between the induced, collective hierarchical models and the user-specific induced models. Initial evaluation results provided illustrate the behavior of the proposed methods in both approaches of classification. We have observed promising results that suggest further investigation towards the direction of user classification in folksonomies.

Future plans include multi-label classification experiments, the application of the method to larger datasets with more users and for specific communities of users and experimentation with various probability matching schemes regarding the user classification based on the hierarchy comparison method.

References

1. A. Budura, D. Bourges-Waldegg, and J. Riordan. Deriving expertise profiles from tags. In *CSE (4)'09*, 2009.
2. W. Buntine. Estimating Likelihoods for Topic Models. In *ACML*, 2009.
3. F. Carmagnola, F. Cena, L. Console, O. Cortassa, C. Gena, A. Goy, I. Torre, A. Toso, and F. Vernero. Tag-based user modeling for social multi-device adaptive guides. *User Modeling And User-Adapted Interaction*, 18(5):497–538, 2008.
4. J. Diederich and T. Iofciu. Finding communities of practice from user profiles based on folksonomies. In *EC-TEL 2006 Workshop Proceedings*, 2006.
5. S. Golder and B. Huberman. Usage patterns of collaborative tagging systems. *Journal of Information Science*, 32(2):198–208, 2006.
6. H. Halpin, V. Robu, and H. Shepard. The dynamics and semantics of collaborative tagging. In *SAAW'06*, 2006.
7. P. Haymann and H. Garcia-Molina. Collaborative creation of communal hierarchical taxonomies in social tagging systems. In *Technical Report*, 2006.
8. H. Kim, S. Hwang, and H. Kim. Fca-based approach for mining contextualized folksonomy. In *Proc. of the ACM Symposium on Applied Computing*, 2007.
9. X. Li, L. Guo, and Y. Zhao. Tag-based social interest discovery. In *WWW*, 2008.
10. A. Maedche and S. Staab. Measuring similarity between ontologies. In *EKAW*, 2002.
11. A. Mathes. Folksonomies - cooperative classification and communication through shared metadata, December 2004.
12. P. Mika. Ontologies are us: A unified model of social networks and semantics. *Journal of Web Semantics*, 5(1):5–15, 2007.
13. M. Sanderson and B. Croft. Deriving concept hierarchies from text. In *ACM Conf. of the Special Interest Group in Information Retrieval*, pages 206–213, 1999.
14. C. Schmitz, A. Hotho, R. Jschke, and G. Stumme. Mining association rules in folksonomies. In *Data Science and Classification*, pages 261–270, 2006.
15. P. Schmitz. Inducing ontology from flickr tags. In *Collaborative Web Tagging Workshop*, 2006.
16. M. Szomszor, H. Alani, I. Cantador, K. O'Hara, and N. Shadbolt. Semantic modelling of user interests based on cross-folksonomy analysis. In *ISWC*, 2008.
17. Y. Teh, M. Jordan, M. Beal, and D. Blei. Hierarchical dirichlet processes. *Journal of the American Statistical Association*, 2006.
18. H. Wu, M. Zubair, and K. Maly. Harvesting social knowledge from folksonomies. In *Proc. of the 17th Conference on Hypertext and Hypermedia*, 2006.
19. E. Zavitsanos. *Learning Ontologies from Text Collections and Evaluating them Against Gold Standards.* PhD Thesis, University of the Aegean, 2009.
20. E. Zavitsanos, G. Paliouras, and G. A. Vouros. Gold standard evaluation of ontology learning methods through ontology transformation and alignment. *TKDE*, `http://doi.ieeecomputersociety.org/10.1109/TKDE.2010.195` (to appear).
21. M. Zhou, S. Bao, X. Wu, and Y. Yu. An unsupervised model for exploring hierarchical semantics from social annotations. In *ISWC/ASWC2007*, 2007.