Characterizing Travel Time Reliability and Passenger Path Choice in a Metro Network

Lijun SUN
Future Cities Laboratory, Singapore-ETH Centre
lijun.sun@ivt.baug.ethz.ch
National University of Singapore (NUS)
Background

• Service Reliability
• Travel time
• Train load
• Transfer convenience
• Service disruption
• Flow assignment
• Service Reliability
• The reliability of metro systems is higher than other transit modes. However, travel time variability still shows accumulative effect.
Background

• Most metro systems are closed environments, which only register transactions when passengers enter and leave the system

• Crucial questions:
  • Predicting travel time (and reliability)
  • Inferring route choices
  • Inferring train load
  • Identifying critical transfer location
  • Building sophisticated flow assignment models
  • All linked together......
Background

• **Data**
  - Operation log?
  - Transfer demand?
  - Link flow?
  - Route choice?
  - Trajectory?

• **Smart card (Boarding station, Alighting station, Travel time)**
Metro System

T_a

T_b
Metro System

• What we know? (Observed)
  – Network configuration
  – Boarding station B
  – Alighting station A
  – Travel time $t$

• What we do not know? (Unknown)
  – Travel time on each link
  – Reliability of each link
  – Other time cost (waiting at platform, walking between fare gantry and platform)
  – Route choice
Metro System

- Research question
- Given observations to infer unknowns

- Boarding station B
- Alighting station A
- Travel time $t$

- Travel time on each link
- Reliability of each link
- Other time cost
- Route choice
Metro System

- **Methodology:** Bayesian inference
- \[ P(\text{unknown} | \text{observed}) \propto P(\text{observed} | \text{unknown}) \times P(\text{unknown}) \]

- Likelihood \times Prior

- **Boarding station B**
- **Alighting station A**
- **Travel time t**

- **Observed**
- **Unknown**
  - Travel time on each link
  - Reliability of each link
  - Other time cost
  - Route choice
Unknown

- Travel time on each link
- Travel time variation on each link (in-vehicle / transfer links)
  - Assuming that link cost follows a normal distribution, which has a constant coefficient of variation (linear mean v.s. std)
    \[ x_a \sim \mathcal{N}(c_a, (\alpha c_a)^2) \]
  - Assuming that all links are independent
  - Then the travel time on a particular route \( r \) follows
    \[ t \mid r \sim \mathcal{N}\left(\sum_{a \in r} c_a, \alpha^2 \sum_{a \in r} c_a^2\right) \]
• Other time cost
  – Assuming that other cost follows a normal distribution, with
    \[ y \sim \mathcal{N}(m, \sigma_y^2) \]
  – Assumed to be consistent for all OD pairs
• Route choice
Which route to take?
Using brute-force search
• Which route to take?
• Using brute-force search
• Which route to take?
• Using brute-force search
• Which route to take?
• Using brute-force search
• Which route to take?
• Using brute-force search
Unknown

- Which route to take?
- Using brute-force search
• Which route to take?
• Using brute-force search
Unknown

- Which route to take?
- Using brute-force search
• Which route to take?
• Using brute-force search
Unknown

- Which route to take?
- Using brute-force search
More than 30 routes in total

- Which route to take?
- Using brute-force search
Unknown

- Route choice (in general)
- Multinomial Logit (MNL)
- Representative Utility
  \[ V_i = \sum_k \theta_k X_{ik} \]
- Choice probability
  \[ P_i = \frac{\exp(V_i)}{\sum_{j \in A} \exp(V_j)} \]
- Parameters \( \theta_k \) are unknown
• Route choice (for the metro network)

• Multinomial Logit (MNL)

• Utility \( V_r = \theta_1 \times \sum_{a \in r \setminus r_t} c_a + \theta_2 \times \sum_{a \in r_t} c_a \)

  - in-vehicle time
  - transfer time
• Route choice (for the metro network)
• Multinomial Logit (MNL)

Utility \( V_r = \theta_1 \times \sum_{a \in r \setminus r_t} c_a + \theta_2 \times \sum_{a \in r_t} c_a \)

- in-vehicle time
- transfer time

Choice probability

\[
f_w(r \mid c, \alpha, \theta) = \frac{\exp(V_r)}{\sum_{r \in R_w} \exp(V_r)}
\]

For each OD pair \( w \)

\(|R_w| \geq 1|
Observed

- Observing travel time $t = t_b - t_a$ for OD pair $w = (a, b)$
- The probability observing $t$ on route $r$

$$t \mid r \sim \mathcal{N}\left(\sum_{a \in r} c_a + m, \alpha^2 \sum_{a \in r} c_a^2 + \sigma_y^2\right)$$

- The probability observing $t$ on OD pair $w$

$$p_w(t \mid c, \alpha, \theta, m) = \sum_{r \in R_w} h(t \mid r) f_w(r \mid c, \alpha, \theta, m)$$
The likelihood of observation all smart card transactions
(travel time)

\[ \mathcal{L}(c, \alpha, \theta, m | T) = \prod_{w \in W} p(T_w | c, \alpha, \theta, m) \]

\[ = \prod_{w \in W} \left( \prod_{t \in T_w} \left( \sum_{r \in R_w} h(t | r) f_w(r | c, \alpha, \theta, m) \right) \right) \]
Observed

