

THE RESEARCH UNIVERSITY PRESS

International
Computational Law Journal

VOLUME 2 / ISSUE 3 / CURATED SOURCE EDITION

Automating Risk Analysis in Corporate Insurance Portfolios with Computable Contracts

Raphael Ancellin

MIT Computational Law Report · Originally published 2022-08-22

Independent reader edition under CC BY 4.0. This attribution cover was added; original source PDF pages follow unchanged. Original authorship and source identity are retained. MIT, Stanford University, Stanford Law School and CodeX have not sponsored or endorsed this republication.

Source, license and version evidence accompany the online article.

MIT Computational Law Report

Automating Risk Analysis in Corporate Insurance Portfolios with Computable Contracts

Raphael Ancellin

Published on: Aug 22, 2022

URL: https://law.mit.edu/pub/6b9nz4w5

License: Creative Commons Attribution 4.0 International License (CC-BY 4.0)

What is the problem?

Data analysis is a key tool for effectively and profitably managing an insurance business. Insurance policy contracts, or "products," a central component of the insurance business, are a primary target for such analytics, with many types of analyses possible. In this paper, I focus on risk analytics applied to contracts at the portfolio level (active contracts grouped together for administration/management purposes), a topic at the very heart of the insurance business. The goal of this inquiry is eventually to automatically answer questions such as, "*What is the cumulative risk in US dollars for peril X in portfolio Z if event A and/or B happens?"*.

Insurance should be one of the best industries at analyzing or simulating the risk exposure on its portfolios. However, in traditional insurance practice, only pre-defined data points are reported in policy manager systems during contract issuance. In a world of paper and natural language, this is primarily an offline process that involves multiple players (insurers, brokers, customers, etc.) and is difficult to standardize and govern. The reporting system records and transmits only limited aspects of the transaction, which are not always accurate nor standardized. Therefore, accuracy and granularity of the data are often insufficient to support complex forms of analysis, such as coverage calculations.

Traditionally, when insurers seek to perform in-depth risk analysis, they must go back to "ground truth," aka the signed paper contract – where all the data resides. Accurate extraction of these details is time-consuming and expensive.

Contract automation can greatly improve insurance data

Traditionally, insurance risk analysis requires experts to process millions of documents. Insurance companies are trying to automate the process to get answers faster at lower costs. Currently, a complete analysis of a corporate contract can take up to two days. One strategy has been to rely on Machine Learning (ML) and Natural Language Processing, provided that natural language policy documents, and other solutions based on ML have been tested previously and used in other circumstances. Despite some progress, there are a number of obstacles that must still be overcome, including the following:

> **Challenge #1: Accessing and cleaning the data.** This first phase begins with identifying the latest versions of the documents that have been stored in a variety of formats across several enterprise Content Management Systems (CMS) in multiple countries. The second step consists of connecting all the documents required for the analysis of a single contract (e.g., general conditions, specific conditions, etc.). Currently, categorizing documents requires human analysis and cannot be fully automated because it involves complex reasoning. Then, the documents are "OCRized"[1] to convert the PDF format into a computable data format.

> Even with the most sophisticated solutions, data extraction (parsing and name entity recognition) can only approach 70-80% accuracy on average. This excludes documents containing elements such as infographics and tables that can significantly lower accuracy.

> **Challenge #2: Identifying relevant clauses for analysis.** After natural language and numerical data have been converted from PDF to a computable format, analysis requires that the AI/ML tool identifies key clauses and information, such as coverage definitions, limits, and exclusions.
>
> Currently, ML-driven solutions can help analysts by identifying the recurrence of certain key words (e.g., "cyber," "data") and combining this with metadata (e.g., customer industry sector such as healthcare or energy) to retrieve contracts, and prioritize portions of the contract, that ought to be checked first by human analysts. Prototypes developed at AXA over the past four years have shown that a human expert training an ML solution could help further automate this triage task over time.

