Construction and Evaluation of Data Driven Hybrid Tag Trees

Abstract

We present a data-driven approach for the construction of an ontological tag tree for a set of image tags obtained from a large corpus of images, where each image in the corpus is annotated with one or more tags. We treat each possible tag as a node in a tree, and formulate the tag tree construction as an optimization problem over the space of trees defined over the set of tags. A preliminary tree is constructed based on the WordNet hierarchy and used as an initialization for the tree construction. While such a preliminary tree captures semantic relations between tags, it fails to encode the information present in a given corpus of images pertaining to interaction between the tags. To refine this preliminary tree and adapt it to a given corpus, we propose a novel local search based approach utilizing the co-occurrence statistics of the tags in the corpus. We use this approach to construct trees over tags for two corpora of images, one from Flickr and one from a set of stock images. To evaluate the tag trees, we propose a novel task involving tag prediction in images. We show that for predicting unseen tags from a partially observed set of tags of a given image, using the proposed tag tree is more effective than tag trees obtained using WordNet, or similar tag graphs built using standard approaches.

1 Introduction

The consumer electronic revolution and the Internet have led to the availability of almost limitless amounts of 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, Flickr and YouTube. 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. Therefore, the massive scale of the data, accompanied by the lack of useful metadata makes the automatic processing of such data an enormous challenge.

A commonly used strategy to organize a collection of data is to group it into categories and specify the relationships among the various categories. Ontologies (Fensel 2001) are often employed to specify predefined relations between categories. The conventional way of building an ontology (Gruber 1995) involves significant manual effort. The concepts or categories of the ontology have to be specified, and the relations between the categories defined, all manually. Furthermore, the ontology has to be updated when data belonging to hitherto unseen categories becomes available. Once the ontology has been specified, data samples must be annotated, again manually, to assign them to one or more categories in the ontology so that rules or classifiers can be learned for that category. Therefore, manual techniques for ontological or taxonomic organization of data become especially challenging and cumbersome when there are large amounts of data. Also, ontologies built for one setting are rarely reusable even in other closely related domains. This necessitates the building of an ontology afresh for each new setting. As data could be from one of an ever increasing pool of knowledge domains, manually constructing ontologies for data in each domain is infeasible. Furthermore, when the data obtained is noisy, as is the case for user generated content on the Internet, the problem is accentuated as more manual effort might be needed to clean up the data followed by ontological organization.

The challenges associated with the manual construction of ontologies have led to efforts that use semi-automatic (Jaimes and Smith 2003) and fully automatic techniques (Buitelaar, Cimiano, and Magnini 2005) in domains such as multimedia and text based ontologies respectively. Most existing automatic approaches to ontology construction use text mining techniques to identify the concepts and then define relations between the concepts based on their semantic similarity as obtained from lexical databases such as WordNet (Miller 1995). While the semantic graph available in WordNet is an important resource for linguistic and machine learning related problems, it fails to capture the information that is characteristic of an available corpus. Consider for instance a corpus of annotated images from Flickr (e.g publicly available MIR Flickr dataset (Huiskes and Lew 2008)). Each image is associated with a set of *tags* that are applied by users to describe the image. An image typically consists of one or more tags. The co-occurrence of

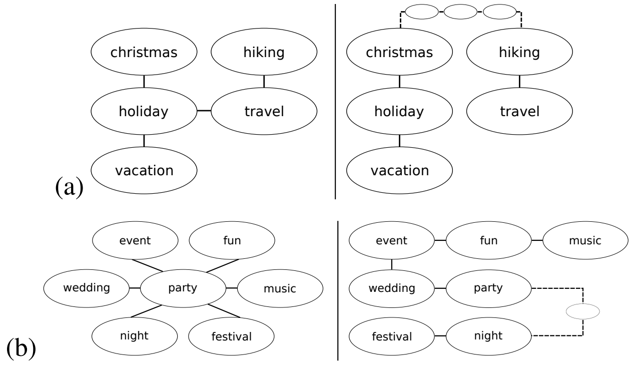

Figure 1: Two examples of subgraphs built using (left) the proposed data-driven approach and (right) corresponding sub-graphs obtained using WordNet. In example (a), 'holiday' and 'travel' are directly connected using our approach but are separated by multiple hops in the Wordnet hierarchy. In example (b), the proposed approach is able to identify 'party' as the central node that connects several other party-related tags.

tags in a given image 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, ODP (ODP online)) because the semantic relations in the above hierarchies are pre-defined, and do not account for a connection between the two tags. To address this gap between information available in a corpus of data, and manually constructed semantic ontologies, one also needs to account for data specific interactions between the concepts that may not be inferred from prior domain knowledge.

