Claims Frequency Modeling using Telematics Car Driving Data

Mario V. Wüthrich
RiskLab, ETH Zurich

joint work with
Guangyuan Gao & Shengwang Meng
Renmin University, Beijing

Insurance Data Science Conference
London, July 16, 2018
Telematics Car Driving Data

driver 21, trip number 1

time in seconds
speed / angle / acceleration
acceleration / change in direction / speed

driver 20, trip number 7

time in seconds
speed / angle / acceleration
acceleration / change in direction / speed
Available Car Driving Data

Find structure (driving styles) in features

\[ \{x_1, \ldots, x_n\} \subset \mathcal{X}, \]

of \( n \) insurance policies in a given feature space \( \mathcal{X} \).

- **Data**: 12'076 drivers with
  
  - classical features like age, gender, type of car, prize of car, etc.,
  - telematics data of all trips including GPS location (sec by sec), time stamp, speed, acceleration (in all directions), engine revolutions per minute,
  - claims data,

  from 2014-2017 (1GB per day, 1.5TB in total).
Two Different Approaches for Driving Styles

• Score individual trips.

• Build summary statistics per driver (law of large numbers) and score those.
Normalized $v$-$a$ Heatmaps

- Calculate $v$-$a$ heatmap of all trips in speed bucket $[5, 20)$km/h for all $n$ drivers.
- These heatmaps measure the amount of time spent in a $(v, a)$ location.
- Normalization gives (discrete) probability distributions $x_i$ for drivers $i = 1, \ldots, n$.

$v$-$a$ heatmaps of drivers $i = 3, 44, 1001$ in speed bucket $[5, 20)$km/h.
Autoencoders for Data Compression

• **Encoder:**
  \[ \varphi : X \rightarrow Z, \]
  where \( Z \) is low-dimensional.

• **Decoder:**
  \[ \psi : Z \rightarrow X. \]

• **Goal:** Choose functions \( \varphi \) and \( \psi \) such that
  \[ \pi(x) = \psi \circ \varphi(x) \]
  is close to input \( x \).

\( \varphi(x) \in Z \) is used as low-dimensional representation for \( x \in X \).
Principal Component Analysis (PCA)

- Consider the design matrix $X = (x_1', \ldots, x_n')' \in \mathbb{R}^{n \times d}$ of rank $d \leq n$.

- Singular value decomposition (SVD) provides (an) optimal approximation $X_q$ of design matrix $X$ of (smaller) rank $q \leq d$ (in Frobenius norm).

SVD result of driver $i = 3$ for ranks $q = 1, 2$ (true heatmap on the left).
• Calibrate bottleneck neural network such that inputs $x_i$ and outputs $\pi_i = \pi(x_i)$ are close in Kullback-Leibler (KL) divergence

$$\mathcal{L}_{KL} ((x_i)_i, (\pi_i)_i) = \frac{1}{n} \sum_{i=1}^{n} d_{KL}(x_i \| \pi_i).$$

• Signals at the bottleneck are the $Z$-representations of drivers $i = 1, \ldots, n$. 
SVD vs. Bottleneck Network for $q = 2$

KL divergence of the SVD with $q=2$

KL divergence deep net with $(p,q,p)=(7,2,7)$

KL divergences of SVD and the bottleneck neural network

$(\text{drivers } i = 3, 44, 300, 1001; 642, 1645)$. 
• Predictive Power of $\nu - \alpha$ Heatmaps?
Poisson GAM Regression Models

Assume for $i = 1, \ldots, n$

$$Y_i \overset{\text{ind.}}{\sim} \text{Poi} (\lambda(x_i)v_i),$$

with exposures $v_i > 0$ and regression function $\lambda : \mathcal{X} \to \mathbb{R}_+$ given by

Model 0: $\log \lambda(x) = \beta_0 + s_1(\text{age driver}) + \beta_2 \cdot \text{age car},$

Model 1: $\log \lambda(x) = \beta_0 + s_1(\text{age driver}) + \beta_2 \cdot \text{age car} + \beta_3 \cdot \text{PCA(heatmap)},$

Model 2: $\log \lambda(x) = \beta_0 + s_1(\text{age driver}) + \beta_2 \cdot \text{age car} + \beta_3 \cdot \text{BN(heatmap)}.$

<table>
<thead>
<tr>
<th>Model</th>
<th>Cross-validation out-of-sample loss</th>
<th>Std. dev. error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 0 (GAM classic)</td>
<td>1.4806</td>
<td>0.0240</td>
</tr>
<tr>
<td>Model 1 (PCA)</td>
<td>1.4573</td>
<td>0.0266</td>
</tr>
<tr>
<td>Model 2 (bottleneck net)</td>
<td>1.4579</td>
<td>0.0232</td>
</tr>
</tbody>
</table>
Conclusions

- $v-a$ heatmaps allow for low-dimensional representations and approximations.
- Do these heatmaps have predictive power? Preliminary analysis shows “yes”!
- We have central limit theorems and rate of convergence for $v-a$ heatmaps.
- Other speed buckets and claim sizes?