Urban Computing With City Dynamics

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 Focuses

- 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
  - ....
POI Data (2007 – 2012)
People’s location history (trajectories) in Beijing
Check-in data
GPS trajectories of 33,000 taxis in 2009, 2010, and 2011
Heat Maps of Beijing (2011)
Finding Smart Driving Directions

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

Route Construction from Uncertain Trajectories

KDD’12 and ICDE 2012

Anomalous Events Detection

KDD’11 and ICDM 2012

Discovery of Functional Regions

KDD’12

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

Driver A

8:00

Driver A

14:00

Driver B

14:00
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

Human Intelligence + Traffic patterns

Drive behavior
1. Send a query \( Q=\langle q_s, q_d, t, \alpha \rangle \)

2. Route computing

3. Route downloading

4. Logging the real travel with a GPS trace

5. Learning new \( \alpha \)
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
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>
<tr>
<td>Slot 4</td>
<td>7:30pm-7:00am</td>
<td></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

![Skyline Graphs and Patterns](image)

Day 1

- Day 2
- Day 3

Mining skyline patterns

Support=1.0

Patterns

- \( r_1 \rightarrow r_2 \rightarrow r_6 \rightarrow r_4 \)
- \( r_4 \rightarrow r_5 \)

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

2009

2010

2011

Workdays

Holidays
Example 1

2009

Workdays

2010

Workdays
The dotted road was upgraded in August 2009.
Example 2

2009

2010

2011

Workdays
Dec. 2010 Subway Line 15 was launched
Example 3

2010

2011
Discover Regions of Different Functions using **Human Mobility** and **POIs**

In KDD 2012
Goals

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

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

- **POIs indicate the function**
- **Human mobility**
  - Differentiate between POIs of the same category
  - Indicate the function of a region
- **But not enough**
  - Compound
  - Quality
Methodology Overview

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

Infer the topic distribution using a LDA (Latent Dirichlet allocation)-variant topic model
Mobility Patterns

• Transition: $Tr = (r_1 \rightarrow r_2, t_A, t_L)$
• Mobility Patterns
 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)
Annotation of Territories

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

2010

2011

Land use planning (2002-2010)

Results of 2011
Detecting and Diagnosing Anomalies in City Dynamics

In KDD 2011 and ICDM 2012
Anomaly Detection in City Dynamics

• What is an anomaly in city dynamics
  – Traffic accidents, controls, under construction
  – Disasters: downpour, surface collapse, snow storms, fires
  – Celebrations, games, and big events
  – ......

• Data sources
  – Transportation sensor data: GPS data, loop sensors
  – Social media data: tweeters, weibo, foursquare
  – Mobile phone data
  – Web log data: query log
Anomaly Detection in City Dynamics

• Examples
  – Olympic park of Beijing
  – Earthquake in Japan
  – Singapore F-1 race
Anomaly Detection in City Dynamics

• Methods (depending on applications)
  – Spatio-temporal outlier detection methods
  – PCA, DP algorithms

• Publications
Anomaly Detection in City Dynamics

- Traffic modeling
  - Map segmentation
  - Building a region graph
  - Identify three features for each

\(<#Obj, Pct_o, Pct_d>\)

- (a) Road network of Beijing
- (b) Partitioned regions
- (c) Example of traffic among regions
- (d) A graph of regions
Anomaly Detection in City Dynamics

- Anomaly detection
  - Calculate the distance between the corresponding item in the feature vectors of two time bins
    \[
    \langle \#\text{Obj}, \ Pct_o, \ Pct_d \rangle
    \]
  - Identify the minDistort
  - Find out the outlier points as anomalies using Mahalanobis distance
Anomaly Detection in City Dynamics

• Results

• Next step
  – Identify the root cause of the problem
  – From regions to road segments
  – Estimate the impact of an anomaly and effective visualization
Diagnosing the Road Traffic Anomalies

A) Visualization of the anomaly and the root cause paths

B) Traffic flows on L1

C) Traffic flows on L2
Diagnosing the anomalous road traffic patterns

A) A region map

B) Flows (by routes)

C) Paths

D) Links (region graph)

E) Link-route matrix
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(3), UbiComp(3), ICDE(4), 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

Link to the data
Link to the data
Miscellaneous

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

Yu Zheng

yuzheng@microsoft.com

Homepage