Placing images on the world map: a microblog-based enrichment approach

Claudia Hauff & Geert-Jan Houben, TU Delft (NL)
SIGIR 2012
Problem & Motivation

Travel Timeline

User Profile

Personal Archive Organization

Travelogue Illustration

Latitude/Longitude: 48.23/-74.34

Latitude/Longitude: 52.4/-3.2

Latitude/Longitude: 48.23/-74.34

Latitude/Longitude: 44.65/-63.22

Latitude/Longitude: 4.65/-15.19

Latitude/Longitude: 0.44/-12.29

our Flickr data set: 80% of images not geo-tagged

TU Delft

Claudia Hauff, 2012
flickr image with geo-tag

By Perspective II Photography
Nathan Pirtz  + Add Contact

This photo was taken on April 20, 2012 in Portland, Oregon, US, using a Nikon D80.

This photo belongs to

Perspective II Photography's photostream (258)

This photo also appears in

- Personal Favorites (set)
- Oregon (set)

Source: http://www.flickr.com/photos/nathanpirtz/6963996476/
Portland Japanese Garden

Source: http://www.flickr.com/photos/nathanpirtz/6963996476/
flickr image without geo-tag

Source: http://www.flickr.com/photos/29738009@N08/2975466425/
On a visit to the beautiful **Japanese Garden** in **Portland, Oregon #mustsee #pdx**

2:03PM – 18 June 2010

Source: http://www.flickr.com/photos/nido/4737115541/
The past vs. our approach

- Location estimation based on text (mainly image tags)
  - Serdyukov et al., Van Laere et al.

- Location estimation based on visual features
  - Lux et al.

- Hybrid approaches (visual features as backup in the estimation)
  - Kelm et al.

- This work: text-based, merges traces of the user on different social Web streams (cross-system exploitation)

Hypothesis: enriching the image’s textual meta-data with the user’s tweets improves the accuracy of the location estimation.
Why do people tweet?

Why do we think our hypothesis holds?

• Tweet categories, Java et al.
  • Daily chatter
  • Shared information and hyperlinks
  • Conversations
  • News

• Majority of users (~80%) focus on themselves, Naaman et al.

• Users’ view on the why, Zhao et al.
  • Keeping in touch
  • Collecting information (work & spare-time related)
From images to documents

Formulating an information retrieval problem

- Given a set of training images with known latitude/longitude
  - Start with a grid cell spanning the world map
  - Iteratively training images
  - Split dense cells
    ➞ cells of small size in regions with large amounts of training data

- Each cell is transformed into a "region document"
  - The textual meta-data across the images is concatenated into one document
From documents to location estimation

Formulating an information retrieval problem

- A language model is derived from each world region (document)

  \[
  T_I = \{t_1, t_2, \ldots, t_n\}
  \]

- The possible regions where test image \( I \) with textual metadata may have been taken are ranked according to:

\[
P(\theta_R | T_I) = \frac{P(T_I | \theta_R)P(\theta_R)}{P(T_I)} \propto P(\theta_R) \times \prod_{i=1}^{n} P(t_i | \theta_R)
\]

- Assign \( I \) the location of the top ranked training image
Eliminating noisy terms

Geographic spread filtering

• Not all image tags/terms are equally useful

<table>
<thead>
<tr>
<th>bowling</th>
<th>london</th>
</tr>
</thead>
<tbody>
<tr>
<td>baby</td>
<td>british</td>
</tr>
</tbody>
</table>

• Spread of training images on the world map is a good indicator
Eliminating noisy terms

Geographic spread filtering

UK, British Columbia, the British Virgin Islands, British restaurants in the US, places with historic battles against the British

<table>
<thead>
<tr>
<th>Term</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>bowling</td>
<td>3.237</td>
</tr>
<tr>
<td>baby</td>
<td>1.809</td>
</tr>
<tr>
<td>east</td>
<td>0.695</td>
</tr>
<tr>
<td>british</td>
<td>0.363</td>
</tr>
<tr>
<td>lakepukaki</td>
<td>0.049</td>
</tr>
<tr>
<td>london</td>
<td>0.010</td>
</tr>
<tr>
<td>sydney</td>
<td>0.007</td>
</tr>
</tbody>
</table>

\[ \theta_{geo} \]
Adding additional knowledge

\[ P(\theta_R \mid T_I) = \frac{P(T_I \mid \theta_R)P(\theta_R)}{P(T_I)} \propto P(\theta_R) \times \prod_{i=1}^{n} P(t_i \mid \theta_R) \]

- **Region prior**: instead of a uniform probability, add knowledge about the world and the different regions of the world
  - Population density, climate
- **Set of terms**: the bag-of-words that describe an image can be extended by including terms from the user’s traces on the social Web
  - Tweets within \( D \) days of the image being taken

"New York City" 3,869,086 results
"Great Victoria Desert" 131 results
Experimental setup

- Training data: MediaEval data set, **3.2M** geo-tagged images
  - Lack of usable Twitter resources: few geo-tagged tweets