In this paper, we use the term ontological tag tree or simply tag tree to denote an undirected weighted tree of concepts where the relationships between the concept nodes in the tree are defined only in terms of a scalar weight i.e. explicit relationship types are not specified. Once the connections between concepts are determined through an ontological tag tree, their relationship types can be acquired from other sources such as ODP, Wikipedia (Wikipedia Online) or WordNet (Miller 1995). The focus of this paper is on the construction of ontological tag trees. The proposed construction approach is illustrated using two large image corpora - one corpus obtained from Flickr, and the other obtained from a set of stock images. 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 edges and their weights in the graph. Note that we do not attempt to give semantic interpretations to the relations between tags like ontologies do. To solve the graph learning problem, we formulate an optimization problem on the space of spanning trees of a suitably constructed similarity graph. We solve it using a "local search" paradigm by constructing a simple edge exchange based neighbourhood on the space of candidate trees. We initialize our approach using a preliminary tag tree built purely based on semantic relations from the WordNet hierarchy. The proposed local search based approach is then used to refine the preliminary tag tree based on the co-occurrence statistics of the corpus .

The evaluation of structures capturing the relationships between different 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. Most approaches to evaluation are *qualitative* and use manual assessment or expert judgement to evaluate ontologies (Buitelaar, Cimiano, and Magnini 2005). Some *quantitative* approaches have been proposed to evaluate an ontology by comparing the performance for a specific task to that of a predefined gold standard ontology (Porzel and Malaka 2004). Both approaches therefore require some manual intervention. In this work, we also propose a novel fully automatic framework to evaluate ontological tag trees based on the task of inferring all the tags of an image given an incomplete subset of tags.

The key contributions of this paper are as follows:

1. We propose a framework to construct an ontological tag tree of tags in a given corpus. The objective of the formulation is to obtain a tag tree where the connections between the tags in the corpus are built using both semantic and data specific means. To obtain this we first build a preliminary tag tree using the WordNet hierarchy. This preliminary tree is then refined to incorporate data specific relations by performing a novel local search operating on local neighborhoods in the space of spanning trees over a similarity graph defined over the tags.
2. We propose a novel framework for evaluating ontological tag trees using a tag prediction task.
3. We evaluate the constructed tag trees on image corpora, using the above evaluation framework and show that it outperforms tag trees built using semantic ontologies and other commonly used corpus based approaches.

Figure 1 illustrates a couple of examples for which the proposed approach leads to qualitatively better connections between tags as compared to the tree obtained using WordNet alone. We first start with a brief discussion on the related literature.

1.1 Related Work

There have been several works addressing the general topic of ontology building. For a good review of ontology learning from text see (Buitelaar, Cimiano, and Magnini 2005). Most automatic approaches use some form of clustering to combine similar terms or keywords together to form concepts. First, a similarity metric is defined between tags, words or concepts, and then a hierarchical clustering algorithm is utilized to form a dendrogram with the concepts as leaves of the formed tree. The hierarchical clustering algorithm can be either bottom-up (agglomerative) or top-down (divisive). Such a procedure creates auxiliary entities in the tree representing combination of multiple entities of interest. For example (Neshati et al. 2007) uses hierarchical clustering based on a compound similarity measure between words. The similarity score is obtained by using a neural network model on syntactical information and corpus based similarities. However such techniques can only provide grouped entities, instead of modeling data semantics in a structural form. In (Dietz, Vandic, and Frasincar 2012), given a corpus corresponding to a domain, relations between important concepts are learned with the help of WordNet or by using a search engine. Hierarchical clustering is employed to construct a domain specific dendrogram as mentioned above.

