Urban Computing With City Dynamics


Dr. Yu Zheng
Lead Researcher
Microsoft Research Asia
My Research Background

Location-Based Social Networks

Urban Computing

Computing with Spatial Trajectories
Urban computing is emerging as a concept where every sensor, device, person, vehicle, building, and street in the urban areas can be used as a component to sense city dynamics to enable a city-wide computing to tackle the challenges in urban areas as so to serve people and cities.
Key Features

• Sensing city dynamics
  – Unobtrusively, automatically, and constantly
  – A variety of sensors: Mobile phones, vehicles, cameras, loops,…
  – **Human as a sensor:** User generated content (check in, photos, tweets)

• Computing with heterogeneous data sources
  – A synergy of various data and discovery of collective knowledge
  – **Geospatial** is a unique and strategically important dimension
  – Geospatial, temporal, social, text, images, economic, environmental,…

• Blending the physical and virtual worlds
  – Serving both people and cities (virtually and physically)
  – **Hybrid systems:** Mobile + Cloud, crowd sourcing, participatory sensing…
City Dynamics

Scope

- Traffic flow
- Human mobility
- Energy consumption
- Environment
- Economic
- Populations
- …..

Data available

- Mobile phone signal
- GPS traces of vehicles and people
- Ticketing data in public transportation systems
- User-generated content
- Transportation sensor networks
  - Camera and loop sensors
  - Parking lots
- Environmental sensor network
  - Air quality
  - Temperature
  - Radiation
- Transaction records of credit cards
- ….
Beijing road networks 2009-2011

2011: 121,771 nodes and 162,246 segments, 19,524km
People’s location history (trajectories) in Beijing
Check-in data
Heat Maps of Beijing (2011)
Finding Smart Driving Directions

ACM SIGSPATIAL GIS’10 best paper runner-up, KDD’11

Route Construction from Uncertain Trajectories

KDD’12

Discovery of Functional Regions

KDD’12

Anomalous Events Detection

KDD’11

Passengers-Cabbie Recommender system

Ubicomp’11

Urban Computing for Urban Planning

Ubicomp’11 Best paper nominee
Finding Smart Driving Directions
Driving Direction Based on Taxi Trajectories

A *time-dependent*, *user-specific*, and *self-adaptive* driving directions service using

- GPS trajectories of a large number of taxicabs
- GPS log of an end user

Physical Routes + Traffic flows + Drive behavior

ACM SIGSPATIAL GIS 2010 best paper runner-up award and a publication on KDD 2011
Driving Direction Based on Taxi Trajectories
Driving Direction Based on Taxi Trajectories

Log user B’s driving routes for 1 month
Motivation

- Taxi drivers are *experienced* drivers
- GPS-equipped taxis are *mobile sensors*
- GPS logs imply the *drive behavior* of a user
System Overview

1. Send a query
   \[ Q = (q_s, q_d, t, \alpha) \]

2. Route computing
   - Normal weather
   - Severe weather
   - Weekday
   - Weekend

3. Route downloading
   - Historical trajectories and weather
   - Landmark Graphs

4. Logging the real travel with a GPS trace

5. Learning new \( \alpha \)

Knowledge discovery

- Offline mining
- Online inference
- Real-time taxi trajectories

Cyber world

Physical world

Real-world taxi trajectories
Offline Mining

Challenges

- Intelligence modeling
- Data sparseness
- Low-sampling-rate
Offline Mining

Detecting landmarks
- A landmark is a frequently-traversed road segment
- Top k road segments, e.g. k=4

Building landmark edges
- Number of transitions between two landmark edges > $\delta$
- E.g., $\delta = 1$

A) Matched taxi trajectories
B) Detected landmarks
C) A landmark graph
Mining Taxi Drivers’ Knowledge

- Learning travel time distributions for each landmark edge
  - Traffic patterns vary in time on an edge
  - Different edges have different distributions
Reported by MIT Technology Review Twice, featured once
Results

• More effective
  • 60-70% of the routes suggested by our method are faster than Bing and Google Maps.
  • Over 50% of the routes are 20+% faster than Bing and Google.
  • On average, we save 5 minutes per 30 minutes driving trip.

