TransRe

We value risk.

AI and Machine Learning in Re/Insurance series: Setting the Stage

In Summary

Reinsurance is a complex industry with multiple opportunities to apply artificial intelligence.

Greater computational power and new machine-learning architectures expand the possibilities,

but

A lack of data quality and standardization hinders efficacy and adoption.

This brief overview introduces the concepts, with explorations of topics and challenges to follow.

The history of artificial intelligence in one paragraph

In 1950 Alan Turing asked whether we could build machines that think. The goal of imbuing sentience has a long history in fiction, from Shelley's Frankenstein (1818) to Fritz Lang's Metropolis (1927) and Karel Capek's play R.U.R. (1932) in which he coined the term "robot". Scientist John McCarthy is credited with creating the term 'Artificial Intelligence' in 1955 as part of a proposal for an Artificial Intelligence symposium that would take place at Dartmouth College a year later. Fast forward decades, include the creation of perceptron and backpropagation algorithms, sprinkle in some Artificial Intelligence 'winters' when the kinetic warmth of the hype drained away and research funding dried up, then add in more recent, massive increases in computation ability and new model architectures (such as convolutional nets and transformers) and today the increase in the number of model features in Large Language Models far outpaces Moore's Law.

The AI/ML Decision

Machine learning doesn't actually learn. Artificial intelligence isn't actually intelligent, at least not in the human sense. These semantics serve as aspirational anthropomorphic benchmarks. In fact the goal of ML and AI is ever improving optimization. 'Artificial Intelligence' is an umbrella term that includes Computer Vision, Robotics, Automation and Machine Learning. While "AI" and "ML" are used almost interchangeably, most (all?) subsets of Artificial Intelligence employ machine learning, where algorithms (a set or series of actions or steps) use data to optimize for some objective function. By adding more, better, cleaner data to fuel and train the model, the algorithms "learn" and improve their optimization. The decision to embark on using machine learning at any company should include identifying and understanding concrete use cases, assessing the overall data quality landscape, and hiring people that have a demonstrable track record of framing business problems into ML questions and then building, monetizable ML projects.

Types Of Machine Learning Models

Some machine learning models learn by being shown examples of the types of predictions desired (supervised learning).

Some models learn on their own by clustering (unsupervised learning).

Some models mimic the neural pathways of the brain (deep learning, so called because of the multiple layers of neural nets).

Some models reward agents' correct behavior (reinforcement learning).

Generalization, Overfitting And Regularization

ML models seek to make broad generalizations with their predictions. Not too hot, not too cold, just right. If you use data that is too specific and siloed, you run the risk of overfitting that data. Train a model to predict stock behavior only on tech company data and the model will fit to that data, learn industry-specific patterns but will not work on bank or oil stocks. Train a model to predict pictures of dogs based on pictures of cats, and your Rottweiler may come back a Donskoy. Machine Learning cleverly attacks overfitting by adding certain methods that penalize an increase in model complexity (which may lead to overfitting) through methods called regularization. Regularization helps nudge machine learning models back towards more generalized predictions.

Model Explicability/Opacity

Machine learning models are magic. They can be intimidating. Their goal is to predict the future. One resistance to machine learning adoption stems from a perceived lack of linear interpretability. Explainable AI will be a prerequisite for AIML adoption within the industry. The answers are there, buried in the equations for things like gradient decent and sigmoid function, in the component pieces of the equations and algorithms. ML developers need to build, then sell/market and explain their model/application to the target users. Some model types such as decision trees are more intuitively explainable, while other model types such as neural nets continue to develop new methods towards explainability.

Fetch Me A Beer from the Fridge

AI/ML can solve for certain narrow tasks much better than humans, but they still trail human ability when tasks are grouped together. For instance, machine learning can identify a beer bottle in a group of objects but fails when you ask it to "fetch me a beer" from the fridge or a "popsicle from the freezer". The combinatory nature of such tasks is remains too complex for AI/ML to handle. Known as Moravec's paradox, machines are able to do with ease certain things humans can't (massive computation and decision scaling) yet humans can do basic tasks such as balance and perception which AI/ML struggles with.

