The new era of insurance analytics

Driven by technology, toolkits and talent

By Claudine Modlin and Graham Wright

Claudine Modlin
Specializes in P&C predictive analytics.
Willis Towers Watson, San Diego

Graham Wright
Specializes in P&C predictive analytics.
Willis Towers Watson, Reigate

Advanced analytics is helping some insurers offer innovative products and solutions. What do insurers need to know about the changing nature of analytics and whether it is worth the investment? Claudine Modlin and Graham Wright discuss technology, toolkits and talent – topics that may help you decide.

Q: What has led to the increased attention and high expectations for advanced analytics in insurance?

Wright: While many factors contribute to this trend, it boils down to emerging technologies, a broadening toolkit and an influx of talent.

Vastly increased computational power coupled with cost-effective and scalable data storage has made statistical and machine-learning methods, which have existed for decades, a practical reality. For example, a computational model for neural networks based on algorithms was created as early as the 1940s, but only recently has computing power enabled its use in insurance.

Technology is also enabling the analysis of large volumes of data in a wide variety of data types. We've seen an explosion in types of data from location tracking to web clickstream (what web pages the customer clicked) to the Internet of Things. New types of databases and the ability to store and process data over clusters of servers make analysis of such data possible. So, for instance, insurers can analyze web clickstream data and traditional risk characteristics obtained through the quote process to better understand the customer quote journey.

Modlin: Toolkits are ever growing, largely because of open-source software environments, programming languages and analytical libraries that support a range of analytics techniques. Open-source software simply means advanced analytical methods are more accessible to every member of the analytics community, which is growing rapidly. Moreover, members of this community can expand or add to capabilities through user-created packages.

We're also seeing an influx of quantitative talent into the insurance industry. In addition to actuaries, insurers are hiring statisticians, data scientists, marketing scientists and behavioral scientists. The industry is challenging these professionals to solve a wider range of problems across the customer value chain. This means quantitative experts are working closely with professionals from various functional and business backgrounds.

Q: How important is domain expertise in this new era of advanced analytics in insurance?

Wright: The expertise of data scientists is greatly touted, and many are entering the industry from adjacent sectors such as banking, finance and retail. They bring a wealth of skills related to data manipulation and advanced analytical methods, but not necessarily a detailed understanding of the insurance industry. In fact, some fail to recognize that the insurance industry has a long history of using data to estimate risk.

Data scientists working alone may miss important nuances in the problems facing insurers or the numerous challenges with implementation. However, data scientists can help domain experts tap into a broader array of data sources and leverage machine-learning techniques that reveal new insights about the data. Put simply, we need a balance of data science and industry experience for the best result.

Modlin: Clearly, domain expertise is critical – without this, we run the risk of creating a customer journey that simply won't work. Take the example of pricing for household composition in personal lines. A pure data science approach may recognize that having a disproportionately high number of drivers compared to cars represents higher risk, and domain experts would hardly disagree. However, domain experts would recognize that reflecting this directly in pricing may manifest in situations when a policy drops a vehicle and the policy premium goes up. As insurers uncover increasingly complex relationships that drive risk and attempt to operationalize these findings, a new type of analysis around investigating unintended consequences in the customer journey is developing.

Q: Insurance pricing is clearly one area where even small improvements can reap large benefits for insurers. Do you see any one analytical method used most commonly because it outperforms all others?

Modlin: There is no single method that will be most suitable for every type of data or business problem. It's important to explore different methods and consider how a method will be used.