> **Challenge #3: Transforming content into quantifiable exposures.** This step involves a qualitative analysis of the whole contract to understand if a certain loss, such as business interruption, is covered. What are the exclusions, limits, etc. in a given situation? Handling rules[2] are also included to calculate claim payments.
>
> This type of analysis must take into account each clause as well as the relationship between clauses. For example, an exclusion previously mentioned can change the output of another clause in the contract. After developing multiple prototypes, AXA technical teams concluded that there is currently no ML-driven solution that automates the task with "business-ready" accuracy. This conclusion is the growing consensus in the industry.

> **Challenge #4: Aggregating risk exposure (on selected perils) at portfolio level for further analysis**. It would be tempting to skip Challenge #3, and instead extract maximum coverage limits from each contract and add them across a portfolio to calculate risk. However, this calculation is insufficient because exclusions and handling rules are not included. Moreover, the proper automation of this phase is necessarily linked to overcoming aforementioned Challenges #2 and #3, which, as discussed, is not an easy undertaking.

Currently, NLP and ML simply cannot offer a reliable risk assessment.[3] Even if an 80% accuracy is achievable with data extraction and analysis in the near term, a markedly impressive outcome for data science, it would remain insufficient for risk management.

Automated Computable Contracts could be the solution

Alternatively, prototypes of "insurance contracts as code" show promising results in calculating risk automatically and instantly in large portfolio of contracts, with an accuracy approaching 100%.

As discussed, NLP approaches rely on interpreting a "paper-first" contract, and often with limited accuracy. Computable products, by contrast, are code first, and their reasoning relies on logic programming.[4] Provided the data on the contract are properly entered to begin with, the data and outcome are fully accessible for analysis as they are already set out in structured formats, with the logic of the agreement fully realized in the computer code embodying the contract.

In addition to improved risk analysis, other potential benefits include improved operations. For example, a computable contract can be plugged into an API to then "feed" many other insurance systems (e.g., the claim manager, policy manager, and/or call center software, etc.) with the data source derived directly from the contracts. The approach can work with both legacy and natively computable contracts alike:

- **For legacy contracts/products,** our work at AXA has demonstrated that it typically takes two days for an insurance agent to convert one traditional insurance contract into code, leveraging a "no-code" interface developed at our company. AI/ML solutions are already supporting this effort – extracting data points to be verified by a human. Ongoing prototyping efforts leveraging cutting-edge technologies, such as GPT3, are encouraging. These technologies could pave the way to automatic conversion from text to "insurance product as code".
- **For new products/contracts,** technology solutions exist for designing insurance products as code. AXA, for example, leverages computable clauses associated with certain risk and pricing models. A human agent can assemble these clauses using a no-code interface to build the computable insurance product, incorporating the associated model(s). The tool then generates the contract document to be signed by the customer. This approach has been tested in production for a direct-to-customer distribution model for personal insurance. However, this is not the only distribution scenario. For instance, in corporate insurance, brokers are typically involved in the process, and there is no standard way to design a contract. We could envision working with brokers to co-build a solution allowing them to maintain a central role in customization, while still maintaining the benefits of automation. The standardization of this design process (business, data and tech standards) will bring benefits cross-industry including for purposes of actuarial analysis to product comparison, and potentially towards a streamlined reinsurance market.

How to move forward?

Here are steps that can help insurance companies accelerate automation of risk analytics at portfolio level – and that will also contribute to the work on computable products:

- **Acknowledge that fully automating analysis of the paper-first contract is some distance off, and focus, instead, on machine-assisted human analysis.** This feedback loop requires the agent to do most of the tasks at the beginning to train an AI/ML solution. Then the AI/ML will take over some tasks, such as clause identification, and accelerate other tasks, such as classification of the contract to be checked by humans in

Extracting the reasoning of the contract itself – transforming the content into quantifiable exposures – will likely remain human-driven until a substantive technological breakthrough. And, when it does occur, insurance executives should be prepared to accept that results of these analyses are an approximation based on a machine-led interpretation that may be difficult to explain *a posteriori*.