Works such as (Hearst 1992) and (Cimiano, Hotho, and Staab 2005) utilize natural language based grammar rules to learn hierarchies between text entities. These works cannot be applied outside of the domain of natural language, since they depend at least in part upon grammatical speech. Semi-automatic techniques for ontology construction such as Text2Onto (Cimiano and Völker 2005) assist the user in constructing ontologies from a given set of text based data. Similar techniques have been attempted in the domain of annotated multimedia content, such as images and videos (Jaimes and Smith 2003). Fully automatic techniques such as OntoLearn etc. (Velardi et al. 2005; Navigli, Velardi, and Gangemi 2003; Mani et al. 2004) use natural language processing and machine learning to extract concepts and relations from data.

The construction of tag ontologies and taxonomies specifically for image corpora such as Flickr has also been explored. For annotated images, (Schmitz) proposed the application of a co-occurrence based subsumption model from (Sanderson and Croft 1999), to learn whether a tag *subsumes* another. (Griffin and Perona 2008) uses the category confusion matrix to cluster similar categories together in a hierarchical manner. To construct an ontology for a set of tags, (Djuana, Xu, and Li) maps the tags to WordNet and leverages WordNet's hierarchy. In addition to the above, tag graphs have been utilized for various applications such as tag ranking (Liu et al. 2009) to represent the pair-wise similarities or distances between tags. While several works use tag graphs as complete graphs on the set of tags, others choose sets of edges that have their distance lower than a heuristically chosen threshold (Heymann and Garcia-Molina 2006). In general, tag graphs have $O(N^2)$ edges with correspondingly large storage requirement for large values of N.

In most of the works discussed above evaluation of a constructed ontology is either done manually or by comparing it to a gold standard ontology that is obtained from existing ontologies such as ODP (Open Directory Project) and WordNet or manually constructed. Manual evaluation of ontologies makes the process subjective and time consuming. While several applications pertaining to annotated multimedia content utilize tag graphs, they usually are not evaluated quantitatively.

The works discussed above utilize either corpus based techniques or manually created ontologies such as WordNet to derive structural relatioships between tags or entities. The proposed work is different from other approaches as we formulate an automated *hybrid* approach by building an ontological tag tree with $O(N)$ edges using WordNet followed by a data-driven refinement. To our knowledge, the formulation of ontological tag tree construction as an optimization problem on the space of spanning trees and its solution using the "local search" paradigm is completely novel. We use a variant of the edge-exchange method to construct the neighborhood on the solution space.

The use of local search methods in combinatorial optimization has a long history (Aarts and Lenstra 1997). The paradigm has been extensively studied (Johnson, Papadimitriou, and Yannakakis 1988)(Aarts and Korst 1988) due to its practical success on many NP-Hard problems and also for the insights it provides on the structure of discrete optimization problems. We next discuss the problem statement addressed in this paper, followed by the proposed tree construction approach.

2 Problem Statement

We assume that we are given a corpus $\mathscr{C}$ of annotated images, where each image is associated with a variable number of tags. Let $\mathcal{I} = \{i_l\}$ denote the set of images, where $l = 1$ to $| \mathscr{C} |$. Let $\mathcal{T} = \{t_j\}$ denote the set of tags where $j = 1$ to N.

We define an ontological tag tree as an undirected weighted tree on the set of tags $\mathcal{T}$. This implies that the tag tree is connected and has no simple cycles. The task is to arrange the set of tags $\mathcal{T}$ in an ontological tag tree.

3 Construction of Ontological Tag Tree

In order to construct the tag tree, we propose an approach that starts with a preliminary tag tree obtained using the semantics encoded by the WordNet hierarchy. We follow this by a tag tree refinement based on the co-occurrence statistics of the tags in the corpus. Construction of the WordNet based preliminary tree is described below.

