Construction of Tag Ontological Graphs by Locally Minimizing Weighted Average Hops

[Extended Abstract]

Chetan Verma
University of California San Diego
cverma@ucsd.edu

Vijay Mahadevan
Yahoo Labs
vmahadev@yahoo-inc.com

Nikhil Rasiwasia
Yahoo Labs
nikux@yahoo-inc.com

Gaurav Aggarwal
Yahoo Labs
gaggarwa@yahoo-inc.com

Ravi Kant
Yahoo Labs
rkant@yahoo-inc.com

Alejandro Jaimes
Yahoo Labs
ajaimes@yahoo-inc.com

Sujit Dey
University of California San Diego
sdey@ucsd.edu

ABSTRACT

We present a data-driven approach for the construction of ontological graphs on a set of image tags obtained from annotated image corpus. We treat each tag as a node in a graph, and starting with a preliminary graph obtained using WordNet, we propose the graph construction as a refinement of the preliminary graph using corpus statistics. Towards this, we formulate an optimization problem which is solved using a local search based approach. To evaluate the constructed ontological graphs, we propose a novel task which involves associating test images with tags while observing partial set of associated tags.

1. INTRODUCTION

The growing popularity of Internet has led to availability of almost limitless amounts of multimedia data, such as on Flickr, YouTube. Since the data is usually user generated, it is often noisy and has little to no textual information, making its automatic processing an enormous challenge.

A commonly used approach to organize a collection of data involves grouping it into categories and specifying the relationships between the categories. Ontologies have been employed in several domains to do so for defined relations between categories [5]. The conventional way of building an ontology [8] involves significant manual effort and this has led to efforts that use semi-automatic [9] and fully automatic techniques [2] in domains such as multimedia and text respectively. For instance, [11] derives subsumption relations from corpus of annotated images while [7] uses hierarchical clustering to derive ontology among categories. Several approaches such as [3] use existing lexical ontologies such as WordNet [10] to derive relations between the categories. While the set of semantic relations available in WordNet is an important resource for linguistic and machine learning related problems, it fails to capture the information that is characteristic to a corpus. To address this gap, data specific interactions need to be accounted for, that cannot be inferred from prior domain knowledge.

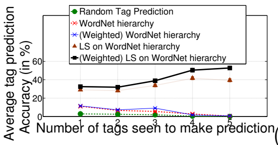
In this work, we propose a novel automatic hybrid approach to building an ontological graph over the domain of image tags. We do not attempt to give semantic interpretations to the relations between tags, like ontologies do. We first build a preliminary graph using WordNet, then model the data-driven ontology refinement as an optimization problem on the space of trees on the set of tags. We solve this using the "local search" paradigm and illustrate our approach using two large image corpora - one from Flickr, another from a set of stock images. There are often no clear quantitative metrics to compare different ontologies that can be built for the same corpus of data. Most previous attempts at evaluation are *qualitative* and use manual assessment to evaluate ontologies [2]. Some *quantitative* techniques compute similarity between an ontology and a gold standard. Both approaches therefore require some manual intervention. Hence, in this work, we also propose a novel fully automatic framework to evaluate ontological graphs over tags through a novel tag prediction task.

2. GRAPH BUILDING

2.1 Preliminary Graph Building from WordNet

We modify the approach outlined in [4] in order to derive a preliminary ontological graph for set of tags $\mathcal{T}$ using Word-

WWW'14 Companion, April 7–11, 2014, Seoul, Korea.
ACM 978-1-4503-2745-9/14/04.
http://dx.doi.org/10.1145/2567948.2577359.

(a) 

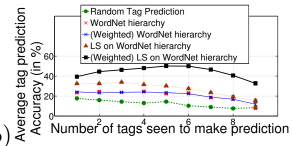
(b) 

Figure 1: Average Tag Prediction Accuracies marginalized over total number of tags in image, for various methods for (a) Flickr corpus, and (b) Stock images corpus. Graphs corresponding to Weighted LS and Weighted WordNet have jaccard similarity similarity from [1] as edge weights respectively.

Net. First, the "is-a" and "part-of" relations between tags are extracted from WordNet. Since we are interested in undirected relations between tags, the directions are ignored and the disconnected components are joined in a greedy manner based on WordNet distances as obtained from [1]. The resulting preliminary graph $\mathcal{G}(W)$ is a tree with N nodes corresponding to N tags and $(N-1)$ edges.

