Using Google’s Aggregated and Anonymized Trip Data to Estimate Dynamic Origin-Destination Matrices for San Francisco

TRB Applications Conference 2017

Bhargava Sana, Joe Castiglione, Dan Tischler, Drew Cooper
OD Data Collection

- Conventional methods
  - License plate surveys
  - Roadside interviews

- Emerging passive data collection methods
  - Bluetooth detectors
  - Cell phone Call Detail Records (CDR) data
  - GPS, Wi-Fi detectors
Google’s Better Cities Program
Aggregated and Anonymized Trip (AAT) Data

- Minimize congestion, improve safety and reduce infrastructure spending
- Aggregated and Anonymized Trip (AAT) information from location reports
  - Extract data from moving users
  - Clean data and snap to road network
  - Aggregate OD trip counts
  - Apply differential privacy filters and minimum trips threshold
Google AAT Dataset

- Hourly AAT data for six months (Apr-Jun and Sep-Nov 2015)
- Flow data provided as relative trips as opposed to absolute counts
- Convert relative flows to trips using HH travel survey?
Relative Flow Conversion Model

\[ PT_{odt} = \beta_t(RF_{odt}) + \epsilon ; \quad 0 \leq t \leq 23 \]

- \( PT_{odt} \): Person trips between o-d for hour-of-day t
- \( \beta_t \): Coefficient of AAT relative flow for t
- \( RF_{odt} \): Avg AAT flow between o-d for hour-of-day t
- \( \epsilon \): Error term

- No geographic constants applied
- District- and County-level regression models estimated
- 20% sample used for validation
Predicted vs Observed

Origin-Destination-Hour
Summary

- AAT relative flow magnitudes correlated with actual trips
- AAT geographic coverage significantly higher
- Simple linear regression model may be used for conversion
- Could support measuring longitudinal variation
- Further studies
  - Better/smooth survey data
  - Compare with cell CDR data