Contextualizing Tags

Raul Guerra, Darya Filippova, Richard B. Johnson
{rguerra,dfilippo,rbjohnso}@cs.umd.edu

Abstract

Art objects in museums across United States get annotated with tags by museum visitors. We obtained the tag dataset and analyzed major themes addressed by the tags. We focused on ten most frequently occurring tags and constructed a theme map for each. We then analyzed the overlap between the themes each of the tags describe as well as the overlap between different contexts for the same tag.

1 Introduction

Steve Museum Project is a joint effort between museums, libraries, and several universities across the United States [4]. The project seeks to improve both public access and the public engagement with art museum collections. To do so, participating museums are having the public annotate art images with tags. Various university research groups analyze the tags to improve image access and to understand what the public sees in works of arts.

One of the interesting tasks in analyzing tags is figuring out how museum visitors use tags. Unlike individual words in text, tags lack context (as in sentence, paragraph, section, etc) and meaning disambiguation is harder. However, tags do have an indirect context: the tags that co-occur with a given tag, as well as images that they annotate and people who have used the tag. We are interested in exploring these indirect tag contexts and comparing them.

We believe that the relationships found in the literature for tags on Flickr [11] and del.icio.us [3] in different contexts will vary from the relationships in the Steve Museum dataset for different contexts. Flickr and del.icio.us users tend to put user posts, news, and images into various categories rather than summarize the contents of the object. Tags on the art objects differ from Flickr and del.icio.us tags in that they mostly describe the contents of the art piece, its style and implementation.

There are several works in the literature [12], [13], [3], [5], [1] showing that co-occurring tags tend to have similar meanings. If semantic relationships do occur in the Steve museum dataset, we want to compare and contrast with the relationships described in the literature.

We also believe that co-occurrence is more than a semantic relatedness indicator, especially in the Steve Museum dataset. From manual inspection of the Steve Museum dataset, we have seen that semantically unrelated tags can co-occur and these relationships are not noise. In the Steve Museum dataset; tags that are not semantically related and that co-occur tend to be popular themes or entities in the artworks. For example “woman” and “tree” for are two entities that occur in artworks, but are not semantically related.

This paper is an exploratory paper that seeks to better understand how tags are used for art annotation. We explore tag co-occurrences in different contexts, what can be inferred from these different contexts, and come up with a machine learning approach to validate our results.

2 Related work

Work has been done to find similar tags in the area of information retrieval [11],[12]. Tag recommendation and query expansion processes both depend on similarity relationships within the tag set to produce lists of related tags.

To establish similarity relationships, one has to define a tag context and decide on the similarity metric. Cosine similarity is widely used in the literature [3],[1],[10],[11] and is generally defined as a cosine between two vectors describing those tags:

$$\frac{\sum_{i=1}^{n} a_i b_i}{\sqrt{\sum_{i=1}^{n} a_i} \sqrt{\sum_{i=1}^{n} b_i}}$$
where vectors $a, b$ vary with the tag context. Second-order co-occurrence takes into account co-occurrences between a tag and all tags that it co-occurs with [12]. Sigurbjornsson and Zwol [11] define two measures based on Jaccard similarity; first one is symmetric:

$$ J(t_i, t_j) = \frac{|t_i \cap t_j|}{|t_i \cup t_j|} $$

and the second one asymmetric:

$$ J_{asym}(t_i, t_j) = \frac{|t_i \cap t_j|}{|t_i|}, $$

asymmetric Jaccard similarity measures subsumption. Markines and colleagues [10] evaluate several similarity measures; they report that mutual information was able to capture similarities better than the rest of measures on their data.

Similarity depends on the context and specific measures may be more appropriate within a given context. Cattuto, et al. [3] single out tag context, resource context, and user context. According to [3] different contexts highlight different relationships among tags such as synonyms vs. concept hierarchy.

Sigurbjornsson and Zwol [11] developed a Flickr tag recommendation system. When user adds a tag to an image, the recommendation system looks up the tags that were used with the new tag by other users, merges and ranks them according to a combination of measures. The three measures described in the paper are designed in a way to demote tags that occur too often and promote ‘descriptive’ tags that do not occur often, but would potentially provide a useful context.

