Learning To Tag

Lei Wu, Linjun Yang, Nenghai Yu, Xian-Sheng Hua

University of Science and Technology of China
Microsoft Research Asia
We did not find results for a man with an apple on his left hand and textbook on his right hand. Try the suggestions below or type a new query above.

Suggestions:

- Change your search settings.
- Check your spelling.
- Try more general words.
- Try different words that mean the same thing.
- Turn SafeSearch off.
Image Search

Query → Annotations → Images → Search Result

Text Based:
Annotation by surrounding text

Content Based:
Annotation by the content of images

Social Based:
Annotation by users in a social network

Automatic → Automatic → Manual
Why Need Tag Recommendation?

Three Issues: Tags are

1. Ambiguous
2. Incomplete
3. Noisy

Recommendation:

Sea
Ocean
Water
Sky
Cloud
Island
Coast
Why Need Tag Recommendation?

Three Issues: Tags are
1. Ambiguous
2. Incomplete
3. Noisy

Possible Tags:
- Apple
- Fruit
- Red
- Corporation
- Logo
- Products

Ambiguous
Incomplete
Problems with Social Tagging

- Different keywords
- Ambiguous keywords
- Ignore facts
- Typing error
Figure 1: Tag distribution over a collection of 640 million images from Flickr.com. There are totally 1,300 million tags. Around 1% of the tags appearing more than 20,000 times, which contain little information. Around 5.82% of the tags have appeared more than 5,000 times, which are considered as popular tags. 33.21% of the tags appear more than 50 and less than 5,000 times, which are defined as specific tags. 60% of the tags have appeared less than 50 times.
Tag Recommendation

What is Tag Recommendation

- Given one or more initial tags for an image
- Provide a list of possible tags automatically (ordered by relevance scores)

Advantages of Tag Recommendation

- Enable fast tagging
- Enable high-quality tagging
  - Higher Relevance/Accuracy
  - Wider Coverage
  - Less noises
Recommendation based on collective knowledge [www2008]

\[ J(t_i, t_j) := \frac{|t_i \cap t_j|}{|t_i \cup t_j|} \]

\[ P(t_j|t_i) := \frac{|t_i \cap t_j|}{|t_i|} \]

**Drawback:**

- Cannot deal with **synonym**
  - i.e. “soccer” == “football”

- Cannot deal with **polysemy**
  - i.e. “apple” fruit != “apple” logo

- Cannot deal with **meronymy**
  - i.e., car vs. wheel

- Target image is not taken into account (**image independent**)
  - Different images with the same initial tags will get the same recommended tag list
Learning To Tag

Correlation Measure

Tag

Visual

Visual Language Model

Flickr Distance /Tag Content Correlation

Textual

Tag Co-occurrence

Image Conditioned Similarity (ICS)
Three Features to Tag Correlations

- **Tag Concurrence (TC)**
  - The same as previous work

- **Tag Content Correlation (TCC)**
  - To solve the first three issues (synonym, polysemy & meronymy)

- **Image-Conditioned Tag Correlation (ITC)**
  - To solve the fourth and second issue (image independent, & polysemy)
Learning To Tag

[Diagram showing a flow of processes related to image tagging, including Tag Selection, Tag Invert Table, Visual Language Model, Tag Co-occurrence, Visual Correlation, Tag + content, RankBoost recommendation, and Social Tagging.]
Learning To Tag

Correlation Measure

Tag

Visual

Visual Language Model

Flickr Distance /Tag Content Correlation

Textual

Tag Co-occurrence

Image Conditioned Similarity (ICS)
Similarity in Visual Domain

- Visual Language Modeling (Lei et al. MIR07)
Learning To Tag

- Correlation Measure

- Visual
  - Visual Language Model
    - Flickr Distance /Tag Content Correlation

- Textual
  - Tag Co-occurrence
    - Image Conditioned Similarity (ICS)
Tag Content Correlation

Based on “Flickr Distance” or “PicNet Distance” – ACM MM 09 Best Paper Candidate
Can handle concurrency, synonym, polysemy & meronymy