- Test data: starting with an 11 months Twitter data set of **20,000** users, we searched for corresponding Flickr accounts
  - A crawl of friendfeed.com profiles
  - Manual assessment of posted tweets

Nov'10 – Sept'11
27,879 images
7477 geo-tagged
30,951 images

252 users
1.89M tweets
0.15M images
## Results

<table>
<thead>
<tr>
<th></th>
<th>Percentage of test images within</th>
<th>Median error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1km</td>
<td>10km</td>
</tr>
<tr>
<td><strong>Baseline</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BaseLine</td>
<td>7.5%</td>
<td>27.9%</td>
</tr>
<tr>
<td>BaseLine_geo</td>
<td>7.2%</td>
<td>35.0%</td>
</tr>
<tr>
<td><strong>Population</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BaseLine</td>
<td>7.1%</td>
<td>34.7%</td>
</tr>
</tbody>
</table>

BaseLine: 5600 unique terms
BaseLine_geo: 466 unique terms
# Results

<table>
<thead>
<tr>
<th>Method</th>
<th>Median error</th>
<th>1km</th>
<th>10km</th>
<th>50km</th>
<th>1000km</th>
</tr>
</thead>
<tbody>
<tr>
<td>BaseLine</td>
<td>61km</td>
<td>7.5%</td>
<td>27.9%</td>
<td>34.9%</td>
<td>42.3%</td>
</tr>
<tr>
<td>BaseLine geo</td>
<td>61km</td>
<td>7.2%</td>
<td>35.0%</td>
<td>48.6%</td>
<td>61.4%</td>
</tr>
<tr>
<td>Population</td>
<td>62km</td>
<td>7.1%</td>
<td>34.7%</td>
<td>48.4%</td>
<td>70.4%</td>
</tr>
<tr>
<td>+/-2 days</td>
<td>1974km</td>
<td>4.3%</td>
<td>16.9%</td>
<td>25.2%</td>
<td>41.8%</td>
</tr>
<tr>
<td>+/-2 days geo</td>
<td>22km</td>
<td>9.0%</td>
<td>38.2%</td>
<td>54.7%</td>
<td>71.2%</td>
</tr>
<tr>
<td>+/-20 days</td>
<td>27km</td>
<td>8.3%</td>
<td>36.7%</td>
<td>53.6%</td>
<td>70.8%</td>
</tr>
<tr>
<td>+/-2 days pop</td>
<td>21km</td>
<td>9.0%</td>
<td>37.9%</td>
<td>54.6%</td>
<td>76.0%</td>
</tr>
</tbody>
</table>

BaseLine: 5600 unique terms
BaseLine geo: 466 unique terms
## Results

Image location estimation based on user traces across social Web platforms decreases the median error distance by up to 67%.

<table>
<thead>
<tr>
<th></th>
<th>1km</th>
<th>10km</th>
<th>50km</th>
<th>1000km</th>
<th>61km</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BaseLine</strong></td>
<td>7.5%</td>
<td>27.9%</td>
<td>54.9%</td>
<td>61.4%</td>
<td>61km</td>
</tr>
<tr>
<td><strong>BaseLine</strong>&lt;sub&gt;geo&lt;/sub&gt;</td>
<td>5600 unique terms</td>
<td>466 unique terms</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The population density prior improves the accuracy (in the long range).

<table>
<thead>
<tr>
<th></th>
<th>1km</th>
<th>10km</th>
<th>50km</th>
<th>1000km</th>
<th>62km</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Population</strong></td>
<td>7.1%</td>
<td>34.7%</td>
<td>48.4%</td>
<td>70.4%</td>
<td>62km</td>
</tr>
<tr>
<td><strong>Population</strong></td>
<td>66x175</td>
<td>+/-2 days</td>
<td>9.0%</td>
<td>38.2%</td>
<td>54.7%</td>
</tr>
<tr>
<td><strong>Population</strong></td>
<td>66x175</td>
<td>+/-20 days</td>
<td>8.3%</td>
<td>36.7%</td>
<td>53.6%</td>
</tr>
</tbody>
</table>

BaseLine: 5600 unique terms

BaseLine<sub>geo</sub>: 466 unique terms
What images benefit from Twitter enrichment?

Number of tags

Test set split according to the number of tags after geo-filtering.
What images benefit from Twitter enrichment?

Number of tags

The Twitter stream is particularly useful in cases of little or no textual meta-data.

Test set split according to the number of tags after geo-filtering.
What images benefit from Twitter enrichment?

Distance to home location (4515 test images)
What images benefit from Twitter enrichment?

Distance to home location (4515 test images)

Locations further away from home are recognized with higher accuracy when using the Twitter stream.
Conclusions & future work

- Image location estimation based on user traces across social Web platforms outperforms the single-source baseline

- The Twitter stream is particularly useful in cases of little or no textual meta-data

- The population density prior improves the accuracy (in the long range)

- Future work
  - Tweet filtering (personal experiences vs. news)
  - Improved combination of data gathered from social Web streams
  - Turning the task around: user account matching
Thank you!

c.hauff@tudelft.nl