Annotate Wikipedia with Flickr Images: Concepts and Case Study

Present by Jie Xiao

Joint work with Dr. Qi Tian

Dept. of Computer Science
Univ. of Texas at San Antonio
Outline

- Problem and Motivation
- System
- Approach
- Experiment
- Conclusion
Problem and Motivation

- Wikipedia as an open editable resource, provides reliable knowledge and taxonomy.

- Can we describe a Wikipedia concept visually with its corresponding knowledge?
Problem and Motivation (Cont.)

- A straightforward approach: annotate Wikipedia with labeled images
  - Time-consuming
  - Expensive

- Tagged images are good candidates.
Benefit of the data from social network:

- **Cheap:**
  - users labeled images with tags
  - Public API

- **Useful community information:**
  - Interest group collects images with the same theme

- **Useful user information:**
  - The users tag images with their habit or preference.
Observation:

- An image can be collected by multiple communities who share common interest of the same topic.
- The more times images are collected by the same groups the more likely they are topic related.
Simplified by focusing on Wikipedia entries with geographical information.

Objectives:
- Make use of communities’ effort in harvesting data from social network.
- Visually describe Wikipedia entries according to taxonomy.
- Obtain diverse and representative results for a given city.
WAS system consists of three components:
- Community contributed data collection
- Top-down geo clustering
- Bottom-up merging for ranking
System Overview

Three components in the system.

DATA COLLECTION
- Text Information: Parse wiki by taxonomy
- Visual Information: top 10 groups on 1000 seed images, co-occurrence

TOP-DOWN CLUSTERING
- Textual Rank:
  \[ w_y = \text{tf}_y \times \log_2 \frac{N}{n} \]
- TF-IDF cosine distance
- Visual Rank
- Page Rank
- Ranked

BOTTOM-UP Merging

jxiao@cs.utsa.edu
Data Collection

- Obtain 1000 seed images from Flickr API
- Parse webpage for each image in Flickr to get its group names
- Obtain the top 10 groups referred as trustable groups which collect the largest number of seed images
- Identify the authority seed images which are collected by trustable groups.
- Estimate the target city’s geo coordinate
- Filter the noisy seeds by the estimated geo location
- Expand the image dataset with all the images from the trustable groups.

jxiao@cs.utsa.edu
Data Collection (Cont.)

Text Information

Parse wiki by taxonomy

Visual Information

top 10 groups on 1000 seed images co-occurrence

Group info

Group: USA Favorites
Group: New York: landmarks - the best of...
Group: New York life
Group: New York, New York
Group: I Love NY
Group: New York, New York
Group: New, york, manhattan, culture
new, york, geography
Top-down geo clustering

Top-down clustering

1

2

3

4

Textual Rank

Visual Rank

Page Rank

TF-IDF cosine distance

\[ w_{ij} = tf_{ij} \times \log_2 \frac{N}{n} \]
Algorithm 1: Top-down geo clustering under density constraint

**Input:** $S = \{\langle \text{lat}_i, \text{lng}_i \rangle \}, \ i = 1, \ldots, \ N_s$, a set of geo coordinates.

- $R$, root node of current sub-tree

**Output:** $R$, geo coordinate tree, where image number of each leaf node should be no greater than the given density threshold

1. **Initialize** $C$, a set of child partition nodes by current geo range
2. For $i = 1, \ldots, N_s$ in $S$, do
3. Assign $\langle \text{lat}_i, \text{lng}_i \rangle$ to the corresponding child node
4. For $j = 1, \ldots, N_c$ in $C$, do
5. If the number of geo coordinate $> \text{density threshold}$
6. Recursively do Alg 1. with input $C_j$
7. Else
8. Do visual rank for all images in $C_j$

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rank</td>
<td>Combined visual and textual rank</td>
</tr>
<tr>
<td>RankV</td>
<td>Visual rank for a set of images in a leaf node</td>
</tr>
<tr>
<td>RankT</td>
<td>Textual rank for a set of images</td>
</tr>
<tr>
<td>R</td>
<td>Root node of current sub-tree</td>
</tr>
<tr>
<td>C</td>
<td>A set of children nodes of $R$, partitioning the geo area</td>
</tr>
<tr>
<td>S</td>
<td>Set of geo coordinates associated with images in current region</td>
</tr>
<tr>
<td>Nc</td>
<td>Number of children nodes of current node</td>
</tr>
<tr>
<td>Ns</td>
<td>Number of images in current geo region</td>
</tr>
<tr>
<td>Np</td>
<td>Number of images propagated up from child nodes</td>
</tr>
</tbody>
</table>
Bottom-up merging for ranking

