Placing Flickr Photos on a Map

Fateme Shirazi

Saarland University
Outline

- Introduction
- Representing Locations on the Map
- Modeling Locations
- Experimental Setup
- Results
- Error analysis
- Conclusion
Outline

- Introduction
- Representing Locations on the Map
- Modeling Locations
- Experimental Setup
- Results
- Error analysis
- Conclusion
Introduction

- GPS-enabled cameras and mobile phones
  - latitude and longitude

- Geo-referenced photos
  - organized in a browsable taxonomy
  - or pin-pointed on a map
Introduction

Objective of the paper:

- Automatically placing photos uploaded in Flickr on the world map, using tags provided by the user.
Frame work of the paper:

- Constructing an $m \times n$ grid based on the longitude/latitude.
- Using a set of images whose locations are known.
- Placing images in its corresponding grid cell.
- Using the description of the images given by the Flickr users.
- Finally, investigating how to incorporate external resources into the model.
Outline

- Introduction
- Representing Locations on the Map
- Modeling Locations
- Experimental Setup
- Results
- Error analysis
- Conclusion
Locations as bags-of-tags

- Each photo has a FlickrID, a geographic co-ordinate, and a set of tags.
- Map each geographic co-ordinate on the map.
- Derive a language model from tags of photos at this location.
Representing Locations on the Map

- Locations as a graph
  - Spatial relationship to build an undirected graph.
    - Where the link exists only if situated close enough on the grid.
  - Locations found within a predefined distance are linked and considered to be neighbors.
Representing Locations on the Map

- Locations as *pseudo-documents*.
- Linked locations more likely have *similar* tags.
- Locations relevant to an image being close in the graph.
Outline

- Introduction
- Related work
- Representing Locations on the Map
- **Modeling Locations**
- Experimental Setup
- Results
- Error analysis
- Conclusion
Modeling Locations

- Ranking list of locations $L$, for a given tag set $T$ belonging to an image taken within the bounds of $L$:

$$P(L|T) = \frac{P(T|L)P(L)}{P(T)}$$

- $P(T|L)$ : the probability to generate the tag set of the image.

$$P(T|L) = \prod_{i=1}^{\mid T \mid} P(t_i|L)$$
Modeling Locations

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

- \( |L| \) : the size of the location \( L \) in tags.
- \( P(t|L) \) : maximum likelihood estimates of tag generation probabilities for the location’s language model.
- \( P(t|G) \) : maximum likelihood estimates of tag generation probabilities for the global language model.
Smoothing

1. Tag-based smoothing with neighbors
2. Smoothing cell relevance probabilities
3. Boosting geo-related tags
4. Spatial ambiguity-aware smoothing
1. Tag-based smoothing with neighbors

To overcome data sparseness.

Tags indicate an area exceeding the bounds of a location.
1. Tag-based smoothing with neighbors (con’t)

- Considering that each tag is generated by
  - location’s language model,
  - or by language models of neighboring locations:

- \( \text{NB}(L) \) : all locations included in the neighborhood of location \( L \).

\[
P(t|L) = \mu \frac{|L| \cdot P(t|L)_{ML}}{|L| + \lambda} + (1-\mu) \frac{P(t|NB(L))}{|L| + \lambda} + \frac{\lambda \cdot P(t|G)_{ML}}{|L| + \lambda}
\]

\[
P(t|NB(L)) = \sum_{L' \in NB(L)} \frac{|L'|}{|L'| + \lambda \frac{P(t|L')_{ML}}{(2d + 1)^2 - 1}}
\]
2. Smoothing cell relevance probabilities

- “Good” locations come from “good” neighborhoods.
- Some relevance should be propagated through the links between close locations.
- The tag set probability augmented with the probabilities of neighboring locations:

\[
P(T|L) = \alpha P(T|L) + (1 - \alpha) \sum_{L' \in NB(L)} \frac{P(T|L')}{(2d + 1)^2 - 1}
\]
2. Smoothing cell relevance probabilities
Propagation in direction of higher relevance

- Propagating relevance from locations having lower scores than the location to be smoothed.

