IMPROVING PUBLIC TRANSPORTATION THROUGH CROWD-SOURCING


January 6, 2015
Motivation

- But public transportation is very unpredictable
- We seek to improve predictability through better availability of real-time information
For example, queries such as "When is the bus numbered 422 expected at my bus stop?" can be answered
**Related Work**

- *EasyTracker*: arrival time prediction, requires on-vehicle sensors
- Zhou et. al.: bus arrival time prediction using crowd-sourcing, assumes specific bus ticketing system
- *Moovit, Tiramisu*: closest to our goal, crowd-sourcing specific to public transit
  - Design, evaluation not available for comparison
  - Additional aspect: traffic in dense Indian cities chaotic, potential for high variability, static schedules not of much use
- Our work: naming framework for crowd-sourcing, aim to develop and evaluate in Indian conditions
We need a naming scheme for name-based storage and retrieval of information.

Access operations on the content easier through a hierarchical structure.

The framework supports two classes of entities: *public transportation* and *locations*.

Separate naming hierarchy for each class.
Public Transportation Hierarchy

- Route
- Fare
- Static schedule

Public Transportation System

- Mode of travel
  - (road / rail)

Attributes

Service ID #1
- Attributes

Service ID #2
- Attributes

Service ID #3
- Attributes

Vehicle ID #1
- Attributes

Vehicle ID #2
- Attributes

Vehicle ID #3
- Attributes

- Real-time location
- Real-time occupancy
- Real-time schedule
Location Hierarchy

- Geographic coordinates
- Commonly used name(s)
- Type of transit station

Region R

Sub Region A

Sub Region B

Sub Region C

- Road
- Rail

- Start and end locations
- Real-time traffic & history
- Road condition
Example Query

- **When is the bus numbered 422 expected at my location?**
  - From the public transportation hierarchy, we gain:
    - Route of the service provided by 422
    - Real-time location of the bus
    - Real-time occupancy - useful for commuter to decide whether to board or not
  - From the location hierarchy, we gain:
    - The bus stop nearest to commuters location
    - Traffic on the route (from public trans. hierarchy)
    - Road condition - useful to infer travel time
  - Using information gained above, we can infer the time at which bus 422 may arrive
  - A query such as **What is the best way to reach B from A?** will require information from all layers in the hierarchies
Field Experimentation

- Conducted a preliminary collection of field data to validate feasibility
- Used an android application which collects data from GPS sensors on the smart-phone

Data Collection
- Collected GPS data along specific routes in two cities: Mumbai and Chandigarh
- The data collection would start once the bus starts and would end at the destination
- Application contained options to mark ground-truth information for bus stops and traffic lights
Cannot rely on GPS connectivity alone for data collection and analysis
Mean dwell time: Mumbai - 5.9 sec and Chandigarh - 5.1 sec

Such small numbers imply that ignoring dwell time in bus arrival time prediction algorithm would not result in significant error.
Bus Stop Detection

- Examined feasibility of automated bus-stop detection
- Given a collection of GPS traces over multiple trips along the same route, is it possible to deduce locations of bus-stops?
- Assumption:
  - Instance of stationary bus as bus stop: movement under 20 m within 5 sec
  - We undertake this assumption to account for inherent GPS error
- To combine information across trips, we use two approaches: Fuzzy intersection and Fuzzy union
- "Fuzzy": Consider two locations marked as bus stops to be same, if they are less than 30 m apart
Fuzzy Intersection and Union

Distance along the route (in metres)

Trip 1
- A1
- A2
- A3

Trip 2
- B1
- B2

Intersection
- A1
- A3

Union
- B1
- B2
- B3
## Field Experimentation

### Bus Stop Detection

#### Results

<table>
<thead>
<tr>
<th>Number of Trips considered</th>
<th>Precision</th>
<th>Recall</th>
<th>Recall for traffic lights</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fuzzy-Intersection, Mumbai</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.20</td>
<td>0.61</td>
<td>0.36</td>
</tr>
<tr>
<td>2</td>
<td>0.31</td>
<td>0.65</td>
<td>0.36</td>
</tr>
<tr>
<td>3</td>
<td>0.40</td>
<td>0.26</td>
<td>0.09</td>
</tr>
<tr>
<td>4</td>
<td>0.40</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td><strong>Fuzzy-Union, Mumbai</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.20</td>
<td>0.61</td>
<td>0.36</td>
</tr>
<tr>
<td>2</td>
<td>0.11</td>
<td>0.74</td>
<td>0.73</td>
</tr>
<tr>
<td>3</td>
<td>0.08</td>
<td>0.87</td>
<td>0.73</td>
</tr>
<tr>
<td>4</td>
<td>0.06</td>
<td>0.91</td>
<td>0.73</td>
</tr>
<tr>
<td><strong>Fuzzy-Intersection, Chandigarh</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.30</td>
<td>0.72</td>
<td>0.58</td>
</tr>
<tr>
<td>2</td>
<td>0.34</td>
<td>0.66</td>
<td>0.58</td>
</tr>
<tr>
<td>3</td>
<td>0.37</td>
<td>0.60</td>
<td>0.46</td>
</tr>
<tr>
<td>4</td>
<td>0.38</td>
<td>0.51</td>
<td>0.36</td>
</tr>
<tr>
<td><strong>Fuzzy-Union, Chandigarh</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.30</td>
<td>0.72</td>
<td>0.58</td>
</tr>
<tr>
<td>2</td>
<td>0.12</td>
<td>0.90</td>
<td>0.75</td>
</tr>
<tr>
<td>3</td>
<td>0.08</td>
<td>0.90</td>
<td>0.75</td>
</tr>
<tr>
<td>4</td>
<td>0.06</td>
<td>0.90</td>
<td>0.75</td>
</tr>
</tbody>
</table>
Results

- Precision increases with number of trips considered, in fuzzy intersection, but recall drops
- The behaviour is reversed in the case of fuzzy union
- Neither method has high value of precision and recall for any number of trips
- Need to somehow distinguish traffic lights from bus stops
- Overall implication - Bus stop detection appears to be a tough problem
Developed an intuitive and hierarchical naming framework which will lead to a flexible and scalable implementation in practice.

Using GPS alone might be error-prone due to loss of GPS connectivity.

Dwell times are mostly within 5-10 sec range and do not contribute a lot to the travel time.

Results from the automated bus-stop detection task points towards challenges brought out by the noise in the system through GPS errors, bus-stops getting confused with traffic-related stops, etc.