User Profile Enhanced Geolocation Suggestion for Social Images

Yuan Quan
SCE & IMI

Supervisor:
Asst. Prof. Gao Cong, School of Computer Engineering

Co-supervisor:
Prof Nadia M. Thalmann, Institute for Media Innovation

24/04/2012
Outlines

• Motivation & Objective
• Related Work
• Geotagging Behavior For Social Images
• Proposed Method
• Experiments & Results
• Conclusion
Motivations & Objectives

Social image sharing services (Flickr, Zooomr) have accumulated a huge number of photos. Photos can be associated with geolocation information (coordinates).

- Better organize and browse photos.
- Enable location-aware queries.

Two sources of geolocations:
- GPS-enabled cameras.
  - Not very popular.
- Manual geotagging.
  - Tedious and inaccurate.
Motivations & Objectives

• Develop an effective and efficient method to suggest geolocations for images using textual tags.
Related Work

- **Geolocation Suggestion for Images**
    - Visual features, time-consuming.
  - Placing flickr photos on a map, P. Serdyukov et al. SIGIR'09
    - Textual tags, many images have no tag.
  - Mapping the world's photos, D. Crandall et al. WWW'09
    - Textual & visual features, difficult to scale

- **Geolocation Suggestion for Other Contents**
  - You are where you tweet: a content-based approach to geolocating twitter users, Z. Cheng et al. CIKM'10
    - For users
  - Spatial variation in search engine queries, L. Backstrom, WWW'08
    - For queries
  - Geolocating blogs from their textual content, C. Fink et al. AAAI'09
    - For blogs
Geotagging Behavior - Dataset

• Collect a dataset of Flickr images uploaded by a set of randomly selected users

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Textual Tagged</th>
<th>Geotagged</th>
<th>Both Tagged</th>
</tr>
</thead>
<tbody>
<tr>
<td>Images</td>
<td>221,801,183</td>
<td>99,649,530</td>
<td>17,355,876</td>
<td>13,268,992</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(44.9%)</td>
<td>(7.8%)</td>
<td>(5.9%)</td>
</tr>
<tr>
<td>Users</td>
<td>2,252,758</td>
<td>468,555</td>
<td>106,289</td>
<td>97,061</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(20.7%)</td>
<td>(4.7%)</td>
<td>(4.3%)</td>
</tr>
</tbody>
</table>

• More than half of the images have no tag
• A small number of users have geotagged images
Geotagging Behavior

• We analyze the Cumulative Distribution Function (CDF) of distance between two images with closest taken time by the same user.
• Most images are close to the preceding ones.
• The geolocation of the preceding image can be used to geotag the new image.
• However…
Geotagging Behavior

- Turn to textual tags
- Generate user profile using the tags of historical images
  - Close locations share similar tags
  - Most images are spatially close to preceding image
  - Query image is likely to be taken within locations represented by user profile.

- Treat historical tags equally?
  - Close in taken time → Close in space
  - Analyze relation between distance and taken time gap.
  - Tags of recent images are more valuable.
Proposed Method
—Location Representation

• Place a grid over the world map, and define a location as a cell on it.
• Different granularity: 1, 10, 100km over latitude.
• Map the photos to the corresponding cells.
Proposed Method
— Using Image Tags

• **k-Nearest Neighbors Method (kNN)**
  – Each image is represented by a vector on tags. Predict location of an image using k nearest neighbors.

• **Language Model Based Method (LM)**
  – The textual tags of photos are used to derive a language model that represents a location.

\[
P(l | T) = \frac{P(T | l)P(l)}{P(T)} \propto P(T | l) = \prod_{t \in T} P(t | l)
\]

\[
P(t | l) = \frac{|l|}{|l|+\lambda} P(t | l)_{ML} + \frac{\lambda}{|l|+\lambda} P(t | G)_{ML}
\]

• **Naïve Bayes Method (NB, NB+RT, NB+CT)**
  – Expand a query image’s tag set by adding with its owner's history tags until |T|=K.

\[
P(l | T) = \frac{P(T | l)P(l)}{P(T)} \propto P(T | l)P(l) = \prod_{t \in T} P(t | l)P(l)
\]
Proposed Method
—Example of Language Model

\[ P(l | T) = \frac{P(T | l)P(l)}{P(T)} \propto P(T | l) = \prod_{t \in T} P(t | l) \]
\[ P(t | l) = \frac{|l|}{|l| + \lambda} P(t | l)_{ML} + \frac{\lambda}{|l| + \lambda} P(t | G)_{ML} \]

<table>
<thead>
<tr>
<th>tag</th>
<th>L</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>NTU</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>SG</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>UK</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>

