

DAV/DGVFM Herbsttagung – ADS

ML, LLMs and Statistics in migration

Axel Helmert (msg life), 21. November 2023, Hannover



Sam Altman, CEO? von OpenAI

The reason to develop AI at all, in terms of impact on our lives and improving our lives and upside, this will be the greatest technology humanity has yet developed.

Agenda

1	Abstract	2
2	Rahmenbedingungen	3
3	Zielbild	4
4	Modelle	5
4.1	Blackbox-Modelle	5
4.2	Whitebox-Modelle	6
4.3	Einsatz von LLMs	7
5	Source Code	9
6	Demo	10
7	Umfangreiche maschinelle Tests	11
8	Qualitätssteigerung und Effizienz	12
9	Some Math	13
9.1	Using basic actuarial knowledge	13
9.2	Multiple linear regression	15
9.3	Symbolic regression and genetic programming	16
10	Resume, Discussion, Questions	17

1 Abstract

Bestandsmigrationen sind herausfordernd - als Bedingung für moderne und effiziente Methoden bei Entwicklung, Verkauf und Verwaltung von Versicherungsverträgen aber unabdingbar.

Der Vortrag zeigt am Beispiel der LV wie mit innovativen KI-basierten Ansätzen, durch Kombination mit allgemeinen mathematischen- und aktuariellen Lehrbuchmethoden erklär- und anpassbare aktuarielle Funktionen maschinell aus dem Quellsystem erzeugt, getestet und ins Zielsystem zur operativen Nutzung transferiert werden können.

Der zusätzliche Einsatz von LLMs (Transformer) verleiht dem Verfahren Sicherheit und bietet eine Verbesserung der Qualität der Abbildung verglichen mit dem herkömmlichen Ansatz.

Das Vorgehen wird anhand einer kleinen Demo auch live gezeigt.

Modifikation: Aus technischen Gründen ist eine live Demo bei der Tagung leider nicht möglich. Ich füge ein paar Screen-Shots ein. Über [diesen Link](#) kann auch ein kurzes Video abgespielt werden.

2 Rahmenbedingungen

- Im ersten Schritt wird der Transfer des Wissens (also der Funktionalität des Quellsystems) bearbeitet.
- Dies ist im Durchschnitt der Migrationen der langwierigste und teuerste Aspekt ($\sim 35\%$).
- Das Vorhaben läuft im Rahmen eines Forschungsprojektes (Förderung vom Freistaat Bayern!), gemeinsam mit der LMU und dem Munich Center for ML (Hr. Prof. Dr. Bernd Bischl).
- Eine Anmeldung beim europäischen Patentamt ist unterwegs.
- Der Transfer geht einerseits von den tatsächlich durchgeführten Verarbeitungen im Quellsystem aus.
- Es muss also auf den technischen Vertragsdaten aufgesetzt werden.
- Diese bilden die Realität der Verarbeitungen und zugleich die vertrags- und steuerrechtlichen Anforderungen ab.
- Andererseits muss die aufsichtsrechtliche Sicht berücksichtigt werden. Diese findet sich (hoffentlich) in den entsprechenden Dokumenten des Quellsystems (Geschäftspläne etc.).
- Ziel ist es, beide Sichtweisen möglichst umfassend maschinell zu unterstützen (Reduktion von Aufwand und Zeit und Entlastung der knappen Ressource Aktuar),
- Frühzeitiger maschineller Abgleich beider Sichtweisen (zusätzliche Steigerung der Qualität).

4 Modelle

4.1 Einsatz von Neural Networks und XAI

- Ein mathematisch und technisch aufwendiger, aber allgemeiner Ansatz ergibt sich durch die Anwendung von Neural Networks.
- Die Fragen nach der Erklärbarkeit können dabei durch XAI und durch Verweis auf die Quelldokumente beantwortet werden.
- Zusätzlich müssen umfangreiche Tests durchgeführt werden.
- Das Verfahren konnte bereits erfolgreich produktiv eingesetzt werden, ist aber aufgrund der hohen aufsichtsrechtlichen Anforderungen nach dem derzeitigen Stand der Technik aufwendig.
- Ist die Komplexität der funktionalen Abbildung niedrig können auch Neural Trees eingesetzt werden.

