Folksonomy-Based Tag Recommendation for Collaborative Tagging Systems

Frederic Font, Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain
Joan Serrà, Artificial Intelligence Research Institute (IIIA-CSIC), Spanish National Research Council, Bellaterra, Spain
Xavier Serra, Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain

ABSTRACT

Collaborative tagging has emerged as a common solution for labelling and organising online digital content. However, collaborative tagging systems typically suffer from a number of issues such as tag scarcity or ambiguous labelling. As a result, the organisation and browsing of tagged content is far from being optimal. In this work, the authors present a general scheme for building a folksonomy-based tag recommendation system to help users tagging online content resources. Based on this general scheme, the authors describe eight tag recommendation methods and extensively evaluate them with data coming from two real-world large-scale datasets of tagged images and sound clips. Their results show that the proposed methods can effectively recommend relevant tags, given a set of input tags and tag co-occurrence information. Moreover, the authors show how novel strategies for selecting the appropriate number of tags to be recommended can significantly improve methods performances. Approaches such as the one presented here can be useful to obtain more comprehensive and coherent descriptions of tagged resources, thus allowing a better organisation, browsing and reuse of online content. Moreover, they can increase the value of folksonomies as reliable sources for knowledge-mining.

Keywords: Collaborative Tagging, Folksonomies, Intelligent System, Online Digital Content, Recommendation System

1. INTRODUCTION

Collaborative tagging has emerged as a common and successful solution for labelling and organising huge amounts of digital content, being adopted by many well-known sites such as Youtube, Flickr, Last.fm, or Delicious (Marlow, Naaman, Boyd, & Davis, 2006). In collaborative tagging, users assign a number of free-form semantically-meaningful textual labels (tags) to information resources. These tags can be then used for many purposes, including retrieval, browsing and categorisation (Bischoff, Firan, Nejdl, & Paiu, 2008). For

DOI: 10.4018/jswis.2013040101
instance, they can be used for matching user queries with resources tags, or for building tag clouds to navigate across resources. Such usages are of special importance for platforms that share multimedia content such as videos, images, or audio, since such contents cannot be so directly and straightforwardly indexed as it would be done with textual data like books or web pages (Bischoff, Firan, Nejdl, & Paiu, 2008). Because of this importance, collaborative tagging systems have been widely researched in the last few years. In particular, a focus has been given to collaborative tagging dynamics and user behaviour (Marlow, Naaman, Boyd, & Davis, 2006; Halpin, Robu, & Shepard, 2006; Golder & Huberman, 2006; Farooq, Kannam-pallil, Song, Ganoe, Carroll, & Giles, 2007) and to automatic tag classification methods based on user motivations ([5, 6]). Nevertheless, collaborative tagging systems suffer from a number of well-known issues (Halpin, Robu, & Shepard, 2006; Cantador, Konstas, & Jose, 2011), which include tag scarcity, the use of different labels to refer to a single concept (synonymy), the ambiguity in the meaning of certain labels (polysemy), the commonness of typographical errors, the use of user-specific naming conventions, or even the use of different languages. One strategy for trying to overcome these problems, and thus to obtain more comprehensive and consistent tag assignments, is the use of tag recommendation systems to help users in the tagging process (Jäschke, Marinho, Hotho, Schmidt-Thieme, & Stumme, 2007). In that case, when users are labeling online resources, tag recommendation systems automatically suggest new tags that can also be meaningful or relevant for the resource being described. This way, tag recommendation serves the purpose of consolidating the tag vocabulary among users in a collaborative tagging system (Jäschke, Marinho, Hotho, Schmidt-Thieme, & Stumme, 2007). In addition, tag recommendation systems can be used, in an off-line mode, to extend the descriptions of information resources by automatically adding new tags.

Here we describe a general scheme for tag recommendation in large-scale collaborative tagging systems. Our approach is folksonomy-based, meaning that we do not perform any content analysis of the information resources for which we perform tag recommendations, but uniquely rely on the tag co-occurrence information that can be derived from the folksonomy itself. A particularly interesting aspect of our tag recommendation scheme is a step focused on automatically selecting the number of tags to recommend given a list of candidates with assigned scores. Other tag recommendation methods found in the literature generally do not consider this aspect and evaluate their solutions at different values of $K$ recommended tags (see Related work). Moreover, as the scheme we describe only relies on tag information derived from a folksonomy, it is rather domain-independent and could be easily adapted to other collaborative tagging systems, either alone or as a complement of more specific content-based strategies. We believe that a tag recommendation method such as the one we propose here can be useful to obtain more comprehensive and coherent descriptions of tagged resources, and help the emergence of less noisy and more consistent folksonomies. This can greatly benefit organisation, browsing and reuse of online content, and also leverage the value of folksonomies as reliable sources for knowledge-mining (Al-Khalifa & Davis, 2007; Limpens, Gandon, & Buffa, 2009).

We propose eight tag recommendation methods which are based on the aforementioned general scheme. The proposed methods, jointly with several baselines, are evaluated with data coming from Freesound$^1$ (an online audio clip sharing site with more than two million registered users and 150,000 sounds (Akkermans, Font, Funollet, De Jong, Roma, Togias, & Serra, 2011) and Flickr$^2$ (a well known photo sharing site that, according to Wikipedia (Wikipedia, 2012), has more than 50 million registered users and six billion photos). For the best scoring methods, we also analyse the impact of their configurable parameters. Overall, we have
Related Content

Data Mining, Validation, and Collaborative Knowledge Capture
[www.irma-international.org/chapter/data-mining-validation-collaborative-knowledge/65692/](http://www.irma-international.org/chapter/data-mining-validation-collaborative-knowledge/65692/)

Knowledge Based Business Intelligence for Business User Information Self-Service
[www.irma-international.org/chapter/knowledge-based-business-intelligence-business/65697/](http://www.irma-international.org/chapter/knowledge-based-business-intelligence-business/65697/)

An Approach to a Semantic Recommender System for Digital Libraries
[www.irma-international.org/chapter/approach-semantic-recommender-system-digital/52866/](http://www.irma-international.org/chapter/approach-semantic-recommender-system-digital/52866/)

An Enhanced Semantic Layer for Hybrid Recommender Systems: Application to News Recommendation
[www.irma-international.org/article/enhanced-semantic-layer-hybrid-recommender/55391/](http://www.irma-international.org/article/enhanced-semantic-layer-hybrid-recommender/55391/)

SEMCON: A Semantic and Contextual Objective Metric for Enriching Domain Ontology Concepts
[www.irma-international.org/article/semcon/153665/](http://www.irma-international.org/article/semcon/153665/)