Analytics Offer Unlimited Opportunities to Actuaries

BY GARY WANG, PINNACLE ACTUARIAL RESOURCES, INC.

About the Author
Gary Wang is a senior consulting actuary with Pinnacle Actuarial Resources, Inc. in their Bloomington, Illinois office. He is a Fellow of the Casualty Actuarial Society and a Member of the American Academy of Actuaries. He has been with Pinnacle for 12 years and has worked in the property/casualty industry for 18 years. At Pinnacle, Mr. Wang has worked extensively on the application of modeling techniques to the insurance process.

Abstract
The actuarial practice is analytical in nature. As such, the continuing advances in analytics and increased availability of data will profoundly affect the actuarial profession. This article discusses the opportunities and challenges the current analytics environment poses to the traditional actuarial practices, and how the actuarial profession has responded.

“99% OF THE ECONOMIC VALUE CREATED BY AI TODAY IS THROUGH ONE TYPE OF AI, WHICH IS LEARNING A TO B, OR INPUT TO OUTPUT MAPPINGS.”

-Dr. Andrew Ng, at EmTech 2017[1]
Insurance product pricing is difficult since production costs are unknown at the time prices or premiums are quoted and bound for insurance contracts. Furthermore, most insurance products cover rare events. For the majority of insureds, typically for any year with adverse claims that causes the product to be a net loss to the insurance company, there will be several years during which little to no claim activity occurs and the product is a net profit to the company. The actuary is tasked with creating a pricing plan such that, in this rare events environment, each insured is priced according to their risk potential. To this end, actuaries rely on data, and models that make sense of that data.

When it comes to predicting risk potential, current advances in algorithms and model structures offer strong alternative designs.

The current modeling design standard for developing an insurance product pricing plan is Generalized Linear Models (GLM), a generalization of the classic linear regression technique. In addition to reasonable predictions in practice, several features of this model structure are aesthetically pleasing to companies and regulators. The model’s structure fits neatly into insurance carriers’ traditional multiplicative pricing structure, in which one can look up relativities associated with each rating element and directly calculate the overall premium through simple arithmetic. The model also provides an intuitive interpretation whereby one can study the relativities along each dimension and assess the impact of changes along that dimension on the overall premium charged. This transparency in how each variable affects overall premium makes for relatively easy discussions with state regulators on the factors considered in rating plan design, and the impacts various rating elements have on premium.

It is worth noting that even in the GLM-based modeling environment, transparency can be illusory, particularly in situations where multiple rating elements are derived from the same or highly-correlated raw information. In these cases, one must take into account all relevant dimensions so that all effects are discussed and a complete picture is presented.

When it comes to predicting risk potential, current advances in algorithms and model structures offer strong alternative designs. Consider the recently-completed Kaggle competition, Porto Seguro’s Safe Driver Prediction, where the aim is to predict whether a driver will file an insurance claim within the next year. Given the infrequent nature of multiple claims for a given vehicle in a given policy year, the modeling exercise is essentially a classic frequency modeling problem an actuary may undertake as part of a pricing analysis. On the discussion board, the winner of the contest described his winning submission as an ensemble of one Light Gradient Boosted Model and five Neural Networks. The superior performance of his design is notable, especially if you compare it with the typical business practice of utilizing GLM here in the United States. However, the level of complexity in the relation between the rating elements and the likelihood of a future claim is equally notable, and the increased complexity poses challenges from a business perspective.

The perception among most of the actuarial community is that, given the lack of intuitiveness regarding the relationships between the various rating elements and premium, the current regulatory environment is highly unlikely to accept a filing directly based on this ensemble model solution. The community expects similar difficulty in educating agents to properly present and discuss premium with potential insurance product buyers. Today, presented with the two options, most companies will choose to stay with the GLM-based one, given the ease of overcoming implementation hurdles in systems, marketing, agent training and regulatory approval.

It is not hard to envision how having both models can help an insurer improve its operations. Suppose, for practical reasons, the company chooses to go with the traditional GLM design despite having an alternative model with more predictive power. It can analyze the two sets of predictions to understand where forecasts diverge, and in these situations whether the implemented plan is presenting too low a price or too high a price compared to the filed pricing scheme. In either situation, if segments are sizeable and can be reasonably defined, the analyst can engineer new variables to properly segment these potential policyholders and feed that new information back into the GLM model to specifically differentiate these risks. Even if the company chooses not to refine the GLM-based pricing plan, the company has guidance on profit expectations associated with writing these risks and can make business decisions accordingly.

This is consistent with today’s trend, whereby companies accept the necessity of a regression structure to the model for business considerations. They focus efforts on utilizing feature engineering to improve how existing variables are handled in the rating plan, while constantly seeking out promising new variables to incorporate. In the personal and commercial auto insurance arena, usage-based insurance data has great potential. The driver generates data, as does the vehicle as it is used, and raw information regarding when, where and how the vehicle is driven is then processed, analyzed and scored. This type of intelligence represents the first time in the insurance pricing arena that the volume of data and the frequency with which data is generated comes close to what other industries generally consider Big Data. Moreover, this data has great intuitive appeal in that the public generally accepts driving patterns as having a causal relationship with accident potential. On the regulatory front, many departments accept the intellectual property nature of the information and its analysis, creating an environment in which greater complexity of model design can be utilized and tested. Given the voluntary nature of allowing a company to capture this information, it remains to be seen if the insurance industry can work with the necessary partner agents, such as the government and consumer groups, to create a future in which data capture becomes the norm rather than the exception. This would provide analytics the opportunity to showcase its full array of insurance capabilities.

