A Semantic Tag Recommendation Framework for Collaborative Tagging Systems

Conference Paper · October 2011

DOI: 10.1109/PASSAT/SocialCom.2011.170 · Source: DBLP

3 authors, including:

Konstantinos N. Vavliakis
Aristotle University of Thessaloniki

Pericles A. Mitkas
Aristotle University of Thessaloniki

2011 IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing

A Semantic Tag Recommendation Framework for Collaborative Tagging Systems

Zinovia I. Alepidou*, Konstantinos N. Vavliakis*,† and Pericles A. Mitkas*,†
*Electrical and Computer Engineering, Aristotle University of Thessaloniki
†Informatics and Telematics Institute, CERTH
Thessaloniki, Greece
Email: zalepidou@issel.ee.auth.gr, kvavliak@issel.ee.auth.gr, mitkas@eng.auth.gr

***Abstract*—In this work we focus on folksonomies. Our goal is to develop techniques that coordinate information processing, by taking advantage of user preferences, in order to automatically produce semantic tag recommendations. To this end, we propose a *generalized tag recommendation framework* that conveys the semantics of resources according to different user profiles. We present the integration of various models that take into account content, historic values, user preferences and tagging behavior to produce accurate personalized tag recommendations. Based on this information we build several Bayesian models, we evaluate their performance, and we discuss differences in accuracy with respect to semantic matching criteria, and other approaches.**

***Keywords*-Folksonomy; tagging; recommendation; personalization; semantic evaluation.**

I. Introduction

Folksonomies allow users to describe a set of objects using freely chosen keywords. Formally, folksonomies can be regarded as a tripartite graph of three disjoint sets of vertices: *a)* users, $U = \{a_1, ..., a_n\}$, *b)* resources $R = \{r_1, ...r_n\}$ and *c)* tags, $T = \{t_1, ..., t_n\}$, thus defined by a set of annotations $F \subseteq U \times R \times T$.

The selection of keywords depends on user's comprehension of the concept/entity being tagged. Different attributes of the resource, i.e. title, content, author, creation date, as well as individual user attributes, i.e. preferences, profile and background, in correlation with the tagging time and date, contribute towards creating user's individual perception on a particular concept/entity. Our main goal is to produce user-centric, yet generalized, algorithms which simulate human behavior towards tagging.

Three are the main contributions of our work: 1) We provide a modular tag recommendation framework, built on user-centric Bayesian models. 2) We combine and evaluate different recommendation algorithms, and we show that different techniques should be used, depending on the available information. 3) We provide a linguistic approach to tag recommendation, indicating that various algorithms may yield different results, depending on the evaluation method.

The paper is organized as follows: Section II presents related work, while Section III outlines the preprocessing steps, and the Bayesian models developed. Section IV discusses semantic and string based evaluation and the last section presents our conclusions and future work.

II. Related work

Numerous methodologies, such as statistical models, multilabel classifiers and collaborative systems that combine information from multiple resources have been proposed for tag recommendation. These methodologies are based on resource attributes and have been applied in different fields, such as YouTube, Flickr and music resources. In all cases the goal is to accurately specify the resource's content, in order to elicit words semantically related to it. The majority of tag recommendation systems are based on statistical models. These systems propose tags based on TF-IDF weights (term frequency - inverse document frequency) built from the information contained in the content of the resource, the user profile, and the resource history.

Other techniques focus on eliciting latent concepts included in the document. LDA (Latent Dirichlet Allocation) [1], ACT (Author-Conference-Topic) [2] and NMF (Non-Negative Matrix Factorization) classify a resource based on its main concepts. These algorithms attribute resources to appropriate categories and then propose a corresponding set of tags. Furthermore, well accepted machine learning techniques, like clustering, logistic regression, and classification trees [3] have been extensively applied in order to produce tag recommendations. Many methodologies concentrate only on the content of resources. In this context, [4] uses FDT (Feature-Driven Tagging), a fast method that indexes tags by features and then creates a list of weighted tags that form the pool of possible suggestions. Moreover, [5] calculates the importance of words derived from the resource and combines information sources with barely-used tags.

A group of recommendation techniques propose tags based on similarities among posts. These similarities may refer to the users of posts, the documents they are about to mark, or the tags they assign to the resources [5]. In [6], Marinho et al. introduce relational classifiers that take into consideration information derived from all the resource's attributes, as well as latent semantic relations between posts. For example, a scientific publication may be related to another one if both have been written by the same author or have the same citations. Finally, Lipczak et al. in [7], integrate various recommendation mechanisms and generate a final tag recommendation set with high accuracy.