Methods such as generalized linear models (GLMs) have long been the global industry standard for ratemaking analytics because of their robust statistical framework, high level of transparency and algorithmic form that aligns well with insurance rating engines. Carriers are exploring machine-learning methods in ratemaking, but it's hard to imagine them completely replacing this market-tested primary method because of transparency and deployment issues, particularly in regulated markets such as the U.S. It's more likely in the near term that machine-learning methods will be used to augment GLMs in ratemaking to:

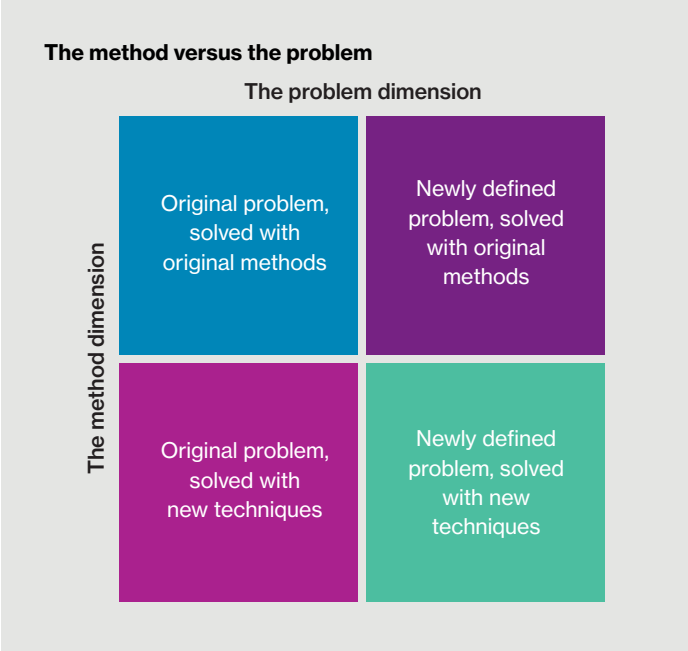
- Create new variables or combinations of variables
- More efficiently whittle down a list of hundreds or thousands of predictor variables
- Identify model shortcomings
- Identify opportunities to build hierarchies of models (i.e., a large number of well-defined, simple models rather than a small number of seemingly complex models)

Wright: Many U.K. practitioners remember a time not too long ago when motor personal injury costs started to increase significantly, and models were not explaining the phenomenon well. It was not a machine but rather humans who observed that injury claims that also had substantial physical damage to the vehicle had very different claim experience than those that did not. Building a hierarchy of models according to this criteria enabled actuaries to much more effectively understand what was driving the claim costs.

We might naturally expect machine-learning methods to detect patterns that other methods might miss. However, it hasn't been easy to identify and explain the unique insights that these methods reveal. Even so, we've made progress in teasing out and visualizing what these methods are learning.

It's also worth noting that we've successfully automated methods like GLMs so they more closely mimic machine-learning results, especially when the modeler is willing to relinquish control and allow even statistical methods to find patterns we may not understand. And this is really the crux of method choice: Practitioners' choices vary depending on whether the ability to interpret or understand are critical requirements of their models.

Clearly, domain expertise is critical – without this, we run the risk of creating a customer journey that simply won't work.



Q: But is the choice of method or model form what really matters in insurance analytics?

Wright: On its own, no – it's one dimension. I like to think of it as asking the machine to do the rote activities and allowing humans to do the clever thinking. The value of our role is in defining the problem to better use the analytics. Consider organizing every analysis into four quadrants (see figure).

For a long time we've lived in the top left corner – focused on doing what we've always done but doing it better each time. The hype around methods and the likes of crowdsourced analytics competitions such as Kaggle drive down the left-hand column, competing over which method is best. On its own, this single component misses the point: The problem must be correctly defined before we throw new methods at it.

Modlin: I agree. Defining the problem may relate to the response we're trying to predict. Claim managers are looking for advanced analytics to help identify potentially complex claims early in the life cycle so they can be managed with appropriate resources and protocols. The challenge is how to define "complex," which often tends to be equated with a high expected claim size or a high probability of being large. However, there is marginal benefit in a model that informs claim managers that fatalities will result in large claims. A more suitable model response for identifying complexity may be related to how much the claim escalates (i.e., the difference between settlement value and early case reserve).

An influx of quantitative talent is strengthening insurers' analyses of data, but a balance of data science and industry experience is needed for the best result.

