MOBILITY IN THE AGE OF BIG DATA

Assoc. Prof. Ivana Semanjski
DATA ANALYTICS 2018, The Seventh International Conference on Data Analytics,
November 18 - 22, 2018 - Athens, Greece
INTRODUCTION

The Seventh International Conference on Data Analytics DATA ANALYTICS 2018 November 18-22, 2018 - Athens, Greece
TTASC: Transport and Traffic Analytics in Smart Cities
Keynote: Ivana Semanjski - Mobility in the age of big data
ADDING VALUE TO SOCIETY
Innovation Centre For Intelligent Information Processing
WHAT WE DO
Spin off companies
Spin off companies

Progeno
Identifies genetic potential

Design a breeding Strategy

BED
Believe in 5D!

A revolutionary marketing display

VISUAL INTELLIGENCE FOR YOUR WEB APPLICATION
COSCALE
Industrial cooperation and partnerships

Barco
Philips
NXP
Gent
Trafficon
TOMTOM
Flow
Volvo
Proximus
AviaGIS
Delhaize
telenet
BEMobile
NMBS
Lijn
Cambio
Enovates
Synopsys
Grontmij
Traject
Altera
Transportation planning

prediction of usage demand in future travel

ensuring necessary facilities and services to cater to that demand
HOW DO WE DO IT?

Four-step models

- Trip generation
- Trip distribution
- Mode choice
- Route assignment

The Seventh International Conference on Data Analytics DATA ANALYTICS 2018 November 18-22, 2018 - Athens, Greece
TTASC: Transport and Traffic Analytics in Smart Cities
Keynote: Ivana Semanjski - Mobility in the age of big data
How do we collect data for transportation planning

- Travel diaries
- Interviews
78% of European citizens live in cities

2014 54% of the total global population lived in urban areas

average each person produces 1 GB of content daily
The Washington Post

Exabytes: Documenting the 'digital age' and huge growth in computing capacity

THE WORLD'S CAPACITY TO STORE INFORMATION

This chart shows the world's growth in storage capacity for both analog data (books, newspapers, videocassettes, etc.) and digital (CDs, DVDs, computer hard drives, smartphone drives, etc.)

In gigabytes or estimated equivalent

- 2000
  - ANALOG STORAGE
  - DIGITAL

1986 ANALOG 2.62 billion

1993 DIGITAL 0.02 billion

COMPUTING POWER

In 1986, pocket calculators accounted for much of the world's data-processing power.

Percentage of available processing power by device:

<table>
<thead>
<tr>
<th>Year</th>
<th>Pocket calculators</th>
<th>Personal computers</th>
<th>Videogame consoles</th>
<th>Servers, mainframes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1986</td>
<td>41%</td>
<td>33%</td>
<td>9%</td>
<td>17%</td>
</tr>
<tr>
<td>2007</td>
<td>90%</td>
<td>10%</td>
<td>1%</td>
<td>0%</td>
</tr>
</tbody>
</table>

*By Brian Valente
Online Opinion Post (September 11, 2013 - 11:17 PM)

2007 ANALOG 18.86 billion gigabytes

- Paper, film, audiotapes and vinyl 0.2%
- Other digital media 0.8%
- Portable media players, flash storage 2%
- Portable hard disks 2.4%
- CDs and minidisks 6.8%
- Computer servers and mainframe hard disks 0.6%
- Digital tape: 11.6%
- DVD/Blu-ray: 22.8%

2007 DIGITAL 276.12 billion gigabytes

*Other includes chip cards, memory cards, flash drives, mobile phones (PDAs), smartphones, camcorders, video games

ATHENS, Greece

The Seventh International Conference on Data Analytics DATA ANALYTICS 2018 November 18-22, 2018 - Athens, Greece

TTASC: Transport and Traffic Analytics in Smart Cities

Keynote: Ivana Semanjski - Mobility in the age of big data
Big data sources for mobility studiers

- CDR & signalisation
- GNSS
- Mobile phone
CDR & NETWORK SIGNALIZATION DATA

The Seventh International Conference on Data Analytics DATA ANALYTICS 2018 November 18-22, 2018 - Athens, Greece

