BIG DATA EUROPE

TRANSPORT PILOT:
INTRODUCING
THESSALONIKI

Josep Maria Salanova Grau CERTH-HIT
Thessaloniki on the map

~ 1.400.000 inhabitants & ~ 1.300.000 daily trips
~450.000 private cars & ~ 20.000 motorcycles
1 (+1) public transport operator for urban trips & 1.950 taxis
~35 public transport operators for extra-urban trips
6.433 kms of streets - 8,8 kms of dedicated bus lanes - 89,4 kms of ring road
197.696 parking places
Probe data in Thessaloniki

- Static sensors network: Point to point tracking of MAC ids along the network through Bluetooth detectors (43 devices).
- Dynamic sensors fleet: Floating Car Data provided in real time by a professional fleet (more than 1,200 vehicles).
- Cooperative technologies (COMPASS4D and COGISTICS): RSU is a static sensor and OBU is a dynamic sensors (CAM message).
- Social media (Twitter & Facebook)
How do we use Probe Data?

- **Stationary sensors network**: Point to point tracking of MAC ids along the network through 43 Bluetooth device detectors.
  - Travel time estimation
  - Route choice model calibration
  - Origin – Destination matrix estimation / **Mobility patterns estimation**
  - Traffic flow extrapolation

- **Dynamic sensors fleet**: Floating Car Data provided in real time by a professional fleets composed of 1,200 taxis and 600 buses
  - Traffic status estimation (average speed)
  - Origin – Destination matrix estimation / **Mobility patterns estimation**
  - Taxi/bus performance indicators

- **Social media** (geolocated tweets & Facebook check-in events)
  - Activity patterns estimation
  - Events / incidents detection
  - Attraction models estimation
Point to point BT network

- 43 detectors (EEA, SEE-ITS & EASYTRIP)
  - 4 million detections per week (peak period)
  - 25,000 unique devices detected per day (one intersection)
  - 1 million “tracked” trips per week
  - 20,000 “tracked” trips per day (one path)
- More detectors installed in other cities and in Bulgaria (SEE-ITS & EASYTRIP)
Point to point BT network

- Real time travel time provision to drivers (VMS, internet, smart device)
Floating Car Data

- More than 1,200 vehicles (one taxi fleet)
  - Circulating 16-24 hours per day
  - Pulse generated each 100 meters (10-12 seconds)
  - 500-2,500 pulses per minute

- 600 vehicles generating CAM each second
Floating Car Data

- Real time traffic conditions information (speed)
Social media

44,000 check-in events per week (750 locations)

Up to
- 50 check-in events per minute (in the 136 locations tagged as bar)
- 17 check-in events per minute (in the 150 locations tagged as restaurant)
- 12 check-in events per minute (in the 32 locations tagged as outdoor)
- 10 check-in events per minute (in the 125 locations tagged as cafe)
- 10 check-in events per minute (in the 55 locations tagged as nightlife)

Up to
- 1,265 check-in events during the “peak hour”
- 920 check-in events in bars (Sunday 01.00)
- 300 check-in events in restaurants (Saturday 22.00)
Social media

BAR

CAFE

NIGHTLIFE
Social media
Social media
Real-time traffic conditions estimation
Processing of big data in Thessaloniki

- Traffic flow estimation from stationary probe data
- Travel time estimation using stationary probe data
- Travel time estimation using floating probe data
- Traffic flow estimation based on travel time
- Short-term traffic flow prediction
- Spatial expansion of traffic flows
- Real-time traffic conditions estimation
BDE pilot in Thessaloniki

- **Probe data that is used**
  - Floating Car Data (500-2,500 locations per minute)
  - Bluetooth detections (millions of daily detections in 43 locations)

- **Services that are being implemented**
  - Improved topology-based map matching
  - Mobility patterns recognition and forecasting
BDE pilot in Thessaloniki

Input: Link data, Node data, Positioning data, Turn restriction data

(a) Initial MM

1. Identify a set of the candidate links
2. Identify the correct link from the set of the candidate links
3. Determine vehicle position on the identified link

(b) MM on a link

- The correct link is the previously selected link
- Determine vehicle position on the identified link

Check whether vehicle speed =0

Yes
- The correct link is the previously selected link
- Determine vehicle position on the identified link

No

Check whether the next positioning fix is near to a junction

Yes
- Identify a set of the candidate links
- Identify the correct link from the set of the candidate links
- Determine vehicle position on the identified link

No

(c) MM at a junction
BDE pilot in Thessaloniki

**Start**

### Historical

- **Historical Link Traffic State BT**
  \[ I_{t}^{bt}(Q_{bt}, V_{bt}) \]
- **Historical Link Traffic State (FCD)**
  \[ I_{t}^{fcd}(Q_{fcd}, V_{fcd}) \]
- **Historical Link Traffic State (Loop Detectors)**
  \[ I_{t}^{l}(Q_{l}, V_{l}) \]

**Historical Link Traffic State Classification**
\[ I_{t}^{l}(Q_{fcd}, V_{fcd}, Q_{bt}, V_{bt}, Q_{l}, V_{l}) \]

**Define Current Link Traffic State**
\[ I_{te}^{l}(Q_{fcd}, V_{fcd}, Q_{bt}, V_{bt}, Q_{l}, V_{l}), t_{e} = t_{n-k}, t_{n-k+1}, \ldots, t_{n} \]

