Mobile Sensing and Machine Learning for Identifying Driving Safety Profiles
#18-01416

Eleni I. Vlahogianni
Assistant Professor
National Technical University of Athens

Transportation Research Board 97th Annual Meeting, January 7-11, 2018 – Washington, D.C.
Data driven Era in Transportation

Data

Improve the understanding and solutions of existing problems

Create new research questions.
The Problem

How do we improve the system? (safety, mobility, economy)

Monitor, Identify and control driving behavior
The Potential

Smartphone sensors and communications features, allow for continuous, inexpensive, multipurpose and non intrusive data collection.

The data collected can be used for monitoring driving behavior.

Research questions:

*Is it possible to produce a meaningful characterization for every trip in relation to the driving behavior?*

*Is it possible to extend trip characterization to driver characterization and produce a consistent manner to distinguish different drivers on the road?*
Scope

Identify different driving profiles

*Data are being collected a smartphone application*

*Unsupervised learning techniques are applied to identify comprehensive clusters of driving behaviors per trip*

Define an indicator of the volatility of driving behavior for different drivers

*Extend the trip information to meaningful information for the driver*
Methodology: Data collection

Cloud-based service + mobile app
  automatic detection of the end of the trip
  data recording with no user involvement from accelerometer, gyroscope and GPS sensors
  data uploaded and stored in an anonymized way for further processing
  Machine learning and computational intelligence to provide driving analytics

10212 trips made by 129 unique drivers from June 2016 to April 2017
  The data are continuously growing
  External information and smartphone data are also fused
Methodology: Variables

Aggressiveness
Harsh accelerations/brakes per km
   The number of harsh accelerations/brakes per kilometer traveled

Smoothness indicator
   Kinetic energy during acceleration

Standard deviation of acceleration
   The standard deviation of acceleration in each trip

Distraction
Percent of mobile usage
   The percentage of trip duration in which the driver interacts with the mobile phone

Risk taking
Percent of speeding
   The percentage of trip duration in which the vehicle travels over the speed limit
Methodology: Trips’ characterization

Ranked by importance to driving safety:

Safe behavior

Aggressive behavior (harsh accelerate and harsh brake)

Risky behavior (speed limit violations)

Distracted Behavior (mobile phone usage)

Aggressive and Risky behavior

Aggressive and Distracted behavior
Methodology:
Unsafe driving detection procedure

1st stage of clustering

- Aggressive driving
- Non-Aggressive driving

2nd stage of clustering

- Aggressive driving
- Aggressive and risky driving
- Aggressive and distracted driving
- Non-aggressive, safe driving
- Risky driving
- Distracted driving

Processed Sensor Data
Methodology: K-means clustering

Divide a set of samples $N$ into several groups ($k$) so that the samples within the same group are similar by minimizing an objective function:

$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \| x_n - \mu_k \|^2$$

The silhouette index is used to find the optimum value of $k$. 

*it measures how similar an object is to its own cluster (cohesion) compared to other clusters (separation).*
Results: 1st stage clustering

<table>
<thead>
<tr>
<th>Variable/Driving cluster</th>
<th>Harsh Acceleration/km</th>
<th>Harsh Brake/km</th>
<th>Smoothness Indicator</th>
<th>Standard Deviation of Acceleration</th>
<th>Number of trips</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non–aggressive</td>
<td>0.13</td>
<td>0.07</td>
<td>0.32</td>
<td>1.15</td>
<td>7534</td>
</tr>
<tr>
<td>Aggressive</td>
<td>0.72</td>
<td>0.19</td>
<td>0.45</td>
<td>1.62</td>
<td>2678</td>
</tr>
</tbody>
</table>

Some interesting results:
74% of trips are not featured by aggressiveness

Although average acceleration is almost equal, in case of aggressive trips the number of harsh acceleration events is almost 6 times more than in case of non-aggressive trips.
### Results: 2nd stage clustering

<table>
<thead>
<tr>
<th>Variable/Driving cluster</th>
<th>Percent of mobile usage</th>
<th>Percent of speeding</th>
<th>Number of trips</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Aggressive</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aggressive</td>
<td>0.03</td>
<td>0.07</td>
<td>1837</td>
</tr>
<tr>
<td>Distracted</td>
<td>0.44</td>
<td>0.11</td>
<td>226</td>
</tr>
<tr>
<td>Risky</td>
<td>0.02</td>
<td>0.41</td>
<td>615</td>
</tr>
<tr>
<td><strong>Non-aggressive</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Safe</td>
<td>0.02</td>
<td>0.07</td>
<td>5183</td>
</tr>
<tr>
<td>Distracted</td>
<td>0.52</td>
<td>0.06</td>
<td>566</td>
</tr>
<tr>
<td>Risky</td>
<td>0.02</td>
<td>0.29</td>
<td>1785</td>
</tr>
</tbody>
</table>

Some interesting results:
In 69% of aggressive trips, the driver exhibits no other unsafe behavior (risk taking or distracted driving)

In both aggressive and non-aggressive trips only 8% of the trips are characterized as “distracted trips”
Results:
Distribution driving states

- Safe trips (6): 4534
- Aggressive trips (5): 1704
- Risky trips (4): 1617
- Distracted trips (3): 470
- Risky/aggressive trips (2): 548
- Distracted/aggressive trips (1): 200
Results:
Driver’s behavior volatility

Gain/loss ratio per trip is estimated, indicating gain or loss driver receives between successive trips:

$$ r_{t,i} = \ln\left(\frac{S_{t,i}}{S_{t-1,i}}\right) $$

$S_{t,i}$ : driving state of trip $t$

The standard deviation of the gain/loss ratio is the driver’s behavior volatility:

$$ Volatility = \sqrt{\frac{\sum_{t=1}^{n}(r_{t,i} - \bar{r}_{i})^{2}}{n - 1}} $$
Results:
Network Level Impacts (preliminary findings)

HARSH and ECO model perform better than MIX in terms of mobility
- Homogeneity of traffic
- Reduced micro-variations of speed
- Increase of road capacity and reduce of pollutants

Project Luxembourg SUMO Traffic Scenario 2.0 (LuST)
- 24h Simulation
- 5959 edges, 929.5 km
- 4477 junctions
- 203 traffic lights
- 3155 Induction Loops
- Demand Peak: max.5000 vehicles/h
Conclusions

2-stage unsupervised learning models to extract driving profiles from real world naturalistic data

6 driving profiles are identified

Aggressiveness does not mean risk taking or distracted driving

Drivers exhibit risk taking driving behavior, either more often or rarely!

Drivers do not have a stable driving profile, but they change the way they drive on every trip
Future research

Impact of external factors (traffic, road conditions, adverse weather)

Identification of additional unsafe behavior while driving (Inappropriate lane changing, Overtaking, Abnormal steering)

Linking trip characterization to driver characterization

Driving profiles can be used to develop personalized recommendation systems
Mobile Sensing and Machine Learning for Identifying Driving Safety Profiles
#18-01416

Eleni I. Vlahogianni
Assistant Professor
National Technical University of Athens

Lectern Session 841
Machine Learning Methods for Crash Prediction and Safety Analysis
Sponsored by ABJ70