Credibility Theory Augmentation Through Gradient Boosting: AI-Enhanced Actuarial Modelling Frameworks for Insurance Pricing and Reserving

Dr. Jingxuan He, Professor of Data Science, Peking University, China

1. Introduction to Actuarial Science and Insurance

Actuarial science is the discipline that applies mathematical and statistical methods to analyze risk in insurance, finance, and related fields. In like manner, actuarial science evaluates past data to assess the probability of a future trauma and design plans that help alleviate the financial effect of trauma based on the evaluation of the probability of the trauma. Ramifications of the evaluation of the probability of any trauma, and in insurance operations, non-quantitative factors such as public regulations and the degree of public recognition of traditional actuarial techniques together with the analysis of quantitative factors will provide a more valid operational risk evaluation. AI-enhanced actuarial techniques will be the new generation of actuarial techniques. The main goal of insurance risk evaluation is measuring the value of the financial loss caused by the exposure of its policyholders from economic or physical trauma. In simple terms, the purpose of insurance is to cover the periodical, stochastic loss expectation of the insureds with the relatively deterministic projection of industry-wide average loss frequency. In defending a company's insolvency, the employees and management often argue that the failed company suffered from unusually severe loss shocks associated with unpredictable peril. To reassure regulators and investors that the subject company is solvent and stable before the injury occurs, the actuary must quantify in economic terms the probability of an economic loss exceeding the entity's capital and surplus. In other words, the actuary must evaluate the industry-wide potential damage caused by the frequency of a stochastic insurance loss exceeding the assumed deterministic loss projection.

1.1. Overview of Actuarial Science

Actuarial science is the discipline that applies mathematical and statistical methods to evaluate risk in the insurance and finance industries. As a profession, actuarial science has existed for more than two centuries, with actuaries traditionally focusing on non-trivial problems such as pricing insurance policies, reserving for future insurance claims, and helping life and pension providers to define and manage their products. Traditionally, the information available to the actuary was scarce, gradually expanding to more data than could be processed by a spreadsheet or another popular tool. Today, the scope of information required to be processed by actuaries is vast, as insurance companies are sitting on a mountain of data, and industry-wide data are available through various public and private repositories. A medium-sized insurer may have detailed information about, for example, the precise description of past claims or policy renewals. Property-level data may specify how policyholders can make their assets resilient against flood-related losses. Various sets of demographic, socio-economic, and environmental data can be appended to this information. Non-traditional data such as social media footprints, credit scores, GPS positions, or publicly available risk scenarios are expensive, but they can be purchased and used to enhance traditional models such as demographic or BED variables. Furthermore, advanced analytics, big data, and machine learning techniques are already being utilized by several insurers both in the front- and back-office domains.

1.2. Importance of Risk Prediction and Financial Forecasting in Insurance

Underwriting is the core business function of an insurance company. It is the process by which an insurer evaluates the risks of a particular customer based on individual characteristics and decides whether to accept the risk. Traditional actuarial models, which are based on statistical techniques and economic theories, have been used for decades to assess an applicant's risk level primarily by quantifying a "price tag" in terms of premium adjustment for risk acceptance or by flagging prospective bad risks for rejection. The concept of underwriting, however, has been changing since insurers have been collecting substantial amounts of operational and external data as a result of advances in information technology. External data includes demographic, geographic, and social media data in addition to applicants' self-reported data. The new data types capture policyholders' historical behaviors and personal lives.

An updated underwriting process can take advantage of AI techniques, including machine learning approaches. These approaches are trained to link the observed individual data with an outcome to allow non-linear and interaction effects among the input variables. Some machine learning models in general prediction problems do not even require that inputs are strictly related to an outcome in order to achieve model performance. Their advantage is that prediction results go beyond drawing inferences to making predictions. Insurers and reinsurers face challenging tasks in making internal risk forecasts and financial statements. Regulatory and rating agency concerns require clear insights regarding the holding company's ability to absorb losses from its affiliates and ultimate affiliates. Although traditional risk forecasting techniques are conceptually clear, their omitted risk lays out heavy burdens on both the parents and ultimate affiliates for proposing strong risk mitigation strategies. Using alternative data and AI techniques, a large amount of unmeasured risk can be revealed. AI forecasts extend traditional techniques by providing tailored extra information in AI-generated credible scenarios. These can capture embedded unique interactions among both the insurance entity and its transfers and lay out pertinent details of the extra risk for the pricing details.

