Capturing Urban Mobility using Crowdsourced Data from Smartphones

Dr. Yang Wang and Dr. David Philip McArthur
Introduction

• Project funded by InnovateUK

• The aim is to encourage innovation and new commercial ventures

• Our project is to produce detailed mobility data
Some current data sources and limitations

• **Surveys**
  • Small samples
  • Expensive
  • Hard to complete
  • May lack detail

• **Sensor networks**
  • Requires (expensive) infrastructure
  • Lacks origin/destination info
  • Not multimodal
(Some) Big data
Some limitations

• Hard/expensive to access
• Pre-processing of data can be a black box
• May be tied to one mode
• Spatial coverage may be patchy
Smartphone data

- Smartphones offer the chance to collect rich data
  - Independent of operator
  - Information for all travel modes
  - Full door-to-door OD and route data
  - Potentially real-time
Catch! Project

• The Catch! (citizens at the city’s heart!) app is a journey planning app
• It passively collects GPS trajectories
• The phones’ sensors and machine learning are used to infer travel mode
• Users get to contribute data to improve transport planning in their city
• Insight from the data can feed back into better journey suggestions
The Consortium

SMEs

- TravelAi, The Behaviourlist, Elgin Roadworks, Placr
- *App development, data sources, citizen messaging, impact assessment*

Research organisations

- University of Glasgow (UBDC), University of Leeds (CDRC), Transport Systems Catapult
- *Data cleaning, anonymization, aggregation, analysis*

Local Authorities and cities

- Coventry, Ipswich, Leeds, Newcastle, Oxfordshire
- *Citizen access, sounding boards, pilots, data sources, advocates*
The App

Arrive at 10:12 (36 minutes)
Catch! App

- Includes real time information on roadworks (Elgin) and public transport performance (TransportAPI)
Data from the app

- Person identifier
- Latitude, longitude
- Time
- Inferred mode
- Collected every 5 seconds
Our part of the project

- Clean and enhance the data
- Map match the data
- Detect and annotate stops
- Aggregate the data into a useful, privacy-protecting format
- Improve mode detection with post-processing
Semantic Trajectory

An Illustration of the Method

[Map showing locations for Home, Shopping, and Move with timestamps]
Semantic Trajectory Framework
Stop/Move Table Structure

**Move**

<table>
<thead>
<tr>
<th>Id</th>
<th>tripId</th>
<th><em>CellId</em></th>
<th>inTime</th>
<th>outTime</th>
<th>Duration</th>
<th>Road</th>
</tr>
</thead>
</table>

**Stop**

<table>
<thead>
<tr>
<th>Id</th>
<th>startTripId</th>
<th>endTripId</th>
<th><em>CellId</em></th>
<th>inTime</th>
<th>outTime</th>
<th>Duration</th>
<th>Place</th>
</tr>
</thead>
</table>

**Move**

<table>
<thead>
<tr>
<th>Id</th>
<th>tripId</th>
<th>ts</th>
<th>Duration_lag</th>
<th>Longitude</th>
<th>Latitude</th>
<th>Distance_lag</th>
<th>Geometry</th>
<th><em>CellId</em></th>
</tr>
</thead>
</table>

**Stop**

<table>
<thead>
<tr>
<th>Id</th>
<th>startTripId</th>
<th>endTripId</th>
<th>ts</th>
<th>Duration_lag</th>
<th>Longitude</th>
<th>Latitude</th>
<th>Distance_lag</th>
<th>Geometry</th>
<th><em>CellId</em></th>
</tr>
</thead>
</table>

**Raw**

<table>
<thead>
<tr>
<th>Id</th>
<th>ts</th>
<th>Duration_lag</th>
<th>Longitude</th>
<th>Latitude</th>
<th>Distance_lag</th>
<th>Geometry</th>
</tr>
</thead>
</table>

---

Urban Big Data Centre

**catch!**

---
Travel Mode Detection

- **DeviceID**: Device_1
- **Email**: dummy@travelai.co.uk
- **Start Date**: 2016/10/27 08:48:19
- **End Date**: 2016/11/16 01:24:50
- **System**: System_1
- **Datasource**: Source_1

### Mode Counts (4048)

- **WALK**: 2066 (51.04%)
- **BUS**: 197 (4.94%)
- **METRO**: 26 (0.65%)
- **CAR**: 1033 (25.52%)
- **BICYCLE**: 8 (0.20%)
- **RUN**: 1 (0.02%)
- **TRAIN**: 1 (0.02%)
- **STATIONARY**: 282 (6.97%)

### Mode Distance (approx. 141.96 km)

- **WALK**: 20.35 km (14.33%)
- **BUS**: 11.77 km (12.59%)
- **METRO**: 25.65 km (18.07%)
- **CAR**: 52.43 km (36.93%)
- **BICYCLE**: 17.70 km (5.42%)
- **RUN**: 0.10 km (0.07%)
- **TRAIN**: 0.00 km (0.00%)
- **STATIONARY**: 17.85 km (12.58%)
Map Matching to Road Network
Semantic Annotation

Example of Annotated Trip

Moves with roads

Stops with surrounding information
Privacy Issue – A Real Case

Focus on two schools

1. App adopted by pupils and parents
2. Data analysis:
   - Quantification of transport modes
   - Correlation with geography
   - Correlation with congestion
   - Correlation with air quality
   - Identification of related journeys
3. Long term benefits
   - Reduce congestion and improve air quality
   - Shift to healthier and eco-friendly transport
   - Input into school travel plans
   - Improved town planning
   - Car sharing and child safety apps
   - Better transport routing
   - Base data for traffic modelling
Anonymization - Sharing Aggregated ‘Moves’

- Trip count within temporal window

Example of extracting home related trip counts for Glasgow area on Sundays (Left) and Mondays (Right) symbolized in trip counts against the whole raster coverage (RasterIndex) in blue.
Anonymization - Sharing Aggregated ‘Moves’ -2

- $k$-anonymity within temporal window

Shared ‘move’ cells confirming to 2-anonymity in Glasgow within 17:00-18:00-time window for use.
Anonymization - Sharing Aggregated ‘Moves’ -3

Aggregation of GPS attributes on roads within temporal window

- Aggregated GPS information on roads for 120 users.
- Left is aggregated travel mean speed (km/h) on road segments symbolized from yellow-red.
- Right is aggregated wait time (a) on road intersections.
Anonymization - Sensitive Locations - 1

- **Grid-masking**

  Grid-masking result with different cell sizes
Anonymization - Cloaking Sensitive ‘Stops’

- Blurring stops into the environment

Blur the stops in the ‘home/work’ category into the 10 nearest buildings and the stops in the ‘other’ category into the 10 nearest points of interest.
Anonymization - Cloaking sensitive ‘moves’

- Zoomed into specific purpose of trips;
- Zoomed into temporal window;
- Information loss for
  - selection of methods;
  - selection of thresholds.
(Some) limitations of the project

• Will the app be used by enough people?
• The data are biased; but how? Will this change over time?
To download the app

• Not always easy to find in the App/Play Store due to the name

iOS

Android
Thank you for your attention.