In the homeowners and commercial property insurance space, image analytics offers similar potential. Drone utilization enabled large amounts of data imagery to be gathered, and advances in image analysis are, coincidentally, among today’s hottest research and development topics. Consider the recent annual ImageNet competition, where competitors were tasked with designing models to locate and categorize objects within images. In the 2017 ImageNet competition, the leading model achieved an impressive classification error rate of 2.25%. The convergence of availability of data, coupled with improvements in image recognition models, makes for a...
...it behooves us as actuaries to deliberately consider the current state of analytics vis-à-vis what we do, and how our jobs may evolve as analytics do.

ripe environment for business utilization like property inspection or claims investigation.

In a reserving exercise, building the case to disrupt the traditional process is a major hurdle. For credibility and computational considerations, actuaries have historically aggregated claims development data before analyzing patterns to make ultimate loss predictions. They aggregate, typically to accident year, pertinent information such as incurred and paid claims counts and loss amounts, maturity in twelve-month increments and lines of business. Today, we have sufficient computational power to leave the data at the individual claims level, then model on more detailed information. Furthermore, doing so allows some nuances to naturally flow through, to the extent information is captured as part of the claims detail. For example, the categorization of medical claims by severity of injuries is intuitively an indicator of likely differing development patterns. When aggregating data, a shift in the claims mix over the years would be hidden by the summarization and require an insightful actuary to note and take into account the changing mix.

But the reserving analysis is not exposed to competitive pressures in the same manner as the pricing analysis. As such, more latitude is given to the relative accuracy of the point estimate. In a stable environment, most traditional techniques put us in the right ball park, requiring nudges year to year, but generally within reason. This creates a disincentive for companies to disrupt the existing processes to improve the reserving methodologies in place, since the current reserving process is of sufficient rigor to meet regulatory and accounting requirements. While recent emphasis in the range and variability of the estimates may serve as part of the claims detail. For example, the categorization of medical claims by severity of injuries is intuitively an indicator of likely differing development patterns. When aggregating data, a shift in the claims mix over the years would be hidden by the summarization and require an insightful actuary to note and take into account the changing mix.

Interestingly, a case for improvements to the reserving process may eventually result as a byproduct of other current insurance company trends. Of note, companies have exhibited great interest in understanding claims development and ways to predict and manage potentially explosive claims that blow up through unreported complexity and/or litigation involvement. Such supervised training models are likely to include predictions of expected ultimate claim amounts, and this information can then be leveraged as part of the reserving process.

Supervised learning challenges are not limited to classic pricing and reserving issues. In addition to the aforementioned claims-handling opportunities, there are other often-posed questions to which one versed in supervised learning problems can become a valuable resource:

- Which company policyholders are most likely to renew?
  - Are they a profitable segment of the portfolio?
  - If not, how will corrective adjustments to the rating plan affect retention?
- Which quoted policies are more likely to convert?
  - Are they consistent with the target market?
- Which properties are more likely to have unmitigated hazardous conditions?
- Which claims are more likely to be fraudulent?

Given the current state of analytics, an increase in available tools is making algorithms more accessible to interested business analysts. An actuary should not lose sight of the fact that they are not alone in the analytics arena. For example, many insurance company research labs currently staff a non-negligible number of statisticians and data scientists, clearly signaling actuaries have no monopoly on an insurer’s analytics. As more tools are built to improve methodology and processing efficiency, the pool of professionals qualified to manage them will grow.

As tools become more sophisticated, greater data processing and model iterating automation will become the norm rather than the exception. This is not a bad thing. Just as monitoring reports evolved from manual creation to template updating to automated batch processing, one should embrace increasing analytics automation. There are many repetitive data processing steps that are generally variations of a theme when it comes to assembling the data for modeling and, to some extent, for exploratory data analysis. As automation takes over these tasks, it behooves us to ask ourselves how we can best support it – and be supported by it.

Does all of this spell doom for actuaries? I am optimistic for the profession in this environment. Our primary focus is in the domain expertise, with sufficient grounding in statistics to support the typically-expected pricing and reserving work. Actuaries often expand their skillsets over time to move beyond these classic exercises and become core members of teams handling other issues companies may face.

Given an actuary’s limited time resources, they should not expect to become a master of all relevant insurance industry domain knowledge areas, the relevant programming and coding skills and statistical and machine learning. Each of these three key categories contains information that would take lifetimes to learn, and each continues to evolve by the minute. As analytics-oriented actuaries, a good grounding in each of the three areas is to be expected. The Casualty Actuarial Society recognizes this need, and its subsidiary, the CAS Institute, created a Certified Specialist in Predictive Analytics specialty track to specifically recognize practitioners who possess just such a foundation. Once an actuary has built this knowledge base, many variations of specialization are possible and the sky’s the limit.