Bottom – up Merging

jxiao@cs.utsa.edu
Algorithm 2: bottom-up merging

**Input:** R, node of pre-built geo information tree by Alg 1.

**Output:** Rank, ranked list for images in Ci

1. **If** R is a leaf node
2. **Load** RankT, pre-calculated rank on textual similarity.
3. **Load** RankV, rank on visual similarity, calculated in Alg. 1
4. **Rank** = merge (RankT, RankV), propagate up by proportion
5. Set status as ready to merge

**Else**
6. **For** j = 1, ..., Nc in C, children nodes of R, **do**
7. **If** Cj is not ready
8. **Recursively** do Alg 2 with input Cj
9. **For** i = 1, ..., Np in S, **do**
10. Re-rank images and update Rank

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rank</td>
<td>Combined visual and textual rank</td>
</tr>
<tr>
<td>RankV</td>
<td>Visual rank for a set of images in a leaf node</td>
</tr>
<tr>
<td>RankT</td>
<td>Textual rank for a set of images</td>
</tr>
<tr>
<td>R</td>
<td>Root node of current sub-tree</td>
</tr>
<tr>
<td>C</td>
<td>A set of children nodes of R, partitioning the geo area</td>
</tr>
<tr>
<td>S</td>
<td>Set of geo coordinates associated with images in current region</td>
</tr>
<tr>
<td>Nc</td>
<td>Number of children nodes of current node</td>
</tr>
<tr>
<td>Ns</td>
<td>Number of images in current geo region</td>
</tr>
<tr>
<td>Np</td>
<td>Number of images propagated up from child nodes</td>
</tr>
</tbody>
</table>
Visual Feature

- A scalable invariant feature (SIFT) is used to represent each local salient patch.

- 10% of extracted SIFT feature from all images are used to build a hierarchical clustering tree.

- A vector quantization step is proceeded to obtain a visual word histogram for each image.
Visual Similarity

\[
\text{Similarity} = \frac{\text{number of shared interest points}}{\text{average number of all the features in two images}}
\]

Hierarchical clustering on visual feature
Textual Feature

- Articles are distinguished by token and built as a dictionary.

- An article or paragraph is represented as a histogram over the dictionary considering the frequency a word occurred in a document and the log inverted frequency it appears in the whole document set. --- TF-IDF
Textual Similarity

- Cosine distance between any given two textual feature.
Geographical Feature

- The latitude and longitude of a geographical coordinate.

- Obtained by the meta data attached in images while the users upload images to Flickr.
In general, the highly rated images among the top 10% seed images are collected by over 30 groups. The most popular images are collected and favored by over 100 groups.
Case Study: New York

Top: Top 10 images by Flickr API
Bottom: Top 10 images by our approach
Relationship between seed images and corresponding groups
Images from top 10 groups and 1000 seeds before filtering.
Case Study: New York (Cont.)

**Geography**

Main articles: Geography of New York City and Geography of New York Harbor

New York City is located in the Northeastern United States, in southeastern New York State, approximately halfway between Washington, D.C., and Boston. The location at the mouth of the Hudson River, which feeds into a naturally sheltered harbor and then into the Atlantic Ocean, has helped the city grow in significance as a trading city. Much of New York is built on the three islands of Manhattan, Staten Island, and Long Island, making land scarce and encouraging a high population density.

The Hudson River flows through the Hudson Valley into New York Bay. Between New York City and Troy, New York, the river is an estuary. The Hudson separates the city from New Jersey. The East River—a tidal strait—flows from Long Island Sound and separates the Bronx and Manhattan from Long Island. The Harlem River, another tidal strait between the East and Hudson Rivers, separates Manhattan from the Bronx.
Conclusion

- We propose a community contributed data collection framework to harvest social media data.

- We design and build a prototype system to visually annotate geo-related entry in Wikipedia based on its taxonomy.
Thank you!