Crowdsourcing Pedestrian and Bicyclist Activity Data

TRB UTC Spotlight Conference
December 1, 2016

Alex Rixey
@alexrixey
Overview

1. Crowdsourced Data Typology
2. Data Source Examples
3. Data Challenges
4. Applications
5. New Crowdsourced Data Opportunities
6. Better Data → Better Safety Analysis?
1 | Crowdsourced Data Typology

In-Situ

Thematic

Thumbtack

Spatial Inventory
1 | Additional Crowdsourced Data Characteristics

**Explicit**
Defined problem communicated directly to participants

**Implicit**
Participants may be unaware of secondary use of their data

Misra, Gooze, Watkins, Asad and Le Dantec (2014)
1 | Additional Crowdsourced Data Characteristics

**General Purpose**

Does not require specialized knowledge from participants

**Domain-Specific**

Data collected from participants with existing expertise

Misra, Gooze, Watkins, Asad and Le Dantec (2014)
1 | Additional Crowdsourced Data Characteristics

Misra, Gooze, Watkins, Asad and Le Dantec (2014)
2 | In-Situ Data (Explicit)
2 | In-Situ Data (Implicit)

Source: Joe Wergin
2 | In-Situ Data (Implicit)
2 | Thematic Data
2 | Thumbtack Data

http://visionzero.ddot.dc.gov/VisionZero/
2 | Thumbtack Data
2 | Spatial Inventory Data
2 | Spatial Inventory Data

Bicycle Stress Map

What is Traffic Stress?
When cyclists travel on roadways, they encounter varying levels of stress from traffic. A quiet residential street with a 25-mile-per-hour speed limit is considered a very low-stress environment for cyclists. But a six-lane suburban highway with a 40-mile-per-hour speed limit represents a high-stress environment for cyclists who must share the roadway with traffic. As a result, fewer people are likely to cycle on the highway. More...  

Video Examples

- HIGH STRESS
  - Shady Grove Road

- MODERATE-LOW STRESS
  - Security Lane

- LOW STRESS
  - Muddy Branch Road

Stress Tolerance Levels
- Very High: (very few adults will bicycle)
- High: (some adults will bicycle)
- Moderate High: (some adults will bicycle)
- Moderate Low: (many adults will bicycle)
- Low: (most adults will bicycle)
- Very Low: (everyone will bicycle)

CREATE YOUR OWN ANALYSIS
- Map My Biloxi
- Map My Route

CONNECTIVITY ANALYSIS
- Great Lanes and sticky point for analysis

- Rail: Metrorail
- Purple Line
- Silver Line (planned)
- Corridor Cities Transitway (planned)

Public Schools
- Elementary
- Middle
- High

Public Facilities
- Libraries
- Recreation Centers
3 | (Some of the) Current Data Challenges

- Privacy concerns
- Proprietary vs. Open data
- Biases, e.g:
  - Underreporting of collision data
  - Self-selection
  - Demographic
  - Honesty and validity
- Interpretation
4 | Modeling CaBi Ridership

- **Capital Bikeshare Station**
- LTS on Shortest Path Streets
  - 1
  - 2
  - 3
  - 4
  - Other Streets

Line weight indicates relative volume of bikeshare ridership.
4 | Estimating Crossing Risk

Schneider, Henry, Mitman, Stonehill and Koehler

TABLE 4 Preferred San Francisco Pedestrian Volume Model

<table>
<thead>
<tr>
<th>Model Variables</th>
<th>Coefficient</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total households within 1/4 mile (10,000s)</td>
<td>1.81</td>
<td>2.12</td>
<td>0.043</td>
</tr>
<tr>
<td>Total employment within 1/4 mile (100,000s)</td>
<td>2.43</td>
<td>2.27</td>
<td>0.082</td>
</tr>
<tr>
<td>Intersection is in a high-activity zone</td>
<td>1.27</td>
<td>3.79</td>
<td>0.000</td>
</tr>
<tr>
<td>Maximum slope on any intersection approach leg (100s)</td>
<td>-0.40</td>
<td>-3.07</td>
<td>0.004</td>
</tr>
<tr>
<td>Intersection is within 1/4 mile of a university campus</td>
<td>0.035</td>
<td>1.45</td>
<td>0.154</td>
</tr>
<tr>
<td>Intersection is controlled by a traffic signal</td>
<td>1.16</td>
<td>4.08</td>
<td>0.000</td>
</tr>
<tr>
<td>Constant</td>
<td>12.9</td>
<td>33.29</td>
<td>0.000</td>
</tr>
<tr>
<td><strong>Overall Model</strong></td>
<td><strong>50</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusted R²-Value</td>
<td>0.804</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F-Value (Test value)</td>
<td>34.4</td>
<td>(p &lt; 0.001)</td>
<td></td>
</tr>
</tbody>
</table>

