Learning to Recommend Tags for On-line Photos

Zheshen Wang and Baoxin Li

Computer Science and Engineering
Arizona State University

Mar. 31, 2009
Virtual Social Network or Community

Make friends ...

Share your story ...

Explore the world ...

facebook

shutterfly

flickr

Kodak Gallery

myphotoalbum
Tag – Compact Semantic Description

animals architecture art asia australia autumn baby band barcelona beach berlin bird birthday black blackandwhite blue bw california cameraphone canada canon car cat chicago china christmas church city clouds color concert cute d80 dance day de dog england europe fall family fashion festival film florida flower flowers food football france friends fun garden geotagged germany girl girls graffiti green halloween hawaii holiday home house india ireland island italia italy japan july kids la lake landscape light live london love macro me mexico mountain mountains museum music nature new newyork newyorkcity night nikon nyc ocean old paris park party people photo photography photos portrait red river rock rome san sanfrancisco scotland sea seattle show sky snow spain spring street summer sun sunset taiwan texas thailand tokyo toronto tour travel tree trees trip uk urban usa vacation vancouver washington water wedding white winter yellow york zoo

From Flickr
Related Work (1)

- Image annotation/categorization (e.g. [1])
  - Sub-regions and a mapping between keywords
  - Not so effective for generating tags with higher-level semantics that often link to the image as a whole

Related Work (2)

- Tag recommendation based on tag co-occurrence [2]


- Pure text based
- Require at least one tag from the user
Our Idea

Tag popularity score $a_1$ + Image-tag correlation score $a_2$ = Overall tag ranking score

Exploiting collective intelligence in social computing!

KCCA [W_x, W_y]

Test Image

Office uwo lab heal
Laslenas ski skiing snow southamerica argentina
Sun sunset lamps
Text Feature
Image Feature

Exploiting collective intelligence in social computing!
Canonical Correlation Analysis (CCA)

- **CCA**
  - To find basis vectors for two sets of variables such that the correlation between the projections of the variables onto these basis vectors is mutually maximized.

  \[
  \rho = \max_{W_x, W_y} \text{corr}(F_x \cdot W_x, F_y \cdot W_y)
  \]

  - The optimization problem can be formulated as a standard Eigen problem.

- **Kernel-CCA**
  - Non-linear correlations considered
  - Gaussian kernel used in our experiments
Image–Tag Correlation Score

Training

- Text and image features extraction
  \[ F_x = [f_x^1, ..., f_x^n]^T \quad F_y = [f_y^1, ..., f_y^n]^T \]
- Kernel mapping
  \[ F_x \to F'_x \quad F_y \to F'_y \]
- KCCA – basis discovery
  \[ [W_x, W_y] = KCCA(F'_x, F'_y) \]
- KCCA – projection
  \[ F''_x = F'_x \times W_x^k \quad F''_y = F'_y \times W_y^k \]

Test

- Image feature \( f_{y_0} \)
- Kernel mapping \( f'_{y_0} \)
- KCCA projection \( f''_{y_0} \)
- Image-tag correlation score
  \[
  corr_{t_i} = \frac{\text{cor}(f''_{y_0}, f''_{y_i})}{\max_{i=1...n}(\text{cor}(f''_{y_0}, f''_{y_i}))}
  \]
  \[
  S^corr_{t_j} = \max_{i=1...n}\{corr_{t_i}\}
  \]
Features

- **Text features**
  - Document-term (DT) matrix of normalized counts of appearances of each term (non-English words removed and stemming performed)
  - Only terms with high $\text{TFICF}$ score considered
    
    $$
    \text{TFICF}(T_k, C_i) = \text{TF}(T_k, C_i) \times \text{ICF}(T_k)
    $$
    
    $$
    \text{ICF}(T_k) = \log(|C|/\text{CF}(T_k))
    $$

- **Image features**
  - Color: HSV color histograms
  - Texture: Garbor gradient energies
  - Multi-resolution representation using spatial pyramid
  - 756-d feature vector (normalized) for each image
Tag Popularity Score

- Normalized counts of term appearances

\[ S_{t_j}^{\text{pop}} = \frac{c_{t_j}}{\max_{k=1...m}\{c_{t_k}\}} \]

- Input-independent score for each tag
- How likely a word is used as a tag based on the training set
- All terms appearing in the training set are considered.
Semantic Image–Tag Correlation Analysis

- **Image-tag correlation score**
  \[ S_{t_j}^{corr} = \max_{i=1...n} \{ corr_{t_j}^i \} \]

- **Tag popularity score**
  \[ S_{t_j}^{pop} = c_{t_j} / \max_{k=1...m} \{ c_{t_k} \} \]

- **Overall tag ranking score**
  \[ S_{t_j} = (1 - a) \cdot S_{t_j}^{corr} + a \cdot S_{t_j}^{pop} \]
  - \( a \in [0,1] \), a constant weight which provides a flexible control of the contributions from the two scores.
Class selection based on Flickr statistics [2]
- Locations: office, stadium
- Artifacts/objects: Greatwall, pyramid
- Actions/events: skiing, sunset

