Choice Set Generation and Model Identification for Route Choice using GPS-Data from Smart Phones

Dipl.-Ing. Eileen Mandir

University of Stuttgart
Institute for Road and Transport Science
Department for Transport Planning and Traffic Engineering

email: eileen.mandir@isv.uni-stuttgart.de  •  phone: +49 711 685-82477
What to expect?

Can traveler’s smart phones be used to

- provide a platform for long term recording of GPS data
- determine the size of choice sets in route choice
- help to identify suitable choice models

and which problems need to be addressed to exploit this data source?
Contents

- Data Collection
- Data Processing
- Choice Set Generation
- Model Identification
- Conclusions
Contents

- Data Collection
- Data Processing
- Choice Set Generation
- Model Identification
- Conclusions
Data Collection

- 278 participants
- > 20 miles commute each day
- 8 week period
- Onboard smart phone
- Data transfer to server via GSM mobile phone network
Logging GPS data online with downloadable smartphone applications:

- Benefits:
  - no labor intensive recruiting or instruction needed
  - no investment in survey hardware needed
Data Characteristics

Logging GPS data online with downloadable smartphone applications:

- **Benefits:**
  - no labor intensive recruiting or instruction needed
  - no investment in survey hardware needed

- **Drawbacks:**
  - data loss possible
  - data more difficult to interpret
Data Characteristics

Logging GPS data online with downloadable smart phone applications:

- **Benefits:**
  - no labor intensive recruiting or instruction needed
  - no investment in survey hardware needed

- **Drawbacks:**
  - data loss possible
  - data more difficult to interpret

- Several trips included in one track
- Inaccurate logging at trip ends
Data Volume

<table>
<thead>
<tr>
<th>Data Volume</th>
<th>Total over 278 participants</th>
<th>Per person</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total time of detection</td>
<td>12,044 hours</td>
<td>43.3 hours</td>
</tr>
<tr>
<td>Number of detected trajectories</td>
<td>20,000</td>
<td>71</td>
</tr>
</tbody>
</table>
Contents

- Data Collection
- **Data Processing**
- Choice Set Generation
- Model Identification
- Conclusions
Identification of Trip Ends

- Gaps in track due to loss of GSM network connection or unrecorded trips:
  - jump in time stamp of subsequent data points
  - jump in position of subsequent data points
→ Data loss or unrecorded trip?
Identification of Trip Ends

- Gaps in track due to loss of GSM network connection or unrecorded trips:
  - jump in time stamp of subsequent data points
  - jump in position of subsequent data points
    $\rightarrow$ Data loss or unrecorded trip?

- Data points with speed equal to zero due to loss of GPS signal or stop of vehicle
  $\rightarrow$ Congestion or activity?
Identification of Trip Ends

- Gaps in track due to loss of GSM network connection or unrecorded trips:
  - jump in time stamp of subsequent data points
  - jump in position of subsequent data points
  → Data loss or unrecorded trip?

- Data points with speed equal to zero due to loss of GPS signal or stop of vehicle
  → Congestion or activity?

- Procedure needed to
  - connect gap with shortest path
  - or identify trip end
Identification of Trip Ends

Data Volume

<table>
<thead>
<tr>
<th></th>
<th>Total over 278 participants</th>
<th>Per person</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of GPS trajectories</td>
<td>20,000</td>
<td>71</td>
</tr>
<tr>
<td>Number of identified trips</td>
<td>25,000</td>
<td>89</td>
</tr>
</tbody>
</table>
Identification Activity Location

- Origin and destination points needed for choice set generation
  - Position match
  - Time match
  - Point of Interest match

Usual arrival: 8:30 am
Usual departure: 6:00 pm
→ activity: work
Identification Activity Location

<table>
<thead>
<tr>
<th>Data Volume</th>
<th>Total over 278 participants</th>
<th>Per person</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of trips</td>
<td>25,000</td>
<td>89</td>
</tr>
<tr>
<td>Number of trips between identified activity locations</td>
<td>17,500</td>
<td>63</td>
</tr>
</tbody>
</table>
- Data Collection
- Data Processing
- **Choice Set Generation**
- Model Identification
- Conclusions
Chosen Routes from home to work

Person 185

Period: 61 days
Trips in total: 135
Trips home-work: 27
### Level of Service experienced

#### Direct Distance [km]

#### Direct Speed [km/h]

#### Trips of all participants

<table>
<thead>
<tr>
<th>Direct Distance [km]</th>
<th>Direct Speed [km/h]</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>70</td>
<td>70</td>
</tr>
</tbody>
</table>

#### Trips of participant 185

- A
- B
- C
- D
- E
- F

##### Routes from home to work

<table>
<thead>
<tr>
<th></th>
<th>Total over 278 participants</th>
<th>Per person</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of chosen routes (GPS)</td>
<td>610</td>
<td>2.2</td>
</tr>
<tr>
<td>Number of known routes (interview)</td>
<td>806</td>
<td>2.9</td>
</tr>
</tbody>
</table>
- Equilibrium assignments usually find >30 routes per OD pair for similar size networks
- Choice set generation needs to reflect hierarchical structure of decision process
Contents

- Data Collection
- Data Processing
- Choice Set Generation
- **Model Identification**
- Conclusions
### Classical Discrete Choice Models

- **C-Logit by Cascetta**

\[
P_r = \frac{f^C \cdot \exp(-\beta \cdot G_r)}{\sum_r f^C \cdot \exp(-\beta \cdot G_r)}
\]

- $\beta$: Parameter
- $f^C$: Cascetta coefficient
- $G$: Generalized cost
- $P$: Probability
- $r$: Route

![Graph](image)
Best possible Knowledge of Current Travel Time

Travel Time A9 Munich

- Travel time estimation (traffic flow model)
- Current route travel time (ANPR detectors)

- Peak
- Off-peak
- Night
- Modified Logit with low elasticity for small deviations by Gobiet

\[ P_r = \frac{f_C^r \cdot \exp(-\beta \cdot G_r)}{\sum_{r'} f_C^{r'} \cdot \exp(-\beta \cdot G_{r'})} \]

\( \beta \) Parameter
\( f_C \) Cascetta coefficient
\( G \) Generalized cost
\( P \) Probability
\( r \) Route

![Graph showing the choice probability of route 1 and route 2 against the generalized cost of route 1 and route 2. The graph illustrates how the choice probability changes with the cost for both routes.]
Contents

- Data Collection
- Data Processing
- Choice Set Generation
- Model Identification
- Conclusions
Conclusions

- Push for developing applications for smart phone GPS tracking
- Profound revealed preference data which can be fused with traffic state and traffic information data for detailed estimation
- Data valuable for providing empirical foundation for choice set generation and model identification
- Data valuable for calibration of equilibrium assignment models on chosen routes
Thank you!

Authors:
- Juliane Pillat, University of Stuttgart
- Eileen Mandir, University of Stuttgart
- Markus Friedrich, University of Stuttgart
- Christian Schiller, German Aerospace Center (DLR)