Workshop on Smart Mobility, Luxembourg June 2nd 2017
Session 6 - Demand Estimation using Big Data

Big data to improve the understanding of mobility

Prof. Martin FELLENDORF
Michael CIK, Christopher HORN (invenium)
Graz University of Technology
Inst. f. Transport Planning and Highway Engineering
www.isv.tugraz.at

www.tugraz.at
OD matrices – always needed, never (precisely) available

- trip agents are aggregated in traffic zones
- trips are generated between centroids
- limitations due to spatial aggregation
OD matrices – always needed, never (precisely) available

- trip agents are aggregated in traffic zones
- trips are generated between centroids
- limitations due to spatial aggregation

Emperical problems
- survey with low sample size (travel diary)
- travel demand model
**OD-matrices – Floating Phone Data**

- **conventional:**
  - travel survey with reported trip chains
  - projection of low sample size
  - travel demand model

- **New data source:**
  - Floating Phone Data (FPD)
  - Big Data
The data set

- A1 Mobile Traffic Data Stream (A1MTDS)
- Heatmap of activities within mobile network
A1MTDS – anonymized traffic data stream

Monitoringsystem
- anonymized ID (no IMSI)
- event types
- coordinates (50m*50m)
- time stamp

Event types
- mobile phone site/cells
- text messages
- location area updates
- handover cell updates
- mobile data
- roaming information of other networks

Big data
- ~ 4.5 SIM-cards in AT by A1
- Continuous data since 2013
- ~ 60 GB – 150 GB per day
Quality Control of A1TMDS

- Cross check download rates for different mobile networks
- Quality improvement by LTE (more events)
Validation

- Verification by single GPS tracks matched with A1 raw data

red day A

blue day B

raw data ≠ A1MTDS

Mobile phone events of single phone ID
Trip along various cells, event identification

- Large cells in UMTS

**Trips:** \( F_4 \rightarrow F_2, \ F_2 \rightarrow F_1 \)
System architecture FPD

QZTool Frontend (Data selection, filtering and visualization)

Generate trajectories from FPD

Mode detection (road, rail) by feature extraction (4 steps)

Regional dependent weighting factor

A1MTDS
A1 Mobile Traffic Data Service

Monitoring platform collects & anonymize single events

Mobile network signal data
GSM-Net (2G)
UMTS-Net (3G)
LTE-Network (4G)

OD_{ij} road

OD_{ij} rail

rail boarding

GIP/OSM road and railway network

Public transport real-time time table

Backend – Hadoop Cluster
Algorithm: Mode detection of FPD

- Input phone events of single ID´s (Floating Phone Data)
- Separate rail trips from road trips
- Match mobile phone trajectory on rail and road network (Map Matching)
Feature Extraction Method

1. Distance to suitable links
   • Shortest projected distance of all events between two activities?

2. Match with time table
   • Correlation btw. event time stamps and departure & arrival of real-time time table

Trajectory from Wr. Neustadt to Wien

Number of trajectories

0:00 23:59

0:00 23:59
3. Stationary stop / activity at transit stop?

4. Reasonable travel time?
**OD_{ij} Road,Rail - methodology**

FPD trajectories

Mode detection (road/rail) by feature extraction & Regional weighting factors (market penetration)

OD_{FPD} main zones

OD_{FPD} main zones – road & rail

Disaggregation of main zones

Split trips of main zones to TAZ’s

verify statistical independency of matrices

OD_{TransportModel} vs. OD_{FPD}

OD_{TransportModel} TAZ’s

aggregate TAZ’s to main zones based on mobile phone sites
- Compare individual OD\(ij\) – values
  - Correlation
  - Hypnothis testing
- Preparation work
  - Normalize both matrices
  - Distance matrix

