Applying multiple Big Data sources to improve public transport planning and operations

dr. ir. N. van Oort

Assistant professor public transport
Smart Public Transport Lab
Delft University of Technology
www.SmartPTLab.TUDelft.nl

UITP INDIA
SEMINAR ON URBAN RAIL NETWORK – BUILDING SUSTAINABLE CITIES

December 2018

@Niels_van_Oort
Demand Responsive Transit
Passenger flows
Choice modelling
Network Vulnerability
Network assignment
North-South metro line
CO₂ modelling
Prediction & Information

http://smartptlab.tudelft.nl

Challenge the future
Improving public transport

Data + Trends → Understanding → Predicting → Improving
The challenge

- New methodologies
- Proven in practice

- Data
- Information
- Knowledge

- Evaluation
- Forecasts
- Real time

- Improvements
Promising data sources

Vehicle monitoring
- Traditional: Manual countings / Selection of vehicles equipped with automatical vehicle location (AVL)
- New: Board computers (AVL)

Passenger monitoring
- Traditional: Manual countings / Selection of vehicles equipped with automatical passenger counting (APC)
- New: Smart card data

Mobility monitoring
- GSM/Mobile phone data

Door to door monitoring
- Traveller app data
1) AVL data
AVL data: The Dutch approach: NDOV

NDOV is a nationwide initiative to make transit data available to authorities and the public.

Focus on dynamic traveler information

Timetable and AVL data available from the majority of the transit vehicles.

(source: GOVI)
Cooperation of 15 transit authorities in NL

Two platforms to share all data (real time and offline)

Displays

Standard data format: [http://bison.connekt.nl/](http://bison.connekt.nl/)
The NDOV data architecture

Data interfaces defined in the BISON standard
  e.g. KV1 – timetable, KV6 – AVL, KV15 – free text msgs

(source: GOVI)
Real time positions: www.OVradar.nl
The NDOV data architecture

All AVL data is publicly available without restrictions.

Example:

<table>
<thead>
<tr>
<th>Time</th>
<th>Message type</th>
<th>Operator</th>
<th>Line</th>
<th>Journey</th>
<th>Stop</th>
<th>Punctuality</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:29:00</td>
<td>INIT</td>
<td>...</td>
<td>B120</td>
<td>7001</td>
<td>99990140</td>
<td></td>
</tr>
<tr>
<td>08:29:00</td>
<td>ONSTOP</td>
<td>...</td>
<td>B120</td>
<td>7001</td>
<td>99990140</td>
<td>60</td>
</tr>
<tr>
<td>08:29:22</td>
<td>DEPARTURE</td>
<td>...</td>
<td>B120</td>
<td>7001</td>
<td>99990140</td>
<td>82</td>
</tr>
<tr>
<td>08:31:28</td>
<td>DEPARTURE</td>
<td>...</td>
<td>B120</td>
<td>7001</td>
<td>99990290</td>
<td>88</td>
</tr>
</tbody>
</table>

... 

<table>
<thead>
<tr>
<th>Time</th>
<th>Message type</th>
<th>Operator</th>
<th>Line</th>
<th>Journey</th>
<th>Stop</th>
<th>Punctuality</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:51:04</td>
<td>ONROUTE</td>
<td>...</td>
<td>B120</td>
<td>7001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>08:52:37</td>
<td>ARRIVAL</td>
<td>...</td>
<td>B120</td>
<td>7001</td>
<td>99990500</td>
<td>-202</td>
</tr>
<tr>
<td>08:52:37</td>
<td>END</td>
<td>...</td>
<td>B120</td>
<td>7001</td>
<td>99990500</td>
<td></td>
</tr>
</tbody>
</table>

AVL(1)

Many early trips

Holding regime
AVL (2)
AVL(3)
AVL(4)
Bottleneck detection

Brands, T., N. van Oort, M. Yap (2018),

Automatic bottleneck detection using AVL data: a case study in Amsterdam,

Conference on Advanced Systems in Public Transport and TransitData (CASPT), Brisbane, Australia.
Bottleneck definition

- Contract requirements, literature and expert judgement

- Dwell time
- Variation dwell time (85 and 15 percentile)
- Early departure
- Late departure
- Variation departure time: (85 and 15 percentile)
- Punctuality change compared to previous stop
- Speed
- Trip time compared to free flow (15th percentile of the travel time on Sundays)

Results: bottlenecks

<table>
<thead>
<tr>
<th>Line</th>
<th>Direction</th>
<th>Period</th>
<th>Stop</th>
<th>Stop number</th>
<th>Average dwell time</th>
<th>Dwell time variation</th>
<th>Punctuality 50%</th>
<th>Variation departure time</th>
<th>Punctuality change wrt previous stop</th>
<th>Average speed</th>
<th>Difference with free flow travel time</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 Back</td>
<td>Evening</td>
<td>14Elandsgracht</td>
<td>6028</td>
<td>20</td>
<td>13</td>
<td>-28</td>
<td>217</td>
<td>65</td>
<td>30</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>10 Back</td>
<td>Saturday</td>
<td>13Leidseplein</td>
<td>6061</td>
<td>31</td>
<td>25</td>
<td>66</td>
<td>182</td>
<td>18</td>
<td>14</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>10 Back</td>
<td>Sunday</td>
<td>13Leidseplein</td>
<td>6061</td>
<td>29</td>
<td>22</td>
<td>32</td>
<td>163</td>
<td>14</td>
<td>15</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>10 Back</td>
<td>Sunday</td>
<td>14Elandsgracht</td>
<td>6028</td>
<td>20</td>
<td>17</td>
<td>-42</td>
<td>177</td>
<td>73</td>
<td>32</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>2 Forth</td>
<td>AM peak</td>
<td>09Hoofddorpplein</td>
<td>7049</td>
<td>25</td>
<td>18</td>
<td>73</td>
<td>199</td>
<td>45</td>
<td>13</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>2 Forth</td>
<td>AM peak</td>
<td>10Amstelveenseweg</td>
<td>7034</td>
<td>22</td>
<td>15</td>
<td>95</td>
<td>223</td>
<td>22</td>
<td>15</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>2 Forth</td>
<td>AM peak</td>
<td>13Van Baerlestraat</td>
<td>7322</td>
<td>23</td>
<td>11</td>
<td>50</td>
<td>303</td>
<td>2</td>
<td>18</td>
<td>37</td>
<td></td>
</tr>
</tbody>
</table>
19

