Expert Systems With Applications 42 (2015) 9587–9602

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Construction and evaluation of ontological tag trees

Chetan Verma [a,b,*], Vijay Mahadevan [a], Nikhil Rasiwasia [a], Gaurav Aggarwal [a], Ravi Kant [a], Alejandro Jaimes [a], Sujit Dey [b]

[a] Yahoo Labs, Bengaluru, India
[b] University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093, United States

ARTICLE INFO

Keywords:
Tag tree
Local search
Optimization
Tag prediction
Classification

ABSTRACT

Several expert systems have been proposed to address the sparsity of tags associated with online content such as images and videos. However most of such systems either necessitate extracting domain-specific features, or are solely based on tag semantics, or have significant space requirements to store corpus based tag statistics. To address these shortcomings, in this work we show how ontological tag trees can be used to encode information present in a given corpus pertaining to interaction between the tags, in a space efficient manner. An ontological tag tree is defined as an undirected, weighted tree on the set of tags where each possible tag is treated as a node in the tree. We formulate the tag tree construction as an optimization problem over the space of trees on the set of tags and propose a novel local search based approach utilizing the co-occurrence statistics of the tags in the corpus. To make the proposed optimization more efficient, we initialize using the semantic relationships between the tags. The proposed approach is used to construct tag trees over tags for two large corpora of images, one from Flickr and one from a set of stock images. Extensive data-driven evaluations demonstrate that the constructed tag trees can outperform previous approaches in terms of accuracy in predicting unseen tags using a partially observed set of tags, as well as in efficiency of predicting all applicable tags for a resource.

1. Introduction

The consumer electronic revolution and the Internet have led to the availability of vast amounts of data including multimedia data such as images and videos. A significant fraction of such data is user generated content, in the form of pictures and videos uploaded onto sites such as Facebook (2015), Flickr (2015) and YouTube (2015). Owing to the fact that there are minimal requirements when uploading the content and that mobile uploads are on the rise, users rarely add any extra information such as a textual description to the content. At best, most images and videos are tagged with certain keywords. As these keywords or tags are sometimes applied to entire albums of images or videos at once, or applied in error, the information provided by such tags is quite noisy.

Some examples of images having incorrect tags (as per human experts) are shown in Fig. 1. The massive scale of data and the lack of useful metadata makes it difficult for users to access data that may be of interest to them (Anand & Mampilli, 2014; Jiang, Qian, Shen, Fu & Mei, 2015; Zheng & Li, 2011).

The social tagging at the above mentioned data sharing websites creates a Folksonomy (Hsieh, Stu, Chen, & Chou, 2009; Kim & Kim, 2014; Sun, Wang, Sun, & Lin, 2011) which mitigates the information overload to some extent by creating non-hierarchical categories or indexes for the retrieval of data. Folksonomies make it scalable to assign labels to large volumes of data in a collaborative manner and are hence more appropriate for such data than traditional taxonomies established by expert cataloguers (Kim & Kim, 2014). At the same time, collaboratively produced folksonomies have several issues, particularly related to incorrect tags and their sparsity (Sun et al., 2011; Uddin, Duong, Nguyen, Qi, & Jo, 2013). While incorrect tags have been discussed earlier, the sparsity in folksonomy arises as a result of lack of incentive for the users to tag the resources comprehensively and completely. As a result, the online resources are typically associated with low number of tags, preventing effective searching and browsing through the available data (Uddin et al., 2013).

In order to address the sparsity in folksonomies, several expert systems have been proposed that recommend or suggest additional tags for a resource based on the tags already associated with the resource (Chen, Liu, & Sun, 2015; Hsieh et al., 2009; Sigurbjörnsson & Van Zwol, 2008; Sun et al., 2011). Most of such works depend on the availability of content-based features such as textual features from

* Correspondence to: 1063 Morse Avenue, Apt 3-306, Sunnyvale, CA 94089, United States. Tel.: +1 858 344 9473.

E-mail addresses: cverma@ucsd.edu (C. Verma), vmahadev@yahoo-inc.com (V. Mahadevan), nikux@yahoo-inc.com (N. Rasiwasia), gaggarwa@yahoo-inc.com (G. Aggarwal), rkant@yahoo-inc.com (R. Kant), ajaimes@yahoo-inc.com (A. Jaimes), dey@ece.ucsd.edu (S. Dey).

http://dx.doi.org/10.1016/j.eswa.2015.07.057
0957-4174/

(a) 'animal', **'car'** **(b)'wedding'**, 'mushroom' **(c)** 'farms', **'snow'**

Fig. 1. Some examples of incorrect tags given by users on www.flickr.com. (a) An image of a 'cat' tagged as 'car', which most likely is a spelling mistake, (b) an image of a 'mushroom' also tagged as 'wedding' and (c) an image of a 'goat' tagged with 'snow'. The Flickr owner and photo ids of these images are (8656572@N04,4670326818), (35468147887@N01,252474171), and (39405339@N00,5835556089), respectively.