Term Frequency

<table>
<thead>
<tr>
<th>tag</th>
<th>L</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>NTU</td>
<td>0.75</td>
<td>0.375</td>
</tr>
<tr>
<td>SG</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>UK</td>
<td>0</td>
<td>0.375</td>
</tr>
</tbody>
</table>

Maximum Likelihood

<table>
<thead>
<tr>
<th>tag</th>
<th>L</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>NTU</td>
<td>0.5625</td>
<td></td>
</tr>
<tr>
<td>SG</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>UK</td>
<td>0.1875</td>
<td></td>
</tr>
</tbody>
</table>

Smoothed

\[ P(l | T) = 0.5625^2 \times 0.25 \times 0.1875 \approx 0.015 \]
Proposed Method
— Using User Profile

- Model User Profile (P, WP)
  - Treat the set of entire historical tags used by a user as her profile, and build a language model U to model it.
  - Two choices:
    - Equal weight for every tag (P)
    - Different weight for different tag based on taken time of an image (WP)

- Measure similarity between the user profile and each location
  - Use KL divergence
    \[ D_{KL}(U \parallel L) = \sum_i U(i) \log \frac{U(i)}{L(i)} \]
Proposed Method
—Combining Image Tags and User Profile

- Unified Framework (NB+P, NB+WP)
  - Combine the evidence of image tags and user profile

\[
S = (1 - b^{-n})S_{\text{tag}} + b^{-n}S_{\text{user}}
\]

\[
S_{\text{user}} = S_{D_{KL}(U \parallel L)} = 1 - \frac{D_{KL}(U \parallel L) - \min D_{KL}}{\max D_{KL} - \min D_{KL}}
\]

\[
S_{\text{tag}} = S_{P(l \mid T)} = \frac{P(l \mid T) - \min P}{\max P - \min P}
\]
Experiments

- **Dataset:**
  - Training set: 3,491,429 images uploaded before March 1st 2011
  - Tuning set: 10,000 images uploaded after March 1st 2011
  - Testing set: 10,000 images uploaded after March 1st 2011

- **Metric:**
  - Accuracy (Acc)
  - Fraction of images predicted correctly within k-cell distance (Acc@k)
  - Fraction of images predicted correctly among top-k locations (Top-k)
Experiment Results