2. Fundamentals of Machine Learning in Actuarial Models

Given the advent of new data sources and powerful mathematical tools, the potential applications of machine learning to the actuarial field are unlimited. In the context of this paper, machine learning can facilitate and expedite the estimation and combination of the complex stochastic procedures that are at the heart of many actuarial models, substituting for currently used deterministic approximation methods. While we desire machine learning to be able to replace the long-established methods, by using deterministic approximations as a method to train our learning algorithms, a machine learning approach can iron out the rough edges of existing methods, even if the rough edges are due to non-fundamental causes such as poor convergence or interest rate discontinuities. Finally, there are some unique actuarial challenges that machine learning was built to handle. These can occur in many steps of an actuarial process, including parameter estimation, selection of predictive features, and in the process of combining stochastic claims processes at the heart of loss reserving, pricing, and capital modeling.

Our goal is not to develop new AI technologies per se, potentially scaring users with intimidating algorithms. Instead, we hope to demonstrate the potential of currently available machine learning tools in illustrative situations that resonate with actuarial audiences. Models have many parameters, making it difficult to accurately estimate through linear methods. The machine learning methods demonstrated in this paper have been proven to handle a very wide variety of predictive problems and have been shown to be competitive with other predictive methods. Further, machine learning methods excel in automating model performance assessment, predictive factor identification, and in automating the outputs of various stochastic processes in the models. Our ultimate goal is to build a predictive modeling framework designed for tasks with actuarial purposes.

2.1. Basic Concepts and Terminology in Machine Learning

The goal of machine learning is to develop a model that can accurately make predictions in the context of existing data. Although most machine learning models make predictions for a known quantity of response, forecasts involved in the insurance application can also be considered as predictions for future events. Therefore, a model is generally trained using historical data with responses available and then applied to a data set with no corresponding responses to generate predictions using feature attributes for which information is available. The machine learning model should have high forecastability, but its understanding is often opaque and hence hard to justify, which makes traditional predictive models important.

In the process of constructing a machine learning model, feature selection is also a critical step in addition to data preprocessing and model specification. Results from traditional predictive models built for insurance analysts to understand what led to certain predictions can be beneficial, while the choice of model and feature engineering exerts a less significant effect on the uncertainty quantification. In addition, we delineate the performance of many predictive models for right-censored response that is commonly seen in life and health insurance. Some key concepts with the accompanying terminology and their relevance to insurance data will be briefly introduced in the upcoming subsections. We will then review various traditional models in which actuarial work was involved typically in data pretreatment and efficiency evaluation only. By leaning upon these historical experiences, we examine if the AI-enhanced models are of any value in insurance practice, and if the time-dependent models have certain advantages for insurance applications over the existing ones.

2.2. Types of Machine Learning Algorithms Used in Actuarial Models

While introducing the components of an actuarial model, we mentioned that an actuarial model can be viewed as subsets of the various techniques used in machine learning-based quantitative modeling. Simple models, rules, expert judgment, exogenous data, quality assurance and control, internal reporting, risk management, model output, human interpretation and review, and the model update control loop are all important aspects. There have been many types of machine learning algorithms used in recent actuarial literature. Identifying and distinguishing between them through a single criterion may be non-trivial. The input is that they distinguish between distant learning and predictive analytics. More detail will be used to dissect the types of machine learning algorithms used in the paper. There are four types of machine learning algorithms distinguished in recent actuarial literature: unsupervised learning, supervised learning, semi-supervised learning, and reinforcement learning.

Supervised learning is the domain of machine learning where we attempt to map a set of input data to labels. The aim is to derive a model using the training data so that when presented with unknown data, it can predict a label. Supervised learning can be divided into two different types of problems, namely regression and classification. The principal function of regression is to model the output based on one or more input variables using a continuous or scaling output. The principal function of classification is to model the input data with a finite set of labels using a discrete output. Since supervised learning involves the use of a labeled dataset to train a model, it is capable of capturing meaningful patterns in the data. Notwithstanding, supervised algorithms require that the input data be processed by an experienced professional and may sometimes require large amounts of labeled data to make strong and reliable predictions.