3.1 Constructing WordNet-based Preliminary Tag Tree

We follow the approach outlined in (Djuana, Xu, and Li) to derive the semantic relations between the set of tags $\mathcal{T}$. This is done in two stages. In the first stage, disambiguation for the meaning of the tags is done by selecting the most popular concept (synset) for every tag. For example a tag "turkey" can be mapped to the bird, or the country. In WordNet, since the synset corresponding to turkey, the bird, has a higher frequency count than the synset corresponding to the Republic of Turkey, the former synset would be selected to map to the tag "turkey". Then in the second stage, in order to find the relationships between different tags, all links between the mapped concepts are found through the WordNet hierarchy for semantic relationships "is-a" or "part-of". Since we are only interested in a tag tree which has undirected edges between the tags, we ignore the directions of the edges in the obtained hierarchy, which could otherwise help distinguish more generic concepts or tags from more specific ones. The resulting undirected graph in general has cycles and is usually disconnected, forming disjoint clusters of tags. In order to construct a tree from the above undirected graph, we first connect disjoint segments in a greedy manner using inter-tag distances from the WordNet Library (WordNet Library) defined as:

$$\frac{\text{MHP}}{\text{HPR + MHP}} \tag{1}$$

where MHP:= Minimum hops to common parent, and HPR:= Hops from common parent to Root of hierarchy. From the resultant graph, any cycles are broken by removing the weakest edge in the cycle based on distances obtained from the WordNet Library (WordNet Library). Once the preliminary tree based on WordNet, T_W, defined over the vertex set $V = \{v_1, \ldots, v_N\}$ where the node v_j corresponds to the tag t_j, is constructed as described, we refine it using a data-driven approach based on the co-occurrence statistics of the tags. We describe this below.

3.2 Co-occurrence based tree refinement

We refine T_W by accounting for those tags that strongly co-occur in the corpus but are not linked in the WordNet based tag tree. To achieve this, we first define $J_{\mathcal{T}}$ the **Jaccard Matrix** of the set of tags $\mathcal{T}$ where $J_{\mathcal{T}}(i,j)$ is the jaccard similarity between tags t_i and t_j, defined as:

$$J_{\mathcal{T}}(i,j) = \frac{\text{Number of images containing } t_i \text{ and } t_j}{\text{Number of images containing } t_i \text{ or } t_j}. \tag{2}$$

We augment the preliminary WordNet based tree, T_W, to construct a *similarity graph* as follows. We start with T_W. Given a threshold τ, $0 \leq \tau \leq 1$, we join all pairs of nodes v_i, v_j with an edge if $J_{\mathcal{T}}(i,j) \geq \tau$, and the edge weight of edge (v_i, v_j) is set to be $J_{\mathcal{T}}(i,j)$. We call the resulting graph $\mathcal{G}_{\mathcal{T}}$, the **Similarity Graph** of $\mathcal{T}$, and it represents a hybrid tree where the edges in T_W are based on semantic similarity, while the additional edges introduced are based on jaccard similarity derived from co-occurrence statistics of the corpus.

Given the similarity graph $\mathcal{G}_{\mathcal{T}}$, the objective of our refinement stage is to find a tree T in the space of spanning trees of the similarity graph $\mathcal{G}_{\mathcal{T}}$ which minimizes the following objective function:

$$Minimize \sum_{(i,j)} w_{i,j} \mid J_{\mathcal{T}}(i,j) - S_T(i,j) \mid, \tag{3}$$

where $S_T(i,j)$ represents the similarity between tags t_i and t_j estimated using tag tree T. A very close problem is the problem of approximating a given distance matrix through spanning trees, which has been established to be NP hard (Eckhardt et al. 2005). The weights $w_{i,j}$ are taken to be the co-occurrences counts of tags t_i and t_j. While $S_T(i,j)$ can be calculated in several ways, we define $S_T(i,j)$ as

$$S_T(i,j) = \prod_{e \in P_{i,j}} S(e), \tag{4}$$

where $P_{i,j}$ is the path in tag tree T connecting tags t_i and t_j and $S(e)$ is equal to the Jaccard similarity between the tags that edge e connects. Such a definition for $S_T(i,j)$ ensures that it lies between 0 and 1 and no rescaling is required in order to compare $S_T(i,j)$ values with $J_{\mathcal{T}}(i,j)$. The local-search based approach to minimize the above objective functions is described next.