**Compare Traffic States**
(ML, NN)

**Predict**
(ARIMAX | NN)

**Validate Prediction**

**Store in Historical States**
BDE pilot in Thessaloniki

Future plans (next 2 pilots)
- Improve the 2 algorithms (historical data)
- Replace the R components
- Add the BT data source
- Add other data sources (conventional and SM)
- Include more datasets (PuT)
- Use OSM data
- Improve other processes (travel time estimation from BT)
SESSION 2: TECHNICAL REQUIREMENTS AND ADDITIONAL TRANSPORT USE CASES

Josep Maria Salanova Grau CERTH-HIT
Real-time traffic conditions estimation

Diagram:
- **Real world**
  - Traffic flows
  - Filter
  - Travel times

**Off-line processes**
- Historical traffic data
- Macroscopic traffic model
- Traffic scenario

**Real-time processes**
- Traffic flow estimation from SPD
- Short-term traffic flow prediction
- Spatial expansion of traffic flows

**Real-time indicators**
- Travel time estimation
- Real-time traffic conditions estimation
- Traffic flow estimation based on TT
Processing of big data in Thessaloniki

- Traffic flow estimation from stationary probe data
- Travel time estimation using stationary probe data
- Travel time estimation using floating probe data
- Traffic flow estimation based on travel time
- Short-term traffic flow prediction
- Spatial expansion of traffic flows
- Real-time traffic conditions estimation
Traffic flow estimation based on stationary probe data

<table>
<thead>
<tr>
<th>Time interval used for data filtering</th>
<th>Without filtering</th>
<th>5min filter</th>
<th>15min filter</th>
<th>60min filter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation coefficient</td>
<td>0.3412</td>
<td>0.2179</td>
<td>0.1972</td>
<td>0.0442</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.9166</td>
<td>0.9193</td>
<td>0.9337</td>
<td>0.8594</td>
</tr>
<tr>
<td>Largest differences (absolute value and percentage ranges)</td>
<td>-401 / 623</td>
<td>-410 / 437</td>
<td>-336 / 389</td>
<td>-536 / 767</td>
</tr>
<tr>
<td></td>
<td>-26% / 75%</td>
<td>-23% / 61%</td>
<td>-22% / 57%</td>
<td>-35% / 79%</td>
</tr>
</tbody>
</table>

- $y = 0.3682x$
- $R^2 = 0.8947$
Travel time estimation based on stationary probe data
Traffic flow estimation based on travel time

Conversion from route travel time to link travel time

\[
\min \delta_1 \sum (A \cdot x - b) + \delta_2 \sum (x - v0)/v0 \\
\text{s.t.} \quad x_i > t_{0i} \forall i \in I \\
x_j = t_j \forall j \in J
\]
Short-term traffic flow prediction

Linear autoregressive (AR) model

\[ \varphi_{k}^{i+1} = \overline{\varphi} + \sum_{j=1}^{N} \beta_{k}^{j} \cdot (\varphi_{k}^{i+1-j} - \overline{\varphi}) \]  \hspace{1cm} (4)

Spatial expansion of traffic flow

Data Expansion Algorithm (DEA, [Lederman and Wynter 2009])

Optimize \((D - p.* C_{n}) \cdot x - (p.* C_{c} * c)\)  \hspace{1cm} (5)

s.t. \(x_{i} > lb_{i} \forall i \in I\)  \hspace{1cm} (6)
\(x_{i} < ub_{i} \forall i \in I\)  \hspace{1cm} (7)
\(x_{j} = f_{j} \forall j \in J\)  \hspace{1cm} (8)
FCD data and average speeds on the road network
Bluetooth Sensors data and estimated travel times on the road network
BDE Components integration with the legacy system
BDE Components

BDE Cluster

R

R Map Matching

R Classification

Flink – Kafka Consumer 1

Flink – Kafka Consumer 2

ElasticSearch

Kibana

Kafka HTTP Producer

Kafka cluster

Kafka FTP Producer

RDBMS W/T Spatial Support
QUESTIONS

- What are the pros and cons of the technical implementation of the platform offered by BigDataEurope?
- How easy is to implement it to transport use case?
- Lessons learnt from the first pilot implementation?
- How adaptable / usable is it?
QUESTIONS

- Any non-technical barriers to be considered? (legal, open data)
- Does the open data flow initiative pose any threat/opportunity?
- In which transport use case can we reproduce the pilot?
- Which are the characteristics of the transport data that had to be considered in the design of the architecture?
Any non-technical barriers to be considered? (legal, open data)

- Privacy (the driver IDs are modified every 24 hours)
- Data owner is a private entity (we rely on their willing to share the data)
- Updated maps are needed (OSM can be a solution)
- Telecommunication costs (in our case are covered by the private company since is crucial for their professional activity)
QUESTIONS

- Does the open data flow initiative pose any threat/opportunity?
  - ++ data standardization
  - ++ data availability
  - ++ up-to-date datasets
  - -- data quality validation
In which transport use case can we reproduce the pilot?

- In any city having similar data sets
- In other transport modes (PuT)
http://opendata.imet.gr/dataset

**itravel-traveltimes**
Current travel times for selected paths
- JSON
- XML
- CSV

**fcd-compass4d**
Floating car data along 2 arterials (zones)
- JSON
- XML
- CSV
- KML
- MAP

Dr. Josep Maria Salanova Grau

[jose@certh.gr](mailto:jose@certh.gr)  +30 2310 498 433

---

Centre for Research and Technology Hellas (CERTH)**

Hellenic Institute of Transport (HIT)**

6th km, Chaniaou-Thermi Rd.
P.O. Box 60361, 57001 Thessaloniki, Greece
Tel.: +30 2310.498433  •  FAX: +30 2310.498269
E-mail: jose@certh.gr  •  www.hit.certh.gr