Clough, Joho, and Sanderson [5] used a subsumption model for term similarity. They applied it to terms extracted from image captions. They found that their model elicited unobvious relationships “due to unknown concepts (e.g. names of unfamiliar places or domain-specific words and phrases)”. In particular, they found situations in which terms or concepts are not related lexically, but together describe visual properties of an image.

[1] built document and tag co-occurrence graphs, assigned edge weights based on node similarity and then clustered the networks. They compared the resulting clusters to manually selected groups of tags and inferred whether the clusters and the groups were similar or not.

3 Preprocessing the Data

The total number of tags with repetitions in the Steve Museum dataset is 49,767 (17,102 unique tags). We removed multi-word tags and that reduced the tag count to 39,259. We removed punctuation and ran the nltk’s lemmatizer [2] and were left with 7,810 unique tags.

In Steve Museum dataset, each art object belongs to a category such as painting, metalwork, etc. We aggregated related categories that contained only a few art objects into bigger groups. Our assumption is that images with different object type would be described by different sets of tags (i.e. would display different themes). We selected 7 most abundant object types from the original 168, manually merged them with less frequently used object types that we thought would share themes, and gave each class a label (see table 1).
4 Methods

We use notation from [1] to describe our data. Steve Museum is a folksonomy $F$ that is a tuple $F = (U, T, I, A)$, where $U$ is a set of users, $T$ is a set of tags, and $I$ is a set of images, $A \subseteq U \times T \times I$ is a set of annotations; $|U| = 2,047$, $|T| = 7,810$, $|I| = 1,784$, $|A| = 38,866$.

The Steve Museum is a tripartite hypergraph: $H = < V, E >$ where $V = (U \cup T \cup I)$ and $E = \{(u, t, i)|(u, t, i) \in A\}$. For this project we focused on different subnetworks generated from this hypergraph. It is important to note that the subnetworks derived in the following sections are built for a particular tag $t$.

We focused on top 10 most frequently used tags from the set $T$ - they appear with most images, are used by most users, and, therefore, would produce the most interesting co-occurrence graphs. The top ten tags are: $\text{TopTen} = \{\text{blue, flower, gold, landscape, portrait, red, sculpture, tree, woman, wood}\}$.

4.1 Defining tag context

Tag context determines the theme, or sense, of a tag. To disambiguate tag sense, we had to put the tag into context. Based on the results of [3], we look at tag context, resource context, and user context for each tag in $\text{TopTen}$.

Given a tag $t$, we define $U_t$, user context, as the set of users who have used the tag $t$ on one or more images. Formally:

$$U_t = \{u|\exists i \in I, (u, t, i) \in A\}$$

$I_t$, the resource context, is a set of images that have been annotated with $t$:

$$I_t = \{i|\exists u \in U, (u, t, d) \in A\}$$

Lastly, $T_t$ is a tag context, or a set of tags that have been used together with $t$ on the same images by the same users:

$$T_t = \{t'|\exists (u, i) \in U_t \times I_t, (u, t, i) \in A\}$$

4.2 Tag context similarity networks

To identify tag themes for $t$, we build a co-occurrence graph $G_t$ of all tags that appear together with the tag $t$ and clustered the graph. We considered clusters as the ‘themes’ for that particular tag.

For each of the tags in $\text{TopTen}$ and each context $U_t$, $I_t$ and $T_t$, we created a co-occurrence graph, where we considered two tags related if they were both used to annotate the same image, were used by the same user, or appeared in the same image and were used in for that image by the same user.

Tag co-occurrence definition varies with context. For tag context, tag $t^*$ co-occurs with tag $t$ if the tag were used to annotate the same image by the same user. For user context, tags co-occur if they were used by the same user, i.e. user’s dictionary is a clique of co-occurring tags. For resource context, two tags co-occur if they were used to annotate the same art object (maybe by different users).

