Recent Achievements and Perspectives in Actuarial Data Science

Risk Day, ETH Zurich
13th September 2019

Dr. Jürg Schelldorfer, Actuary SAA

Senior Analytics Professional, Swiss Re
Chair of the «Data Science» working group of the Swiss Association of Actuaries (SAA)
Disclaimer

The opinions expressed in this presentation are those of the author only. They are inspired by the work that the author is doing for both Swiss Re and the SAA, but they do not necessarily reflect any official view of either Swiss Re or the SAA.
Machine Learning in the insurance industry

Dr. Tobias Büttner, Head of Claims, Munich Re, mentioned the following:\[1\]:

Property claims were assessed using images.

But later the reserves had to be increased significantly. **Damages below/hidden in the roofs have not been appropriately estimated.**

Implications of the use of Machine Learning (ML) in insurance:

- **ML can affects operations**, which impact the data actuaries use (i.e. claims, underwritten risks,...)
- **ML can affect** the underlying risks
- ML can be used to strengthen the core skills
- **Automation** (not necessarily ML) can help to improve efficiency

---

1 SZ-Fachkonferenz: KI und Data Analytics in der Versicherungsbranche; Data Analytics im Management von Großschäden, Büttner T. (2019), Munich Re
Table of Content

1. Recent achievements
2. Perspectives
3. Non-quantitative aspects
4. Summary
Recent achievements
1 – Factor embeddings in neural networks

- In insurance pricing, factor variables (i.e. vehicle brand, region,...) consist of many levels and are often encoded as dummy variables (or one-hot encoding), i.e. the levels are orthogonal in the feature space.

- With neural networks, we use (factor) embeddings which make the fitting of neural networks with many factors and levels feasible and natural.

2 – Combined Actuarial Neural Networks (CANN)\(^1\)

Advantages:
- Extension of GLM
- GLM as starting point for optimization
- Enables uncertainty quantification

- Linear predictor with regression parameter \(\beta = (\beta_0, \ldots, \beta_q)^T \in \mathbb{R}^{q+1}\)
  \[x \in \mathcal{X} \subset \mathbb{R}^q \mapsto \theta_{GLM}(x; \beta) = \langle \beta, \bar{x} \rangle \in \Theta.\]

  ▶ Feature pre-processing is done by the actuary/statistician.

- Choose network of depth \(d \in \mathbb{N}\) with network parameter \(w = (W_{1:d}, w_{d+1})\)
  \[z \in \mathbb{R}^{qd} \mapsto \theta_{NN}(z; w_{d+1}) = \langle w_{d+1}, \bar{z} \rangle \in \Theta,\]
  with neural network function (feature pre-processing \(x \mapsto z\))

- Choose linear predictor with parameter \((\beta, w) = (\beta, W_{1:d}, w_{d+1})\)
  \[x \mapsto \theta_{CANN}(x; \beta, w) = \langle \beta, \bar{x} \rangle + \left\langle w_{d+1}, \bar{z}^{(d+1)}(x) \right\rangle.\]

\(^1\) Paper: \[\text{https://doi.org/10.1017/asb.2018.42}\]
3 – Portfolio bias in neural networks

GLM provide unbiased estimates on a portfolio level, and the GLM provides exactly the same unbiased estimated portfolio average as the homogeneous model.

Due to early stopping in neural networks model calibration, the model has a bias on the portfolio level!

Extract from an example for claims frequencies:

Remedies are proposed in the corresponding papers.
4 – Random Forest and Boosting

Achievements:

- rfCountData: random forest for Poisson distribution (GitHub)
- Review of most relevant boosting algorithm (AdaBoost, LogitBoost, XGBoost)

Perspectives:

- Usage of random forest for claims severities (L2 is not a good loss function) and total loss amounts?
- rfSeverityData package?
- How to make random forest (better) interpretable?
- Are random forest / boosting appropriate for uncertainty quantification?