2.2 Corpus based graph refinement

We refine $\mathcal{G}(W)$ by accounting for those relations between tags that appear strongly in the corpus but are not present in $\mathcal{G}(W)$. The objective of our refinement stage is to find a tree which minimizes the following objective function:

$$\sum_i \sum_{j, j \neq i} J_{\mathcal{T}}(i,j) d_{i,j}, \quad (1)$$

where $d_{i,j}$ represents number of hops between tag t_i and tag t_j in the tree, and $J_{\mathcal{T}}(i,j)$ represents the jaccard similarity between sets of images tagged with tag t_i and tag t_j respectively. The above objective function is equal to the weighted average number of pair-wise hops between tags, in the tree. We add a constraint that prevents an edge to exist between tags t_i and t_j if $J_{\mathcal{T}}(i,j) \leq \tau$ where $0 \leq \tau \leq 1$. The motivation for such an objective function is that in order to obtain a low value for (1), tag pairs (i,j) corresponding to high values of $J_{\mathcal{T}}(i,j)$ should be separated by less number of hops. For a general graph, the problem of minimizing the weighted average number of hops has been established to be an NP hard problem [6]. We propose an approach to obtain local minima for (1) through the Local Search (LS) paradigm. LS moves from one solution to another in the search space of candidate solutions until objective function doesn't improve further. For our problem, a tree on $\mathcal{T}$ is a candidate solution. Consider two trees T_1, T_2 on the set of tags $\mathcal{T}$. We say that T_2 is a neighbour of T_1 iff T_2 can be obtained by adding one edge to, and removing one edge from T_1, the two edges being distinct.

3. EVALUATION

We provide experimental results for two corpora - Flickr corpus, and corpus of professionally annotated stock images. Ontological graphs are constructed through the proposed approach as discussed in Section 2. τ is chosen as median of the pair-wise jaccard similarity values. We describe below the data-driven task used to evaluate the constructed ontological graphs.

3.1 Tag Prediction

In this evaluation, the task is to observe a subset of tags associated with a test image, and predict the unobserved tags. Specifically, let a test image i be tagged with the set of tags $\mathcal{T}_i$. Assume that the tags in the subset $\mathcal{T}_{i,seen}$ are observed such that $\mathcal{T}_{i,seen} \subset \mathcal{T}_i$. The remaining tags $\mathcal{T}_i \backslash \mathcal{T}_{i,seen}$ are to be predicted using an ontological graph. Predictions are made by ranking the tags based on average distance from tags in $\mathcal{T}_{i,seen}$ and predicting the $\mid \mathcal{T}_i \backslash \mathcal{T}_{i,seen} \mid$ number of tags that are closest to $\mathcal{T}_{i,seen}$. Tags in $\mathcal{T}_{i,seen}$ are not predicted. The distance between two tags in an ontological graph is calculated as the sum of weights of edges connecting the two tags. Experimental results are provided with edge weight equal to 1 (non-weighted graph) and equal to jaccard similarity between the connecting tags (weighted graph). For a certain number of tags in image i, and a certain number of tags observed (or seen), the performance for i is measured as Tag Prediction Accuracy i.e., the fractions of predictions that were correct. For the Flickr corpus, top 117 tags are chosen, that are also available in WordNet and are not stopwords. For Stock images corpus, top 30 such keywords (tags) are chosen. The corresponding average Tag Prediction Accuracies marginalized over total number of tags in image, are shown in Fig. 1a and Fig. 1b respectively. It can be seen that the ontological graphs obtained by the proposed methods outperform those obtained from WordNet. The weighted graphs perform better than non-weighted counterparts and as expected, the random tag prediction performs the worst.

4. REFERENCES

[1] RiTa WordNet Library. http://www.rednoise.org/rita/wordnet/documentation.

[2] P. Buitelaar, P. Cimiano, and B. Magnini. *Ontology learning from text: methods, evaluation and applications*, volume 123. IOS press, 2005.

[3] E. Dietz, D. Vandic, and F. Frasincar. Taxolearn: A semantic approach to domain taxonomy learning. In *Web Intelligence and Intelligent Agent Technology (WI-IAT)*, pages 58–65, 2012.

[4] E. Djuana, Y. Xu, and Y. Li. Constructing tag ontology from folksonomy based on wordnet. In *IADIS International Conference on Internet Technologies and Society, 2011.*

[5] D. Fensel. *Ontologies*. Springer, 2001.

[6] M. R. Garey and D. S. Johnson. *Computers and intractability*, volume 174. Freeman New York, 1979.

[7] G. Griffin and P. Perona. Learning and using taxonomies for fast visual categorization. In *IEEE conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2008.

[8] T. R. Gruber. Toward principles for the design of ontologies used for knowledge sharing? *International journal of human-computer studies*, 43(5):907–928, 1995.

[9] A. Jaimes and J. R. Smith. Semi-automatic, data-driven construction of multimedia ontologies. In *International Conference on Multimedia and Expo*, volume 1, pages I–781. IEEE, 2003.

[10] G. A. Miller. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41, 1995.

[11] P. Schmitz. Inducing ontology from flickr tags. In *Collaborative Web Tagging Workshop at WWW, 2006.*