



Article from

Actuary of the Future

May 2019

Issue 44

Applying Predictive Analytics for Insurance Assumptions Setting

By Lenny Shteyman

Isn't it amazing that online retailers seem to know so much about us? I recently bought a new phone, and the next time I logged online, the retailer showed me samples of phone cases. As my niece's birthday approached, the site knew I might be looking for another gift online and showcased a few ideas. This is not your old-fashioned customer profiling of the 20th century. This is a fully automated and *individualized* customer-level calculation, designed to predict short-term spending interests *in real time*. For millions of customers, this process miraculously refreshes again and again, based on the new information it receives.

Insurance companies are now looking to catch up with big tech to enable them to make better, faster and bolder decisions. Actuaries, responsible for pricing or assumption setting, are evaluating opportunities to integrate predictive analytics into their work. For those who have not yet switched from traditional actuarial techniques to a 21st-century toolkit, let's first start with professional geography:

- **Short-term versus long-term products.** Elaborate predictive techniques became a staple for Property & Casualty (P&C) and health-care insurance companies, largely due to data availability and projection horizons. Unlike life insurance companies, which experience fewer transactions but often focus on long-term risks and predictions, P&C and health-care companies have many more transactions per client.
 - **Company size.** Larger companies typically have more mature governance models and could experience a larger financial impact from an assumption switch. If the in force is small, on the other hand, it's easier to accept material changes in methodology. All else equal, there will be fewer challenges to make a switch in a smaller company, although own data of a smaller company might be insufficient for building a predictive model.
 - **Pricing versus valuation.** A product cannot be priced and launched without new assumptions, which can make it easier to adopt a new technique for pricing. The financial reporting function, on the other hand, already has an established assumption for its in force. Switching away from established assumption will require a justification and business case.
- There are many benefits of using advanced techniques: using own data more efficiently, consistently capturing insights into main drivers, discovering new predictors, developing a more granular view, reducing future reliance on business experts, repeatability and reproducibility of research, and so on. With that said, for an actuarial assumption-setting purpose specifically, I believe it is reasonable to suggest exhausting “small data” solutions before jumping into “big data” ones. Not every actuarial problem merits a build of a statistical model, although the most interesting and complex problems are certainly good candidates. Whether tackling the advanced techniques on your own or collaborating with IT and data scientists, consider this high-level list of questions:
- **Success definition.** What will make this project successful?
 - **Data availability.** Do you have enough clean and relevant data?
 - **Granularity.** Will you need a granular answer? (e.g., customer or producer level)
 - **Problem complexity.** Are you solving a complex multi-dimensional problem?
 - **Prediction period.** Will you need a short-term prediction or classification? (versus long-term prediction)
 - **Calculation frequency.** How often will you be repeating this calculation?
 - **Business impact.** Are you ready to justify the cost and duration of a predictive analytics project?
 - **Implementation constraints.** Are you free to implement the solution as you see fit (versus how actuarial projections software recommends)?
 - **Transparency.** Can decision makers accept complexity or reduced transparency when the model is first developed or enhanced?
- If you answered yes to at least six of these questions, you should consider adopting a predictive model.

Let's elaborate on these high-level dimensions.

SUCCESS DEFINITION

Data scientists are not magicians; they are business professionals and success definition is critical to all projects, not just predictive ones. Even though deep learning allows one to search for relationships not previously known, and artificial intelligence sounds like a self-improving magical golem, success of a predictive project needs to be specific and measurable. What do you want to learn more about? Are you looking to make predictions for probability of a customer buying a product? Assign an underwriting class? Determine a price? Determine expected success rate of a conservation effort? Achieve a certain minimum financial impact? In addition to success being measurable, it is crucial to communicate what insights already exist, why they are insufficient for the big question and most important, what will you be able to accomplish with the future predictive findings and insights?

DATA AVAILABILITY

Data scientists cannot provide quality insight without a large amount of relevant data. Further, quantity of data is not the only dimension to consider. For example, male data would not be relevant for predictions made regarding females, and nonsmoker data would not be relevant for predictions made regarding smokers. Another example—some interest-rate-sensitive products use dynamic lapse formulas that depend upon the level of

interest rates. Since we have not observed 10 percent—or zero percent—interest rates during the lifetime of such products, even if general amount of lapse data is satisfactory, the amount of lapse data in the high and very low interest-rate environment is nonexistent. Any model making extrapolation predictions based on nonexistent data will be weak.

