Road Danger Prediction - Classic models, AI models and Data Challenges

Presentation by Richard Owen, Craig Smith and George Ursachi, Agilysis, UK

Roundtable on Artificial Intelligence in Road Traffic Crash Prevention
10-12 February 2021
AGENDA

• Classic models
• AI models
• Data requirements
• New developments in AI – Explainable AI
• New datasets and missing datasets
• Pilot RAPIER
CLASSIC MODELS

“PAIR” VARIANCE OF THE DEPENDENT VARIABLE WITH VARIANCE OF THE INPUT VARIABLES

Advantages

• Very strong and straightforward with “perfect” data
• Expose coefficients for each selected variable
• Expose the decision process for variable inclusion/exclusion
• Allow for interrogation

Disadvantages

• What can’t be “paired” is assigned to constant
• Does not always fit best to the actual situation
• Work with user defined limitations and assumptions
• Sometimes assign (pair) effects “wrongly” to other variables (collinear or co-dependent with missing data or information)
Advantages

• Usually exhibit better results than classic models
• Find the best fit, without pre-set limitations
• Allow for more interaction between input variables

Disadvantages

• Do not expose the coefficients or the decisions processes
• More difficult to interrogate and therefore spot warnings
• Also sometimes assign effects “wrongly” to other variables (collinear or co/inter-dependent)
The unpleasant part of using ANN is that you can get more accurate predictions, but you can’t understand why, or which variable is influencing the model in what way.

With GLM on the other hand, individual variables contribution can be assessed from the model.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Density (PC)</th>
<th>Rate (PC)</th>
<th>Density (Ped)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dual Carriageway</td>
<td>-0.049</td>
<td>-0.081</td>
<td>-0.045</td>
</tr>
<tr>
<td>Shared Use Carriageway</td>
<td>-0.081</td>
<td>-0.178</td>
<td>-0.195</td>
</tr>
<tr>
<td>Carriageway Width</td>
<td>+0.200</td>
<td>+0.260</td>
<td>+0.158</td>
</tr>
<tr>
<td>Garden</td>
<td>-0.014</td>
<td>-0.027</td>
<td>-0.026</td>
</tr>
<tr>
<td>20 mph Limit</td>
<td>+0.014</td>
<td>+0.018</td>
<td>+0.021</td>
</tr>
<tr>
<td>Average Speed</td>
<td>+0.160</td>
<td>+0.394</td>
<td>+0.285</td>
</tr>
</tbody>
</table>

***ANN = Artificial Neural Network

***GLM = Generalised Linear Model
DATA REQUIREMENTS

• Relevant
• Reliable
• To cover the complete picture of variables that can exhibit influence on the dependent variable (driver characteristics?)
• Appropriate granularity
• Clean, no systematic errors

• For both classic models and AI the following is valid:

  Garbage in  →  Garbage out

• Classic models allow sometimes to point out that there is a problem, AI might cover it through fitting better to whatever information it has been fed with
DATA REQUIREMENTS

• Until we are sure the data fulfils the mentioned requirements at the minimum acceptable at least, AI should not be seen as a better solution, but better as to be used together with classic models

• Data is still the Holy Grail

• First use of AI → to improve data:
  • Quality
  • Quantity (granularity)
  • Availability

• Then use AI together with classic models to improve predictions
NEW DEVELOPMENTS IN AI

CRAIG SMITH
The limitations of AI and its inherent “black box” nature, paired with the increasing use of AI to make significant real-world decisions, leads to a demand for explainability.

Model-Agnostic Explainable AI (XAI) is becoming more popular and more accessible.

Bridges the gap between the more accurate/flexible AI models and the classical models that are more easily interrogated.
One of the leading techniques is to build local approximations to complex AI models out of simpler, more explainable models.

Can be done by looking at how the AI model reacts to perturbations to the data around specific datapoints, and fitting a local linear model.

Or using game-theoretic Shapley values to assess the local contributions of input variables.