Perspectives
We have written the following six tutorials:

2. Insights from Inside Neural Networks: Guidance how to fit neural networks for insurance data
3. Nesting Classical Actuarial Models into Neural Networks: Embedding of GLM’s into neural networks
5. Unsupervised Learning: What is a Sports Car?: Unsupervised learning techniques applied in P&C

We are working on the following:

- Natural Language Processing and RNN’s
- Segmentation using decision trees
- Mortality forecasting, Part II

Further topics and ideas:

- Missing data and data imputation
- Dissimilarity measures for categorical variables
- Convolutional Neural Networks and images
- Explainability / Interpretability of machine learning models
- Graphical Models / Causality?
- GAN?
- Performance measures and visualizations?
- Spatial modeling and random (Gaussian) fields?
Selected L&H business application
(including my personal biases)

Network-based approach to medical health used for underwriting$^{1,2}$.

Individual-based Mortality Forecasting$^3$

1. Swiss Re, Understanding medical risk: a network-based approach (Link)
2. SZ-Fachkonferenz: KI und Data Analytics in der Versicherungsbranche; Expore your health, schnell und smart durch die Gesundheitsfragen, Dannenberg T. (2019), RISK-CONSULTING Prof. Dr. Weyer GmbH
3. Euroforum: Rethinking Insurance; Big Data – Mehrwerte durch Data Analytics generieren, Caro G. (2019), Swiss Re
Selected P&C business application
(including my personal biases)

2. Driving data for automobile insurance: will telematics change ratemaking?, Monserrat Guillén, SAV Mitgliederversammlung 2019, Lucerne

Behavioural and situational data for the vessels in marine insurance\(^1\).

Satellite imagineries in agroculture insurance\(^1\).

There is a move from..
• ...pure claims modeling to..
• ...claims + behavioural + lapse modelling\(^2\)

2. Driving data for automobile insurance: will telematics change ratemaking?, Monserrat Guillén, SAV Mitgliederversammlung 2019, Lucerne
Non-quantitative aspects
Model Risk Management


• «...that model governance techniques and frameworks that exist today do not need to be fundamentally altered, but can be enhanced and adjusted to meet the evolving needs of complex tools and machine learning developments»

• Model Management Framework

• Ethical Framework
Ethics and Company-internal Training Guide

Ethics:
Publications on Ethics in ML/AI applications:
- Ethical Codex for Data-Based Value Creation, Swiss Alliance for Data-Intensive Services, 2019
- Ethics Guidelines for Trustworthy AI, European Commission, 2019
- Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore’s financial industry, MAS, 2019
- Ethically Aligned Design, IEEE, 2019

These papers raise some questions w.r.t. to the role and responsibility of the actuaries:
- Should an actuary be fulfilling the relevant ethical codex for a Data Scientist? Or is he already doing it?
- What should be expected from an actuary w.r.t. to ethics?

Company-internal Training Guide
For already fully qualified actuaries in industry (demand from smaller actuarial associations and companies) we have summarized the topics to start with ADS...
Summary
Conclusions

• Statistical learning methods and neural networks allow to fit dependency structures naturally beyond the (currently used) GLM.

• CANN provide the framework for extending the GLM’s, allowing to improve the accuracy of the model as well as providing a framework to assess the uncertainties.

• Model risk management needs to be addressed carefully for machine learning models

• There are many business challenges ahead which require machine learning skills.

And yet, a very well calibrated GLM may still be as good as an advanced machine learning model in terms of accuracy.
Visit

www.actuarialdatascience.org

Article, data and code of the tutorials
References to literature
Acknowledgements

People:
• All members of the SAA working group
• Dr. Alexander Noll
• Dr. Simon Renzmann
• Ron Richman

Institutions:
• Swiss Association of Actuaries (SAA)
• RiskLab at ETH Zurich
• MobiLab for Analytics at ETH Zurich