- Support documents with enough probability to be relevant, rather than relevant documents support poor ones.

- Edges between cells become directed (from lower scored to higher scored cells)
3. Boosting geo-related tags

- Some tags are more popular near certain locations even without analysis of their spatial distribution.

\[
P(t|L)_{ML} = P(t|L)_{ML} (1 + \beta P(\text{Loc}|t) / Z)
\]

- \(\beta\): boosting coefficient
- \(P(\text{Loc}|t)\): probability of the tag \(t\) to be location-specific.
- \(Z\): normalisation coefficient.
4. Spatial ambiguity-aware smoothing

- Some tags are specific for more than one location:
  - Their scope exceeds the bounds of a single cell.
  - or due to their ambiguity (for example bath and Bath, UK).
  - or because they are instances that are typically spotted at a few specific locations, such as Elephants.
4. Spatial ambiguity-aware smoothing (con’t)

- The smoothing coefficient $\lambda(t)$ in Equation 3 will be tag-specific and proportional to the ambiguity of a tag:

$$\lambda(t) = \lambda + \gamma(\sigma_{lat}(t) + \sigma_{lon}(t))$$

- $\gamma$: weight coefficient to control the influence of ambiguity level on smoothing.

- Standard deviation of its latitudes and longitudes
Outline

- Introduction
- Representing Locations on the Map
- Modeling Locations
- Experimental Setup
- Results
- Conclusion
Experimental Setup

- The Flickr data set
  - A randomly sampled set of 397,000 Flickr images with the associated tags.
  - A filter applied to remove the effect of bulk uploads by users.
    - This reduces the set of photos to 140,000.
  - For better understanding the data, all images are georeferenced with the GeoNames gazetteer.
  - Overall, the collection contains photos from about 180 different countries.
Evaluation measures

- *(Acc)* the percentage of correct predictions over all test examples.

- *(MRR)* the ability of the system to find the actual location of a photo among its top recommendations.

\[
\frac{1}{N} \sum_{r=1}^{N} \frac{1}{r}
\]

- \( r \) : first rank with correct position.
- \( N \) : number of locations.
Experimental Setup

- **Acc@K** whether the actual location is within a K-cell distance from the predicted location.
  - Cell-based distance

- **PAcc** whether the predicted location belongs to the same parent with the correct location (for instance, 100 km cells are parents for 50 km cells, 50 km cells - for 10 km cells, etc.).
Location belongs to the same parent
Results

- Baseline language model LM
- Tag-based smoothing TS (1)
- Cell-based smoothing CS (2)
- Cell based smoothing with score propagation in the direction of higher relevance CSR (2’)
- Toponym based boosting TB (3)
- Ambiguity-aware tag specific smoothing AS (4)
Performance of the baseline LM method

<table>
<thead>
<tr>
<th>Division</th>
<th>Acc</th>
<th>MRR</th>
<th>Acc@1</th>
<th>Acc@2</th>
<th>Acc@3</th>
<th>PAcc</th>
</tr>
</thead>
<tbody>
<tr>
<td>1km</td>
<td>0.067</td>
<td>0.073</td>
<td>0.125</td>
<td>0.152</td>
<td>0.170</td>
<td>0.122</td>
</tr>
<tr>
<td>5km</td>
<td>0.140</td>
<td>0.155</td>
<td>0.226</td>
<td>0.248</td>
<td>0.259</td>
<td>0.177</td>
</tr>
<tr>
<td>10km</td>
<td>0.181</td>
<td>0.197</td>
<td>0.261</td>
<td>0.278</td>
<td>0.291</td>
<td>0.247</td>
</tr>
<tr>
<td>50km</td>
<td>0.256</td>
<td>0.277</td>
<td>0.332</td>
<td>0.354</td>
<td>0.378</td>
<td>0.289</td>
</tr>
<tr>
<td>100km</td>
<td>0.288</td>
<td>0.309</td>
<td>0.370</td>
<td>0.410</td>
<td>0.435</td>
<td>-</td>
</tr>
</tbody>
</table>
Results

- The accuracy increases,
  - when increasing the grid size, from 0.067 to 0.288.