Data Quality

Machine learning models are most effective when fed large amounts of clean, standardized data. The re/insurance industry has multiple data-transfer waypoints, where data is 'adjusted', 'scrubbed','pro-forma'd' and then delivered in structured, semi-structured, and unstructured ways (PDFs, Emails, Excel sheets, Word documents). The data can also vary over time, both in how and what is delivered, and how it is structured, like Matryoshka dolls filled with potentially critical information just waiting to be extracted.

We value risk.

Automated Data Cleaning and Outlier Detection

Preparing for machine learning is like preparing a meal, so let's get our ingredients in place and establish our mise en place. Our raw ingredients (policy level/facultative/bordereaux data) arrive from various sources in varying conditions. The ingredients need to be prepped. The raw data are run against a series of algorithms to identify the shape of the data, match the names of columns to the standardized column headers, and automatically scrub the data for mistakes and outliers. Cleaning algorithms remove empty cells, NA's, column values that differ by orders of magnitude, and data types that don't make sense for a given column (a numeric data typed in a text field such as "Insured Entity" column). The cleaning algorithm may offer up questionable data for human interpretation. Cleaner and more standardized data makes prepping (aggregating and sorting) easier.

Once cleaned, outliers examined/removed, the data is ready for name mapping.

Name Matching and Entity Recognition

Different data sources may name an insured "Disney", "Disney Co.", "Walt Disney Company" etc. Solving the puzzle known as entity recognition appears in many aspects of re/insurance underwriting, actuarial science, and claims. To solve the puzzle, we start with an agreed ground truth, a "Rosetta Stone" for each entity. The ground truth can come from 3rd party vendors (Bloomberg, FactSet, Dun & Bradstreet) or be developed in house. Your machine learning team can also deploy algorithms to "learn" from all name variants and better identify ground-truth matches.

Prediction

Having spent much time cleansing, validating, and preparing the data, 'cooking' algorithms are employed to generate a range of predictions. Our use cases for predictive analytics include the spread of cyber events and wildfires, the likely impact of class action lawsuits (to assist early claims settlement recommendations) and policy pricing proposals. In addition to the quantity, quality and standardized nature of the training data, the accuracy of the prediction relies on which algorithm is selected, how the hyperparameters are tuned as well as feature selection. These all combine to form the 'secret sauce' of production-line machine learning.

Cycle Management/Portfolio Selection/Optimization

Portfolio optimization drives re/insurance success, from risk/return optimization, volatility dampening to driving exposure changes through the market cycle. The classic Markowitz method of mean variance optimization has been around since 1952. It has acknowledged practical drawbacks but ML now offers complementary tools built on massively greater computational power and new model architectures. If your choice is Spreadsheet vs. multiple GPUs and millions of iterations vs billions, .are you sure you are better off with the "way it's always been done?"

Large Language Models

If GPT recently updated your resume, wrote your cover letter or drafted your presentation then you are familiar with the hype and attention around Large Language Models (LLMs) like Open AI's GPT and Google's PaLM. You may also be worried about your job. Google and Microsoft will soon offer products that allow companies to train chat bots on their own internal proprietary data, allowing for multiple use cases that include identifying profitable contract language, bucketing historical client performance, and creating new parameterized insurance products. As of this writing GPT is on their 4th version of the model. What will version 11 look like? The non-linear growth of LLMs make it difficult to predict a ceiling to their future expansion and use case adoption.

TransRe™
We value risk.

Conclusion

Reinsurance applications of AI/ML learning remain in their infancy. The lack of clean and standardized data represents both obstacle and opportunity. There is an AI/ML sprint among re/insurers. The re/insurance race to adopt AI/ML may not start as a zero-sum game, but successful deployments will crown clear "zero something" winners and losers by monetizing their data insights. Don't slow down. Don't get left behind.

Otakar G. Hubschmann
May 2023

Otakar G. Hubschmann leads TransRe's Applied Data team, which researches, develops and deploys artificial intelligence and machine learning applications and datasets to help underwriters, actuaries and claims teams expand their understanding of client information, and improve client service. Contact Otakar with questions, comments, or ways the Applied Data team may help you.