Companies:
• Swiss Re
References

- [www.actuarialdatascience.org](http://www.actuarialdatascience.org)
- *Nesting Classical Actuarial Models into Neural Networks*, Schelldorfer J. and Wüthrich M.V. (2019), SAA
- *Editorial: Yes, we CANN!*, Wüthrich, M.V., Merz, M. (2019). ASTIN Bulletin 49/1
- [rfCountData](http://rfCountData), Pechon F. (2018), GitHub
- SZ-Fachkonferenz: KI und Data Analytics in der Versicherungsbranche; Data Analytics im Management von Großschäden, Büttner T. (2019), Munich Re
- SZ-Fachkonferenz: KI und Data Analytics in der Versicherungsbranche; Expore your health, schnell und smart durch die Gesundheitsfragen, Dannenberg T. (2019), RISK-CONSULTING Prof. Dr. Weyer GmbH
- Euroforum: Rethinking Insurance; Big Data – Mehrwerte durch Data Analytics generieren, Caro G. (2019), Swiss Re
- *Driving data for automobile insurance: will telematics change ratemaking?*, Monserrat Guillen (2019)
Appendix
ADS basics: Articles and repositories

The following articles/repositories are fundamental for entering the topic of Actuarial Data Science (ADS):

- **Data Analytics for Non-Life Insurance Pricing**, ETH Zurich, M.V. Wüthrich and C. Buser
- **AI in Actuarial Science**, R. Richman, SSRN, 2018
- **ADS Tutorials**, SAA, 2018-present
- **Insurance Analytics – A Primer**, International Summer School of the Swiss Association of Actuaries, 2018
- **Insurance Data Science: Use and Value of Unusual Data**, International Summer School of the Swiss Association of Actuaries, 2019

And do not forget the fundamentals of Statistics vs. Machine Learning:

- **To explain or to Predict?**, G. Shmueli, Statistical Science 25/3, 289-310, 2010
## ADS basics: R packages\(^1,2\)

<table>
<thead>
<tr>
<th>ML meta packages:</th>
<th>Neural Networks:</th>
<th>Machine/Statistical Learning excl. NN</th>
</tr>
</thead>
<tbody>
<tr>
<td>caret</td>
<td>keras</td>
<td>rpart</td>
</tr>
<tr>
<td>mlr</td>
<td></td>
<td>ranger, randomForest, rfCountData</td>
</tr>
<tr>
<td></td>
<td></td>
<td>xgboost, gbm</td>
</tr>
<tr>
<td></td>
<td></td>
<td>cluster, clusterR, tsne, umap, kohonen</td>
</tr>
<tr>
<td></td>
<td></td>
<td>glmnet</td>
</tr>
<tr>
<td><strong>data:</strong></td>
<td><strong>Visualisations:</strong></td>
<td></td>
</tr>
<tr>
<td>tidyverse</td>
<td>ggplot2</td>
<td></td>
</tr>
<tr>
<td>data.table</td>
<td>DataExplorer</td>
<td></td>
</tr>
<tr>
<td></td>
<td>esquisse</td>
<td></td>
</tr>
<tr>
<td><strong>Insurance data:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CASdatasets</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Interpretability:</th>
<th>Others:</th>
</tr>
</thead>
<tbody>
<tr>
<td>iml</td>
<td>Rmarkdown</td>
</tr>
<tr>
<td></td>
<td>Rshiny</td>
</tr>
</tbody>
</table>

---


\(^2\) CRAN Task View: Machine Learning & Statistical Learning, T. Hothorn, 2019
Outlook and Call for Action

**Outlook Working Group:**
- Additional tutorials
- Offering an ADS block course
- Dedicated SAA working group on ethics and providing a structure for Data Scientists to become member of the SAA.

**Call for Action:**
- Insurance analytics and actuarial data science should be strengthened at actuarial education and research institutions.
- Foster research and developments in actuarial data science between companies and universities.
- Synthetic data generation (Simulation Machine, GAN?,...) techniques to allow collaborations with research institutions and actuarial associations.
- How to generate publicly available and yet well calibrated actuarial data sets for machine learning?