Full-Day Accessibility Evaluation of Transit Systems Using GPS-Based Location Data

Ying Song
Assistant Professor
Geography, Environment and Society
University of Minnesota
Background

- **Automatic vehicle location (AVL)**
- **Automatic passenger counters (APC)**
- **Traveler Information System (TIS)**
- **Electric fare payment**
- **Transit signal priority**
- **Security systems**
- **Computer-aided dispatch**

Source: GAO analysis of Department of Transportation documents GAO-16-638
Background

- **Using AVL-APC data to improve transit performance & management**
  - **Automatic vehicle location (AVL)**
    ✓ primarily for real-time monitoring/control
    ✓ not typically processed for off-line analysis
  - **Automatic passenger counters (APC)**
    ✓ provide passenger-activity data compatible with AVL operating data
  - **Off-line analysis has substantial promise**
    ✓ provide insights on transit performance
    ✓ improve service planning and scheduling
    ✓ promote transit ridership

- **reliable services & accurate information**

---

Background

- **Improve transit accessibility**
  - Fast and frequent services
  - Sufficient service coverages
  - First- and last- mile connection

- **Reliable schedule**
  - Better scheduled arrival times
  - Accurate real-time estimations

---

**Portland TriMet**

TriMet’s real-time arrival information system

**Chicago Transit**

Your official source for up-to-the-minute arrival info.
Methods

- **Model Components**
  
  1. **Linear-referencing**
     - allocate AVL and APC data along transit routes
  
  2. **Visual exploration**
     - spatio-temporal patterns of bus delays along routes
  
  3. **Semi-Markov process**
     - model vehicle movements as “jumping” processes from one stop to its next stop
     - calibrate “holding times” for the “jump” between stops

- **Flowchart**

  - **AVL & APC database**
    - **Select** records along a route during specific time periods
    - **Separate** records into trips and **sort** records along each trip by datetime
    - **Linear-reference** records along each trip to its corresponding route
    - **Clean up** linear-referenced results (time intervals, speed limits...)
    - **Visualize** linear-referenced trajectories in space and time by hour(s) of the day, day of the week, month, season, special time periods...
    - **Spatial-temporal patterns of delays**
    - **Calibrate** holding time density function between consecutive bus stops (sensitive to spatial-temporal patterns of delays)
    - **Points cover entire route?**
      - Yes
Methods

✈ Method Highlights

1. **Linear-referencing**
   - √ discrete GPS coordinates ⇒ *an ordered sequence of placement along route*
   - √ eliminate artificial movements due to device precision etc.

2. **Visual exploration**
   - √ where and when are the delays ⇒ *at and between bus stops*
   - √ traffic congestions, dwelling times, ...  

3. **Semi-Markov process**
   - √ recognize *ripple impacts of delays* along the trip
   - √ sensitive to *traffic situation and ridership* during specific time periods

Use **AVL and APC data** to calibrate delays between each pair of bus stops.
A Pilot Study

- **Original Dataset**
  - **A-Line**
    - ✓ every 10 minutes
    - ✓ pay ticket before boarding
  - **AVL and APC data**
    - ✓ Oct. 1 to Oct. 8, 2016
    - ✓ 12 vehicles
    - ✓ 136,883 messages
Results

1. Linear Referencing

\[(\text{longitude, latitude, timestamp, …}) \Rightarrow (\text{route, distance from first stop, time stamp, …})\]
Results

2. Visual exploration – day of the week

- Weekday
- Monday
- Tuesday
- Wednesday
- Thursday
- Friday

less variation
less delays
⇒
more reliable

Macalester College & University Ave W

where lead to
more delays

Python & ArcPy
Results

2. **Visual exploration – hour of the day**

- **00-03 am**: Macalester College & University Ave W
  - 9am – 12pm
- **03-06 am**: Near Cleveland Ave S & Ford Pkwy cross
  - 6pm – 9pm

Consistent with afternoon peak hours for traffic

Where & where has more delays
Results

3. Calibrate delays between stops

1) kurtosis (variability) and squared skewness (location) of empirical delays

![Cullen and Frey graph]

- Kurtosis: measure of heavy-tailed distribution
- Skewness: measure of symmetry

Theoretical distributions:
- normal
- uniform
- exponential
- logistic
- beta
- lognormal
- gamma

(Weibull is close to gamma and lognormal)
3. Calibrate delays between stops

Lognormal & Gamma distribution

- Patterns are slightly different along trips between two stops

- **Beginning and end** of the trip tend to have **low kurtosis**
  ⇒ less outliers (long delays)

- **Towards the middle** of the trip, the **skewness is larger**
  ⇒ not symmetrically distributed
Results

3. Calibrate delays between stops

2) Compare empirical and theoretical fitted distribution \(<\text{maximum likelihood}>\)

Mean = 56.26 seconds  
Std. = 2.01 seconds
Discussion

❖ Results Highlights

• Visualizing linear referenced trips can reveal spatial-temporal patterns of vehicle delays along the routes, and identify locations and/or times that lead to delays.

• Along A-Line route, delays during weekdays fit best with lognormal and gamma distribution for movements between any two consecutive stops; yet the patterns are slightly different with respect to location and variability of empirical delays.

❖ Practical Merits

• Adjust bus schedules to reflect the revealed delay patterns.

• Provide more accurate arrival time estimations to transit users that can
  - recognize that the delays between two stops may not be the same
  - capture the ridership and traffic “congestions” during different time periods

• Refine current transit accessibility measures to consider expected/mean delays at stops in addition to the scheduled service time and frequency.
Ying Song  yingsong@umn.edu

This material is based upon work supported by CTS Seed Fund

The supplementary data is provided by Metro Transit
Semi-Markov Process (Song et al. 2016)

1. Jump Rate

\[ p_{ij} = \begin{cases} 
1 & t_{ij} < t_j^+ - t_i^- \\
0 & \text{otherwise}
\end{cases} \]

2. Holding Time– reach \( e_{ij} \) from \( v_i \) at time \( t \)

\[
p_i(t) = \begin{cases} 
0 & , \quad t \in [t_0, t_i^-) \\
\int_{t_i^-}^t \lambda e^{-\lambda \tau} d\tau = e^{-\lambda t_i^-} - e^{-\lambda t}, & , \quad t \in [t_i^-, t_j^+ - t_{ij}) \\
\int_{t_i^-}^{t_j^+ - t_{ij}} \lambda e^{-\lambda \tau} d\tau = e^{-\lambda t_i^-} - e^{-\lambda (t_j^+ - t_{ij})}, & , \quad t \in [t_j^+ - t_{ij}, t_D]
\end{cases}
\]

3. State probability

\[ Prob(e_{ij}, t) = p_{ij} * p_i(t) * p_j(t) \]
Root Mean Square Error ("fitdistrplus" package in R)

Based 1: the estimated NTP

\[ RMSE^1(t) = \frac{\sum_{(e_{ij})^t} \left( EmpP(e_{ij}, t) - SimP(e_{ij}, t) \right)^2}{Cnt(t)} \]

\{e_{ij}\}^t : edges within NTP at time \( t \)

\{SimP(e_{ij}, t)\} simulated visit probabilities

\{EmpP(e_{ij}, t)\} empirical visit probabilities

Based 2: the empirical NTP traces

Further select from \{e_{ij}\}^t : edges have been used by at least one GPS trace