Knowledge-Based Systems 67 (2014) 131–142

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

# Class-based tag recommendation and user-based evaluation in online audio clip sharing

Frederic Font [a,*], Joan Serrà [b], Xavier Serra [a]

[a] Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain
[b] Artificial Intelligence Research Institute (IIIA-CSIC), Spanish National Research Council, Bellaterra, Spain

ARTICLE INFO

*Article history:*
Received 8 November 2013
Received in revised form 22 April 2014
Accepted 2 June 2014
Available online 14 June 2014

*Keywords:*
Collaborative tagging
Tag recommendation
User study
Folksonomy
Freesound

ABSTRACT

Online sharing platforms often rely on collaborative tagging systems for annotating content. In this way, users themselves annotate and describe the shared contents using textual labels, commonly called tags. These annotations typically suffer from a number of issues such as tag scarcity or ambiguous labelling. Hence, to minimise some of these issues, tag recommendation systems can be employed to suggest potentially relevant tags during the annotation process. In this work, we present a tag recommendation system and evaluate it in the context of an online platform for audio clip sharing. By exploiting domain-specific knowledge, the system we present is able to classify an audio clip among a number of predefined audio classes and to produce specific tag recommendations for the different classes. We perform an in-depth user-based evaluation of the recommendation method along with two baselines and a former version that we described in previous work. This user-based evaluation is further complemented with a prediction-based evaluation following standard information retrieval methodologies. Results show that the proposed tag recommendation method brings a statistically significant improvement over the previous method and the baselines. In addition, we report a number of findings based on the detailed analysis of user feedback provided during the evaluation process. The considered methods, when applied to real-world collaborative tagging systems, should serve the purpose of consolidating the tagging vocabulary and improving the quality of content annotations.



## 1. Introduction

Free-form semantically-meaningful textual labels, called tags, are extensively used in online sharing platforms for describing and annotating contents. Systems that provide the functionality for making these annotations are normally referred to as collaborative tagging systems. Several problems arise when users annotate shared and/or online resources [9]. The most typical ones are tag scarcity, the use of different tags to refer to a single concept (synonymy), the ambiguity in the meaning of certain tags (polysemy), the commonness of typographical errors, the use of user-specific naming conventions, or the use of different languages. To minimise some of these problems, tag recommendation systems can be employed to suggest potentially relevant tags during the annotation process [14]. As users are exposed to the suggestions of the system, the annotation process partially shifts from the creation of textual labels to the recognition of tags in a list [23], and thus all users receive a certain common influence from the system. Hence, tag recommendation serves the purpose of consolidating the vocabulary of collaborative tagging systems [13].

In general, tag recommendations are either based on content analysis of online resources or in the other tags that users introduce during the annotation process. In the case of content-based recommendations, a typical approach consists in, given a resource to be described, defining a neighbourhood of other resources (based on some similarity measure) and then recommending tags that are used to annotate resources in this neighbourhood [12,24]. Another approach is the use of machine learning techniques to learn mappings between tags and content features [15,25,26]. On the other side, there are tag recommendation strategies which are based on the tags that users introduce during the annotation process itself, prior to the moment of the recommendation. Disadvantages of these strategies compared to content-based recommendation methods are that they require the existence of at least one tag to provide recommendations, whereas content-based recommendation systems can provide recommendations to resources with no associated tags or other metadata. Nevertheless, tag recommendation methods based on the tags that users introduce during the annotation process have the advantage of

* Corresponding author. Tel.: +34 935422101.
*E-mail address:* frederic.font@upf.edu (F. Font).

http://dx.doi.org/10.1016/j.knosys.2014.06.003
0950-7051/

not requiring any specific processing of the content of the resources being annotated, thus being typically less expensive in terms of computation resources and being more easily generalisable to other multimedia domains. These methods usually consider the *folksonomy* (i.e., the set of associations between tags, users and content resources) of a collaborative tagging system to estimate tag similarity from their resource co-occurrence. In this way, candidate tags can be selected according to their similarity to the introduced tags, and a sorting algorithm can rank them in terms of estimated relevance [4,8,14,22]. In previous work, we described and evaluated a general scheme for folksonomy-based tag recommendation in collaborative tagging systems [7]. Out of that scheme, eight particular methods were proposed which form the basis of the method presented in this work.

Besides content-based and folksonomy-based tag recommendation systems, other approaches have been described in the literature. Anderson et al. [1] describe a tag recommendation system for Flickr,[1] a well known photo sharing site, which combines both content-based recommendations (by training a predictive model that learns the mapping between tags and extracted content image features) with folksonomy-based recommendations (following an strategy very similar to [22]). Naaman and Nair [19] describe another tag recommendation system for Flickr, which takes advantage of the geolocation metadata attached to images and recommends tags that other users employed in close areas. Chen et al. [3] describe a tag recommendation system for video resources which crawls the web for information about these videos and identifies keywords to recommend as tags.

