Road operating speed estimation

David Fiedler
Artificial Intelligence Center, Department of Computer Science, Faculty of Electrical Engineering, CTU in Prague, Czech Republic
Outline

- Problem definition
- Applications
- Model Classification
- Model examples
- Obtaining data
- Our Research area
  - Problems
  - SOTA
- Our solution
  - Ideas
  - Preliminary results
Road speed classification

- **Design speed**
  - Determined before the building of the infrastructure
  - Does not change in time

- **Posted speed**
  - Speed limit
  - Limited by the law and safety concerns

- **Operating speed**
  - The real speed of the vehicle on the road
Problem: Estimating the operating speed on the road segment
Estimating speed

- **Design speed**
  - We can look up the designed speed in the project documentation
  - Can be inferred from curvature, lane width, grade
- **Posted speed**
  - Is determined by law and traffic signs
- **Operating speed**
  - ??
  - Can we extract the operating speed from design speed and/or posted speed?
Extracting operating speed

Can be greater than design speed

Can be greater than posted speed

Can be lower than design speed

Can be lower than posted speed
Applications - Routing

We need the road operating speed to find the shortest time path, which is the main task for navigation applications.
Applications - Safety

Car navigation can recommend the operating speed for the upcoming segments to driver.
Applications - Traffic simulations

Traffic simulations need to configure the operating speed in order to deliver high fidelity.

- This was our motivation to accept this challenge
- Currently, posted speed is used as operating speed in simulations
Classification is non-hierarchical, almost all combination exists. Categories:

- per road segment vs parametric
- Free-flow vs congested flow
- Online vs offline
Models - per road segment

We measure the operating speeds on each road segment and build a speed model for each segment

- Google maps, waze
- Usually combined with online data
- Requires a sufficient amount of data for each segment
Models - parametric

We measure the operating speed on part of the segments and build some abstract model on top of that

- Traffic theory
- Can be based on many different parameters: physical, psychological, spatial, temporal
- Does not require data for each road segment

\[
S_f = (25.6 + 0.47S_{pl} + f_{cs} + f_A) \times f_L
\]

\[
V_{85} = 31.565 + 6.491 \text{ lane.num} - 0.101 \text{ roadside} - 0.051 \text{ driveway} - 0.082 \text{ intersection} + 3.01 \text{ curb} - 4.265 \text{ sidewalk} - 3.189 \text{ parking} + 3.312 \text{ land.use1} + 3.273 \text{ land.use2}
\]

\[
S_f = S_{pl} + 5
\]

\[
V_{85} = 51.520 + 1.567 \text{ ST} - 2.795 \text{ MT} - 4.001 \text{ PT} - 2.150 \text{ AG} + 2.221 \ln (LC)
\]
Online vs Offline estimation

Online: we can use current car data
- Uses speed on previous segment to predict current speed
- Speed can be inferred from driver behaviour

Offline: we can use only average car data
- Speeds on consecutive segments do not have to be related
- No individual driver behaviour
Models - free flow

Operating speed is not affected by traffic flow (number of cars using the road network)

- Samples have to be measured during the periods of low traffic flow (at night)
- Same free-flow speed during the whole day
Models - congested flow

- Operating speed is affected by traffic flow - number of cars using the road network
- Has to be sampled with some rate to cover the traffic flow variation
Model examples - Google maps

- Per road segment
- Congested flow
- Online
Model examples - Highway Capacity Manual

- Parametric
- Free-flow
- Offline

\[ S_f = S_0 + f_{CS} + f_A \]

\[ S_0 = 25.6 + 0.47S_{pl} \text{ where } S_{pl} = \text{posted speed limit (mi/h)}. \]

\[ f_{CS} = 1.5 \rho_m - 0.47 \rho_{curb} - 3.7 \rho_{curb} \rho_m \text{ where } \rho_m = \text{proportion of link length with restrictive median (decimal) and } \rho_{curb} = \text{proportion of segment with curb on the right-hand side (decimal)}. \]

\[ f_A = -0.078 \frac{D_a}{N_{th}} \text{ with } D_a = 5,280 \frac{(N_{ap,s} + N_{ap,o})}{(L-W)} \text{ where } D_a = \text{access point density on segment (points/mi); } N_{th} = \text{number of through lanes on the segment in the subject direction of travel (ln); } N_{ap,s} = \text{number of access point approaches on the right side in the subject direction of travel (points); } N_{ap,o} = \text{number of access point approaches on the right side in the opposing direction of travel (points); and } W_i = \text{width of signalized intersection (ft)}. \]
Model examples - Austrian Institute of Technology

