OD MATSim: Aggregated Mobile Phone Data Transport Simulations

PRESENTS
Cuauhtemoc Anda

FCL
Dr. Pieter Fourie
Dr. Sergio Ordonez
Cuauhtemoc Anda

DataSpark
Dr. Shixin Luo
AGENDA

1. The Task

2. Aggregated OD matrices
   1. Data processing

3. OD MATSim
   1. Strategy
   2. Replanning
   3. Spatial disaggregation
   4. Mode Choice
   5. Time disaggregation
   6. Setup
   7. Results

4. Markov MATSim
• Run a MATSim scenario using DataSpark’s aggregated OD matrices
• 1st milestone in the DataSpark/FCL collaboration
Aggregated OD matrices
• Spatial resolution: URA's subzones boundaries
• Temporal resolution: Hourly
• April 2017
DATA PROCESSING

- Filter 18\textsuperscript{th} of April (Tuesday)
- Clean by travel time
  - Bus drivers
  - Taxi drivers
  - Delivery people
  - Stay location errors
- Keep until 3 hours travel time
- Island Subzone: Semakau
Aggregate by start_time (travel time: 0,1,2,3)
OD MATSim
• 1 Trip 1 Agent
• For every count in the OD matrix create an agent that performs that trip
• Tasks: spatial and temporal disaggregation, mode choice
1st iteration – Shortest path

Agents are scored based only on their trips (not activities)

In each iteration:
  • 10% agents are allowed to re-schedule their departure times
  • 10% agents are allowed to re-route
SPATIAL DISAGGREGATION

- Choose a random facility within the subzone for the origin and destination
## MODE CHOICE

- Gradient Boosting using HITS 2012

<table>
<thead>
<tr>
<th>ptid</th>
<th>o_subzone</th>
<th>d_subzone</th>
<th>CAR ORIGIN</th>
<th>CAR DEST</th>
<th>PT ORIGIN</th>
<th>PT DEST</th>
<th>TEL ORIGIN</th>
<th>TEL DEST</th>
<th>DISTANCE</th>
<th>time_int</th>
<th>matsim_mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>009040AR1_1_1</td>
<td>TELOK BLANGAH RISE</td>
<td>0.079602</td>
<td>0.089109</td>
<td>0.731343</td>
<td>0.712871</td>
<td>0.189055</td>
<td>0.198020</td>
<td>0.000000</td>
<td>17.0</td>
<td>car</td>
</tr>
<tr>
<td>1</td>
<td>009040AR1_2_1</td>
<td>TELOK BLANGAH RISE</td>
<td>0.079602</td>
<td>0.164869</td>
<td>0.731343</td>
<td>0.676425</td>
<td>0.189055</td>
<td>0.158706</td>
<td>7938.276065</td>
<td>7.0</td>
<td>car</td>
</tr>
<tr>
<td>2</td>
<td>009040AR1_2_2</td>
<td>ALJUNIED</td>
<td>0.165891</td>
<td>0.089109</td>
<td>0.688372</td>
<td>0.712871</td>
<td>0.145736</td>
<td>0.198020</td>
<td>7938.276065</td>
<td>7.0</td>
<td>car</td>
</tr>
<tr>
<td>3</td>
<td>009040AR1_3_1</td>
<td>TELOK BLANGAH RISE</td>
<td>0.079602</td>
<td>0.155340</td>
<td>0.731343</td>
<td>0.757282</td>
<td>0.189055</td>
<td>0.087379</td>
<td>11213.648937</td>
<td>7.0</td>
<td>teleport</td>
</tr>
<tr>
<td>4</td>
<td>009040AR1_3_2</td>
<td>JURONG GATEWAY</td>
<td>0.188119</td>
<td>0.089109</td>
<td>0.752475</td>
<td>0.712871</td>
<td>0.059406</td>
<td>0.198020</td>
<td>11213.648937</td>
<td>7.0</td>
<td>pt</td>
</tr>
</tbody>
</table>
MODE CHOICE

Data: \( \{x_i \in \mathbb{R}^p, y_i \in \{\text{car}, \text{pt}, \text{teleport}\}\}^N_{i=1} \)

Problem:
Find a function: \( F[x; \beta] \rightarrow [\text{car}, \text{pt}, \text{teleport}] \) (Classification)

Probabilistic Output: \( F[x; \beta] \rightarrow \{p_{\text{car}}, p_{\text{pt}}, p_{\text{teleport}}\} \)

Estimation/Learning:
\[ \beta_{\text{estimate}} = \beta \rightarrow \arg\min_{\beta} \sum_{i=1}^{N} \text{Loss}(\beta; x_i, y_i) \]