- **Build standardized clause libraries.** When we compare insurance contracts of different industry players, we quickly notice that many clauses are quite similar. Short term, having a standardized clause library at the enterprise level, and possibly at the industry level, will support better NLP-driven text search and clause comparisons. In the longer term, standardized clauses will also help insurers build faster and safer contracts by assembling clauses that have been pre-approved. There is a balance to be found between formalization and flexibility in contract design in order to preserve sufficient creativity or negotiation. Finally, these standardized clauses could be progressively converted from "text first" to "code first" to lay the foundation for computable contracts.
- **Structure the data for improved analytics during the subscription phase[5] of the contract.** This is a challenging technological and business transformation involving multiple players (insurers, brokers, customers, etc.). Antiquated tools and habits must be changed. Replacing email and word processing solutions with contract lifecycle management systems appears to be an efficient way to streamline the subscription process, including contract drafting and clause comparison; collect the right data, including mapping contract clauses and their limits, and supporting basic automation of analytics.
- **Start to create computable products and contract portfolios in selected lines of business.** Low-cost and/or ultra-tailored products, as well as product lines with problematic claims experience, can provide suitable initial targets for portfolio conversion into computable products for marketing and claims operations. The automation of analytics will come along naturally as the conversion goes forward. This trailblazing exercise will hopefully pave the way for further innovation.

Conclusion

Risk analysis and simulation for insurance contracts at the portfolio level is a critical area to automate. Despite progress, AI/ML approaches cannot yet provide a solution that sufficiently accommodates the demands of the industry. Namely, the accuracy of ML and NLP models may continue to fall short of results needed for a reliable actuarial and risk analysis for the foreseeable future.

On the other hand, computable contracts offer promise. It is time to change how we represent insurance contracts, using software to make them computable; improving not only their interpretation and data production, but providing benefits to claim administration, sales, and many other aspects of the insurance business, while improving consumer choice and experience.

In order to fully capture these benefits, the industry should work together on a shared specification that can be used between and across insurance enterprises.

The CodeX Insurance Initiative has invited leaders from industry, academia, and the regulatory community to contribute short papers describing the author's views on important issues relating to the application of computable contracting in the insurance industry. The development of computable contacting for insurance is still a work in progress, and the sharing of ideas and approaches within the community of interest is a major goal of the Insurance Initiative. As a part of this conversation, these papers present the views of their authors, and do not necessarily reflect the views of CodeX, of the Insurance Initiative, or of any of its participants.

Raphael Ancellin is a Fellow at CodeX — The Stanford Center for Legal Informatics and a Product Manager at AXA

"Circuit Board" by bravenewtraveler is marked with CC BY 2.0.

Footnotes

1. Using optical character recognition technology ↩

2. Obligations of insurance companies when managing claims (e.g., duty to act in good faith, reasonable standard of prompt investigation and claim processing, etc.) ↩

3. *See* for example, prior research on the limitations of NLP, particularly in the ability to draw contextual meaning. Emily M. Bender, "Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data," available at: https://aclanthology.org/2020.acl-main.463/. Currently, state-of-the-art models are able, with sufficient training, to seemingly infer answers from relevant texts. Nevertheless, in a field as nascent and expert-driven as insurance, there remains gaps in the interpretation of particular terms-of-art. This is similar to the discussions on the differences between machine and human interpretation of legal language. See Emily M. Bender in her response, "A reply: On the differences between human and machine processing of legal language," to Reuben Binns' article, "Analogies and Disanalogies between Machine-Driven and Human-Driven Legal Judgment," available at: https://journalcrcl.org/crcl/article/view/5.

↩

4. *See* for example discussion by Michael Genesereth, *The Legacy of Hammurabi*, MIT Computational Law Report Release 3.2 (2021), available at: https://law.mit.edu/pub/thelegacyofhammurabi/release/3. [Insert reference to Goodenough primer on computable contracts] ↩

5. Defined as designated period for prospective clients to enroll/purchase insurance. ↩