Challenge the future

Type of bottleneck

1. Average dwell time
2. Early departure
3. Variation departure
4. Worse punctuality
5. Low speed
6. Large travel time
2) Smartcard data
Smartcard data

The Netherlands
- OV Chipkaart
- Nationwide (since 2012)
- All modes: train, metro, tram, bus
- Tap in and tap out
- Bus and tram: devices are in the vehicle

Issues
- Privacy
- Data accessibility via operators

- Several applications of smartcard data: Pelletier et. al (2011). Transportation Research Part C
Task: introduce and maintain smartcard system in NL

- Owned by all PT operators in NL
- >25 million smart card issued
- > 2 billion transactions annually
Examples: Monitoring and predicting passenger numbers
Examples: Monitoring and predicting passenger numbers
Examples: Monitoring and predicting passenger numbers
Examples: Monitoring and predicting passenger numbers
Examples: Monitoring and predicting passenger numbers
Examples: Monitoring and predicting passenger numbers
Examples: Monitoring and predicting passenger numbers
Spatiotemporal Load Profiles

- Spatiotemporal load profiles of transit vehicles can be constructed using multiple data sources.
- Combination of data sets: AFC/smartcard, AVL and GTFS.


Acknowledgement: HTM and Stichting OpenGeo provided the AFC and AVL datasets, resp.
Challenge the future
Occupancy pattern clustering and recognition (Artificial intelligence)

**Legend:**
- Almost empty
- Sit alone
- Sit next to someone
- All seats occupied

Heydenrijk-Ottens et al. (2018), Supervised learning: Predicting passenger load in public transport, CASPT conference, Brisbane
Passenger service reliability

AVL+Smartcard

Dixit et al. (2019).
TRB
What-if analysis with smart card data
PT modelling

Traditional (4-step) model
- Multimodal (~PT)
- Network
- Complex
- Long calculation time
- Visualisation
- Much data
- Detailed results

Simple calculation
- PT only
- Line
- Transparent
- Short calculation time
- Only numbers
- Little data
- Assessments

Short term predictions

Elasticity method based on smartcard data
Combining models and smartcard data

Connecting to transport model
- Evaluating history
- Predicting the future

- What if scenario’s
  - Stops: removing or adding
  - Faster and higher frequencies
  - Route changes

- Quick insights into
  - Expected cost coverage
  - Expected occupancy

New generation of transport models: data driven

Van Oort, N., T. Brands, E. de Romph (2015), Short-Term Prediction of Ridership on Public Transport with Smart Card Data, Transportation Research Record, No. 2535, pp. 105-111.
What if: increased frequencies

fictitious data
What if: increased speed
3) GSM/cell phone data
Data pre-processing

**GSM data**

**Value:** Amount of visitors detected in a zone, on a specific day, per hour

- Data from one network operator
  - Algorithm to increase sample data to total population
- Distinction inhabitants or visitors
  - Place of residence based on overnight stays per month
- Spatial level of detail: zone level
  - Antennas have overlapping reach
- Difference occupancy between subsequent hours is a net change
Rotterdam late evenings & early mornings

Working day scenario
06:00-07:00
Cost benefit analysis

Optimizing planning and real time control
Van Oort, N. and R. van Nes (2009), Control of public transport operations to improve reliability: theory and practice, Transportation research record, No. 2112, pp. 70-76.


Optimizing synchronization multimodal transfers
Lee, A. N. van Oort, R. van Nes (2014), Service reliability in a network context: impacts of synchronizing schedules in long headway services, TRB


Improved scheduling


Passenger behavior and modelling
Literature

http://nielsvanoort.weblog.tudelft.nl/

Brands, T., N. van Oort, M. Yap (2018), **Automatic bottleneck detection using AVL data: a case study in Amsterdam**, Conference on Advanced Systems in Public Transport and TransitData (CASPT), Brisbane, Australia.

De Regt K., Cats O., van Oort N. and van Lint H. (2017). **Investigating Potential Transit Ridership by Fusing Smartcard Data and GSM Data.** *Transportation Research Record: Journal of the Transportation Research Board*, No. 2652


Heydenrijk-Ottens, L. et al. (2018), Supervised learning: **Predicting passenger load in public transport**, CASPT conference, Brisbane


Van Oort, N., T. Brands, E. de Romph (2015), **Short-Term Prediction of Ridership on Public Transport with Smart Card Data**, *Transportation Research Record*, No. 2535, pp. 105-111.

Questions / Contact

N.vanOort@TUDelft.nl
@Niels_van_Oort

Publications:
http://nielsvanoort.weblog.tudelft.nl/

www.smartPTlab.tudelft.nl

Amsterdam
May 2019

http://ppts-course.org/