We put an edge between a pair of tags if the two tags have appeared together given the context. The edge weight is defined by the cosine [3] and Jaccard [10] similarity. For cosine similarity, tag vector definition depended on the context:

- **user** : $w_{U_t, t_j} = |\{u|\exists i_a, i_b, (u, t_i, i_a) \in A \land (u, t_j, i_b) \in A\}|$ - the number of users that used both $t_i$ and $t_j$
- **resource** : $w_{I_t, t_j} = |\{i|\exists u_a, u_b, (u_a, t_i, i) \in A \land (u_b, t_j, i) \in A\}|$ - the number of images tagged by $t_i$ and $t_j$
- **tag** : $w_{T_t, t_j} = |\{(u, i)|\exists (u, t_i, i) \in A \land (u, t_j, i) \in A\}|$ - the number of the user-image pairs annotated by both $t_i$ and $t_j$

We limited our experiments to tags that we used more than once across the whole dataset. Such low frequency tags would be connected to other tags through low weight edges and would add noise to the co-occurrence graph rather than add any useful information. We expect that this constraint would produce smaller and tighter clusters.

To cluster the graphs, we use MCL [7] with default inflation factor. We analyzed clustering results using Coral [8].
Table 1: Comparing clusters for blue built on different contexts

<table>
<thead>
<tr>
<th>Context pair</th>
<th>Jaccard</th>
<th>VI</th>
</tr>
</thead>
<tbody>
<tr>
<td>user, tag</td>
<td>0.21</td>
<td>0.84</td>
</tr>
<tr>
<td>image, user</td>
<td>0.30</td>
<td>0.73</td>
</tr>
<tr>
<td>image, tag</td>
<td>0.32</td>
<td>0.49</td>
</tr>
</tbody>
</table>

5 Results

We performed a manual inspection of the clusters for several tags $t$. Unlike the findings in [1], our clusters described the themes in the art works that were tagged with $t$. In [1], two of the clusters for the tag “wine” were \{ food, shopping, drink, vino, cooking \} and \{ linux, ubuntu, emulation, windows, software \}. These two clusters represent two different contexts for wine. The first is a kind of alcoholic drink and the second is the name of a software application. On the other hand, we found two clusters for the tag “blue”, \{ sail, sailboat, seascape, sea, ocean, wave storm, sailing, boat, . . . \} and \{ Japanese, kimono, traditional, print, noble \}. These two clusters describe themes of the artworks where the tag “blue” is found.

5.1 Comparing clusters

There was some overlap between clusters obtained from different contexts for the same tag. For example, we compare user, resource, and tag context clusters for blue (see table 1, fig. 2). Jaccard similarity between different partitionings was less than 0.32, i.e. only 32% of tag pairs were together in the same cluster in both partitions. Variation of information scores [9] were highest between user and image-user clusters, i.e. clusters for blue within these two contexts were most diverse.

We used Coral [8] to compare clusters obtained in different contexts for blue tag. Given a graph decomposition $C$ into clusters $C_1, C_2, \ldots, C_n$, Coral constructs a co-occurrence matrix $M_C$ where $m_{ij} = 1$ if nodes $i$ and $j$ were in the same cluster $C_k$ in decomposition $C$. When there are several decompositions of the same network or networks that share most of their nodes, Coral constructs a co-occurrence matrix for each decomposition and adds
the matrices together to get a matrix $M$. The entries in $M$ indicate the number of times the two nodes $i$, $j$ were in the same cluster across all decompositions. Contiguous groups of high-value entries point out groups of nodes that stay together in most decompositions - we call these nodes ‘core’, or ‘central’, nodes.

We loaded clusters for blue tag from three different contexts (user, resource, and tag) into Coral. The co-occurrence matrix is block diagonal (fig. 2) which means that tags in the clusters did not move from cluster to cluster too much across different contexts and mostly stayed in the same groups. Here are some of the tight clusters: {kimono, japanese}, {angel, prayer, monk, madonna}, {water, harbor, sea, rudder, ocean, sail, rock, fishing, seascape, sailboat, coast}.

