Using public transport smart card data for large-scale, agent-based transport demand simulation using MATSim: the case of Singapore

Alex Erath, Sergio Ordonez, Artem Chakirov, Pieter Fourie
Future Cities Laboratory, Singapore ETH-Centre
Idea: modelled Big Data for scenario forecast
Using real demand to simulate public transport

Derive travel demand from smart card transactions

- Transactions recorded on Tuesday, 22\textsuperscript{nd} April 2011
- 4 Mio journeys
- Boarding stop (journey level)
- Boarding time
- Alighting stop (journey level)

Possible demand reactions

- New routes (including transfers)
- Walk to other stops
- Mode switch (except for walk)
- Time of day
- Location of start/end stop
- Induced demand
Preparing demand data

Challenges

1. No information about actual arrival time at bus stop
2. No information about actual origin and destination on building level

Basic assumptions

1. Uniform arrival rate between two scheduled services
2. Journey starts and ends at reported public transport stops

Boarding, alighting at stop x, bus line y
Demand: behavioral parameters

**Public transport**
- Value of in-vehicle time: 8 SGD/h
- Value for waiting (start and transfer): 12.89 SGD/h
- Additional penalty for transfer: 0.65 SGD = 5 min in-vehicle time

**On foot (access/egress)**
- Walking speed: 4km/h
- Value of walking time: 16.92 SGD/h

**In future scenarios:**
- Value of a seat/crowdedness
- Preference for bus (anecdotal evidence)
- Agent specific preference
Supply: stochastic nature of travel times

Speed between stops

\[ v \sim N(\mu, \sigma) \]
\[ \mu = f\left(\frac{f}{c}, v_f, tl, m, l \ldots\right) \]
\[ \sigma = f\left(\frac{f}{c}, v_f, tl, m, \ldots\right) \]

Dwell time

\[ d \sim N(\mu, \sigma) \]
\[ \mu = f(b, a, p, t) \]
\[ \sigma = f(\mu) \]

Trip speed

\[ v \sim N(\mu, \sigma) \]
\[ \mu = f\left(\frac{f}{c}, v_f, tl, m, b, a, p, t, \ldots\right) \]
\[ \sigma = f\left(\frac{f}{c}, v_f, tl, m, \mu_d\right) \]

Simulation parameters

Observation
Calibration of simulation

**starting values**

- $v_{bus, trunk} = 26 \text{ km/h}$
- $v_{bus, exp} = 50 \text{ km/h}$
- $\sigma_{bus}(v) = 0.2 \cdot v_{bus}$
- $v_{train} = 72 \text{ km/h}$
- $\sigma_{train}(v) = 0$

Bus stops: sequential operations
Rail: access and waiting time **not included** in MATSim

**calibrated values**

- $v_{bus, trunk} = \frac{22 \text{ km}}{h} \pm f(h)$
- $v_{bus, exp} = 50 \text{ km/h}$
- $v_{bus, art} = 40 \text{ km/h}$
- $\sigma_{bus}(v) = 1.1 \cdot \sigma_{bus, Cepas, h}$
- $v_{train} = 72 \text{ km/h}$
- $\sigma_{train}(v) = 0$

Bus stops: parallel boarding
Rail: access and waiting time **included**
Case study I: Adding a new bus line

Residential new town
- Tidal demand patterns
- Issues with overcrowding during peak hours

New bus line:
- 26 stops
- 10km
- Loop from MRT A to MRT B and back
Ridership: **Line1, Line 2 and new line**

**before**

**after**
Reality check

Ridership per line and time of day

- Line 1
- Line 2
- New Line

Time of day [h]
Load factor: **Line 1**

Before and After comparison of load factor over time of day [h].

<table>
<thead>
<tr>
<th>Time of day [h]</th>
<th>empty</th>
<th>full</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**BEFORE**

**AFTER**
Load factor: **Line 2**

**BEFORE**

**AFTER**

Time of day [h]

- Direction
  - empty
  - full

- empty
  - full
Load factor: New line

AFTER
Case study II: split bus line in two parts

BEFORE SPLIT

95 stops
94 stops

Line length: 37km

AFTER SPLIT

46 stops
43 stops

Line length 18km

50 stops
52 stops

Line length 19km
Reliability before: Simulation vs. Observation
Reliability before: Simulation vs. Observation
Reliability: before and after split

Simulation: BEFORE line split

Simulation: AFTER line split
Reliability: Excess waiting time along line

Excess waiting time [s]

- Before split
- After split

Start | Split stop | End of Line
Computation: hardware and run times

HPC setting
- IBM System x3850 X5 featuring 4 Intel Xeon E7-4870
- 15 threads used for each simulation scenario
- Requires up to 80 Gb RAM for each simulation scenario

1 iteration = simulation of all public transport lines in Singapore
- 6 minutes for simulation
- 2 minutes for finding new routes

Number of iterations required to reach equilibrium
- We computed 200 iterations -> 26h
- 50 iterations probably already sufficient -> 6.5h
Conclusion

**Modeled Big Data**
- Public transport smart card data
- Scenario forecast (rather than pattern analysis)

**Use multi-agent transport simulation software MATSim to simulate CEPAS data**
- Observed demand as input
- Full temporal dynamics
- Demand reactions restricted to route choice

**Next steps**
- Demand: from stops to buildings
- Improving computational performance
- Make such scenario forecast accessible to planning practice