Loss Gradient Boosting:
\[ -\log\left(1 + e^{-y F[x_i]}\right) \]
\[ F = T_0 + \alpha_1 T_1 + \alpha_2 T_2 + \cdots \]
\[ -\log\left(\frac{e^{F y_i [x_i]}}{\sum_{j=1}^{c} e^{F_j [x_i]}}\right) + \lambda \times \Omega(\beta) \]
Generalization:
Estimate/train model with 70% of the data, test for **accuracy** on 30%

**Gradient Boosting**

<table>
<thead>
<tr>
<th>Prediction</th>
<th>car</th>
<th>pt</th>
<th>teleport</th>
</tr>
</thead>
<tbody>
<tr>
<td>car</td>
<td>659</td>
<td>349</td>
<td>325</td>
</tr>
<tr>
<td>pt</td>
<td>3032</td>
<td>11677</td>
<td>2491</td>
</tr>
<tr>
<td>teleport</td>
<td>244</td>
<td>215</td>
<td>436</td>
</tr>
</tbody>
</table>

**Overall Statistics**

- **Accuracy**: 0.6574
- 95% CI: (0.6507, 0.6641)
- No Information Rate: 0.6301
- P-Value [Acc > NIR]: 1.136e-15

**Relative Influence**

<table>
<thead>
<tr>
<th>var</th>
<th>rel.inf</th>
</tr>
</thead>
<tbody>
<tr>
<td>PT_ORIGIN</td>
<td>23.288183</td>
</tr>
<tr>
<td>PT_DEST</td>
<td>18.814545</td>
</tr>
<tr>
<td>DISTANCE</td>
<td>14.522452</td>
</tr>
<tr>
<td>TEL_ORIGIN</td>
<td>10.791773</td>
</tr>
<tr>
<td>CAR_DEST</td>
<td>9.595804</td>
</tr>
<tr>
<td>CAR_ORIGIN</td>
<td>9.311015</td>
</tr>
<tr>
<td>TEL_DEST</td>
<td>7.434257</td>
</tr>
<tr>
<td>time_int</td>
<td>6.241970</td>
</tr>
</tbody>
</table>
• Time resampling in MATSim
• Evolution of agents
• Let all trips begin within at half the hours

<table>
<thead>
<tr>
<th>time</th>
<th>origin</th>
<th>oX</th>
<th>oY</th>
<th>destination</th>
<th>dX</th>
<th>dY</th>
</tr>
</thead>
<tbody>
<tr>
<td>23:30</td>
<td>YUNNAN</td>
<td>354744.887410</td>
<td>148286.490678</td>
<td>YUNNAN</td>
<td>355593.007362</td>
<td>149329.985929</td>
</tr>
<tr>
<td>23:30</td>
<td>YUNNAN</td>
<td>354279.482898</td>
<td>148068.406644</td>
<td>YUNNAN</td>
<td>354405.245586</td>
<td>148099.160148</td>
</tr>
<tr>
<td>23:30</td>
<td>YUNNAN</td>
<td>354797.936648</td>
<td>148870.012315</td>
<td>YUNNAN</td>
<td>355161.310863</td>
<td>149534.294882</td>
</tr>
<tr>
<td>23:30</td>
<td>YUNNAN</td>
<td>355255.874712</td>
<td>149668.363002</td>
<td>YUNNAN</td>
<td>354486.973501</td>
<td>148290.804638</td>
</tr>
<tr>
<td>23:30</td>
<td>YUNNAN</td>
<td>355266.503961</td>
<td>149150.392740</td>
<td>YUNNAN</td>
<td>355396.675166</td>
<td>149492.383747</td>
</tr>
</tbody>
</table>
• Time resampling in MATSim
• Bound by hours
• Time resampling in MATSim
• Disaggregate starting time \textit{a priori}
• For every hour calculate probability weights at each second
• Use exponential interpolation function
MATSIM SET UP

- Cluster of 4 computers, each with 192 GB of RAM and 72 cores
- Average Running time per iteration: 30min
- 100 iterations – 2 days + 2 hours
- Singtel's marketshare assumption: 50% - Capacity links: 50% reduction
- + 5.5 million agents-trips
- Datasets: Singapore Transport Network, PT Schedules, ERP, facilities
- 20% Route replanning only (no time mutation)
MATSim Results
Match against Traffic Counts

(Disclaimer: counts are from a different day)
Next Step: Markov MATSim
• From aggregated mobile phone data disaggregate in realistic agents
• Privacy-by-design approach