- Prior knowledge
- Mean link travel time follows normal distribution
  \[ c_a \sim \mathcal{N}(2,1) \]
  - Travel time between stations / transfer time: around 2 minutes

- Other cost follows a normal distribution
  \[ m \sim \mathcal{N}(4,1) \]
  - Waiting time plus access/egress cost: around 4 minutes in total
• Prior knowledge
• Parameters for MNL: we do not have any information
• In the literature:

<table>
<thead>
<tr>
<th>Attribute</th>
<th>London Underground Parameter</th>
<th>t-Value</th>
<th>Santiago Metro Parameter</th>
<th>t-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>In-vehicle time</td>
<td>-0.121</td>
<td>-0.20</td>
<td>-0.074</td>
<td>-6.30</td>
</tr>
<tr>
<td>Morning peak</td>
<td>-0.084</td>
<td>-5.19</td>
<td>-0.014</td>
<td>-2.52</td>
</tr>
<tr>
<td>Afternoon peak</td>
<td>-0.042</td>
<td>-2.42</td>
<td>-0.025</td>
<td>-5.03</td>
</tr>
<tr>
<td>Morning peak</td>
<td>-0.269</td>
<td>-14.21</td>
<td>-0.083</td>
<td>-3.62</td>
</tr>
<tr>
<td>Time</td>
<td>-0.208</td>
<td>-5.94</td>
<td>-0.094</td>
<td>-2.60</td>
</tr>
<tr>
<td>Walking time</td>
<td>-0.299</td>
<td>-9.32</td>
<td>-0.201</td>
<td>-2.34</td>
</tr>
<tr>
<td>Women</td>
<td>-0.048</td>
<td>-2.06</td>
<td>-0.074</td>
<td>-2.67</td>
</tr>
<tr>
<td>Number of transfers</td>
<td>-1.321</td>
<td>-4.14</td>
<td>-0.662</td>
<td>-4.19</td>
</tr>
<tr>
<td>Ascending transfers</td>
<td>-0.205</td>
<td>-2.53</td>
<td>-0.308</td>
<td>-3.60</td>
</tr>
<tr>
<td>Even transfers</td>
<td>0.610</td>
<td>3.82</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>Descending transfers</td>
<td>0.000</td>
<td>n.a.</td>
<td>0.000</td>
<td>n.a.</td>
</tr>
<tr>
<td>Assisted transfers</td>
<td>0.000</td>
<td>n.a.</td>
<td>0.000</td>
<td>n.a.</td>
</tr>
<tr>
<td>Semi-assisted transfers</td>
<td>-0.271</td>
<td>-5.30</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>Non-assisted transfers</td>
<td>-0.398</td>
<td>-6.33</td>
<td>-0.183</td>
<td>-5.11</td>
</tr>
<tr>
<td>Mean occupancy</td>
<td>-2.898</td>
<td>3.25</td>
<td>-0.935</td>
<td>-5.10</td>
</tr>
<tr>
<td>Getting a seat</td>
<td>0.117</td>
<td>2.22</td>
<td>0.805</td>
<td>3.68</td>
</tr>
<tr>
<td>Not boarding</td>
<td>-0.502</td>
<td>-6.23</td>
<td>-0.358</td>
<td>-2.29</td>
</tr>
<tr>
<td>Angular cost</td>
<td>-0.088</td>
<td>3.89</td>
<td>-0.029</td>
<td>-3.84</td>
</tr>
<tr>
<td>Restrictive purpose</td>
<td>0.049</td>
<td>3.79</td>
<td>0.011</td>
<td>2.70</td>
</tr>
<tr>
<td>Map distance</td>
<td>-0.364</td>
<td>-5.43</td>
<td>-0.278</td>
<td>-4.83</td>
</tr>
<tr>
<td>Number of stations</td>
<td>-0.424</td>
<td>-5.07</td>
<td>-0.158</td>
<td>-3.62</td>
</tr>
<tr>
<td>Turning back</td>
<td>-0.650</td>
<td>-8.85</td>
<td>-0.142</td>
<td>-8.10</td>
</tr>
<tr>
<td>Turning away</td>
<td>-0.943</td>
<td>-7.77</td>
<td>-0.231</td>
<td>-8.87</td>
</tr>
<tr>
<td>Commonality factor</td>
<td>-0.396</td>
<td>-3.74</td>
<td>-0.541</td>
<td>-3.41</td>
</tr>
<tr>
<td>Sample size</td>
<td>17,073</td>
<td></td>
<td>28,061</td>
<td></td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>-6090</td>
<td></td>
<td>-12,881</td>
<td></td>
</tr>
<tr>
<td>Corrected $\rho^2$</td>
<td>0.567</td>
<td></td>
<td>0.383</td>
<td></td>
</tr>
</tbody>
</table>
Observed

• Prior knowledge
• Parameters for MNL: we do not have any information
  – We take uniform priors
    \[ \theta \sim \mathcal{U}(-4, 0) \]