3. Integration of AI and Machine Learning in Actuarial Practices

The advantage of AI is its adaptability of learning mechanisms to any actuarial method through data, enabling actuaries to implement AI in the most suitable actuarial method for the business model with the highest effectiveness. For example, machine learning algorithms can process and analyze huge unstructured actuarial data, creating relationship models between them in a way that conventional actuarial models cannot do. Technically, machine learning algorithms do not make any assumptions regarding the distribution of risks, the credibility formula, a relationship, or even a priori between the different risks to be modeled. Their approach is holistic since they verify the performance of all contributing actors and relate each actor to others through the data process, which helps to capture unseen relationships.

It can go further as these learned relationships have no underlying structural form. They are purely data-driven inferences, enabling the re-evaluation and re-calibration of any standard actuarial model if such a model was initially created using an incorrect assumption, hypothesis, or procedures that might not have been recognized at the time. Specialized AI techniques are even advancing as they can process huge unstructured datasets with a high number of layers, such as large artificial neural networks, which have many advantages over traditional machine learning algorithms. Unlike machine learning algorithms that tend to overfit the model due to irrelevant data, deep learning can generalize and measure the quality of predictions using a measure called cross-validation error.

3.1. Challenges and Opportunities in Implementing AI in Actuarial Science

Artificial intelligence (AI) has been a focal point for multiple industries as a technology expected to disrupt the future. The actuarial function can also benefit from AI to meet future challenges by unlocking new sources of data and reducing errors. The current reliance on localized and linear models in actuarial science could hamper its ability to deliver expected results to businesses, regulators, and policyholders. The text identifies the various challenges of implementing AI in actuarial science and highlights the opportunities to bring better results to users. Actuarial science is going through an evolutionary phase of embracing and adopting machine learning and artificial intelligence for operations within the traditional actuary domains. It is the responsibility of senior actuaries who are engaged in the education and training of junior actuaries not only to fill the technical-practical gap but also to ensure cultural and value coordination as competencies inside the company grow. The actuarial role is changing with the infusion of AI. AI technology has been able to add value to the actuarial function and enhance the data analysis capabilities of the actuarial role. It will significantly impact the actuarial role in the document retention space with enhanced scientific rigor, judgment, and accuracy compared to traditional techniques. These technologies encourage actuaries to adopt a growth mindset.

3.2. Benefits of AI-Enhanced Actuarial Models in Insurance

The fundamental principle for the pricing of insurance risks, i.e., building an appropriate insurance tariff and consequently ensuring the solvency of an insurance company, is the same today as it was over 300 years ago. This long-term fundamental principle results in some restrictions and also regulatory requirements that, first and foremost, reflect the professional ethics of the actuarial profession. The professionals should make use of all vast technological enhancements within the limits of relevant regulatory requirements that will help them to fulfill their fundamental mission. This chapter will try to demonstrate that, in our modern world, AI-enhanced actuarial models can now be considered as the main tool that allows actuaries to price insurance risks at, or close to, the ethical and regulatory limits. In principle, many of the best actuaries today are those with two basic skills: deep domain knowledge and high-quality programming. The combination of these two skills can, with basically all potential artificial intelligence enhancements, produce high-quality actuarial models. AI-enhanced actuarial models in practice can lead to the disaggregation of the process of the insurance tariff into the most rational process sections. It can also demonstrate, with insurance mathematics' support in many cases, that the application of more sophisticated models than normally used is not only possible, but that it may still fulfill its purpose from the abstraction of a target actuarial tariff model. These and other potential benefits lead to the conclusion that AI-enhanced actuarial models can generally be considered as an important stage of the insurance tariff model evolution process from the ill-defined ab initio through the monolithic and hierarchical to the detailed. AI-enhanced actuarial models, with an immense decline in the cost of designing a support preprocessing model, may also allow actuaries to set the building of a tariff model as an optimization process. At clearly higher advantages, for a good part of insurances, actuaries can now generally disseminate a non-guaranteed insurance tariff culture, including the use of AI-enhanced actuarial models.

