Machine Learning
Applied to Insurance Problems

Mario V. Wüthrich
RiskLab, ETH Zurich

Artificial Intelligence in Insurance and Finance
Zurich University of Applied Sciences ZHAW
Winterthur, September 7, 2017
Insurance Technology (InsurTech)\textsuperscript{1} \textsuperscript{2}

- **Artificial Intelligence**: learn and accumulate useful knowledge through data.

- **Big Data Analytics** based on the research of massive data.

- **Cloud Computing** for real-time operations.

- **Block Chain Technology** for a more efficient and anonymous exchange of data (decentralization of information, insured-insurer/insurer-insurer).

- **Internet of Things**: integration and interaction of physical devices (e.g. wearables, sensors) through computer systems to reduce and manage risk.

\textsuperscript{1}innovation and application of advanced technology to the insurance industry

\textsuperscript{2}source: China InsurTech Development White Paper, 2017
Near Future InsurTech Applications

- health and accident insurance
- life and pension insurance
- non-life insurance
- re-insurance

▷ product development and insurance pricing
▷ marketing and distribution of products
▷ administration, accounting and portfolio optimization
▷ claims handling, triage of potentially large claims and claims reserving
▷ risk drivers analysis and risk management
• Section 1: Supervised Learning
Insurance Pricing: State-of-the-Art

Basic Assumption:
There are structural differences which can be explained by a regression function

\[ \mu : X \rightarrow \mathbb{R}, \quad x \mapsto \mu(x). \]

- \( X \) is the feature space, covariate space (containing all potential insurance policies);
- \( x \in X \) is the feature, covariate, explanatory variable of a single policy;
- \( \mu(\cdot) \) is the regression function/pricing functional describing structural differences.

Example of feature:

\[ x = (\text{age, gender, type of car, claims code, income, body weight, smoker, ...}) \]
Supervised Learning

Determine the (unknown) regression function (pricing functional)

\[ \mu : \mathcal{X} \to \mathbb{R}, \quad x \mapsto \mu(x) \]

from (noisy) observations \( \mathcal{D} = \{(Y_1, x_1), \ldots, (Y_n, x_n)\} \) being of type

\[ Y_i = \mu(x_i) + \varepsilon_i \quad \text{with}^3 \quad \mathbb{E}[\varepsilon_i] = 0. \]

This problem is not new in insurance, but \( \mathcal{X} \) is increasingly more complex.

---

^3 Typically, we also assume independence between the \( \varepsilon_i \) for policies \( i = 1, \ldots, n \).
Classical and New Solutions to Pricing Problems

- Classical approaches use generalized linear models (GLMs) and generalized additive models (GAMs), implemented in commercial software like Emblem or SAS.

- New approaches use neural networks, regression trees, random forests, boosting.

⇒ These are still at an experimental stage in the insurance industry:

- Issue 1: standard software does not exist, yet (deviance statistics, big data);
- Issue 2: data warehouses may still bear some issues;
- Issue 3: stability/robustness over time still needs to be explored;
- Issue 4: graphical illustrations are still poor;
- Issue 5: communication to management and customers is still difficult;
- Issue 6: uncertainty measures are completely missing.

We give some examples illustrating the state-of-the-art.
Neural networks are used for sensitivity testing and analysis of non-linearities.

4more than 10 mio. accident claims over 12 accounting years
Prediction Uncertainties in Neural Networks

Process uncertainty from $Y_i = \mu(x_i) + \varepsilon_i$ and parameter uncertainty from

$$
\log \mu(x) = \beta_0 + \sum_{l=1}^{q_2} \beta_i \phi \left( \sum_{k=1}^{q_1} w_{l,k}^{(2)} \phi \left( \sum_{j=1}^{d} w_{k,j}^{(1)} x_j \right) \right).
$$

---

5 deep neural network with two hidden layers and 132 parameters
Regression Trees\textsuperscript{6} for Risk Driver Analysis\textsuperscript{7}

<table>
<thead>
<tr>
<th>time lags $t + 1$:</th>
<th>$Y_1$</th>
<th>$Y_2$</th>
<th>$Y_3$</th>
<th>$Y_4$</th>
<th>$Y_5$</th>
<th>$Y_6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>numbers of leaves</td>
<td>8</td>
<td>11</td>
<td>18</td>
<td>12</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>feature components used for split questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>claim closed</td>
</tr>
<tr>
<td>lawyer involved</td>
</tr>
<tr>
<td>claims code</td>
</tr>
<tr>
<td>claims diagnosis</td>
</tr>
<tr>
<td>reporting delay</td>
</tr>
<tr>
<td>previous payments</td>
</tr>
</tbody>
</table>

Payment process $(Y_t)_{t \geq 0}$: relevant feature information and Markov condition.

\textsuperscript{6}CART, Breiman, Friedman, Olshen and Stone (1984)
\textsuperscript{7}insurance claims cash flows of 90k claims
Stage-wise adaptive learning applied to residuals is known as a Boosting Machine.

8Swiss female mortality data: back-testing the Lee-Carter model
• Section 2: Unsupervised Learning
Feature Extraction: Find patterns in (noisy) features $\mathcal{X}_0 = \{x_1, \ldots, x_n\}$.

GPS location data second by second: 200 individual trips of 3 car drivers.
Normalized $v-a$ Heatmaps

**Unsupervised Learning:** Find patterns in (noisy) features $\mathcal{X}_0 = \{x_1, \ldots, x_n\}$.

Normalized $v-a$ heatmaps of 8 car drivers in speed bucket [5km/h, 20km/h).
Application of $K$-Means Algorithm for $K = 4$
Conclusions should be here ...

... and your remarks!

▷ Lecture notes on SSRN preprint server (first draft):

Data Analytics for Non-Life Insurance Pricing, Manuscript ID 2870308.