- Row/column sums
  - Correlation coefficient

\[ d_{Mk} = p \sqrt{\sum_{i=1}^{d} \left| P_i - Q_i \right|^p} \]

\[
\text{Corr}(d_x, d_y) = \frac{\sum_{i=1}^{r} (d_x(r_i, q) - \bar{d}_x)(d_y(r_i, q) - \bar{d}_y)}{\sqrt{\sum_{i=1}^{r} (d_x(r_i, q) - \bar{d}_x)^2 \sum_{i=1}^{r} (d_y(r_i, q) - \bar{d}_y)^2}}
\]

where \( \bar{d}_x = \frac{\sum_{i=1}^{r} d_x(r_i, q)}{r} \)
$\text{OD}_{ij}^{\text{FPD}} \text{ vs } \text{OD}_{ij}^{\text{TransportModel}}$
Check trip length distribution

Daily Trip Length Distribution in AT
n1=1,475,505 FPD-trips;
n2=145,384 reported workday trips, travel diary 2013/14

- Mobile Phone Data
- Travel Diary All Modes
- Bike
- CarDriver
- PubTransp
- Other
Case study Eastern Austrian regional model

- Comparison $OD_{\text{TransportModel}}$ vs. $OD_{\text{FPD}}$, Road trips from city of Melk to other TAZ´s

Transport model: 1.292,44
FPD: 1.308

Project NaviMOP:
ITS Vienna Region, Know Center, TU Graz
Regional rail based public transport corridors

PAX - FJB nach Wien

<table>
<thead>
<tr>
<th>Date</th>
<th>Wien</th>
</tr>
</thead>
<tbody>
<tr>
<td>21.04.2016</td>
<td>5000</td>
</tr>
<tr>
<td>22.04.2016</td>
<td>7000</td>
</tr>
<tr>
<td>23.04.2016</td>
<td>4000</td>
</tr>
<tr>
<td>24.04.2016</td>
<td>3000</td>
</tr>
<tr>
<td>25.04.2016</td>
<td>8000</td>
</tr>
</tbody>
</table>
Passenger counts at railway stations

St. Pölten Hbf (in Pb)

Gesamt
- Einstieger
- Ausstieger
- Umsteiger

Zeitliche Entwicklung

Richtungen
- Bahnsteige
- Züge

Ankunft
- Wien Meidling
- Wien Westbahnhof
- Linz
- Plochingen
- Tullnerfeld
- Floridsdorf

Abfahrt

Verteilung Wartezeit
Concluding remarks

Privacy issues
- A1TDMS suitable format for transport planners
- A1TDMS Anonomization consideration meets GDPR 2017 regulation
- Trajectories have to remain within data processing center

Applications
- Mode detection, especially rail based transport works IFF real-time time tables are available
- Very good for passenger boardings, passenger interchanges at stations
- Daten-driven ODs substitute model based ODs; validation work in progress
- Better identification of long-distance (rare) trips than travel diaries

Future work:
- Investigation of short trips (< 1km) with more LTE coverage
- Activity estimation by event duration and location
References

- Reiter, T; Cik, M; Haberl, M; Breitwieser, C; Scherer, R; Kröll, M; Horn, C; Müller-Putz, G; Lindstaedt, S; Fellendorf, M: Traffic incident detection with cellular network data. Paper in proceedings of OPTIMUM - International Symposium on Recent Advances in Transport Modelling, Brisbane, Australia, April 21-23, 2013
- Cik, M.; Fellendorf, M.; Vogel, J.: Mobilfunkbewegungsdaten als Erweiterung der Datengrundlage für Verkehrsmodelle. - in: Straßenverkehrstechnik ; 2014, Heft 11, s. 739-744, Bonn
- Horn, C.; Gursch, H.; Kern R.; Cik, M: QZTool—Automatically Generated Origin-Destination Matrices from Cell Phone Trajectories. Advances in Human Aspects of Transportation, 823-833
- Cik, M.; Horn, C.; Fellendorf, M: Floating-Phone-Data für die Berechnung von verkehrsmittelspezifischen Nachfragematrixen. Heureka17, FGSV Stuttgart 22.3.2017