GRANULARITY

A prime example of granularity is predictive underwriting that could lead to customer level pricing. If the business decision is made at a granular level, the prediction must be made at that level as well. In some cases, however, granularity isn't always needed. The assumptions for financial reporting are often set in aggregate. While it might be tempting to set assumptions at the most granular level possible to use as building blocks for all future uses, this could be prohibitive due to data availability and project costs. At the same time, a more granular assumption would be beneficial for a scenario when conditions change, or population mix shifts over time.

PROBLEM COMPLEXITY

There are many dimensions and potential predictors of customer mortality or behavior. If current assumptions are no longer satisfactory, it might be more efficient to capture all such new insights in a predictive model. Another reason why a classical technique might be lacking is due to interdependencies. Size of a policy, for example, could signal a customer's ability



to access higher quality medical care, efficient decision making while utilizing riders, and so on.

PREDICTION PERIOD

As previously mentioned, there is a distinction of short-term versus long-term products. Even with long-term products, some questions focus on the short-term horizon. Predictive underwriting is an example of a classification problem, which identifies the underwriting class and price level to assign to a customer. Similarly, predicting a customer's propensity to react to an in-force management action could work very well for a short-term horizon, provided the model is sufficiently trained, but not necessarily for a long term. The long prediction period challenges could be partially mitigated if the observation data period is long enough.

CALCULATION FREQUENCY

Each year, assumptions are updated for financial reporting. The customer-level pricing engine, however, could be run multiple times per day. For example, companies that have success in automating pricing on smaller contracts, which would be otherwise cost-prohibitive to bid on, demonstrate this is a clear win for predictive analytics.

BUSINESS IMPACT

The timeline of a predictive project is not short. The process includes understanding the data and performing the analysis, performing data cleaning as necessary, understanding known insights and achieving a satisfactory solution for the stakeholder. Obtaining a buy-in from the decision makers can also be challenging and time consuming. Some useful models could be built relatively quickly; however, a complete timeline of a successful project end-to-end could sometimes span nine to twelve months.

The cost estimates of a predictive solution need to factor in data storage, acquisition of new data sources, more powerful computers and talent with subject matter expertise. It's not inexpensive to hire data scientists, either. Only the most important and complex assumptions would likely justify the cost and timeline, with understanding that additional insight and granularity will add incremental business value, not just precision to a financial reporting calculation. One also needs to factor long-term benefits of switching to predictive techniques in decision making, instead of short-term gains only.

IMPLEMENTATION CONSTRAINTS

Once a solution passes all the previous tests, it will need to be "productionalized." Ideally, by the time the research is complete, the testing code can be elevated to production. If the assumption must be implemented in an actuarial software package, however,

a few challenges may arise: modeling resources may have other competing priorities, or the package may not be flexible enough to easily adopt a predictive model. There may also be challenges with input complexity, model validation, runtime or model convergence.

One solution is to simplify the assumption used in the actuarial software package. Another is to perform elaborate calculations outside of the software package, such as in an executable file and feeding calculation output into the software package.

TRANSPARENCY

Some nonparametric predictive methods lack transparency. If the assumption model is a complex black box without a strong validation of business sense, the management team needs to decide if they are comfortable using it for pricing decisions. If only the predictive power is demonstrated without clear-cut first principles logic, such a solution may be acceptable for the management of a proprietary trading firm, but culture constraints will make it difficult to adopt such a model for pricing decisions in insurance companies.

A chief financial officer for a life insurance company who focuses on quarterly earnings may not see the value in a logistic model with 50 predictors. However, a simpler model with five key drivers will capture most of the predictive power and stands a greater chance of providing assurance for the CFO. An assumptions model with unpredictable outputs conflicts with the transparency your CFO is looking for from the actuarial team.

It is easy to see why predictive underwriting application for life insurance became popular. It promises economic gains and its customer-level risk selection and assumption setting is closest to what online retailers benefit from. While some precision may be lost early in the process, new pricing insights and operations efficiencies will make up for it. Yet, not all assumption-setting situations are as clear-cut in terms of having the benefits outweigh the costs. In addition to exploring the data, it's as important to use solid judgment, as well as leverage actuarial standards of practice (ASOPs) for assumptions setting, modeling or use of data. Some situations will require quicker and directionally appropriate decisions, based on limited data. Such an environment is potentially more fit for actuaries than it is for data scientists, as actuaries are accustomed to using their deep business knowledge, sensitivity testing and professional judgment. ■



Lenny Shteyman, FSA, MAAA, is VP and actuary at Prudential Financial. He can be reached at shteyman.lenny@gmail.com.