TTASC: Transport and Traffic Analytics in Smart Cities

Keynote: Ivana Semanjski - Mobility in the age of big data
Mobile phone data

Passive

Active

The Seventh International Conference on Data Analytics DATA ANALYTICS 2018 November 18-22, 2018 - Athens, Greece
TTASC: Transport and Traffic Analytics in Smart Cities
Keynote: Ivana Semanjski - Mobility in the age of big data
<table>
<thead>
<tr>
<th>Literature</th>
<th>OD estimation</th>
<th>OD Accuracy</th>
<th>Vehicle occupancy rates</th>
<th>Occupancy rate Accuracy</th>
<th>Duration test data</th>
<th>Number of users</th>
<th>Number of observations</th>
<th>Dataset</th>
<th>User Validated</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Bohte and Kees, 2009)</td>
<td>No</td>
<td>-</td>
<td>No</td>
<td>-</td>
<td>1 week</td>
<td>1104</td>
<td></td>
<td>GNSS, GIS</td>
<td>Yes</td>
</tr>
<tr>
<td>(Munizaga and Palma, 2012)</td>
<td>Yes</td>
<td>82%</td>
<td>No</td>
<td>-</td>
<td>2 weeks</td>
<td>N/A (74 million observations)</td>
<td></td>
<td>Smartcard, GNSS (PT only)</td>
<td>No</td>
</tr>
<tr>
<td>(Shen and Stopher, 2013)</td>
<td>No</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>3 days</td>
<td>2059</td>
<td></td>
<td>GNSS</td>
<td>No</td>
</tr>
<tr>
<td>(Xiao et al., 2016)</td>
<td>No</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1 week</td>
<td>321</td>
<td></td>
<td>Smartphone</td>
<td>Yes</td>
</tr>
<tr>
<td>(Ge and Fukuda, 2016)</td>
<td>Yes</td>
<td>77%</td>
<td>No</td>
<td>-</td>
<td>1 day</td>
<td>N/A (650,000 observations)</td>
<td></td>
<td>Smartphone</td>
<td>No</td>
</tr>
<tr>
<td>(Dong et al., 2015)</td>
<td>Yes</td>
<td>-</td>
<td>No</td>
<td>-</td>
<td>1 day</td>
<td>N/A</td>
<td></td>
<td>CDR</td>
<td>No</td>
</tr>
<tr>
<td>(Wolf et al., 2001)</td>
<td>Yes</td>
<td>37%</td>
<td>No</td>
<td>-</td>
<td>3 day</td>
<td>13</td>
<td></td>
<td>GNSS, GIS</td>
<td>Yes</td>
</tr>
<tr>
<td>(Lu et al., 2012)</td>
<td>No</td>
<td>-</td>
<td>No</td>
<td>-</td>
<td>13 weeks</td>
<td>N/A (3188 trips)</td>
<td></td>
<td>GNSS, GIS</td>
<td>Yes</td>
</tr>
<tr>
<td>(Feng and Timmermans, 2011)</td>
<td>No</td>
<td>-</td>
<td>No</td>
<td>-</td>
<td>13 weeks</td>
<td>N/A (3188 trips)</td>
<td></td>
<td>GNSS, GIS</td>
<td>Yes</td>
</tr>
<tr>
<td>Literature</td>
<td>Duration of test data</td>
<td>Number of users</td>
<td>Dataset</td>
<td>Accuracy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------------------</td>
<td>-----------------------</td>
<td>----------------------</td>
<td>----------------------------------------------</td>
<td>----------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Fanhas and Saptawati, 2017)</td>
<td>2 months</td>
<td>N/A (16,337 records)</td>
<td>GNSS, taximeter</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Ge and Fukuda, 2016)</td>
<td>1 day</td>
<td>N/A</td>
<td>GNSS</td>
<td>88%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Moreira-Matias et al., 2016)</td>
<td>9 months</td>
<td>441</td>
<td>GNSS, taximeter</td>
<td>79%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Li et al., 2017)</td>
<td>1 month</td>
<td>12,000</td>
<td>GNSS</td>
<td>80%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Ma et al., 2013)</td>
<td>1 month</td>
<td>128000</td>
<td>CDR and network signalization data</td>
<td>78%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Bahoken and Raimond, 2013)</td>
<td>6 weeks</td>
<td><strong>10 millions</strong></td>
<td>CDR</td>
<td>45%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Larijani et al., 2015)</td>
<td>1 day</td>
<td>1.4 millions</td>
<td>CDR and network signalization data</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Bonnel et al., 2015)</td>
<td>10 days</td>
<td>4,1 millions</td>
<td>network signalization data</td>
<td>82%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Alexander et al., 2015)</td>
<td>2 months</td>
<td>2 million</td>
<td>CDR and network signalization data</td>
<td>65%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Gundlegård et al., 2016)</td>
<td>2 weeks</td>
<td>300000</td>
<td>CDR</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Literature</td>
<td>Number of modes</td>
<td>Data</td>
<td>Duration of test data</td>
<td>Number of users</td>
<td>Accuracy</td>
<td>User validated</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-----------------</td>
<td>-------------------------------------------</td>