- Additional performance improvement is observed
  - when analyzing the relaxed accuracy measures to include the direct neighbors of the predicted location.
## Performance of neighborhood smoothing

<table>
<thead>
<tr>
<th>Method</th>
<th>Acc</th>
<th>MRR</th>
<th>Acc@1</th>
<th>Acc@2</th>
<th>Acc@3</th>
<th>PAcc</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 km</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LM</td>
<td>0.067</td>
<td>0.073</td>
<td>0.125</td>
<td>0.152</td>
<td>0.170</td>
<td>0.122</td>
</tr>
<tr>
<td>+TS</td>
<td>0.068</td>
<td>0.074</td>
<td>0.128</td>
<td>0.160</td>
<td>0.180</td>
<td>0.129</td>
</tr>
<tr>
<td>+CS</td>
<td>0.066</td>
<td>0.073</td>
<td>0.13</td>
<td>0.158</td>
<td>0.179</td>
<td>0.126</td>
</tr>
<tr>
<td>+CSR</td>
<td><strong>0.070</strong></td>
<td><strong>0.075</strong></td>
<td><strong>0.141</strong></td>
<td><strong>0.176</strong></td>
<td><strong>0.197</strong></td>
<td><strong>0.140</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>10 km</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LM</td>
<td>0.181</td>
<td>0.197</td>
<td>0.261</td>
<td>0.278</td>
<td>0.291</td>
<td>0.247</td>
</tr>
<tr>
<td>+TS</td>
<td>0.181</td>
<td>0.197</td>
<td>0.260</td>
<td>0.278</td>
<td>0.291</td>
<td>0.245</td>
</tr>
<tr>
<td>+CS</td>
<td>0.183</td>
<td>0.195</td>
<td>0.266</td>
<td>0.285</td>
<td>0.297</td>
<td>0.252</td>
</tr>
<tr>
<td>+CSR</td>
<td><strong>0.187</strong></td>
<td><strong>0.201</strong></td>
<td><strong>0.271</strong></td>
<td><strong>0.288</strong></td>
<td><strong>0.301</strong></td>
<td><strong>0.255</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>100 km</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LM</td>
<td>0.288</td>
<td>0.309</td>
<td>0.370</td>
<td>0.410</td>
<td>0.435</td>
<td>-</td>
</tr>
<tr>
<td>+TS</td>
<td>0.290</td>
<td>0.311</td>
<td>0.371</td>
<td>0.409</td>
<td>0.437</td>
<td>-</td>
</tr>
<tr>
<td>+CS</td>
<td>0.289</td>
<td>0.310</td>
<td>0.387</td>
<td>0.430</td>
<td>0.456</td>
<td>-</td>
</tr>
<tr>
<td>+CSR</td>
<td><strong>0.296</strong></td>
<td><strong>0.314</strong></td>
<td><strong>0.390</strong></td>
<td><strong>0.443</strong></td>
<td><strong>0.470</strong></td>
<td>-</td>
</tr>
</tbody>
</table>
Results

- Marginal improvements,
  - CSR method outperforming the other two smoothing extensions independent of grid size, 1, 10, and 100 km.