III. Tag recommendation

In the preprocessing phase we start by building a lexicon as a repository where words extracted from the resource are connected to a number of tags. Thus, final tag suggestion consists of the *n* tags with the higher co-occurrence probability based on the words that constitute the resource's content. We created the lexicon using a linguistic approach. Linguistic types such as articles, pronouns and conjunctions lack semantic content, thus together with common stop-words were discarded. We used Wordnet to apply stemming in order to acquire the root of each term and enrich our lexicon by adding the capability of hypernym search, alongside word search. Our system was built upon resource's title, history, and other tags. Next, we briefly discuss on each of the three models.

A. Tag recommendation based on resource's title

The title of a resource is a key indicator for its content, thus tags may be extracted from resource's attributes without further elaboration. Nouns, adjectives and verbs are the richest parts of speech, as far as semantic content is concerned. WordNet allows us to retrieve only these parts of speech from text, given the fact that it can recognize the linguistic type of a word. It is worth noticing that some domain specific terms, that convey semantic meaning only within specific subjects or specific communities (e.g. Folksonomy, elearning, bibtex) are not contained in WordNet. These Non-WordNet terms are regarded as candidates for tag recommendations as they hold a significant proportion of the terms contained in the training dataset.

We developed a "TitleRecommender" model, which detaches words from the basic resource's attribute (title or URL), and suggests them as tags. Adjectives, nouns and Non-WordNet terms are extracted from the resource and proposed as tags. Furthermore, tag suggestions may also consist of words extracted from the primary attributes of a resource after a weighting process. The so-called "Bayes (on the title words)" method weights terms by calculating their probability of being suggested as tags provided that they simultaneously occur in the content of the post.

B. Tag recommendation based on history

In case a resource already exists in the folksonomy, its history, that is tags other users have already assigned to it, is a valuable input for building a recommendation model. Likewise, if a user has tagged other resources, the tags he/she has used should be treated as candidates in his/her future tagging actions. If a certain post is already tagged by other users, it may be wise to recommend the tags mostly related to it. Similarly if someone uses specific terms while tagging, it is a strong indication that they have a limited "tagging vocabulary". This model suggests the tags that appear most frequently in the resource's history ("Resource History Tags") or the user's history ("User History Tags") of a post. If both the resource and the user appear for the first time in the system, we can propose the most frequent tags of the training dataset in general ("MostFrequentTags" method).

In an attempt to take advantage of user's personomy (user's past tagging behavior), we created a method called "UserPersonomy", partially presented in Algorithm 1. Whereas "User History Tag" method simply suggests the most popular tags in the user profile, "UserPersonomy" goes one step further, and counts the times a term derived from a resource's primary attribute is also present in user personomy. In other words, this method is a union of "TitleRecommender" and "User History Tags", combining the former's effectiveness, and the latter's subjectivity.

Algorithm 1 Partial pseudo-code for "UserPersonomy".

for $\mathrm{r_i}$ resources in user's personomy **do**
  extract $\mathrm{w_i}$ terms of 'TitleRecommender'
  for $\mathrm{w_i}$ terms in the title **do**
    $Score_{\mathrm{w_i}} = \sum_{\mathrm{i=i}} w_{\mathrm{i}} \cap \ tag_{\mathrm{i}}, user$
suggest w_{i} with the highest $Score_{\mathrm{w_i}}$

C. Tag to tag recommendation

In some cases it may be useful to recommend tags based on other recommended tags, especially when the latter are provided with high probability. Given a tag, other tags can be produced from its etymologically related terms, like synonyms and hypernyms. This type of meta-recommenders may support the basic recommenders by improving the quality and quantity of the recommendation, especially when the main recommender fails to provide a sufficient number of tags. Nevertheless, the main drawback of this technique is that it cannot be used independently, since its results depend on the basic recommenders outcome. To this end, we developed a Bayesian "TagToTagRecommender" which estimates tags occurrence probabilities given other tags occurrence probability (Eq. 1). Hence, suggesting one tag t_j leads to suggesting other tags t_i related to the first one.

$$P(t_i|t_j) = \frac{P(t_j|t_i)P(t_i)}{P(t_j)} \tag{1}$$

IV. Dataset analysis and results

A. Dataset description

We trained and evaluated the set of algorithms described in Section III based on different datasets created from Bibsonomy.org[1] data. Our training dataset contains 64,120 resources and 253,615 tag assignments. The test datasets were provided by the Discovery Challenge of the ECML/PKDD conferences that took place in 2008 and 2009. We tested our

[1] http://www.bibsonomy.org/

framework on three datasets: *a)* dataset D2009_1 contains posts not included in the training dataset, thus recommenders cannot rely on post history, *b)* dataset D2009_2 contains posts also included in the training dataset, as they may have been tagged by other users, and *c)* D2008 contains both previously seen and unseen posts.

With string-based evaluation we can compare our results with other methodologies in an objective and equal basis. On the other hand, thanks to semantic-based evaluation, we evaluate our recommenders' performance from an etymological point of view.