- Parametric
- Congested-flow
- Offline

\[ s_{pi} = c f(i) + \beta f(i) m_{xsp} i + v_i \]

\[
\gamma_i = s_{pi} = c_{f(0)} + \gamma_{i(f(0)} + \beta_{f(0)} m_{xsp} i + u_i \quad \text{and (1a)}
\]

\[
\gamma'_i = \frac{s_{pi}}{m_{xsp} i} = c'_{f(0)} + \gamma'_{i(f(0)} + \beta'_{f(0)} m_{xsp} i + u'_i \quad \text{, (1b)}
\]

\[ \gamma_{t,i} = \gamma_{t,j} \text{ for all } i \in F_j \text{ with } \sum_{i=1}^{96} \gamma_{t,j} = 0 , \ j = 1, ..., g , \quad (2) \]

\[ \hat{y}_i(p | q) = \hat{c}_{f(i)|q} + \hat{\gamma}_{t(i),f(i)|q} + \hat{\beta}_{f(i)|q} m_{xsp} i \quad \text{or (3a)} \]

\[ \hat{y}_i(p | q) = \hat{c}_{f(i)|q} + \hat{\gamma}'_{t(i),f(i)|q} + \hat{\beta}_{f(i)|q} m_{xsp} i , \quad (3b) \]

\[ \hat{r}_q = \sum_{i \in r} \hat{u}_{i|q} \hat{v}_{i|q} \]
Obtaining speed data - field measurement

- Small datasets
- Expensive
- Time consuming
- Can bring speed-enforcement bias
Obtaining speed data - traffic infrastructure

- Acceptable time requirement
- Cheap
- Limited to existing infrastructure
Obtaining speed data - GPS traces

- Cheap
- Almost complete datasets
- Requires map matching
Our Research Area

- We can’t gather data for all roads - parametric model
- We have our own congestion model - free-flow model
- We don’t model drivers behaviour - offline model
- GPS traces as the source data
- Also, we focus on urban models, which brings some difficulties
Our Research Area - Urban Models

Many types of roads
Our Research Area - Urban Models

Lot of roads with small utilization
The main problem is interdisciplinarity

**Traffic researchers** does not have the necessary ML background
- Cannot overcome high-dimensionality
- Chooses field measurements over data-mining

**ML researchers** are not interested
- No article in ML conference or journal even in applications (ICMLA)
- Focus on statistical models

SOTA - limitations
SOTA - limitations

- Only linear regression as a ML model
- Mostly, only one feature at a time is fitted
- Models uses features that have to be measured in the field
  - When we measure the feature in the field, we can measure the speed instead and build the statistical model
- Cherry picking of testing sites - only simple standardized segments are included in most studies
- Radar gun measurement
- Models are evaluated on a very limited - statistically insignificant number of roads
Our model - ideas

- Use only parameters that can be obtained or computed from public resources worldwide (length, curvature, road type, elevation...)
- Use machine learning methods to create models that leverage many different parameters
- Choose relevant parameters based on the efficiency of the model
- Compute geometric features from the map and try if they are related to speed
<table>
<thead>
<tr>
<th>Feature</th>
<th>Expected value</th>
<th>Standard deviation</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Curvature</td>
<td>0.14</td>
<td>0.31</td>
<td>&lt;&lt; 0.001</td>
</tr>
<tr>
<td>Max curvature</td>
<td>0.31</td>
<td>0.7</td>
<td>&lt;&lt; 0.001</td>
</tr>
<tr>
<td>Posted speed</td>
<td>48.8</td>
<td>9.17</td>
<td>&lt;&lt; 0.001</td>
</tr>
<tr>
<td>Length</td>
<td>124</td>
<td>162.5</td>
<td>&lt;&lt; 0.001</td>
</tr>
<tr>
<td>Lane count</td>
<td>1.23</td>
<td>0.91</td>
<td>&lt;&lt; 0.001</td>
</tr>
<tr>
<td>Road type</td>
<td>10.16</td>
<td>1.93</td>
<td>&lt;&lt; 0.001</td>
</tr>
</tbody>
</table>
Our model - Baseline

- Polynomial regression for all parameter combinations

$$\min_w ||Xw - y||_2^2$$
Our model - Ridge regression

- Regularized regression for all parameter combinations

\[ \min_w \|Xw - y\|_2^2 + \alpha \|w\|_2^2 \]
Our model - preliminary results

- With only five parameters, we reached average error \(~7.1\text{ km/h}\) compared to \(~23.9\text{ km/h}\) error when using posted speed as free-flow operating speed (current procedure in most simulations).
- The share of explained variance \((R^2)\) is \(0.48\)
Our model - challenges

- We need to come up with more features/parameters that helps to classify streets
  - Each parameter has significant impact to model accuracy - lastly added road type increased the accuracy by 25%
- We have both continuous and discrete parameters
  - So the model should not be pure regression
- High dimensionality prevent us from visualising the results
Thank you!
References

References