Advanced Computer Graphics CS 525M: Crowds replace Experts: Building Better Location-based Services using Mobile Social Network Interactions

XIAOCHEN HUANG

Computer Science Dept.
Worcester Polytechnic Institute (WPI)
Author

Pravin Shankar  Yun-wu huang  Paul Castro  Badri Nath  Liviu Iftode
Motivation

What is the location-based services?
Motivation
Motivation

Location-based services

- Are growing in popularity
- Face two challenges
  - Collect up-to-date information about places
  - Rank places

Can we build a location-based service by making use of mobile social network interactions?

SocialTelescope
Related work

- Leveraging social interactions
- Crowdsourcing
- Geo-social networks
- Location-based services
- Privacy issues in location sharing
SocialTelescope Design - Motivating Scenario

- A tourist carrying a smartphone reached New York
- Make several queries to a location-based services
- Indulge in some tourism
  - to a good seafood places
  - to a bar that plays good live music
- No friends to ask for, no luxury of time to visit
Figure 1. High-level Architecture of SocialTelescope
SocialTelescope Design - Architecture cont.

Crawling

● Record all public user interactions in mobile social networks, and store them in a repository

● Social network services typically provide well-defined APIs for external applications to crawl their public data

Indexing

● Text entered by a user is converted into a set of tags

● Three entities: locations, tags, users

● Find a user’s mobility profile, interests profile and location’s profile
SocialTelescope Design - Architecture cont.

Query Processing and Ranking Algorithm

- The fraction of total visits by user $u$ to places matching search term $q$ is computed as:

$$ F(q,u) = \frac{n_{q,u}}{\forall i \in Q n_{i,u}} $$

- The relative importance of search term $q$ is computed as:

$$ I(q) = \log \frac{|Q|}{\forall j \in U n_{q,j}} $$

- We define the user expertise score, $S_{q,u}$ as:

$$ S_{q,u} = F(q,u). I(q) $$

Algorithm 1: SocialTelescope Ranking Algorithm

1. Get the list of places $L$ matching search keyword $q$
2. Foreach user $u$:
3. Compute user expertise score $S_{q,u}$
4. Order each place in $L$ by number of user visits weighted by $S_{q,u}$
5. Return the top $k$ results in $L$
## Mobile Social Network Dataset

<table>
<thead>
<tr>
<th><strong>Start date</strong></th>
<th>June 14, 2010</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>End Date</strong></td>
<td>August 20, 2010</td>
</tr>
<tr>
<td><strong>Bounding box for geo-tweets</strong></td>
<td>(40.703, -74.022), (40.879, -73.899)</td>
</tr>
<tr>
<td><strong>Total geo-tweets in the region</strong></td>
<td>198919</td>
</tr>
<tr>
<td><strong>Total number of distinct users</strong></td>
<td>15659</td>
</tr>
<tr>
<td><strong>Total FourSquare checkins corresponding to the geo-tweets</strong></td>
<td>43461</td>
</tr>
<tr>
<td><strong>Total number of distinct FourSquare users</strong></td>
<td>6451</td>
</tr>
</tbody>
</table>

Table 1 details of the SocialTelescope dataset.
Mobile Social Network Dataset cont.

Hourly trends: number of tweets during different hours of day across the entire dataset

Distribution of number of tweets per year

Distribution of active days per user. An active Day refers to a day when a user at least one tweet

Mobility profile per user: number of distinct locations where a user tweeted from
Mobile Social Network Dataset cont.

User tags for locations in the form of a tag cloud

<table>
<thead>
<tr>
<th>Tag</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>bar</td>
<td>385</td>
</tr>
<tr>
<td>coffee</td>
<td>376</td>
</tr>
<tr>
<td>douchebag</td>
<td>355</td>
</tr>
<tr>
<td>pizza</td>
<td>340</td>
</tr>
<tr>
<td>beer</td>
<td>242</td>
</tr>
<tr>
<td>food</td>
<td>223</td>
</tr>
<tr>
<td>restaurant</td>
<td>206</td>
</tr>
<tr>
<td>brunch</td>
<td>181</td>
</tr>
<tr>
<td>gym</td>
<td>166</td>
</tr>
<tr>
<td>brooklyn</td>
<td>162</td>
</tr>
</tbody>
</table>

Trends in user tags for location
Mobile Social Network Dataset cont.

Heatmap of the entire region

Heatmap zoomed in to Central Park to show Trends at a finer granularity

Heatmap of the New York city region based on geo-tweets
Evaluation - Goals and Metrics

- **Goal**
  - Return relevant results efficiently

- **Key measures of the quality**
  - **Coverage**
    - How complete and up-to-date is information about different locations
  - **Relevance**
    - How relevant are the top results to the query
Evaluation - Methodology

- Focus on restaurant search
- Compare between user-review based & page-rank based
- User feedback
### Evaluation - coverage and Relevance Results

<table>
<thead>
<tr>
<th>Query</th>
<th>#Matches</th>
<th>#Experts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barbecue</td>
<td>65</td>
<td>116</td>
</tr>
<tr>
<td>Burger</td>
<td>166</td>
<td>238</td>
</tr>
<tr>
<td>Japanese</td>
<td>237</td>
<td>182</td>
</tr>
<tr>
<td>Indian</td>
<td>70</td>
<td>61</td>
</tr>
<tr>
<td>Seafood</td>
<td>85</td>
<td>60</td>
</tr>
<tr>
<td>Mexican</td>
<td>212</td>
<td>140</td>
</tr>
<tr>
<td>Chinese</td>
<td>165</td>
<td>84</td>
</tr>
<tr>
<td>Steak</td>
<td>78</td>
<td>28</td>
</tr>
<tr>
<td>Thai</td>
<td>102</td>
<td>31</td>
</tr>
<tr>
<td>Italian</td>
<td>332</td>
<td>175</td>
</tr>
</tbody>
</table>

Number of place and expert matches in dataset for the 10 test queries

Distribution of User Expertise Score for Different queries

Details of test queries using Zagat Top Places
Evaluation – Coverage and Relevance Results cont.

Total number of matches in SocialTelescope, for each of the test queries

Relevance of results returned by SocialTelescope, measured as fraction of Zagat Top Places that are contained in the result set
Comparison of ranking results of SocialTelescope, Google Local Search and Yelp, by assuming the Zagat Top Places list to be the ground truth.
Evaluation – user feedback

● Test to 8 users

● Three sets of results
  – SocialTelescope & Google Local Search & Yelp

● Reports from the feedback
  – SocialTelescope performs better for queries that are subjective in nature
  – When a user wanted only places with a speciality cuisine, SocialTelescope did not perform well
Discussion

- Spoofing locations
  Validating a mobile user’s location

- Improvements to the ranking algorithm
  Better infer user preference for a place based on the text they enter

- Cost of building and updating a location-based service
  Cost less than the existing approach on collecting and updating information

- Personalized results
  Reveal no personal information
Conclusion

- Show how a location-based service can be built
- Introduce an algorithm for ranking places
- Present results from an evaluation
- Compare the approach to existing approaches
Thank you!