B. Semantic-based Evaluation

We computed the semantic similarity between proposed and true tags by computing the second order similarity between two input words based on their sets of distributionally similar words. Then, we combined the individual semantic similarities using a Heuristic Matcher. The similarity is computed by dividing the sum of similarity values of all match candidates of both sets (X: true tags, Y: proposed tags) by the total number of set tokens, as in Equation 2.

$$SemanticSimilarity = \frac{2 * Match(\mathbf{X}, \mathbf{Y})}{|\mathbf{X}| + |\mathbf{Y}|} \qquad (2)$$

The evaluation results for each test dataset and for all trained models are provided in Table I. We define the following notation of sets of tags used in this table: **a)** *Title* is the tag set proposed by "TitleRecommender", **b)** *Resource* is the tag set suggested by "Resource History Tags", **c)** *User* is the outcome of "User Personomy", **d)** *TagToTag* is the tags proposed by "TagToTag Recommender", if User and Resource are used as input, **e)** *WeightWords* is the set of tags proposed by the method "Bayes (on title words)", **f)** *Synonyms* contains the synonyms of tags proposed by "TitleRecommender", and **g)** *Hypernyms* contains the hypernym of tags proposed by "TitleRecommender".

The term *Union* indicates that the individual sets are united to one set without further elaboration. The final set consists of the tags contained in each preliminary set. The term *Combine* implies that final suggestions derived from each model undergo through a process so that tags contained in the preliminary sets are valued according to their occurrence frequency and placed into the final set in descending order.

Based on the results of Table I, content-based methods exhibit good results in general. History-based approaches perform poorly when prior information is missing (D2009_1) but may outperform content-based methods when used with resources that have already been tagged by other users, or when using the past "tagging vocabulary" of users (D2009_2). As it turns out, combining these single models provide the best results in both cases that user, resource or tags are contained in the training dataset, and when they are not. This fact is confirmed in the D2008 dataset where both resources with known history and new ones are included.

Table I
SEMANTIC EVALUATION RESULTS.

Recommendation Methods Trained Models	Test Datasets D2008	D2009_1	D2009_2
1.Bayes	0.1416	0.1176	0.2548
2.Hypernym Bayes	0.1291	0.1036	0.2050
3.Title Recommender	0.2657	0.1630	0.2379
4.Bayes (on title words)	0.1950	0.1265	0.2285
5.Resource History Tags	0.0112	0.0071	0.2670
6.User History Tags	0.0030	0.0401	0.0151
7.UserPersonomy	0.0012	0.0357	0.1532
8.TagToTagRecommender	0.0054	0.0045	0.1748
9.TagToTagSynonym/Hypernym	0.1684	0.1185	0.2378
10.Union(Title, Resource)	0.2697	0.1602	0.2689
11.Union(User, Resource)	0.0120	0.0417	0.2975
12.Union(User, Rsc,TagToTag)	0.0112	0.0392	0.2954
13.Union(User, Resource, Title)	0.2691	0.1605	0.2955
14.Union(User, Resource, WeightsOnTitleWords)	0.1977	0.1277	0.2980
15.Combine(Title, User, Rsc)	0.1429	0.1215	0.2717
16.Combine(Title, Synonyms, Hypernyms, User, Resource)	0.1396	0.1201	0.2606

C. String-based Evaluation

Next we compared our system with other approaches. We used the results of the participants for ECML/PKDD Discovery Challenges that took place in 2008 and 2009 that are available online[2,3]. Figure 1 depicts all the results ordered according to the F-Measure value achieved. One can observe that our method outperforms other approaches in the D2008 test dataset, while it comes fourth for both D2009_1 and D2009_2 datasets. It is worth noticing that due to the ability to combine multiple single models our approach manages to have good results for both D2009_1 and D2009_2, while other approaches may perform good in one dataset but fail in others.

Figure 2 provides an overview of both the semantic similarity evaluation and the string-based evaluation of all the proposed models for two of the test datasets. One can notice that the use of hypernyms boosts the performance of the recommender, a fact that string-based evaluation cannot identify. Moreover, "Resource History Tags", and the complex models using it, outperform the rest of the models.

When it comes to posts already introduced to the system, it is wiser to propose tags which have a stronger relation with the user's past behavior than with the resource's semantic content. Nevertheless, even though users tend to use a specific "tagging vocabulary", they will use tags semantically close to the ones assigned to the resource by other users, so that the final set of proposed tags best describes the resource's main concept.

[2]http://www.kde.cs.uni-kassel.de/ws/rsdc08/
[3]http://www.kde.cs.uni-kassel.de/ws/dc09/results/

Figure 1. Comparison of our approach to other approaches.