These techniques provide the model “coefficients” that are so valuable in classical models.
• There are also techniques for global explanations of models
• Like building simpler global “surrogate” models
• Or measuring feature importance (how does accuracy change if input variables are dropped)
• Or feature interactions (Friedman’s H-statistics)

• XAI is a fast-growing field, so new techniques are constantly being developed.
NEW DATASETS

More data!
NEW DATASETS - TELEMATICS

• Global Positioning System (GPS)
  • Vehicle location, speed and movement

• Sensors
  • Capture of data on driver activity, including acceleration, braking and cornering, driver monitoring, vehicle features

• Vehicle diagnostics
  • Post-collision information from sensors
NEW DATASETS - TELEMATICS

Speed

<table>
<thead>
<tr>
<th>Average Speed</th>
<th>AM Peak</th>
<th>PM Peak</th>
<th>Off Peak</th>
<th>Evening</th>
<th>Night</th>
<th>Weekend</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>28</td>
<td>25.2</td>
<td>27.1</td>
<td>23.5</td>
<td>25.8</td>
<td>28</td>
</tr>
<tr>
<td>S</td>
<td>26.5</td>
<td>27.5</td>
<td>25.4</td>
<td>23.8</td>
<td>25.5</td>
<td>21.8</td>
</tr>
<tr>
<td>Both</td>
<td>27.2</td>
<td>26.4</td>
<td>26.3</td>
<td>23.7</td>
<td>25.7</td>
<td>24.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>85th Percentile Speed</th>
<th>AM Peak</th>
<th>PM Peak</th>
<th>Off Peak</th>
<th>Evening</th>
<th>Night</th>
<th>Weekend</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>32.2</td>
<td>28.5</td>
<td>32.2</td>
<td>26</td>
<td>NA</td>
<td>31.9</td>
</tr>
<tr>
<td>S</td>
<td>30.9</td>
<td>30.9</td>
<td>30.9</td>
<td>30.4</td>
<td>NA</td>
<td>31.3</td>
</tr>
<tr>
<td>Both</td>
<td>31.6</td>
<td>29.7</td>
<td>31.6</td>
<td>28.2</td>
<td>NA</td>
<td>31.6</td>
</tr>
</tbody>
</table>
NEW DATASETS - TELEMATICS

Harsh Braking

<table>
<thead>
<tr>
<th></th>
<th>Collisions</th>
<th>Near Collisions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Following Distance</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Red Light</td>
<td>0</td>
<td>18</td>
</tr>
<tr>
<td>Drowsy</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Failed to Keep an Out</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
# NEW DATASETS - TELEMATICS

What metrics do we KNOW cause higher risk?
Are these **road** risks or **user** risks?

<table>
<thead>
<tr>
<th>Harsh Braking?</th>
<th>Following Distance?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Swift Acceleration?</td>
<td>Distraction?</td>
</tr>
<tr>
<td>Unusual Cornering Speed?</td>
<td>Drowsiness?</td>
</tr>
</tbody>
</table>
NEW DATASETS - TELEMATICS

Full access to manufacturer vehicle data!
• Speed
• Flow
• Braking
• Acceleration
• Steering input
• Following distance
• Activated safety measures (AEB)
• Post-collision investigation

agilysis.co.uk
PROJECT RAPIER

IMAGE CAPTURE DATA (GEOCODED)

AI image recognition algorithms

GEOCODED ROAD ATTRIBUTES DATA WAREHOUSE

Other datasets

Road risk assessment platform (by road user type)

AI image recognition algorithms

Predictive models

ROAD DANGER PREDICTIONS PLATFORM (BY ROAD USER TYPE)

Scenarios (1, 2,..., n) Actionable solutions

Predictive models

Predicted outcome 1

Predicted outcome 2

Predicted outcome n
https://s3-eu-west-1.amazonaws.com/agilysis.media/video/Rapier/VIRB0066_annotated.mp4