4.2 Einsatz von erklärbaren Whitebox-Modellen

- Anforderungen, die eine möglichst gute Erklärbarkeit der generierten Formeln nahe legen:
 1. Qualitätssicherung / Überprüfbarkeit,
 2. Regulatorik,
 3. Anpass- und Erweiterbarkeit,
 4. Redundanzfreiheit.
- Idealerweise nutzen die generierten Modelle Begriffe und Semantik aus dem jeweiligen Domänenbereich,
- Dies hat zu einer vollständigen Überarbeitung der bisherigen Ansätze geführt.
- Die Anwendung von Neural Networks bleibt als Option erhalten.
- Wird aber nur eingesetzt, wenn keine sinnvollen fachlichen Annahmen über die Domäne formuliert werden können.

4.3 Einsatz von LLMs

- Durch den Einsatz von LLMs (Transformer) können Formeln:
 1. maschinell ausgelesen
 2. und fachlich zugeordnet werden.
- Aus Dokumenten und / oder Source Code.
- Damit kann ein maschineller Vergleich vorgenommen werden:
 1. Mit bereits implementierten Formeln (Vermeidung von Redundanz)
 2. Mit den aus Daten erzeugten Formeln → Steigerung der Qualität
- Nach diesem Schritt werden die Generierung des Sourcecodes und der Import der Produktdaten angestoßen.

- Als Domänenwissen werden zahlreiche aktuarielle Grundfunktionen (Lehrbuchwissen) und aktuarielle Prinzipien (Äquivalenzprinzip, Thiele'sche Übergangsgleichung, ...) verwendet, die keinerlei VU-spezifische Annahmen über das konkrete Quellsystem enthalten.
- Im Verlauf der Migration können diese erweitert werden, wenn dies zu einer besseren Abbildung führt.
- Die generierten Formeln verwenden nur diese bekannten Formeln und übliche mathematische Operatoren.
- Im Zuge der Optimierung werden intern ggf. komplexe statistische Methoden angewendet:
 1. multiple linear regression,
 2. symbolic regression,
 3. genetic programming,
 4. ...
- Die Lesbarkeit vereinfacht den nächsten Prozessschritt (Generierung von Source Code für das Zielsystem) und
- Die Anwendung von LLMs um weitere Anforderungen umsetzen zu können.

5 Generierung von Source Code

- Die aus den Daten generierten Formeln werden in einem neutralen lesbaren Format im Knowledge-Hub gespeichert.
- Damit ist dieser Schritt unabhängig von der Technologie des Zielsystems.
- Die KI operiert auf Daten und standardisierten Dokumentformaten (PDF, Word, ...). Sie ist damit auch weitgehend unabhängig von der Technologie der Quelle.
- Der Knowledge-Hub kann als Wissensmanagement-Tool dem Zielsystem zur Verfügung gestellt werden.
- Aus den lesbaren Formeln werden dann Source-Code und Dokumentation für das Zielsystem generiert.

6 Demo: Generate formulas from data

Generierung von aktuariellen Formeln aus den Daten des Quellsystems mit Hilfe von Standard-Domänenwissen.
 Tarif: Aufgeschobene Leibrente mit Garantiezeit g und Beitragsrückgewähr

Brutto-Barwerte (inkl. Kosten)	
Leibrente mit Garantie	$bpv_{annuity} = (1 + \gamma_4) \cdot {}_n\ddot{a}_{x::g}$ $1 + \gamma_4 = 1.01$
Beitrag	$ppv = (1 - \beta) \cdot \ddot{a}_{x:t}$ $1 - \beta = 0.965$
Brutto-Tarifbeitrag (inkl. Kosten)	
Tarifbeitrag	$tba = \frac{b_1 \cdot (bpv_{annuity} + bpv_{cost}) + bpv_{death} - reserve}{ppv - ppv_{psum}}$
Brutto - Tarifreserve (inkl. Kosten) und Rückkaufswert	
Tarifreserve	$reserve = b_1 \cdot (bpv_{annuity} + bpv_{cost}) + b_2 \cdot bpv_{death} - tba \cdot (ppv - ppv_{psum})$
Rückkaufswert (Kapital-VS)	$rkw = \text{Max}(0, \text{Min}(VS \cdot {}_mV_{x:n}, -0.02 \cdot ES + 1.02 \cdot VS \cdot {}_mV_{x:n}))$