Data Collection
- Flickr API for downing images with user tags
- No more than 15 photos from the same Flickr ID
- 1800 photos from 993 different Flickr IDs
- 6 classes, 300 images for each class: 200 for training and 100 for test (multiple rounds of random split)

Challenges
- Sparse co-occurrence of tags
  - Only a few tags appear more than 5 times in the user-provided tags for all the training images
mac macpro apple inema30 ergotron Office multiple monitors

office uwo lab heal

london emirates stadium arsenal molenet

ppsnx germany deutschland dortmund bvb borussiadortmund football stadium

hdr pyramid egypt clouds

sky pyramid world heritage cairo Egypt

1755mmf28g nikon d80 china travel mutianyu nikond80 cultures great wall

china great wall

skiing sports snow mtbuller

laslenas ski skiing snow south america argentina

tipaza gouraya sunset alg ie lightpainting light

sun sunset lamps
Evaluation -- Methods

- **Objective**
  - Use user tags as ground-truth
  - Not sufficient. Why?

- **Subjective**
  - Human evaluators
  - Tick all relevant tags in the recommended list for a given image.
  - Three evaluators forming two user sets
Evaluation Metrics (adopted from [2])

- **Hit**: If one of the user tags is among the recommended tag list, we call it a *hit* or one of the recommended tags is deemed as relevant by a human evaluator.
- **≥k-HitRate**: Percentage of images out of all test images that achieve ≥ k *hit*
- **Mean Reciprocal Rank (MRR)**: Where in the ranking the first relevant tag occurs.
- **Success at rank k (S@k)**: The probability of finding a relevant tag among the top k recommended tags.
- **Precision at rank k (P@k)**: The proportion of retrieved tags that are relevant, averaged over all photos.
Results and Observations (1)

Objective evaluations

<table>
<thead>
<tr>
<th>≥k-HitRate (%)</th>
<th>k=1</th>
<th>k=2</th>
<th>k=3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average over 10 fixed random rounds</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a=0.2</td>
<td>97.0</td>
<td>71.8</td>
<td>37.1</td>
</tr>
<tr>
<td>a=0.5</td>
<td>99.9</td>
<td>71.8</td>
<td>34.7</td>
</tr>
</tbody>
</table>

Low hit rate for large $k$:

- Objective evaluation alone cannot sufficiently evaluate the real performance
- For $k=1$, our results are slightly better than [1].
  - Different databases; our database should be more challenging
- No objective evaluation results reported in [2].
Results and Observations (2)

Subjective evaluations

<table>
<thead>
<tr>
<th>≥k HitRate (%)</th>
<th>k=1</th>
<th>k=2</th>
<th>k=3</th>
<th>k=4</th>
<th>k=5</th>
</tr>
</thead>
<tbody>
<tr>
<td>User Set 1</td>
<td>99.2</td>
<td>86.2</td>
<td>64.3</td>
<td>44.0</td>
<td>26.2</td>
</tr>
<tr>
<td>User Set 2</td>
<td>99.8</td>
<td>88.7</td>
<td>72.5</td>
<td>48.5</td>
<td>25.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>a=0.5</th>
<th>MRR</th>
<th>S@1</th>
<th>S@2</th>
<th>S@3</th>
<th>S@4</th>
<th>S@5</th>
<th>P@5</th>
<th>P@10</th>
<th>P@15</th>
</tr>
</thead>
<tbody>
<tr>
<td>User Set 1</td>
<td>1.87</td>
<td>71.5</td>
<td>81.7</td>
<td>88.3</td>
<td>91.8</td>
<td>93.2</td>
<td>64.0</td>
<td>35.7</td>
<td>23.8</td>
</tr>
<tr>
<td>User Set 2</td>
<td>1.66</td>
<td>78.3</td>
<td>84.8</td>
<td>90.5</td>
<td>93.7</td>
<td>95.3</td>
<td>66.9</td>
<td>35.2</td>
<td>23.5</td>
</tr>
</tbody>
</table>

- Better than the best cases reported in state-of-the-art work [1].
- Based on image contents only; not a single user tag required.
Results and Observations (3)

- Visual Results

**User tags**: boardroom, portland, office

**Recommended tags**: office, design, desk, move, new, interior, ika, develop, homeoffice, chair, window, sunset, d300, film, work

**User tags**: sky, pyramid, worldheritage, cairo, egypt

**Recommended tags**: pyramid, egypt, giza, cairo, camel, sphinx, piramid, khafra, africa, cheops, travel, desert, sea, cloud, red
Results and Observations (4)

**User tags:** beijing, greatwall, 2008, spring

**Recommended tags:** greatwall, simatai, greatwallofchina, jinshanlin, travel, wall, beijing, china, muralla, hebei, atk beijingchina, flickr, nikon, d80

- Capable of capturing the underlying semantic correlation between image contents and text tags.
- Tags not listed by the user may also be good recommendations.
Future Work

- Use other available information of photos
  - Such as title, description, comments, geo-location, camera model, etc. ...
- Semantic grouping on tags before creating the DT matrix
- Combine tag co-occurrence strategies (as proposed in [2])
- Customized recommendations
  - Analyze users’ tagging history
  - Consider information from users’ on-line social network
    - Social group/community
    - Social network contacts, e.g., family members, friends, etc.
Thank you!