<table>
<thead>
<tr>
<th></th>
<th>Acc</th>
<th>Acc@1</th>
<th>Acc@2</th>
<th>Acc@3</th>
<th>Top-2</th>
<th>Top-3</th>
<th>Top-4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1K M</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kNN</td>
<td>0.0521</td>
<td>0.0848</td>
<td>0.1054</td>
<td>0.1194</td>
<td>0.0733</td>
<td>0.0852</td>
<td>0.0922</td>
</tr>
<tr>
<td>LM</td>
<td>0.0576</td>
<td>0.0958</td>
<td>0.1173</td>
<td>0.1297</td>
<td>0.0815</td>
<td>0.0935</td>
<td>0.1029</td>
</tr>
<tr>
<td>NB</td>
<td>0.0598</td>
<td>0.0989</td>
<td>0.1217</td>
<td>0.1354</td>
<td>0.0811</td>
<td>0.0947</td>
<td>0.1033</td>
</tr>
<tr>
<td>NB+RT</td>
<td>0.0572</td>
<td>0.0946</td>
<td>0.1197</td>
<td>0.1351</td>
<td>0.0771</td>
<td>0.0906</td>
<td>0.0995</td>
</tr>
<tr>
<td>NB+CT</td>
<td>0.0634</td>
<td>0.1065</td>
<td>0.1352</td>
<td>0.1521</td>
<td>0.0866</td>
<td>0.1011</td>
<td>0.1123</td>
</tr>
<tr>
<td>P</td>
<td>0.0121</td>
<td>0.0270</td>
<td>0.0438</td>
<td>0.0599</td>
<td>0.0187</td>
<td>0.0236</td>
<td>0.0286</td>
</tr>
<tr>
<td>WP</td>
<td>0.0389</td>
<td>0.0737</td>
<td>0.0960</td>
<td>0.1178</td>
<td>0.0565</td>
<td>0.0696</td>
<td>0.0775</td>
</tr>
<tr>
<td>NB+P</td>
<td>0.0577</td>
<td>0.0968</td>
<td>0.1230</td>
<td>0.1430</td>
<td>0.0800</td>
<td>0.0934</td>
<td>0.1046</td>
</tr>
<tr>
<td>NB+WP</td>
<td>0.0637 (+22.3%)</td>
<td>0.1067</td>
<td>0.1334</td>
<td>0.1546</td>
<td>0.0858</td>
<td>0.1007</td>
<td>0.1138</td>
</tr>
<tr>
<td>10K M</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kNN</td>
<td>0.1420</td>
<td>0.1809</td>
<td>0.1953</td>
<td>0.2032</td>
<td>0.1774</td>
<td>0.1904</td>
<td>0.1997</td>
</tr>
<tr>
<td>LM</td>
<td>0.1550</td>
<td>0.2014</td>
<td>0.2187</td>
<td>0.2263</td>
<td>0.1996</td>
<td>0.2135</td>
<td>0.2228</td>
</tr>
<tr>
<td>NB</td>
<td>0.1581</td>
<td>0.2058</td>
<td>0.2245</td>
<td>0.2312</td>
<td>0.2010</td>
<td>0.2218</td>
<td>0.2368</td>
</tr>
<tr>
<td>NB+RT</td>
<td>0.1602</td>
<td>0.1941</td>
<td>0.1985</td>
<td>0.2005</td>
<td>0.2084</td>
<td>0.2311</td>
<td>0.2447</td>
</tr>
<tr>
<td>NB+CT</td>
<td>0.1794</td>
<td>0.2145</td>
<td>0.2193</td>
<td>0.2212</td>
<td>0.2323</td>
<td>0.2575</td>
<td>0.2722</td>
</tr>
<tr>
<td>P</td>
<td>0.0710</td>
<td>0.1214</td>
<td>0.1557</td>
<td>0.1697</td>
<td>0.1004</td>
<td>0.1164</td>
<td>0.1280</td>
</tr>
<tr>
<td>WP</td>
<td>0.1309</td>
<td>0.1982</td>
<td>0.2328</td>
<td>0.2493</td>
<td>0.1734</td>
<td>0.1996</td>
<td>0.2132</td>
</tr>
<tr>
<td>NB+P</td>
<td>0.1685</td>
<td>0.2301</td>
<td>0.2589</td>
<td>0.2717</td>
<td>0.2138</td>
<td>0.2365</td>
<td>0.2503</td>
</tr>
<tr>
<td>NB+WP</td>
<td>0.1811 (+27.5%)</td>
<td>0.2442</td>
<td>0.2721</td>
<td>0.2851</td>
<td>0.2285</td>
<td>0.2515</td>
<td>0.2669</td>
</tr>
<tr>
<td>100K M</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kNN</td>
<td>0.2101</td>
<td>0.2362</td>
<td>0.2510</td>
<td>0.2609</td>
<td>0.2437</td>
<td>0.2565</td>
<td>0.2634</td>
</tr>
<tr>
<td>LM</td>
<td>0.2442</td>
<td>0.2766</td>
<td>0.2976</td>
<td>0.3081</td>
<td>0.2758</td>
<td>0.2898</td>
<td>0.3006</td>
</tr>
<tr>
<td>NB</td>
<td>0.2463</td>
<td>0.2818</td>
<td>0.3056</td>
<td>0.3231</td>
<td>0.2977</td>
<td>0.3300</td>
<td>0.3490</td>
</tr>
<tr>
<td>NB+RT</td>
<td>0.2839</td>
<td>0.3349</td>
<td>0.3644</td>
<td>0.3849</td>
<td>0.3399</td>
<td>0.3685</td>
<td>0.3854</td>
</tr>
<tr>
<td>NB+CT</td>
<td>0.3056</td>
<td>0.3585</td>
<td>0.3912</td>
<td>0.4151</td>
<td>0.3596</td>
<td>0.3870</td>
<td>0.4068</td>
</tr>
<tr>
<td>P</td>
<td>0.2196</td>
<td>0.2828</td>
<td>0.3209</td>
<td>0.3456</td>
<td>0.2729</td>
<td>0.3044</td>
<td>0.3311</td>
</tr>
<tr>
<td>WP</td>
<td>0.2978</td>
<td>0.3643</td>
<td>0.4020</td>
<td>0.4283</td>
<td>0.3560</td>
<td>0.3893</td>
<td>0.4185</td>
</tr>
<tr>
<td>NB+P</td>
<td>0.3023</td>
<td>0.3633</td>
<td>0.3973</td>
<td>0.4189</td>
<td>0.3640</td>
<td>0.3967</td>
<td>0.4214</td>
</tr>
<tr>
<td>NB+WP</td>
<td>0.3277 (+56.0%)</td>
<td>0.3899</td>
<td>0.4254</td>
<td>0.4491</td>
<td>0.3856</td>
<td>0.4193</td>
<td>0.4439</td>
</tr>
</tbody>
</table>

- Combining image tags and user profile together is able to achieve much better result.
- The taken time stamps of historical images play an important role in user profile.
Conclusion

- We focus on the problem of suggesting geolocations for social images.
- For the first time we analyze the user uploading patterns, geotagging behaviors and the relationship between taken-time gap of two images and their distance.
- We proposed a unified framework for geolocation suggestion.
- Experimental results show the effectiveness of the proposed method.
Q & A

Thanks!
Any questions?