Concept A: Airplane

Tag search in Flickr

Concept Model A

LT-VLM

Concept B: Airport

Concept Model B

Flickr Distance (A, B)

Jensen-Shannon Divergence
Tag Content Correlation (TCC)

- Inverse of Flickr Distance

\[
D_{TCC}^s(t_i, t_j) = \frac{1}{2} [KL(L(t_i)||M) + KL(L(t_j)||M)]
\]

\[
M = \frac{1}{2} [L(t_i) + L(t_j)]
\]

\[
R_{TCC}^s(t_i, t_j) = \frac{1}{D_{TCC}^s(t_i, t_j)}
\]
Learning To Tag

- Correlation Measure

Tag

Visual

Visual Language Model

Flickr Distance /Tag Content Correlation

Textual

Tag Co-occurrence

Image Conditioned Similarity (ICS)
Image Conditioned Similarity (ITC)

For test image $x_1$, $x_2$: $t_i = [L_{t_i}^1, L_{t_i}^2, L_{t_i}^3]$

For other image $j$: $t_j = [L_{t_j}^1, L_{t_j}^2, L_{t_j}^3]$

Apple is related to $x_1$, but not to $x_2$ (Pear).
Image-Conditioned Tag Correlation

Based on Visual Language Model (VLM)

Visual Words

VLM of Tag 1  VLM of Tag 2  \cdots  VLM of Tag N

Projection 1  Projection 2  \cdots  Projection N

New Features for Tags (Image Dependant)
Image Conditioned Similarity (ITC)

\[ \mathcal{L}_t^m(x) \propto \prod_{ij} P(w_{ij} | VLM_t^m) \]

\[ \mathcal{L}_t^m(x) \propto \prod_{i=1}^{n-1} P(w_{i0} | VLM_t^m) \prod_{j=1}^{n-1} P(w_{ij} | w_{i,j-1}, VLM_t^m) \]

\[ \Pi_{i=1}^{n-1} P(w_{i0} | VLM_t^m) \prod_{j=1}^{n-1} P(w_{ij} | w_{i,j-1}, VLM_t^m) \]

\[ D_{ITC}(t_i, t_j, x) = \frac{\mathbf{L}_{t_i}(x) \cdot \mathbf{L}_{t_j}(x)}{\|\mathbf{L}_{t_i}(x)\| \|\mathbf{L}_{t_j}(x)\|} \]

\[ \mathbf{L}_{t_i}(x) = [L_{t_i}^1, L_{t_i}^2, L_{t_i}^3] \]
Learning To Tag

Our Approach

- Tag Selection
- Tag Invert Table
- Visual Language Model
- Visual Correlation
- Tag Co-occurrence
- RankBoost recommendation
- Social Tagging
**Recommendation**

- **Ranking Features**

\[
f_l(t_i, t_l) = R_{TC}(t_i, t_l), \quad t_l \in OT, t_i \in UT, \quad l = 1, \ldots, n
\]

\[
f_{n+l}(t_i, t_l) = R_{TCC}(t_i, t_l), \quad t_l \in OT, t_i \in UT, \quad l = 1, \ldots, n
\]

\[
f_{2n+l}(t_i, t_l) = R_{ITC}(t_i, t_l), \quad t_l \in OT, t_i \in UT, \quad l = 1, \ldots, n
\]

- **Loss Function**

\[
r_{loss_D}(H) = \sum_{x_0, x_1} D(x_0, x_1) \delta(H(x_1) \leq H(x_0))
\]

\[
D(x_0, x_1) = c \cdot \text{max}(0, \Phi(x_0, x_1))
\]
Algorithm 1 Cross domain Rankboost training process

Input: Given tags \( t_1, \ldots, t_n \in \mathcal{OT} \), and \( t_1, \ldots, t_m \in \mathcal{UT} \), and distribution \( D \) over \( \mathcal{UT} \times \mathcal{UT} \).

where

- \( \mathcal{OT} \) is the set of the initial tags.
- \( \mathcal{UT} \) is the set of the remaining tags in the database.