- Smoothing was only done with the immediate neighbors ($d = 1$).
## Performance of combinations of methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Acc</th>
<th>MRR</th>
<th>Acc@1</th>
<th>Acc@2</th>
<th>Acc@3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>10 km</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LM</td>
<td>0.181</td>
<td>0.197</td>
<td>0.261</td>
<td>0.278</td>
<td>0.291</td>
</tr>
<tr>
<td>+CSR</td>
<td>0.187</td>
<td>0.201</td>
<td>0.271</td>
<td>0.288</td>
<td>0.301</td>
</tr>
<tr>
<td>+TB</td>
<td>0.198</td>
<td>0.209</td>
<td>0.283</td>
<td>0.303</td>
<td>0.316</td>
</tr>
<tr>
<td>+TB+CSR</td>
<td>0.198</td>
<td>0.210</td>
<td>0.286</td>
<td>0.305</td>
<td>0.319</td>
</tr>
<tr>
<td>+AS</td>
<td>0.190</td>
<td>0.205</td>
<td>0.275</td>
<td>0.292</td>
<td>0.306</td>
</tr>
<tr>
<td>+AS+TB</td>
<td>0.204</td>
<td>0.213</td>
<td>0.295</td>
<td>0.314</td>
<td>0.329</td>
</tr>
<tr>
<td>+AS+CSR</td>
<td>0.194</td>
<td>0.206</td>
<td>0.285</td>
<td>0.303</td>
<td>0.317</td>
</tr>
<tr>
<td><strong>+AS+TB+CSR</strong></td>
<td><strong>0.204 (+13%)</strong></td>
<td><strong>0.213 (+8%)</strong></td>
<td><strong>0.297 (+14%)</strong></td>
<td><strong>0.316 (+14%)</strong></td>
<td><strong>0.332 (+14%)</strong></td>
</tr>
</tbody>
</table>
All methods improve over baseline LM for all measures.

The maximum performance is reached by using all three methods together.

The results for accuracy-based measures were not tested for statistical significance.

- They have a binary outcome (correct or incorrect).
Results
Results

- High relaxed accuracies with K up to 50.
- For regions with larger bounds the performance of the baseline method increases accordingly.
- The benefit from the advanced methods stays the same, probably avoiding especially coarse errors.
Error analysis

- There are two main sources of errors:
  - caused by sparsity and noisiness
  - arising from ambiguity and incompleteness of the tag sets.

- In the first case, the right location is
  - not represented in the data,
  - or poorly represented with tags specific for this location only (e.g. containing no toponyms).
Error analysis

For the second case, images difficult to localize:

1. Images with tags specific to too many locations (e.g. beach coast rocks lovers);
2. Images with toponyms, but with no tags disambiguating them (e.g. michigan cats dogs);
3. Images with a tag falsely indicating to a location (e.g. paris hilton picturing a poster in New York);
4. Images containing a tag specific to a region larger than a chosen grid cell size (e.g. Alaska snow for 100 km cells).
Error analysis

- The first three types of errors can be eliminated in the future by taking some contextual or user-specific evidence:
  - for instance, tags of recently uploaded images or the location of user IP.

- Highly ambiguous tag sets may be successfully mapped by relying on the history of user locations,
  - since such tag sets might be location-specific on the personal level (e.g., people celebrating their birthday in their home location).
Error analysis

- Resolved with additional image content analysis.

- Errors resulting from images containing tags specific to a larger region are more difficult to avoid.
Error analysis

- Photos taken in very popular tourist destinations, are represented better among correct mappings than in the entire data set,
  - tourists almost always describe their photos with location specific tags.
Conclusions

- A generic method for automatically placing photos uploaded in Flickr on the world map.

- Estimate a language model through analysis of the terms people use to describe images taken at a particular location.

- Increasing the accuracy of the predictions
  - by incorporating ambiguity-aware smoothing,
  - cell-based smoothing with score propagation in the direction of highly relevant neighbors.
Improving suggestions

First, to define automatically and appropriate grid division for a tag set.

Minimize interactions between users and the system
- by showing a map view at the optimal zoom level

Second, study the utility of additional evidence coming from a user profile,
- uploads history, social network or IP address.

Finally, images used to build location models can be distinguished
- by using common (e.g. noise ratio) or Flickr-specific (number of views, interestingness) quality measures.