b_1 = vorschüssige Jahresrente, b_2 = Beitragssumme, g = Garantiezeit, n = Aufschubzeit, t = Beitragszahlungsdauer.

7 Umfangreiche maschinelle Tests

- Sofern Generierung des Source Codes und maschineller Abgleich erfolgreich waren, werden für den erzeugten Code umfangreiche maschinelle Tests angestoßen.
- Die Verträge werden z.B. bis zum Ablauf planmäßig entwickelt.
- Zusätzlich können technische Änderungen getestet werden.
- In Abhängigkeit von der Verfügbarkeit von Referenzwerten, werden die Reserveverläufe und die Entwicklung der Garantiewerte vollständig getestet.

8 Maßnahmen zur Steigerung von Qualität und Effizienz

- Die Möglichkeit beide Methoden automatisiert parallel anzustoßen schafft neben der Effizienz auch mehr Qualität.
- Alle Abweichungen werden zum Zeitpunkt der Migration gefunden.
- Dies spart erhebliche Aufwände in der Zukunft.
- Der Einsatz beider Methoden setzt keine vollständige Transformation der Daten voraus. Sie können zu einem frühen Zeitpunkt eingesetzt werden.
- Damit kann ggf. frühzeitig eine Abschätzung über die Qualität des Bestandes und die Kosten der Migration unterstützt werden.

9 Some Math

9.1 Using basic actuarial knowledge

9.1.1 Equivalence Principle

- One special feature of this new method is the use of domain-specific but standardized functionality.
- In life insurance, there is a general principle for many actuarial calculations - the **equivalence principle**.
- Applied to the calculation of reserves (constant premium P and a constant sum insured B):

$$res(x) = B \cdot pvb(x) - P \cdot pvp(x)$$

9.1.2 Elementary Present Values

- In life insurance, a large number of actuarial functions can be represented as a function of standardized elementary present values.
- Standardized here means that it is textbook knowledge. Company-specific differences are not reflected in these functions, but in the combination of them.
- By making these elementary present values available to the algorithm as building blocks, we can significantly reduce complexity.
- Non-linear behavior of the function being searched is often, but not always, already contained in these elementary present values.
- In many situations the function can be represented as a linear combinations of these elementary present values.
- This leads us to the following definition:

Definition 1: Let $k \in \mathbb{N}$. With $Z = \{z_1, z_2, \dots, z_k\}$ we denote the set of elementary present values that can be used as building blocks in the algorithm.

9.2 Multiple linear regression

- Suppose that $n > k$ observations are available, and let y_i denote the i_{th} observed response and x_{ij} denote the i_{th} observation of regressor x_j .
- We may write the sample regression model as:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \epsilon_i \quad (1)$$

$$= \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + \epsilon_i \quad (2)$$

- In the situation discussed here, n is the number of contracts in the portfolio.
- And the z_k are the regressors.

9.3 Symbolic regression and genetic programming

- Non-linear relationships can often be solved by a combination of genetic programming and multiple linear regression,
- Roundings must be treated separately.
- Genetic algorithms are a family of search algorithms inspired by the principles of evolution in nature.
- By imitating the process of natural selection and reproduction, genetic algorithms can produce high-quality solutions for various problems involving search, optimization, and learning.
- At the same time, their analogy to natural evolution allows genetic algorithms to overcome some of the hurdles that are encountered by traditional search and optimization algorithms, especially for problems with a large number of parameters and complex mathematical representations.

10 Resume, Discussion, Questions