3.3 Optimization based on Local-search

Given the Similarity Graph $\mathcal{G}_{\mathcal{T}}$ for a set of tags $\mathcal{T}$, our objective is to construct a spanning tree on $\mathcal{G}_{\mathcal{T}}$ such that (3) is minimized. Towards this, we propose an approach to obtain local optima through the Local Search paradigm.

Local Search – Local Search algorithms provide a local optimum to an optimization problem. This is done by moving from one solution to another, in the search space of candidate solutions.

For the problem of constructing a ontological tag tree, we define a simple edge-exchange neighbourhood on the space of spanning trees of the graph $\mathcal{G}_{\mathcal{T}}$ as follows. Given two spanning trees T_1, T_2 we say that T_2 is a neighbour of T_1 iff it can be obtained from T_1 by the following steps:

1. Select an edge $e_1 \in \mathcal{G}_{\mathcal{T}} \setminus T_1$ and add it to T_1.
2. In the (unique) cycle thus formed in T_1 containing e_1 pick the edge, say e_2 with minimum weight (i.e., jaccard similarity of the tags e_2 connects). Remove e_2 from T_1.

Starting from a spanning tree T_0 of $\mathcal{G}_{\mathcal{T}}$ as an initial solution, we explore all neighbors of T_0 to determine which neighbor minimizes the defined objective function. The winning neighbor is then considered as the next solution and its neighbors are explored. The algorithm stops when no further benefit is seen in the objective function. The output is the locally optimal tag tree T_{opt}.

4 Evaluation

In this section, we describe the experimental setup for construction and evaluation of ontological tag trees. The experiments are conducted on two large corpora of images. For the evaluation, we define a tag prediction task, the details of which are given below.

4.1 Datasets

To test the robustness of our approach to build ontological tag tres for domains with varying tag noise, we use two different image corpora - one from Flickr, composed primarily of user generated content, and one from a professionally curated stock photo agency.

- **Flickr images**: Flickr is a very popular image and video hosting website where users can upload images and associate them with annotations such as titles, tags and descriptions, among others. As Flickr primarily contains user generated content, tags are often noisy, irrelevant to image content or even completely absent. We utilize 500,000 images for training and 100,000 images for testing. All these images are licensed under Creative Commons copyright licenses.
- **Stock images corpus**: To evaluate the proposed approach on a less noisy data, we take a corpus of stock photos

(a) **Seen Tags**: film, france
Unseen Tags: holiday, sky, snow
Predicted Tags: paris, europe, bw

(b) **Seen Tags**: china, family
Unseen Tags: live, summer, usa
Predicted Tags: photography, christmas, photo

(c) **Seen Tags**: canada, ocean
Unseen Tags: red, sky, sunset
Predicted Tags: sea, beach, water

(d) **Seen Tags**: autumn, black
Unseen Tags: light, macro, night
Predicted Tags: white, nature, bw

(e) **Seen Tags**: car, green
Unseen Tags: photo, washington, white
Predicted Tags: red, spring, nature

Figure 2: Example images where the proposed method gave 0% tag prediction accuracy when the first two tags were seen and the next three were unseen and predicted. Also provided are the tags that were predicted by the proposed method. The Flickr owner and photo ids of these images are (king-edward@4061393892) , (familymwr@4928996212) , (alejandroerickson@7730525250) , (wwarby@5145467790) , (1968-dodge-charger@5507716438) respectively.

that are are professionally annotated, and hence are accompanied with a variety of accurate annotations - such as keywords, captions, etc. For this corpus, we use the set of keywords to build the ontological tag tree, and refer to them as "tags". We utilize more than 350,000 images for training and close to 70,000 images for testing.

Training images are used for adapting the WordNet based preliminary tag tree to the given corpus using the local search (referred as LS in performance plots) based approach described in Section 3. Training images are also used for specific required tasks such as training of classifiers. Testing images are used to evaluate the constructed tag trees. There is no overlap between training and test sets. We describe below the task defined to evaluate the constructed tag trees.