We repeated the procedure by loading clusters for all tags in TopTen for a single context at a time (see 3). Co-occurrence matrix for user context is the closest of three to a block-diagonal shape. Some of the highly correlated tag groups are: {cloth, cape, shawl, garment, faded, fringe, weaving, textile}, {temple, procession, scream, spear, carriage, fight, horseman, horse, hunt, rifle, race, windmill}, {wind, marin, shipwreck, clearing, storm}.

5.2 Classifier

We used a Megam classifier [6] to validate the usefulness of tag contexts that we developed by clustering co-occurring tags. Each art object is assigned a type (see sec. 3). As a baseline, we use tag sets associated with every image (dataset features) and object type (class) to train the classifier. Then we break image-tag data into ten equal parts and perform ten fold validation by giving the classifier image tags from the test part of the data set and seeing how many times the classifier would be able to recover the correct tag (see table 2).
Table 2: Classifier recall for various tag contexts and similarity measures

<table>
<thead>
<tr>
<th>Context</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>0.72</td>
</tr>
<tr>
<td>Image</td>
<td>0.70</td>
</tr>
<tr>
<td>Tag</td>
<td>0.69</td>
</tr>
<tr>
<td>User</td>
<td>0.71</td>
</tr>
</tbody>
</table>

(a) Cosine similarity

<table>
<thead>
<tr>
<th>Context</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>0.72</td>
</tr>
<tr>
<td>Image</td>
<td>0.70</td>
</tr>
<tr>
<td>Tag</td>
<td>0.69</td>
</tr>
<tr>
<td>User</td>
<td>0.70</td>
</tr>
</tbody>
</table>

(b) Jaccard similarity

<table>
<thead>
<tr>
<th>Context</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>0.72</td>
</tr>
<tr>
<td>Image</td>
<td>0.70</td>
</tr>
<tr>
<td>Tag</td>
<td>0.70</td>
</tr>
<tr>
<td>User</td>
<td>0.69</td>
</tr>
</tbody>
</table>

(c) Mutual information

For each tag $t$ in $TopTen$, we went through all images that they were used for and looked at overlap between clusters $t_i$ of $t$ and the image tagset. For clusters with maximal overlap, we replaced the tag $t$ in the image tagset with tag $t_i'$. For example, $flower$ becomes $flower2$, $blue$ become $blue6$, etc.

Classifier’s recall did not vary significantly across the contexts nor across the similarity measures (see table 2).

6  Discussion

Through analysis, we discovered that there were similar clusters across all $t \in TopTen$. We believe this happened because the tags in $TopTen$ are the most frequent tags, so they tag a lot of of the same images, are used by the same users and in general share a lot of context.

Understanding the concepts associated with the tag is not straight forward. It is easy in the case of $blue$ and one of its clusters: \{sail, sailboat, seascape, sea, ocean, wave storm, sailing, boat, ...\} - here $blue$ refers to the water color. In another case, $blue$ is related to cluster of \{japanese, kimono, traditional, print, noble\} tags. Unless we look at the art objects related to these tags, we would not understand the relationship. The question is how do we determine what is noise and what is new useful information.

We believe that this set of tags greatly differs from usual tag corpora used for research, Flickr and del.icio.us. Steve Museum users are more preoccupied in describing the content of the image rather than in indexing it. For this reason, we believe that the tags in the Steve Museum data set have relations that are not found in those other tag corpora.

7  Future Work

A better approach would be to get a full view of all the themes across all tags. To do that, we would calculate all possible edges $T \times T$ from the hypergraph $H$ and then repeat the analysis we have done on individual tags: assign weights to the edges based on context and similarity metric, then cluster the graph and inspect the clusters.

It would be interesting to see what clusters, or tag themes, correspond to what art objects. Such analysis would help validate the clusters or dismiss some of them as noise.

Some of the images from the Steve Museum are being tagged in different languages. An interesting way to compare these multilingual tag sets would be to compare the major themes that appear in different languages.
References