4. Case Studies and Applications

Machine learning techniques, especially neural networks, have achieved significant advancements and have been widely applied in the fields of computer vision, natural language processing, and robotics. However, machine learning is still relatively new to actuarial modeling in the insurance industry. When it comes to sophisticated AI techniques, actuarial practitioners are usually undertrained and unfamiliar with them. Moreover, the interpretative capacity is almost the most substantial concern for actuaries who design and use the models and need to show the rationality of the models to regulators. In this paper, three cases of actuarial applications with key features that safeguard the interpretative capacity are introduced, including claims reserving, pricing model design in both customer and reinsurance markets, as well as capital distribution model construction for insurance accounting. Actuarial modeling has long been a tradition in the insurance industry, and the basic concept of actuarial modeling is to predict future losses and quantify the uncertainty for insurance reserve estimation and commercial risk pricing. However, as the complexity and size of the insurance business expand with technological improvements and rigid regulations, traditional actuarial models begin to show their limitations. Firstly, although linear models are simple and thus easily interpretable, they cannot reflect the structural relationship between dependent and independent variables. Nonlinear or generalized linear models are more complicated, but they can capture the structures of client databases more efficiently. Consequently, the performance of these complex actuarial designs is usually limited. Secondly, actuarial modeling often involves a considerable amount of manual expertise, and the process is time-consuming. The lack of an agile model development process will limit the inheritance of industry experience and the potential for improvement of model performance. The traditional actuarial modeling process is usually complicated, and the models are most often company-specific, so an under-studied model development method is used for company-specific data. Thirdly, although traditional actuarial models are relatively robust and parsimonious, they often leave valuable but lost information on the table throughout the model reduction process in order to fit the complex data structure with a relatively small amount of records.

4.1. Real-World Examples of AI-Enhanced Actuarial Models in Insurance

In this section, we present a comprehensive list of actual use cases of AI-enhanced actuarial and underwriting models. Based on our classification, we found that most examples of AI come within the framing of extended actuarial models, with deep actuarial models considered to correspond to few and less common use cases worldwide. Some representative examples of AI-enhanced actuarial models in insurance include:

1. Natural Language Processing for the analysis of customer contracts: intuitively, insurance refers only to some types of contracts; the other ones are a different form of agreements. An insurance policy is a complex contract that contains elements of customer and insurer expectations in different scenarios. AI can understand these customer contracts with NLP, and also, this data can be used as input for risk models.

2. Photo recognition for underwriting in auto and house: users use their smartphones to upload a picture of their vehicle or house, and the AI estimates the appropriate amount. If this customer contracts auto or house insurance, they are more likely to send the photo.

3. Recognition of images to detect fraud: when we receive a claim, it is common to ask for a few photos of the damage. The fraud detection can confirm these damages but can also filter suspicious claims.

5. Future Directions and Ethical Considerations

Will AI take over the work of human actuaries? This is an open question as AI has not yet reached the general AI state, meaning general AI is still not capable of performing all work as well as humans under the same circumstances. Actuaries are still needed to interpret and communicate the relevant technicalities to stakeholders and comply with regulatory requirements for risk reserve. In addition, the AI tools as of now still lack interpretability, which poses huge risks for companies to use. No matter how much AI can assist human actuaries, the responsibility is still on them. What matters in the future might not be what skills actuaries should be equipped with, but the outcome itself—achieve the target or not. Achieving the results is the fundamental purpose of risk modeling. The design of architecture and tested results demonstrate that AI can well represent the complex non-linear dynamics in financial risks under different settings. However, the representation function restriction and vulnerability issues of AI might limit the wider applications in pricing and reserving. Such vulnerabilities need to be researched to accurately capture the nature of AI or to extend the interpretability of AI. Furthermore, bearing the underwriting and reserving will be emerging challenges for AI models due to the material and conclusion responsibility since AI acting as an agent might lead insurance to moral hazard. Additionally, because of the financial significance and market conduct issues, from the perspective of practicality, the integration of economic scenarios, factoring in supply and demand, disaster risks, and other sources of uncertainty into new modeling paradigms will be highlighted themes. Last but not least, the imperfection in data loading, model defaults scrutiny, and broad governance involving different stakeholders will be the frontier for regulatory actions to potentiate or hamper new opportunities.