4.2 Tag Prediction Task

In this evaluation, the task is to predict the unseen tags for images in the test set of the corpus. Let an image i in the corpus be tagged with the set of tags $\mathcal{T}_i$, such that $\mid \mathcal{T}_i \mid = N_{Tags}$. Assume that out of these N_{Tags} tags, only a subset $\mathcal{T}_{i,seen}$ are observed, with $\mid \mathcal{T}_{i,seen} \mid = N_{Seen}$. The objective of the *tag prediction task* is to predict the remaining $(N_{Tags} - N_{Seen})$ tags $\mathcal{T}_i \setminus \mathcal{T}_{i,seen}$. Let P_i be the set of $(N_{Tags} - N_{Seen})$ tags predicted for image i assuming that $\mathcal{T}_{i,Seen}$ is known. Note that the prediction assumes the total number of tags for the image, N_{Tags}, to be known. Performance of tag prediction can be measured by *Tag Prediction Accuracy*, defined as follows:

$$\text{Tag Prediction Accuracy} = \frac{\mid \{\mathcal{T}_i \setminus \mathcal{T}_{i,Seen}\} \cap P_i \mid}{\mid \{\mathcal{T}_i \setminus \mathcal{T}_{i,Seen}\} \mid}. \quad (5)$$

We now discuss the approach we follow to obtain the set of predicted tags P_i when the set of tags $\mathcal{T}_{i,seen}$ is seen, by utilizing a given ontological tag tree.

Utilizing Ontological Tag Tree for Tag Prediction Consider the tag tree T, built over the set of $\mathcal{T}$ tags in a corpus. For image i with N_{Seen} number of seen tags, each tag $t \in \{\mathcal{T} \setminus \mathcal{T}_{i,Seen}\}$ is given a proximity score s_t based on its proximity from the seen tags, as per T. Specifically,

$$s_t = \Sigma_{t' \in \mathcal{T}_{i,Seen}} dist(t, t'), \quad (6)$$

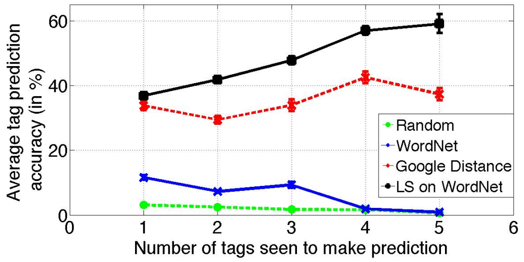

Figure 3: Average Tag Prediction Accuracies marginalized over N_{Tags} for various methods for the tag prediction task on Flickr dataset. Note that Google distance refers to the method corresponding to Google Similarity Distance as outlined in Section 4.2.

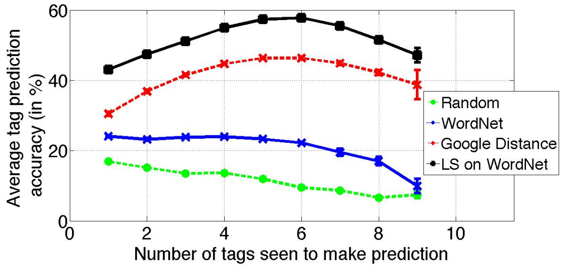

Figure 4: Average Tag Prediction Accuracies marginalized over N_{Tags} for various methods for the tag prediction task on Stock images corpus.

where $dist(t, t')$ is the sum of edge weights along the path connecting tags t and t' in T. A lower proximity score for a tag t indicates that it is closer in a cumulative sense to the set of known tags $\mathcal{T}_{i,Seen}$. The tags are ordered in the increasing order of s_t, and the first $(N_{Tags} - N_{Seen})$ tags, i.e., those corresponding to the least values of s_t, are chosen as the set of predicted tags, i.e., as P_i.