• Coefficient of variation
  – We take a uniform prior
    \[ \alpha \sim \mathcal{U}(0, 1) \]
Posterior

\[
\pi(c, \alpha, \theta, m | T) = \prod_{w \in W} \left( \prod_{t \in T_w} \left( \sum_{r \in R_w} p(t | r) f_w(r | c, \theta) \right) \right) \times \prod_{c \in C} \phi(c; 2, 1) \times \phi(m; 4, 1)
\]
Solution Algorithm

- MCMC (Markov Chain Monte Carlo)
- Variable-at-a-time Metropolis sampling scheme

\[ \delta = (c_1, \ldots, c_N, \alpha, \theta_1, \theta_2, m) = (\delta_1, \ldots, \delta_{N+4}) \]
Solution Algorithm

- MCMC (Markov Chain Monte Carlo)
- Variable-at-a-time Metropolis sampling scheme

\[ \mathbf{\delta} = (c_1, \ldots, c_N, \alpha, \theta_1, \theta_2, m) = (\delta_1, \ldots, \delta_{N+4}) \]

- STEP (0)
- Specify initial sample

\[ \mathbf{\delta}^{(0)} = (c_1^{(0)}, \ldots, c_N^{(0)}, \alpha^{(0)}, \theta_1^{(0)}, \theta_2^{(0)}, m^{(0)}) \]

- Set \( t = 0 \)
Solution Algorithm

• **STEP (1)**
• At step $t$, sample in turn $\delta_i^{(t)}$ (for $i = 1: N+4$)
• Calculate

$$A(\delta_i^*, \delta_i^{(t)}) = \min \left\{ 1, \frac{\rho(T | \delta_i^*, \delta_{-i}^{(t)}) \pi(\delta_i^*, \delta_{-i}^{(t)})}{\rho(T | \delta_i^{(t)}, \delta_{-i}^{(t)}) \pi(\delta_i^{(t)}, \delta_{-i}^{(t)})} \right\}$$

• Where $\delta_{-i}^{(t)} = \left( \delta_1^{(t+1)}, \ldots, \delta_{i-1}^{(t+1)}, \delta_i^{(t+1)}, \delta_{i+1}^{(t)}, \ldots, \delta_{N+4}^{(t)} \right)$
• is the latest updated variables except $\delta_i^{(t)}$

• Accept $\delta_i^{(t+1)} = \delta_i^*$ with probability $A(\delta_i^*, \delta_i^{(t)})$
• Otherwise, set $\delta_i^{(t+1)} = \delta_i^{(t)}$
Solution Algorithm

• **STEP (2)**

• **If** $t < T$ : set $t = t + 1$
  – Go to **STEP (1)**

• **Else:**
  – Stop iteration
Numerical Example: MRT Network in Singapore

- MCMC provides distribution of unknown parameters rather than one value
- Burn-in: 5000 steps
- Effective sample: 25000 (25000+5000 draws in total)

- Standard deviation for Gaussian random walk Metropolis proposals

\[ \delta_i^* = \delta_i^{(t)} + \epsilon_i \quad \epsilon_i \sim \mathcal{N}(0, \sigma_i^2) \]
Numerical Example

- MCMC provides distribution of unknown parameters rather than one value

- $\alpha \ 30.8\%$

- $\theta_1 \ 36.7\%$

- $\theta_2 \ 45.8\%$
Numerical Example

- In this example
- Our prior knowledge is inaccurate
- the large number of travel time observations has corrected it

Transfer time @ Bouna Vista Stn

Travel time from EW1 to EW2
Numerical Example

- Flow assignment based on route choice model

Direction 1
Direction 2
Conclusion

- An integrated statistical model on travel time reliability and route choice behavior
- A metro network in which only travel time is observed
- Bayesian inference framework to formulate posterior probability
- Given the high-dimension of parameters, variable-at-a-time Metropolis sampling algorithm is applied to obtain posterior distribution.
Conclusion

• An integrated statistical model on travel time reliability and route choice behavior
• A metro network in which only travel time is observed
• Bayesian inference framework to formulate posterior probability
• Given the high-dimension of parameters, variable-at-a-time Metropolis sampling algorithm is applied to obtain posterior distribution.
• With this framework, we characterized travel time and its variation on each link. Meanwhile, we also identified contribution of different factors in determining passenger route choice behavior/movement.
Discussion & Outlook

• Most metro systems are closed environments, which only register transactions when passengers enter and leave the system; as a result, route choice (interchange/transfer) and service reliability are not captured in smart card data.

• Our framework does not require a specific route choice model; thus, it can be applied on a more sophisticated model which takes more factors into account, helping us to further understand passenger behavior and build advanced flow assignment models, and further infer individual train load.

• Although Singapore’s network is simple, this framework shows great potential in applying on more complex metro networks, such as London Underground.

• Identifying critical/crowding location/facility in metro network.
Thank you!

Lijun SUN

Future Cities Laboratory, Singapore-ETH Centre

lijun.sun@ivt.baug.ethz.ch