Initialize \( D_1 = D \).

Generate ranking features \( \{ f_{i,j} \}_{i=1}^{3n} \); \( \forall t_i \in \mathcal{UT}, t_l \in \mathcal{OT} \)
- \( f_l(t_i, t_l) = R_{TC}^s(t_i, t_l), \ l = 1, \ldots, n \)
- \( f_{n+l}(t_i, t_l) = R_{ITC}^s(t_i, t_l), \ l = 1, \ldots, n \)
- \( f_{n+l}(t_i, t_l) = R_{ITC}^s(t_i, t_l), \ l = 1, \ldots, n \)

where \( t_i \in \mathcal{UT}, t_l \in \mathcal{OT} \)

for \( k = 1, \ldots, K \) do
- Select pair \( (t_i, t_j) \in \mathcal{UT} \times \mathcal{UT} \) with distribution \( D \).
- Get weak ranking \( h_k \) from ranking features of selected pairs
- Update: \( \alpha_k = \frac{1}{2} \ln \left( \frac{1+r}{1-r} \right) \),
  where \( r = \sum_{t_i,t_j} D_k(h_k(t_i) - h_k(t_j)) \)
- Update: \( D_{k+1}(t_i, t_j) = \frac{D_k(t_i,t_j) \exp(\alpha_k (h_k(t_i) - h_k(t_j)))}{Z_k} \)
  where \( Z_k \) is a normalization factor.
  \( Z_k = \sum_{t_i,t_j} D_k(h_k(t_i) - h_k(t_j)) \)
end for

Output the final ranking: \( H(t) = \sum_{k=1}^{K} \alpha_k h_k(t) \).
<table>
<thead>
<tr>
<th>Tag Concurrence Only</th>
<th>Tag Concurrence &amp; Tag Content</th>
<th>Tag Concurrence &amp; Tag Content &amp; Image-Cond. Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Init Tags</strong></td>
<td><strong>Tag Concurrence Only</strong></td>
<td><strong>Tag Concurrence &amp; Tag Content &amp; Image-Cond. Correlation</strong></td>
</tr>
<tr>
<td>Cruise party boat purple spectrum</td>
<td>Travel sea seaweed water colors</td>
<td>Travel family sea sun beach</td>
</tr>
<tr>
<td>friends fun birthday art girls summer Florida winter snow flower</td>
<td>vacation Asia trip holiday nature city cannon tree Europe building</td>
<td>vacation holiday Europe nature city water trip building Asia light</td>
</tr>
<tr>
<td>friends girls music fun night love art holiday vacation trip</td>
<td>vacation holiday trip Asia Europe nature city fun music friends</td>
<td>vacation trip Asia holiday water Europe nature tree friends sun</td>
</tr>
<tr>
<td>friends dance fun girls night music love men happy laugh</td>
<td>trip ocean sky island nature landscape blue umbrella red men</td>
<td>vacation fun water kids ocean sky holiday sand wave blue</td>
</tr>
</tbody>
</table>
Multi-domain relevance V.S. Tag Cooccurrence

**Evaluation**

\[
Coverage(m_i) = \frac{\sum_{x \in TopN} \delta(x|m_i)}{\sum_i \sum_{x \in TopN} \delta(x|m_i)}
\]

\[
\delta(x|m_i) = \begin{cases} 
1, & \text{given method } i, \text{ } x \text{ is related tag;} \\
0, & \text{otherwise.}
\end{cases}
\]
Rankboost V.S. linear combination

Precision@10

Coverage@10
Parameter influence

![Bar chart showing influence of different parameters](chart.png)
Tagging may be the final solution for visual understanding
- For both search and advertising

Learning2Tag provides an efficient tagging scheme
- Which ensures high-quality tagging results

It can also work as a “tag completion” approach
- Automatically expands tags for multimedia content

It is beyond the use for tagging
- It can also work for general image annotation combined with active learning.
Thanks. Any questions?

For further discussion, please send email to leiwu@live.com or visit http://wuleibig.googlepages.com