Social Computing at MSRA

Chin-Yew LIN cyl@microsoft.com
What is Social Computing?

- Social
  - Living together in communities
    - Live ➔ dynamic
    - Together ➔ more than one person
    - Community ➔ same locality

- Computing
  - To determine by the use of a computer

- Social Computing
  - To live together in communities using computers
  - Creating social contexts online via the use of technology

* American Heritage Dictionary
Why Social Computing?

- Web is increasingly social
  - Wikipedia, blogs, Digg, Flickr, del.icio.us, Amazon, Yahoo! Answers, Live QnA, forums, Facebook, MySpace, LinkedIn, ...

- Social web
  - People + Content + Network

- How do we turn social web into value for people?
  - Assimilation of knowledge → systems pull from data
  - Dissimilation of knowledge → systems to people
  - Elicitation of knowledge → systems pull from people

- “Pay attention to that man behind the curtain!”
Baidu Zhidao (百度知道)

- 17,012,767 resolved questions in two years' operation
- 8,921,610 are knowledge related
- 96.7% of questions are resolved
- 10,000,000 daily visitors
- 71,308 new questions per day
- 3.14 answers per question

Baidu Zhidao Question Types Distribution

- Knowledge: 50.70%
- Life Style: 26.00%
- Entertainment: 17.60%
- Other: 5.70%

Baidu Zhidao Top 10 Question Types

- Internet: 768,668
- Education: 732,976
- Hardware: 579,133
- OS: 574,001
- Language: 500,762
- Relationship: 481,182
- Computer: 468,268
- Software: 409,447
- Music: 359,285
- Cell Phone: 305,000

Stickiness of Baidu Zhidao

- 据正望咨询调查，“百度知道”跟搜索的关系非常紧密，而且对搜索黏性的提高有很大帮助，根据其统计，“百度知道”已成为百度的一个核心产品。“百度的用户中有50%搜索‘知道’，其用户量已经超过百度贴吧，与其MP3搜索可相提并论。”。

- 50% of Baidu users search Baidu Zhidao
- Zhidao search traffic comparable to MP3 search

(http://news.csdn.net/n/20080425/115453.html; 04/25/2008)
Knowledge Distillation & Dissemination

- Social Question Answering and Distillation
  - Highly Structured QnA
  - Structured QnA
  - Semi-structured QnA
  - Unstructured QnA
GeoLife 2.0

Building Social Networks Using Human Location History

Yu Zheng, Xing Xie and Wei-Ying Ma

WSM, Microsoft Research Asia
What is GeoLife 2.0?

- A location-history-based GPS-data-driven social-networking service
  - Enables people to build connections using their GPS trajectories
  - Understand a user and a location
  - Explore the similarity between users and the correlation among locations
GPS Devices and Users

- 60 devices and 138 users
- From May 2007 to present
A Large-Scale GPS Data Set

- 10+ million GPS points
- 260+ million kilometers
- 36 cities in China and a few cities in the USA, Korea and Japan

* Bikely: http://www.bikely.com
* GPS Track Route Exchange Forum: http://www.gpsxchange.com
* GPS Sharing: Http://gpssharing.com
GeoLife Application Architecture

Social Networking Services
- Personalized Friend Recommendation
- Customized Location Recommendation

Individual Service
- Life Experience Sharing
- Life Story Understanding
- Life Blogging (Map-Blog)
  - Multimedia Content Management
  - Sports Activity Analysis
  - Memorizing Personal Past

Individual Dataset

Multi-User Dataset

- Interesting locations
- Classical Travel Sequences
- Experienced Users

GeoLife 1.0

GeoLife 2.0
A Location-History-Based Social Network
Applications

- Understanding People
  - Similar users: Friend recommendation
  - Experienced users: Travel experts recommendation
  - Group users: Community discovery

- Understanding Locations
  - Personalized location recommendation
  - Mining interesting locations
  - Detecting classical travel sequences
Preliminary

- **GPS logs** $P$ and **GPS trajectory**

- **Stay points** $S = \{s_1, s_2, \ldots, s_n\}$.
  - Stands for a geo-region where a user has stayed for a while
  - Carry a semantic meaning beyond a raw GPS point

- **Location history:**
  - Represented by a sequence of stay points with transition intervals

\[
\text{LocH} = (s_1 \rightarrow s_2 \rightarrow \ldots \rightarrow s_n)
\]

<table>
<thead>
<tr>
<th>Latitude, Longitude, Time</th>
<th>p1: Lat1, Lngt1, T1</th>
<th>p2: Lat2, Lngt2, T2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p3: Lat3, Lngt3, T3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>p4: Lat4, Lngt4, T4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>p5: Lat5, Lngt5, T5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>p6: Lat6, Lngt6, T6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>p7: Lat7, Lngt7, T7</td>
<td></td>
</tr>
</tbody>
</table>
Modeling Individual Location History

1. Stay point detection

2. Hierarchical clustering

3. Individual graph building

Shared Hierarchical Framework

- GPS Logs of User 1
- GPS Logs of User 2
- GPS Logs of User i
- GPS Logs of User i+1
- GPS Logs of User n-1
- GPS Logs of User n

- Layer 1
- Layer 2
- Layer 3

- Stands for a stay point  
- Stands for a stay point cluster $c_{ij}$
Architecture (1)

Modeling Location History

A Hierarchical Graph for each individual

Measuring User Similarity

Personalized Friend Recommendation & Location Recommendation
Modeling Multiple Users’ Location Histories

- GPS Logs of User 1
- GPS Logs of User 2
- GPS Logs of User i
- GPS Logs of User n

Stands for a stay point $S$
Stands for a stay point cluster $c_{ij}$

Shared Hierarchical Framework

Inferring User Travel Experience

Mining Interesting Locations

Detecting Classical Travel Sequences
A Hierarchical Graph for each individual

Measuring User Similarity

Modeling Location History

GPS Logs of User 1
GPS Logs of User 2
GPS Logs of User i
GPS Logs of User i+1
GPS Logs of User n-1
GPS Logs of User n

GPS Logs of User 1
GPS Logs of User 2
GPS Logs of User n

Mining Locations Interests and User Experiences

Detecting Classical Travel Sequences

Generic Location & Sequence & Travel Expert Recommendation

Personalized Friend Recommendation & Location Recommendation
GeoLife: Point of Interest

GeoLife: Travel Route
Summary

- Social web
  - People + Content + Network

- Turn **social web** into **value** for **people**
  - Assimilation of knowledge
  - Dissimilation of knowledge
  - Elicitation of knowledge

- “*Pay attention to that man behind the curtain!*”
Yu Zheng, Lizhu Zhang, Xing Xie and Wei-Ying Ma, *Mining Interesting Locations and Travel Sequences from GPS Trajectories*, WWW 2009.

Yu Zheng, Yukun Chen, Xing Xie and Wei-Ying Ma, *GeoLife 2.0: A Location-Based Social Networking Service*, MDM 2009 (Demo).


Yu Zheng, Quannan Li, Yukun Chen, Xing Xie and Wei-Ying Ma, *Understanding Mobility Based on GPS Data*, ACM UbiComp 2008.