Methods Compared We compare the following methods in the tag prediction task:

1. Random: As the name suggests, this baseline method randomly picks $(N_{Tags} - N_{Seen})$ tags from the set of *unseen*

(a) **Seen Tags**: highway, road
Unseen Tags: route, shield, sign

(b) **Seen Tags**: austin, band
Unseen Tags: music, texas, tx

(c) **Seen Tags**: england, europe
Unseen Tags: london, travel, uk

(d) **Seen Tags**: art, dc
Unseen Tags: graffiti, street, washington

(e) **Seen Tags**: concert, england
Unseen Tags: london, music, uk

Figure 5: Example images where proposed method gave 100% tag prediction accuracy when the first two tags were seen and the next three were unseen and predicted. The Flickr owner and photo ids of these images are (dougtone@7975042008) , (elchupacabra@7023118527) , (jeffwilcox@159730021) , (daquellamanera@4678084101) , (martinrp@3832812191) respectively.

tags $\mathcal{T}_i \setminus \mathcal{T}_{i,seen}$.

2. WordNet: This baseline approach uses the semantics based ontological tag tree constructed from WordNet hierarchy using the procedure described in Section 3.1. The edge weights are assigned to be WordNet distances as obtained from (WordNet Library).

3. Google Similarity Distance: Google Similarity Distance (Cilibrasi and Vitanyi 2007) has been used to construct tag graphs in applications such as tag ranking (Liu et al. 2009). As mentioned in Section 1.1, a threshold is used to discard certain edges in tag graphs. We choose a threshold such that for a tag graph with N nodes (or tags), there are exactly $(N-1)$ edges remaining, so that the tag graph thus formed has same number of edges as the tag tree learnt from proposed approach. Edge weights for the tag graph are taken to be the Google Similarity Distance as defined in (Cilibrasi and Vitanyi 2007).

4. LS on WordNet: The tag tree is constructed using the approach outlined in Section 3. The edges in the constructed tree are assigned weights based on the jaccard similarity of the connecting tags.

The prediction task is performed using the approach described in section 4.2. For methods numbered 2, 3 above, $dist(t_i, t_j)$ for (6) is calculated by adding distances of edges in path connecting tags t_i and t_j. For the proposed method, $dist(t_i, t_j)$ is defined as $(1 - S_T(i, j))$ where $S_T(i, j)$ is calculated as per (4).

Flickr: We begin by choosing the top 150 most popular tags in a sample of Flickr images. After selecting only those tags that also occur in the WordNet database, we are left with 117 tags. These comprise the set of tags, $\mathcal{T}$. The total number of tags in an image, N_{Tags} varies from 0 to 6 for most Flickr images. For this task, we vary N_{Tags} from 2 to 6. For each value of N_{Tags}, test images are selected which contain exactly N_{Tags} number of tags. For each such image i, all combinations of N_{Seen} tags are selected to comprise the observed tag set $\mathcal{T}_{i,Seen}$. Predictions are made as discussed in Section 4.2 and performance of tag prediction task is measured using (5). N_{Seen} is varied from 1 through $(N_{Tags}-1)$. Given values for N_{Tags} and N_{Seen}, the Average Tag Prediction Accuracy is the Tag Prediction Accuracy (5) averaged across test images. Ontological tag trees are constructed using proposed approach as outlined in Section 3. The Average Tag Prediction Accuracies marginalized over N_{Tags} are shown in Figure 3. It can be seen that the proposed method outperforms all others. As expected, random tag prediction performs the worst.

Fig. 2 and 5 show some test images from the Flickr dataset that give low and high tag prediction accuracies with corresponding sets of seen, unseen, and the predicted tags.

Stock images corpus: As in the Flickr experiment, we first select the set of most popular tags from the corpus of stock images that are also in WordNet. This produces a set $\mathcal{T}$, of 30 tags. The Average Tag Prediction Accuracies marginalized over N_{Tags} are shown in Fig. 4, plotted as a function of N_{Seen}. If N_{Tags} is kept constant, increasing N_{Seen} reduces the random chance of predicting a correct unseen tag. As a result of this, a drop in performance for random prediction can be seen with increasing N_{Seen}. It is clear that the tag trees built using the proposed approach outperform the baseline trees or graphs constructed using WordNet or Google Distance for all values of N_{Seen}.