1) The dependent variable is the natural logarithm of the annual pedestrian intersection crossing volume at each of the 50 study intersections. This represents the sum of all crossings on each approach leg within 50 feet of intersection. The annual volume estimate is extrapolated from a two-hour manual count taken in September 2009 or July-August 2010. The extrapolation method accounts for variations in pedestrian activity by time of day, day of week, weather, and land use.
2) All distances used to calculate the model variables are straight-line distances rather than roadway network distances.
4 | Estimating Crossing Risk

20 Intersections with Highest Estimated Crossing Risk

Reported Pedestrian Crashes by Intersection (11/2004 to 10/2009)

Schneider, Henry, Mitman, Stonehill and Koehler
5 | More Collision and Near-Miss Data?

CRASH APP

Crowdsourcing Ped+Bike Data
5 | More Collision and Near-Miss Data?
5 | More Infrastructure Data

Source: Google

5 | More Trip Data

Google

UBER

Crowdsourcing Ped+Bike Data
5 | StreetLight Data + Cuebiq

• New mobile device data

• Refining “tripifying” algorithm

• New metrics (Spring ’17?):
  • Zonal O-D data for active modes
  • All-modes data with likely mode split
6 | So what?

Data ↔ Analysis
6 | So what? Imagine we have...

- Better collision and near miss data
- Better trip data
- Better spatial data
6 | How can we improve safety (analysis)?

• Improve **reactive** analysis

• Apply more **proactive** analysis

• Develop a **new** approach?
6 | Current Practice – Reactive Analysis

# Collisions (Injuries, Fatalities)
6 | Current Practice – Reactive Analysis

\[
\text{# Collisions (Injuries, Fatalities)} \div \text{Exposure} = \text{Rate!}
\]
6 | Current Practice – Reactive Analysis

# Collisions (Injuries, Fatalities) / Exposure = Rate!

Identify Countermeasure
Assess Collision Profile
6 | Improving Reactive Analysis

- Easier self-reporting
- More comprehensive professional reporting
- Better trip data

# Collisions (Injuries, Fatalities) / Exposure = *Better! Rate!
6 | Conduct Proactive Analysis

- Where is ped&bike activity suppressed by unsafe conditions? Can short trip data help ID?

- Can we attempt better predictive analysis based on characteristics of known unsafe locations?

- Can we design to prevent injuries and fatalities before they become “hotspots?”
6 | Flipping the script?
6 | Measuring Safety – “Units of Bad per Good”

**Crashes/Injuries/Fatalities**

- per 100 Million VMT
- per Million Entering Vehicles

**per Resident**

**per Employee**

**per Grocery Trip**

**per $ of Goods Delivered**
A multimodal full accounting of the safety costs ("risks") of travel
6 | Safer for whom? Measuring an externality…
6 | Grocery trip
6 | Grocery trip

Logan Circle

- 0.2 miles
- 2 intersections
- 0 crash history locations

Spring Valley

- 1.8 miles
- 20 intersections
- 1 U-turn
- 1 crash history location
6 | Grocery trip safety costs (risks)

**Logan Circle**
- Internal Safety Cost: Low
- External Safety Cost: Zero?

**Spring Valley**
- Internal Safety Cost: Low
- External Safety Cost: Non-Zero
6 | Grocery trip + commute trip + lunch trip + ...
6 | Non-Intersection Countermeasures?

- Can we improve safety by...
  - Designing safer land use + transportation systems?
  - Reducing VMT and vehicle speeds on local roads while helping travelers accomplish their trip purposes?

How else can we use data to improve bike+ped safety?
Thank you!