5.1. Emerging Trends in AI for Actuarial Science

Nowadays, actuaries are increasingly adopting artificial intelligence algorithms driven by the availability of larger databases and measurable computational power. Their adaptability and capability in uncovering complex patterns contribute to boosting predictive accuracy and efficiency in traditional actuarial science practice. Most actuaries focus primarily on pricing models. AI-driven models are more frequently used to considerably improve these models and to examine individual risk for the purposes of underwriting. However, AI-driven tools have demonstrated their applicability in numerous actuarial domains such as reserving, risk exposure, claims analysis, customer relationship management and retention, fraud detection, negotiation, and new product development. However, AI-trained models can sometimes work as a "black box," which concerns the actuarial regulators and some insurer managers. The lack of transparency of the models could lead to hard-to-justify predictions not supported by traditional actuarial principles. Consequently, the purpose of this special issue is to reveal the interesting links that, both statistically and economically, exist between AI and the actuarial sciences to address a wide range of actuarial issues.

5.2. Ethical Implications of AI-Enhanced Actuarial Models

In recent years, the field of AI ethics has grown significantly, with various researchers and practitioners debating the ethical considerations for developing, using, and regulating AI systems. The insurance sector, with its increasing use of AI, is also being drawn into this debate. In particular, there is growing concern about the use of AI-enhanced actuarial models creating unfair bias through discriminatory decision-making and exacerbating inequalities when certain groups are disproportionately impacted or underserved.

Two of the main root causes of these adverse outcomes with AI are biased training data in HR or hiring, educational, lending, policing, or judicial applications, and the specified goals. It is well understood that supervised learning works by analyzing a dataset involving many samples of input data and teaching the system how the input data should lead to a desired output, where the teaching process typically involves adjusting the parameters of the system using an optimization routine to minimize the errors between the desired outputs and the model's predictions. It is also well understood that unsupervised learning analyzes a dataset to identify trends or patterns without known output data, and reinforcement learning teaches a system by forming a feedback loop with the surrounding environment. These premises explain some of the complexities in training data for supervised learning to avoid the adversities of bias, discrimination, or inequality. One prevalent issue is that a machine learning model will reproduce societal prejudices using biased training data to exacerbate discrimination and stereotype biases. This is because, being inherently honest, the model is exploiting the biases incorporated into the training data.

6. Conclusion

The objective of any actuarial rate monitoring process is to identify deteriorating experience quickly enough to take corrective action in premium rates. At issue is how finely the actuaries can identify groups with high future loss ratios from information from the past. Techniques like credibility theory and various forms of experience adjustments are designed to target the underwriting profit problem pre□eminently. These tools address the question, "How likely are we to see losses in group X in the future based on past underwriting results?" The problem of achieving higher levels of consistency in insurance results remains. AI breakthroughs in predictive accuracy in the past few years have led to broader acceptance and application of such tools. AI models matched to an insurer and its internal data are more accurate than external data solutions.

That said, collaboration in AI is still in the early stages among actuaries, data scientists and experienced independent modelers. It is questionable whether collaboration needs to occur from all three groups for as long as the underwriting orientation prevails for actuaries. One thing that is absolutely clear is that it's not too soon for insurers to dive in and start to learn how to apply AI models specifically for their insurance. Insuring and protection businesses in finance are somewhat different in some aspects; one of these is the grouping of complex risks. The known complexity of future financial losses of certain instruments is highly subjective. We believe the reliance on the subjectivity of economic theories and standards is one of the reasons behind the financial crisis. If this subjectivity could have been identified and managed, the financial crisis could have been less severe. The financial risk clearly highlights the importance of its relevance; the main question is how this risk might be quantified and managed. We suggest that by improving the knowledge and understanding of the embedded risks in mortgage securities, the quality of associated models, methods, and processes could improve. The importance of the indirect risk is equal.