<td>-----------------------</td>
<td>-----------------</td>
<td>----------</td>
<td>----------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Reddy et al., 2008)</td>
<td>3</td>
<td>Mobile sensed GNSS, accelerometer</td>
<td>240 min</td>
<td>6</td>
<td>90</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Bohte and Kees, 2009)</td>
<td>4</td>
<td>GNSS, GIS</td>
<td>1 week</td>
<td>1104</td>
<td>70</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Wang et al., 2010)</td>
<td>5</td>
<td>CDR</td>
<td>12 hours</td>
<td>56,715</td>
<td>70</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Reddy et al., 2010)</td>
<td>3</td>
<td>Mobile sensed GNSS, accelerometer</td>
<td>24 hours</td>
<td>16</td>
<td>93</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Manzoni et al., 2010)</td>
<td>7</td>
<td>Mobile sensed GNSS, accelerometer</td>
<td>1 day</td>
<td>4</td>
<td>82</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Hemminki et al., 2013)</td>
<td>4</td>
<td>Mobile sensed accelerometer</td>
<td>150 hours</td>
<td>16</td>
<td>60-85</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Biljecki et al., 2013)</td>
<td>7</td>
<td>GNSS, GIS</td>
<td>1 week</td>
<td>1104</td>
<td>92</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Xiao et al., 2015)</td>
<td>5</td>
<td>GNSS,</td>
<td>5 days</td>
<td>202</td>
<td>86</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Zhou et al., 2016)</td>
<td>3</td>
<td>Mobile sensed GNSS, accelerometer</td>
<td>6 days</td>
<td>12</td>
<td>94</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Semanjski et al., 2017)</td>
<td></td>
<td>Mobile sensed GNSS, GIS</td>
<td>4 months</td>
<td>8000</td>
<td>94</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Literature</td>
<td>Data type</td>
<td>Data Size</td>
<td>Time span</td>
<td>Transport mode</td>
<td>Success rate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td>-----------------------------------------------</td>
<td>------------------------------------------</td>
<td>-----------</td>
<td>----------------</td>
<td>-----------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Marchal et al. 2005)</td>
<td>GNSS, low and high-resolution road data</td>
<td>3 cars (2.5 million points)</td>
<td>3 weeks</td>
<td>Car</td>
<td>Up to 10 m/pt</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Yang et al. 2005)</td>
<td>GNSS, road network</td>
<td>1 car (9500 points)</td>
<td>3 h</td>
<td>Car</td>
<td>100%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Krumm et al. 2007)</td>
<td>GNSS, road network (NAVTEQ)</td>
<td>187 cars (1,351,669 points)</td>
<td>2 weeks</td>
<td>Car</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Chen et al. 2014)</td>
<td>GNSS, road network</td>
<td>10,245 taxis (172,154 points)</td>
<td>15 minutes</td>
<td>Car</td>
<td>90-97 %</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Lou et al. 2009)</td>
<td>synthetic GNSS, GeoLife trajectories, road network</td>
<td>unknown for synthetic data, 28 trajectories for real data</td>
<td>N/A</td>
<td>N/A</td>
<td>Real data: up to 85%, Synthetic data: &gt; 95%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Yuan et al. 2010)</td>
<td>GeoLife trajectories, road network</td>
<td>26 trajectories</td>
<td>30 hours</td>
<td>N/A</td>
<td>66- 84%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Li et al. 2013)</td>
<td>GNSS, OpenStreetMap road network data</td>
<td>121737 trajectories</td>
<td>1 month</td>
<td>Car</td>
<td>85%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
WHERE BIG DATA CAN ASSIST TRANSPORT PLANNING

- Data collection

- Complementing traditional models by:
  - Identifying TAZs properties as population, trip generation/attraction
  - Detecting trip purpose
  - Matching OD pairs
  - Detecting transport mode
  - Reconstructing trip paths….
Sneller in de spits met de fiets
van toepassing in deze zone
Information on popular pedestrian routes and where people like to spend time
TARGET GROUP INSIGHT

- Mobility behavior
  mode, shift, travel time, ...
- Customer metrics
  recency, frequency, duration, ...
- Spatial insight
  O/D, visitor flow, POI, ...

CONSUMER PROFILING

The Seventh International Conference on Data Analytics DATA ANALYTICS 2018 November 18-22, 2018 - Athens, Greece
TTASC: Transport and Traffic Analytics in Smart Cities
Keynote: Ivana Semanjski - Mobility in the age of big data
Thank you for your attention!!!