5 Conclusions and Future Work

We have proposed an approach for the construction of a basic ontological tag tree for a set of image tags using WordNet followed by its refinement using a local search approach that locally optimizes an objective function based on the co-occurrence statistics of the data. By treating each tag as a node in a tree, we constructed a tree that accounts for both semantic similarity and fidelity to the available data in the form of co-occurrence statistics between pairs of tags. We have shown that this approach can be used to build ontological tag trees for tags obtained from two corpora, one composed of noisily annotated Flickr images and the other composed of cleanly annotated stock images. To validate the utility of the constructed ontology, we proposed a novel evaluation task involving tag prediction in images. We have shown that for the task of predicting the unseen tags of a given image with a partially observed set of tags, the proposed ontological tag tree outperforms those obtained using WordNet or similar tag graphs built using standard approaches.

For future work, we plan to construct a hierarchical *taxonomy* where the edges between the nodes are directed, by adopting similar techniques and using the co-occurrence data.

References

Aarts, E., and Korst, J. 1988. Simulated annealing and boltzmann machines.

Aarts, E. E. H., and Lenstra, J. K. 1997. *Local search in combinatorial optimization*. Princeton University Press.

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

Cilibrasi, R. L., and Vitanyi, P. M. 2007. The google similarity distance. *Knowledge and Data Engineering, IEEE Transactions on* 19(3):370–383.

Cimiano, P., and Völker, J. 2005. Text2onto. In *Natural Language Processing and Information Systems*. Springer. 227–238.

Cimiano, P.; Hotho, A.; and Staab, S. 2005. Learning concept hierarchies from text corpora using formal concept analysis. *J. Artif. Intell. Res.(JAIR)* 24:305–339.

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

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

Eckhardt, S.; Kosub, S.; Maaß, M. G.; Täubig, H.; and Wernicke, S. 2005. Combinatorial network abstraction by trees and distances. In *Algorithms and Computation*. Springer. 1100–1109.

Fensel, D. 2001. *Ontologies*. Springer.

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

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

Hearst, M. A. 1992. Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the 14th conference on Computational linguistics-Volume 2*, 539–545. Association for Computational Linguistics.

Heymann, P., and Garcia-Molina, H. 2006. Collaborative creation of communal hierarchical taxonomies in social tagging systems.

Huiskes, M. J., and Lew, M. S. 2008. The mir flickr retrieval evaluation. In *Proceedings of the ACM International Conference on Multimedia Information Retrieval*. New York, NY, USA: ACM.

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

Johnson, D. S.; Papadimitriou, C. H.; and Yannakakis, M. 1988. How easy is local search? *Journal of computer and system sciences* 37(1):79–100.

Liu, D.; Hua, X.-S.; Yang, L.; Wang, M.; and Zhang, H.-J. 2009. Tag ranking. In *Proceedings of the 18th international conference on World wide web*, 351–360. ACM.

Mani, I.; Samuel, K.; Concepcion, K.; and Vogel, D. 2004. Automatically inducing ontologies from corpora. *Corpus* 9(k2):19–024.

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

Navigli, R.; Velardi, P.; and Gangemi, A. 2003. Ontology learning and its application to automated terminology translation. *Intelligent Systems, IEEE* 18(1):22–31.

Neshati, M.; Alijamaat, A.; Abolhassani, H.; Rahimi, A.; and Hoseini, M. 2007. Taxonomy learning using compound similarity measure. In *Web Intelligence, IEEE/WIC/ACM International Conference on*, 487–490. IEEE.

Porzel, R., and Malaka, R. 2004. A task-based approach for ontology evaluation. In *ECAI Workshop on Ontology Learning and Population, Valencia, Spain*. Citeseer.

Rita WordNet Library. `http://www.rednoise.org/rita/wordnet/documentation`.

Sanderson, M., and Croft, B. 1999. Deriving concept hierarchies from text. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, 206–213. ACM.

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

Velardi, P.; Navigli, R.; Cuchiarelli, A.; and Neri, R. 2005. Evaluation of ontolearn, a methodology for automatic learning of domain ontologies. *Ontology Learning from Text: Methods, evaluation and applications* 92–106.

Open Directory Project. `http://www.dmoz.org`.

Wikipedia. `http://en.wikipedia.org/wiki/Main_Page`.