• More efficient
Urban Computing for Urban Planning
Urban Computing for Urban Planning

Goals

- City-wide traffic modeling
- Evaluate city configurations
- Suggest potential improvement to city planners
- Identify root causes of the problem

Datasets


Best paper nominee in UbiComp 2011
City-Wide Traffic Modeling

- Partition a city into regions with major roads
- Regions are root causes of the problem
City-Wide Traffic Modeling

Partition the dataset by time slots (a data-driven method)

<table>
<thead>
<tr>
<th>Time</th>
<th>Work day</th>
<th>Non-Workday</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slot 1</td>
<td>7:00am-10:30am</td>
<td>9:00am-12:30pm</td>
</tr>
<tr>
<td>Slot 2</td>
<td>10:30am-4:00pm</td>
<td>12:30pm-7:30pm</td>
</tr>
<tr>
<td>Slot 3</td>
<td>4:00pm-7:30pm</td>
<td>7:30pm-9:00am</td>
</tr>
</tbody>
</table>
City-Wide Traffic Modeling

- Project taxi trajectories onto these regions
- Building a region graph for each time slot
Finding Problematic Edges

Extracting features from each edge

- $|S|$: Number of taxis
- $E(v)$: Expectation of speed
- $\theta = E(D)/\text{CenDist}(r_1, r_3)$
Finding Problematic Edges

- Select edges with $|S|$ above average
- Detect Skyline edges according to $< E(V), \theta >$
  - Select edges with **big** $\theta$ and **small** $E(V)$
Making Sense of Individual Problematic Edges

- Formulate skyline graphs for each day
- Mining frequent sub-graph patterns across days
  - To avoid false alert
  - Deep understanding

---

**Day 1**
- Skyline Graphs
  - Slot 1
  - **Step (1)** Building skyline graphs
  - **Step (2)** Mining skyline patterns
  - Support = 1.0
    - Patterns
      - $r_1 \rightarrow r_2 \rightarrow r_8 \rightarrow r_4$
      - $r_4 \rightarrow r_5$

**Day 2**
- Skyline Graphs
  - Slot 1
  - **Step (1)** Building skyline graphs
  - **Step (2)** Mining skyline patterns
  - Support = 2/3
    - Patterns
      - $r_2 \rightarrow r_1 \rightarrow r_8 \rightarrow r_4 \rightarrow r_5$
      - $r_3 \rightarrow r_8 \rightarrow r_4 \rightarrow r_5 \rightarrow r_6$
Top 10 most frequent problematic edges

Workdays

2009

2010

2011

Holidays
Example 1

2009

Workdays

2010

Workdays
The dotted road was upgraded in August 2009.
Example 2
Dec. 2010 Subway Line 15 was launched
Example 3
Discover Regions of Different Functions using **Human Mobility** and **POIs**

In KDD 2012
Goals

- Discovery regions of different functions in urban areas
- Identify the kernel density of a functionality
Applications

- Calibrating urban planning
- Business allocation
- Advertising
Motivation and Challenges

• POIs feature the function

• But not enough
  – Compound
  – Quality

• Human mobility
  – Differentiate between POIs of the same category
  – Feature the function of a region
Methodology Overview

- **Mapping from regions to documents**
  - Regions $\rightarrow$ Documents
  - Functions $\rightarrow$ Topics
  - Mobility patterns $\rightarrow$ Words
  - POIs $\rightarrow$ meta data like Key words and authors

Infer the topic distribution using a LDA-variant topic model
Mobility Patterns

• Transition: \( Tr=(r_1 \rightarrow r_2, t_A, t_L) \)

• Mobility Patterns
\[ C_A(1,n,k) = #(r_1 \leftarrow r_n, t_k) \]

Arriving cuboid

\[ C_L(1,n,k) = #(r_1 \rightarrow r_n, t_k) \]

Leaving cuboid
Regard region \( r_1 \) as a document

Authors, affiliation, and key words

<table>
<thead>
<tr>
<th>POIs</th>
<th>( v_1 )</th>
<th>( v_2 )</th>
<th>( v_l )</th>
<th>( v_F )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15</td>
<td>10</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>8</td>
<td>0</td>
<td>6</td>
</tr>
</tbody>
</table>

Arriving Matrix of \( r_1 \)

<table>
<thead>
<tr>
<th>( r_n )</th>
<th>( r_1 )</th>
<th>( r_2 )</th>
<th>( r_j )</th>
<th>( r_l )</th>
<th>( t_1 )</th>
<th>( t_2 )</th>
<th>( t_k )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>30</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Leaving Matrix of \( r_1 \)