The characteristics we consider in our predictions may also define the problem. Techniques such as topic modeling open up entirely new predictors by mining unstructured data such as claim adjuster notes.

Wright: In practice, defining the problem is not as simple as a discrete characterization (as the figure shows), but rather a continuous spectrum. The greatest potential to gain insight is likely to come from addressing newly defined problems with newly defined techniques.

Modlin: It's also important that "deep quants" and domain experts agree on the important metrics used to evaluate different methods for the business community. Often, statisticians will rely on pure statistical diagnostics to demonstrate improvement in one model over another. In fact, this is the metric of choice used in the likes of Kaggle competitions. This diagnostic, however, means very little to the business professional who is more likely interested in key performance indicators such as loss ratio improvement or expense savings.

Additionally, different applications merit different degrees of accuracy. A marketing application that aims to segment customers into three broad categories does not require the same degree of accuracy or effort as pricing segmentation. Lastly, before any analytics project is commissioned, it's important for all stakeholders to discuss transparency/interpretability and deployment requirements.

Q: This is all very well, but why should insurers invest in advanced analytics? What is the value?

Modlin: We've demonstrated in many instances of our work that there is a benefit to incorporating a range of techniques in solving insurance-related problems. In some circumstances, machine-learning methods outperform more traditional statistical techniques in predicting losses or other types of customer behavior. But the realization of that value depends heavily on what is required in the implementation. Some applications, such as pricing and underwriting, require a

high degree of interpretability for a variety of stakeholders (regulator, agent, underwriter and customer). Other applications, such as developing target marketing lists or flagging potentially fraudulent claims, may not require the same level of understanding. Also, many insurers currently struggle with deploying some of the outputs from these new techniques, but that varies by application and will change over time.

Technology triggers have put advanced analytics in insurance near the peak of inflated expectations, as defined by the Gartner Hype Cycle. Most would agree there will be some disillusionment before ultimately reaching a steady state of improved productivity. The uncertainty lies in what this journey will look like, how long it will take and the significance of the sustained productivity gains. The general consensus for the industry is that there is more experimentation and learning with advanced analytics than implementation, but that will change. Learning will encourage implementation that delivers early tangible benefits, and continual research will compound benefits and create an early adopter advantage.

Wright: In the U.K. and the U.S., we have seen a wave of once-in-a-generation system transformations over the last few years, which will gradually enable insurers to overcome some of these implementation challenges. That said, there is still the challenge of interpretability.

In the context of method versus problem (see figure), there is proven value in moving both across and down the grid, but the unknown is the extent to which that additional value compounds toward the bottom right.

Even without implementing any wholesale change in approach, insights from data interrogation techniques offer immense value. Some solutions act as enablers and are not attributable to the bottom line. For example, analytics that exploit large portfolios of data can assist more refined segmental marketing strategies and deliver long-term value.

Q: So what is your key takeaway for insurance practitioners?

Modlin: The analytics talent pool and toolkits are expanding, and the insurance industry will undoubtedly benefit. Our industry is challenged to frame analytics initiatives so that the problem is articulated correctly rather than unduly focusing on trying new methods. Methods are evaluated across statistical and business measures, and implementation requirements are always considered. We will have near-term successes and failures, but even the failures will teach us how to leverage advanced analytics going forward.

Wright: We need a balance of insurance domain knowledge and data science expertise to fully benefit from advanced analytics. New techniques that address existing problems will surely drive some benefit, but the real value will come from discerning the identification of new problems combined with the deployment of machine learning on the right tasks. While the role of the traditional analyst is not at risk, it will change as machine-learning methods are deployed on the right tasks, freeing up analysts to focus on more value-added work.

For comments or questions, call or email Claudine Modlin at +1 805 499 2164, claudine.modlin@willistowerswatson.com; or Graham Wright at +44 1737 274650, graham.wright@willistowerswatson.com.

Our industry is challenged to frame analytics initiatives so that the problem is articulated correctly rather than unduly focusing on trying new methods.