Content-Based Tag Processing for Internet Social Images

Dong Liu
EE Department, Columbia University
Dec 20, 2011
“Tag” has become one of the most popular Internet concepts in the last three years.
Challenge

Social tags are good, but they are

- Lack of relevance information
- Noisy and incomplete
- Annotated only at the image level

Tags need to be processed before using them.
**Tag Ranking**

Liu, Hua, Yang, Zhang, *Tag Ranking*, WWW09

**Basic Idea:**
- Large *tag clusters* should be promoted.
- Semantically close tags should be ranked closely.

---

Exemplar Similarity

Concurrence Similarity

Find

1. bird (0.36)
2. flower (0.28)
3. sky (0.21)
4. tree (0.15)
**Image Retagging**

Basic idea

- Assign **visually** similar images with similar tags.
- Exclude the **content-unrelated** tags.
- Expand the tags with **synonyms** and **hypernyms**.

**Results**

- In terms of average precision, recall, and F1-Measure:
  - 50,000 Flickr images
  - 106,565 unique tags
  - 5000 test images (each tag was judged by human labelers to decide whether it is related to image content.)

After Tag enrichment, the tag quality is further improved.

<table>
<thead>
<tr>
<th>Method</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-Measure</th>
<th>Relevant tag number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before Enrichment</td>
<td>0.71</td>
<td>0.34</td>
<td>0.46</td>
<td>3.09 (4.80 in all)</td>
</tr>
<tr>
<td>After Enrichment</td>
<td>0.90</td>
<td>0.66</td>
<td>0.76</td>
<td>9.34 (10.38 in all)</td>
</tr>
</tbody>
</table>
Issue: image similarity

Whether two images are similar actually depends on what semantic tags we are caring about.

Our Strategy: Learn tag-specific visual representation.
Tag-Specific Visual Vocabulary

Airplane

Noise-Tolerant Learning Algorithm

Visual Vocabulary for airplane

min_f \ Q(f) = \sum_{i=1}^{N} (f_i - q_i)^2 + c_1 \sum_{i,j=1, i \neq j}^{N} o_{ij} (f_i - f_j)^2 - c_2 \sum_{k=1}^{M} P_{ik},

s.t. \ 0 \leq f_i \leq 1, \ i = 1, 2, \ldots, N,
Technical Contributions

- Descriptive visual vocabulary construction.
- Learning with noises.
**Technical Contributions**

- Scalable multi-graph multi-label learning: Multiplicative nonnegative update rule derived from KKT condition of Lagrange function
- *Inter*-graph and *Intra*-graph label propagation.
**Basic Idea:**

- Images with common tags often share similar semantic regions.
- Uncover the region-to-region correspondences for image pairs.
A new research topic in multimedia research community.

Learning with hybrid, unreliable sources.
- Robust, efficient, and scalable solutions.

Data-driven vs. Model-driven.

Interplay of data, user and feature.
Cross-modality tag analysis

Learn an intermediate representation that maximizes the correlation between the visual content and semantic tags.

Visual understanding using tag cues

Infer fruitful contextual information about the visual content from the tags.

Scalable automatic tagging

Develop scalable statistical learning algorithms to handle large scale training data with huge number of tags.
Efficient Manual Tagging System Design

dog, horse, airplane, sky, ...
dog, grass, tree, ....
Manual Album Tagging System

**Batch tagging**
- **Pros:** The manual efforts can be significantly reduced.
- **Cons:** Introduce a lot of imprecise tags to many images.

**Exhaustive tagging**
- **Pros:** Tagging accuracy is relatively high.
- **Cons:** Too labor-intensive and time-consuming.

There is a dilemma between manual efforts and tagging accuracy.
Semi-Automatic Photo Album Tagging

- Dynamically adjust the tagging accuracy
- Visual & temporal information
- Ontology-free
- A good trade-off between manual efforts and tagging accuracy
Discussion: User Interaction

Basic Principles

- Minimize user’s participation
- Maximize system performance
- Efficient User Interface design

Potential directions

- Historic feedback information
- Both textual and visual clues
- Incremental Online Learning