Although it is quite common to personalise tag recommendation systems to the tagging behaviour of particular users by promoting, for example, tags that users introduced in past annotations [2,8,14,16,18,20], most of the current systems do not introduce direct user feedback in the evaluation loop. Thus recommendations are generally evaluated using traditional information retrieval approaches based on the comparison of tag rankings produced by different methods, or using precision and recall metrics computed after a tag prediction task [2,7,8,16,18,20]. To the best of our knowledge, only three studies perform some kind of user-based evaluation. Sigurbjörnsson and Zwol [22] automatically generate tag recommendations for several images from a Flickr dataset and then ask users to rate, in a four-point scale, whether the recommendations are appropriate to a given image. Similarly, De Meo et al. [4] extend the annotations of Delicious' bookmarks[2] and then ask users to evaluate the relevance of every tag/resource association. Jäschke et al. [13] perform a small evaluation based on a real-world scenario where users have to tag bookmarks in BibSonomy.[3] Specifically, precision and recall metrics are computed by comparing tag recommendations performed to every bookmark and the final taglines that users introduced. Due to its subjectiveness and many different ways to be accomplished, tag recommendation is not an easy task to evaluate, and some advantages and disadvantages can be found in both user-based and information retrieval evaluation approaches [8]. However, there is a clear lack of user-based evaluation in previous work, and we believe that every recommendation system should be validated at some point using both evaluation strategies. Proper user feedback should be helpful not only to compare tag recommendation methods but also to better understand the nature of the task and learn how can systems be improved.

The contribution of the present work is twofold. First, we propose an extended version of the best performing tag recommendation method found in our previous work [7]. The main idea behind this extended method is to exploit the automatic classification of the resources to be annotated into a number of predefined classes to further adapt the tag suggestions to the context of these classes. This classification is based on the tags that users start introducing during the annotation process. In this way, instead of personalising recommendations for particular users, we "personalise" them to particular classes of resources. Next, as a second contribution, we perform a comprehensive user-based evaluation through an online experiment were participants are presented with some resources which have to be annotated with the help of a tag recommendation system. These kind of user-based evaluations are very costly and we have seen that they are not very common in the tag recommendation literature. For that reason, we believe our contribution is of great value to the community. In our evaluation, we compare the recommendation method we proposed in previous work and the extended version we describe here along with two random baselines. Moreover, we perform a complementary evaluation based on a tag-prediction task following common information retrieval methodologies. In our previous work [7], the tag recommendation methods were evaluated using a tag-prediction task and compared favourably against four baselines and two state of the art methods [8,22]. For this comparison, we used data from the folksonomies of Freesound,[4] an online audio clip sharing site with more than 3,5 million registered users and 180,000 uploaded sounds [5], and Flickr. Therefore, the recommendation methods were tested in the audio and image domains. Similar results were obtained in both scenarios. In this work, evaluations are carried out in the context of Freesound. Results show that the newly proposed recommendation method brings a statistically significant improvement over the previous method, according to both user-based and prediction-based evaluations. Analysing user-based evaluation results we find that participants which are experienced in working with sound libraries tend to better appreciate the improvements of the new tag recommendation method we describe here. Moreover, we see that the more familiarised the users are with Freesound, the more the number of tag suggestions they accept as valid annotations. User feedback reveals that tag recommendation methods tend to be more useful when recommending broad tags (i.e., referring to generic concepts). Participants also recognise tag annotation as a particularly difficult task, specially if the resources being annotated are not authored by themselves.

The rest of the paper is organised as follows. First, we summarise the steps of the tag recommendation method we proposed in previous work and describe the new approach based on the classification of input tags (Section 2). Then, we describe the online experiment we designed for user-based evaluation (Section 3). Results of the online experiment are reported in Section 4, and the complementary prediction-based evaluation is described and reported in Section 5. We conclude the paper with a discussion about our findings and future work (Section 6).

## 2. Tag recommendation methods

The two tag recommendation methods we describe in this work are based on tag–tag similarities derived from the folksonomy of Freesound. Given a set of input tags $\Gamma_I$, the methods output a set of recommended tags $\Gamma_R$.

### 2.1. General tag recommendation

The general tag recommendation method presented in [7], which we denote by GEN, consists of three steps (Fig. 1):

[1] www.flickr.com.
[2] www.delicious.com.
[3] www.bibsonomy.org.
[4] www.freesound.org.