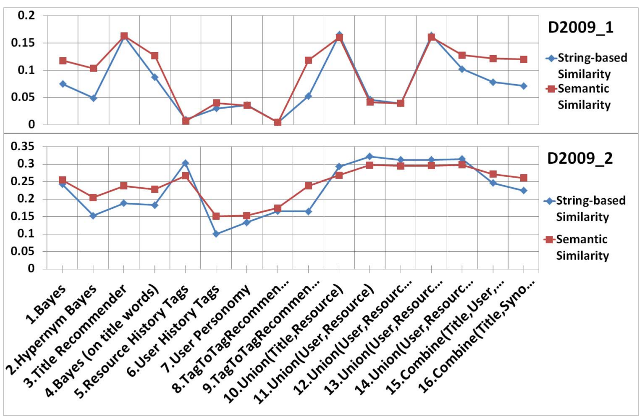

Figure 2. String-based vs Semantic evaluation. String-based similarity values represent F-Measure. while Semantic evaluation values are calculated according to Eq. 2.

V. Conclusions and Future Work

Two are the main conclusions derived from our work: **1)** In order to provide an efficient recommendation technique we should combine different methods, as in different cases different algorithms may perform better. Thus, by creating various simple models, we can combine them and develop complex systems adaptable to various situations. **2)** To provide reliable recommendations one should train the recommendation algorithms based on semantic matching criteria. Thus, one can observe semantic correlations between tags assigned by different users and the resource's content itself.

Future work includes research on self-adaptive techniques for identifying the optimal combinations of different simple methods that best suit each user and each dataset. Further evaluation tests are needed in order to acquire statistically significant results and elaborate on the differences between string-based and semantic-based performance. The determination of the best overall evaluation criterion remains an open question.

References

[1] R. Krestel and P. Fankhauser, "Tag recommendation using probabilistic topic models," in *ECML PKDD DC 2009*, vol. 497. Bled, Slovenia: CEUR Work. Proc., 2009, pp. 131–141.

[2] N. Zhang, Y. Zhang, and J. Tang, "A tag recommendation system based on contents," in *ECML PKDD DC 2009*, vol. 497. Bled, Slovenia: CEUR Work. Proc., 2009, pp. 285–295.

[3] I. Katakis, G. Tsoumakas, and I. Vlahavas, "Multilabel Text Classification for Automated Tag Suggestion," in *Proceedings of the ECML/PKDD 2008 DC*, 2008.

[4] Z. Liu, C. Shi, and M. Sun, "Folkdiffusion: A graph-based tag suggestion method for folksonomies," in *Information Retrieval Technology*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2010, vol. 6458, pp. 231–240.

[5] S. Ju and K.-B. Hwang, *A Weighting Scheme for Tag Recommendation in Social Bookmarking Systems*. CEUR Work. Proc., 2009, vol. 497, pp. 109–118.

[6] L. Marinho, C. Preisach, and L. Schmidt-Thieme, "Relational classification for personalized tag recommendation," in *ECML PKDD DC 2009*, vol. 497. Bled, Slovenia: CEUR Work. Proc., 2009, pp. 7–15.

[7] M. Lipczak and E. Milios, "Learning in efficient tag recommendation," in *Proceedings of the fourth ACM conference on Recommender systems*, ser. RecSys '10. New York, NY, USA: ACM, 2010, pp. 167–174.

[8] M. Lipczak, *Tag Recommendation for Folksonomies Oriented towards Individual Users*. Workshop at 18th Europ. Conf. on Machine Learning (ECML'08) / 11th Europ. Conf. on Principles and Practice of Knowledge Discovery in Databases (PKDD'08), 2008.

[9] Y. Zhang, N. Zhang, , and J. Tang, "A collaborative filtering tag recommendation system based on graph," in *ECML PKDD DC 2009*, vol. 497. CEUR Work. Proc., 2009, pp. 297–305.

[10] S. Rendle and L. Schmidt-Thieme, "Factor models for tag recommendation in bibsonomy," in *ECML PKDD DC 2009*, vol. 497. CEUR Work. Proc., 2009, pp. 235–242.

[11] R. Wetzker, A. Said, and C. Zimmermann, "Understanding the user: Personomy translation for tag recommendation," in *ECML PKDD DC 2009*, vol. 497. CEUR Work. Proc., 2009, pp. 275–284.

[12] M. Tatu, M. Srikanth, and T. D'Silva, *Tag Recommendations using Bookmark Content*. Workshop at 18th Europ. Conf. on Machine Learning (ECML'08) / 11th Europ. Conf. on Principles and Practice of Knowledge Discovery in Databases (PKDD'08), 2008.

[13] J. Mrosek, S. Bussmann, H. Albers, K. Posdziech, B. Hengefeld, N. Opperman, S. Robert, and G. Spira, "Content- and graph-based tag recommendation: Two variations," in *ECML PKDD DC 2009*, vol. 497. Bled, Slovenia: CEUR Work. Proc., 2009, pp. 189–199.