documents or blogs (Chen et al., 2015; Hsieh et al., 2009; Sun et al., 2011), or visual features from images or videos (Xia, Feng, Peng, Wu, & Fan, 2015), and thus cannot be applied to other domains. In addition, extracting and utilizing content based features is known to be computationally expensive and for certain domains, even infeasible (Huang, Fu, & Chen, 2010; Song et al., 2010; Yin, Li, Mei, & Han, 2009; Zanetti, Zelnik-Manor, & Perona, 2008), and so the above works may not be applicable to such domains. Furthermore, expert systems such as Uddin et al. (2013) utilize purely semantic relationships between tags. While semantic relations as obtained from ontologies such as WordNet (Miller, 1995) or Open Directory Project (2015) are an important resource for linguistic and machine learning related problems, such relationships fail to capture the information that is characteristic of an available corpus. Consider for instance a corpus of annotated images from Flickr. The co-occurrence of tags in a given corpus provides interesting insights into the nature of the data. For example, the 2008 Olympics were held in Beijing and as a result, there exist a large number of images in Flickr having '2008' and 'Beijing' as their tags. Such a relation between '2008' and 'Beijing' cannot be obtained from WordNet or similarly formed hierarchies (such as Open Directory Project (2015)) because the semantic relations in the above hierarchies are pre-defined, and do not account for a connection between the two tags. In addition to the above expert systems, works such as Sigurbjörnsson and Van Zwol (2008) capture tag similarities from a given dataset using tag graphs. Tag graphs usually refer to complete graphs representing pair-wise distances or similarities between tags, which are calculated from a given corpus. For certain applications, a threshold is applied and only the most important pair-wise connections are retained. However, storing similarities using tag graphs has several issues. Firstly, the pre-defined threshold value that is chosen to construct them can be arbitrary and there is no clear understanding to what its value should be. The pair-wise edges that have their similarity above the defined threshold are the only ones that are retained in the graph and this leads to completely losing of information of those pairs of concepts or tags that have their similarity below the threshold. Depending on the threshold value, the space requirement of tag graphs can vary as $O(N^2)$ where N is the number of concepts or tags in the tag graph, which can be significantly high for large number of tags. In order to keep a handle on the space requirement, a strict threshold value can be chosen which would result in losing pair-wise similarity information for several pairs of concepts or tags. Lastly, depending on the threshold, it is possible that some concepts or tags are disconnected from the rest. This again implies losing relationship information of the concept or tag with others. Sigurbjörnsson and Van Zwol (2008) estimate the number of tags in Flickr in 2008 to be 3.7 million. Storing each similarity value as a floating point occupying 4 bytes would require more than 27 terabytes just to store the pair-wise relationships.

We attempt to address the above shortcomings in this paper. We use the term ontological tag tree or simply tag tree to denote undirected weighted tree of concepts (or tags) where the relationships between the concept nodes in the tree are defined only in terms of a scalar weight. As compared to tag graphs (Liu, Hua, Yang, Wang, & Zhang, 2009; Sigurbjörnsson & Van Zwol, 2008), ontological tag trees are necessarily trees on the set of tags, i.e., are connected and have no simple cycles. We have chosen a spanning tree to represent the relationships between tags because a spanning tree over the set of tags is necessarily connected and does not lead to losing of information due to possibly disconnected components as in tag graphs. Also, the space requirement of a spanning tree is only $O(N)$ for N tags. For 3.7 million tags (Sigurbjörnsson & Van Zwol, 2008), this implies a significant reduction in the space requirement from 27 terabytes ($O(N^2)$) to less than 50 megabytes ($O(N)$). As a result, expert systems can be implemented even on computing devices that do not have a gigantic memory. Ontological tag trees are constructed using the semantic and the data-driven relations between the tags and hence lead to significantly better performance on data-driven tasks than using solely semantic relationships between tags (Miller, 1995; Uddin et al., 2013). For the constructions of tag trees, we do not utilize content based features, rather we utilize data-driven similarities from tag co-occurrences in the given annotated corpus. As a result, compared to previous expert systems that require extracting and processing content-based (such as visual or textual) features (Chen et al., 2015; Hsieh et al., 2009; Sun et al., 2011; Xia et al., 2015), tag trees can be used to alleviate sparsity in online folksonomies even in domains where extracting domain specific features may be infeasible or inefficient. This also makes the construction approach not married to a single domain such as annotated text documents/blogs or videos or images.

We illustrate the proposed tag tree construction approach using two large image corpora – one obtained from Flickr, and the other obtained from a set of stock images, with the goal of obtaining a tag tree over the set of tags present in these corpora. For these corpora, the co-occurrence count for a pair of tags is defined as the number of images with which both tags are associated. The normalized co-occurrence counts are a measure of how related two tags are. We assume that the concepts or nodes of the tag tree are the tags, and that the tree construction task is to infer the relations between the tags. The task thus becomes a graph learning problem where the nodes of the graph are the tags, and the relations between tags are represented by undirected edges and their weights in the graph. Unlike the relationships given in ontologies, we do not attempt to give semantic interpretations to the relations between tags. To solve the graph learning problem, we formulate an optimization problem on the space of spanning trees of a suitably constructed Similarity Graph that is based on semantic relations between tags, as obtained from WordNet, and on the normalized co-occurrence counts of the corpus. We solve the optimization problem using the 'local search' paradigm by constructing a simple edge exchange based neighborhood on the space of candidate trees. To make the optimization efficient, we initialize our approach using a preliminary tag tree built purely based on semantics from the WordNet hierarchy. The proposed local search based approach is then used to refine the preliminary tag tree based on the corpus statistics.

The evaluation of structures capturing the relationships between different tags or concepts is a difficult task. In the domain of ontologies, there are often no clear quantitative metrics to compare different ontologies that can be built for the same corpus of data. Certain works compare constructed ontologies to a predefined gold standard ontology (Porzel & Malaka, 2004) which is constructed manually. Tag graphs are usually not evaluated explicitly, rather are used in various applications such as tag ranking (Liu et al., 2009). Since manual evaluations are subjective and are not scalable, in this work, we also propose a novel fully automatic framework to evaluate ontological tag trees over tags using the Tag Prediction Accuracy, given an incomplete set of tags for a resource. Furthermore, we also demonstrate that the constructed tag trees can be used to efficiently assign tags to resources in domains where content-based features can be