<table>
<thead>
<tr>
<th>( r_n )</th>
<th>( r_1 )</th>
<th>( r_2 )</th>
<th>( r_j )</th>
<th>( r_l )</th>
<th>( t_1 )</th>
<th>( t_2 )</th>
<th>( t_k )</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>8</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>32</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12</td>
</tr>
</tbody>
</table>
Territory Identification

- Region aggregation
  - Cluster regions according to topic distributions
  - Aggregate individual regions into big territories
Functionality Intensity Estimation

• Functionality varies geospatially
• Human mobility reflects the functionality density
• Using Kernel Density Estimation (KDE)

Diplomatic and embassy areas

Areas of historic interests
Annotation of Territories

Emerging residential areas
Regions under construction
Developed residential areas
Developed commercial/entertainment areas
Areas of historic interests
Nature and parks
Education and science areas
Developing commercial/entertainment areas
Diplomatic and embassy areas

Legend:
- Road
- Emerging residential areas
- Regions under construction
- Developed residential areas
- Developed commercial/entertainment areas
- Areas of historic interests
- Nature and parks
- Education and science areas
- Developing commercial/entertainment areas
- Diplomatic and embassy areas
- Sparse
Evaluation

- Datasets (2010 and 2011, Beijing)

<table>
<thead>
<tr>
<th>Table 3: Statistics of taxi trajectories and road networks</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>year</strong></td>
</tr>
<tr>
<td>-------------------</td>
</tr>
<tr>
<td>#taxis</td>
</tr>
<tr>
<td>#occupied trips</td>
</tr>
<tr>
<td>#effective days</td>
</tr>
<tr>
<td>average trip distance(km)</td>
</tr>
<tr>
<td>average trip duration(min)</td>
</tr>
<tr>
<td>average sampling interval(sec)</td>
</tr>
<tr>
<td>#road segments</td>
</tr>
<tr>
<td>percentage of major roads</td>
</tr>
<tr>
<td>#segmented formal regions</td>
</tr>
<tr>
<td>size of “vocabulary” (non-0 items)</td>
</tr>
</tbody>
</table>
Results

• Baselines
  Only using POI data (TF-IDF)
  Only using mobility data (LDA-based method)
Evaluation

2010

2011

Land use planning (2002-2010)

Results of 2011
Constructing Popular Routes from Uncertain Trajectories
Constructing Popular Routes from Uncertain Trajectories

- Uncertain trajectories
  - check-ins or geo-tagged photos
  - Taxi trajectories, trails of migratory birds
Constructing Popular Routes from Uncertain Trajectories

• **Goal:** Using collective knowledge: The route may not exist in the dataset
  – Mutual reinforcement learning (**uncertain + uncertain** → **certain**)
Constructing Popular Routes from Uncertain Trajectories

• Problem
  – Given a corpus of uncertain trajectories and
  – a user query: some point locations and a time constraint
  – Suggest the top k most popular routes
Impact

High-quality and influential publications
- Best paper runner-up award in ACM SIGSPATIAL GIS 2011
- Best paper nominee in UbiComp 2011
- Best paper award at international conference on advanced data mining applications 2011.
- KDD (4), WWW(2), UbiComp(3), ICDE(2), AAAI(1), TKDE(2), TIST (1)

Media reports on top-tier presses
- MIT Technology Review (3 times, featured twice)
- MSNBC news (1)
- NewScientist (1)
- Many Chinese news...

Potential real-world deployment
- Governments: Beijing urban planning institute
- MS product teams: STB
Datasets Released

GeoLife GPS trajectories
- Generated by 178 users over 3 years
- With transportation mode labels: driving, walking, biking, bus...
- Annual release

T-Drive Taxi trajectories
- Generated by Over 10,000 taxis in one week in Beijing
- 15 million points
- Distance > 9 million km
Miscellaneous

International Workshop on Urban Computing
- In conjunction with KDD 2012 at Beijing China

The special issue on Urban computing at ACM TIST
- Top-tier international journal
- Submission Due: **Oct. 7, 2012**
- [http://tist.acm.org/CFP.html](http://tist.acm.org/CFP.html)

A related text book:
- Computing with spatial trajectories
- Free tutorial slides download ([here](#))
Thanks!

Yu Zheng

yuzheng@microsoft.com