Image Retagging

Dong Liu, Xian-Sheng Hua, Meng Wang and Hong-Jiang Zhang

Harbin Institute of Technology
Microsoft Research Asia
Microsoft Advanced Technology Center
Social Media and the Associated Tags -
Towards Large-Scale Content-Based Multimedia Search

flickr

medici chapel, Firenze, Italy...

Loggia dei lanzi, sword, honeymoon, ...

YouTube

Status, building, sky, Italy, ...

Cathedral, tower, Italy...
Challenge

Social tags are good, but they are

- Noisy
- Ambiguous
- Incomplete
- No relevance information

Two directions to improve tag quality

- Tag Ranking  (Liu, Hua, Zhang. *Tag Ranking*. WWW 09)
- Retagging  (Liu, Hua, Wang, Zhang. *Image Retagging*. MM 10)
Tags associated with social images are imprecise, subjective and incomplete.

- Imprecise Tags

- Subjective Tags

- Missing Tags

  - grass
  - flower
  - cat
  - animal
What we are going to do:

Improve the quality of the tags to better describe content.

To improve:

Tag-based image search

Image annotation (automatic)

top 101
tour
tiger
sweet
big
cloud
dog
house
tree
sky
ground
cloud
But how can we make it? Automatically.
Two Observations

- Similar images ↔ similar tags

- User-provided tags correlate with the image content with high probability
Basic Assumptions

Tag Refinement

- The consistency between visual similarity and semantic similarity should be maximized.
- The deviation from the initially user-provided tags should be minimized.
Tag Refinement

Notations

- a social image collection
- All unique tags
- tag membership
- refinement results
- confidence score vector
- visual similarity
- semantic similarity

\[ \mathcal{D} = \{x_1, x_2, \ldots, x_n\} \]

\[ \mathcal{T} = \{t_1, t_2, \ldots, t_m\} \]

\[ \hat{\mathbf{Y}} \in \{0, 1\}^{n \times m} \]

if \( t_j \) is associated with image \( x_i \), then \( \hat{Y}_{ij} = 1 \)

\[ \mathbf{Y} = \mathbf{Y} \geq 0 \]

confidence:

\[ \mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{im})^\top \] of image \( x_i \)

\[ W_{ij} = \exp\left(-\frac{||x_i - x_j||_2^2}{\sigma^2}\right), \]

\[ S_{ij} = \frac{2 \times IC(lcs(t_i, t_j))}{IC(t_i) + IC(t_j)} \]
Tag Refinement

Modeling the basic assumptions

- Visual and semantic consistency
- User-provided tags are relevant with high probability
- Overall formulation
Optimizing with iterative updating

- Bound the objective function

- Derive the solution

- Iterative updating until convergence

\[ L \leq L' = \sum_{i,j=1}^{n} \left( W_{ij}^2 + \sum_{l=1}^{m} [\tilde{Y} \tilde{S} \tilde{Y}^T]_{ij} \tilde{Y} \tilde{S} \tilde{Y}^T \right)_{ij} \]

\[ -4 \sum_{l=1}^{m} W_{ij} [\tilde{Y} \tilde{S}]_{il} \tilde{Y} \tilde{Y}^T \log Y_{jl} - 2W_{ij} [\tilde{Y} \tilde{S} \tilde{Y}^T]_{ij} \]

\[ + 4 \sum_{k=1}^{m} W_{ij} [S \tilde{Y}^T]_{kj} \log \tilde{Y}_{ik} \]

\[ + C \sum_{j=1}^{n} \sum_{l=1}^{m} \left( Y_{jl}^2 - 2\alpha_j \tilde{Y}_{jl} Y_{jl} (\log \frac{Y_{jl}^2}{Y_{jl}^4} + 1) \right) \exp(\tilde{Y}_{jl}) \]

\[ Y_{jl} = \left[ \frac{-C \exp(Y_{jl}) \tilde{Y}_{jl}^3 + \sqrt{M}}{4[\tilde{Y} \tilde{S} \tilde{Y}^T]_{jl}} \right]^{\frac{1}{2}} \]

\[ \alpha_j = \frac{\sum_{l=1}^{m} \tilde{Y}_{jl} (\log Y_{jl} - \log \tilde{Y}_{jl} + 1)}{\sum_{l=1}^{m} Y_{jl}} \]

- Fix $\alpha$, update $Y$ using the upper equation
- Fix $Y$, update $\alpha$ using the lower equation
Is It Reliable?
Visual Property of Tags

Content-Related Tag

- baby
- night
- cat
- flower
- bike

Content-Unrelated Tag

- fun
- photo
- macro
- my
- old
- Nikon
- science
- best
delete me

Describe the **REAL** visual content of the images. Informative for ALL general users.

Describe the **CONTEXTUAL** information about the images. Only informative to the image owners.
Our Basic Assumption

Similar images have similar tags.

- Involving the content-unrelated tags will
  - Introduce lots of noises.
  - Degrade the algorithmic performance.

Only applicable for “content-related” tags.

These tags should be removed from the automatic learning procedure.
Tag Filtering

Filter out all content-unrelated tags.

Construct a content-related tag dictionary by using the **lexical** and **domain** knowledge.

Traverse along the path until one predefined category is **matched**.

All words

- Non-noun
- Noun
  - content-unrelated
  - content-related

- Organism
- Natural Pho.
- Thing
- Artifact
- Color

organism
animal
artifact
mammal
structure
feline
building
kitty
Is It Enough?
The missing of such tags will degrade the performance of tag-based applications.
Use each tag to perform tag-based image search on Flickr.

The tags with more than 10,000 returned images are retained.
In Summary: Image Retagging

Three-step strategy
In term of average precision, recall and F1-Measure

50,000 Flickr images with 4,556 content-related tags.
2,500 test images.
**Performance of Tag Enrichment**

<table>
<thead>
<tr>
<th>Method</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-measure</th>
<th>Relevant tag num</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>

Tagging quality is further improved rafter the tag enrichment procedure.
Application 1: Tag-Based Image Search

- Use the learnt confidence scores as relevance measure
- Ranking results for query “cat”
Our confidence score based ranking strategy outperforms the other image ranking strategies on Flickr.
Application 2: Auto Tagging

Use top tags of the images after retagging to predict the tags of the unlabeled images.

- water
- flower
- tree
- plant
- cloud
- ocean
- water
- bird
- animal
- nature
- flower
- sky
- cloud
- tree
- landscape
- cat
- tree
- animal
- wildlife
- tiger
Using top tags after image retagging can obtain better results than using the original images directly.
User-provided tags are imprecise and incomplete, which limits the performance of tag-based applications.

We propose an image retagging strategy to solve this problem:
- Tag filtering to remove the content-unrelated tags
- Tag refinement to automatically refine the tags
- Tag enrichment to expand the tags with synonyms and hypernyms.

Image retagging benefits a series of tag-based applications.
Future work

- Extend it to online videos
- Using more fruitful information cues such as image regions and surrounding texts
Thank You