Blinded by Predictive Analytics

By Bryon Robidoux

This article is about a great TED talk that I watched recently titled, "The Human Insights Missing From Big Data" by Tricia Wang. You can watch it yourself at *https://www.ted.com/talks/tricia_wang_the_human_insights_missing_from_big_data*. As I watched her speech, it really occurred to me how important her insights are to actuarial science, modeling and predictive analytics. I thought it would be worthwhile to rehash her main points and apply them to modeling in general.

Dr. Wang starts her lecture by stating that big data is a 122 billion dollar industry, but 73 percent of big data projects are not profitable. Big data and predictive analytics are not giving the breakthroughs that companies are expecting. "Investing in big data is easy, but using it is hard." Her speech focused on why companies are not receiving insights from their big data. She gives the example of Nokia. Before the iPhone came out in 2007, Nokia was the dominant player in the cell phone market, where she was a consultant. As part of her job, she hung out in China with poor Chinese youth in cyber cafes trying to understand their spending habits. She realized, that even though an iPhone or its Chinese knock off cost half a month's salary, the poor would do almost anything to purchase one. After achieving her insights, she took them to Nokia. She explained how she saw a fundamental shift coming in the purchasing habits of the Chinese youth. She pleaded with Nokia to change direction and realize that smart phones are the next market disruption. According to Dr. Wang, Nokia's response was to look at the big data predictive model and state that they have no evidence of her perceived emerging trend. Her 100 diverse data points are not as reliable as their big data models with millions of data points. Intuition and anecdotes are not enough evidence to act upon. Shortly thereafter, Nokia tanked!

The smart phones skyrocketed and, as of today, my wife doesn't even remember Nokia's existence. What if you found yourself in an insurance company that invested multi-millions annually in data science and modeling, how would you handle it if an ethnographer like Dr. Wang said that she had insights into the future of insurance? Would you ignore her and trust your models or would you step back and find a way to double check her findings?

I don't want to sound judgmental of Nokia because it is easy to look back and say management should have been more aware of the signs. What would have happened if Nokia would have listened to Dr. Wang? Let's imagine the reality they would have faced and the questions they would have had to ask themselves.

- Are we really this vulnerable?
- How do we confirm Dr. Wang's theories?
- Who is responsible for seeing the trends?
- What are the deficiencies in our models?
- What are the deficiencies in the data provided to the model?
- How did we miss this?
- What will this mean for our bonuses and our jobs?

- How are we going to retool for the future to compete?
- What will retooling cost?
- How do we explain this to our senior managers or board?
- How do we explain this to our stockholders?

It is actually much easier for Nokia to take comfort in their dominance and their perceived information bias. The harder and scarier scenario is to admit that Dr. Wang was correct and retool accordingly. Given human nature is to follow the path of least resistance and take comfort in our computational bias, can you now see how Nokia's response is exactly what you should expect?

This scenario could theoretically happen to an insurance company or insurance industry. What happens if regulations suddenly changed to allow Facebook and Google to sell life insurance or property casualty insurance? Think about how much personal detailed information people share online and how much that says about their behavior and their risk aversion. What if Facebook or Google could use their data to better predict claims and weed out anti-selection? What if they could more accurately set rates because their data is better at predicting policyholders' behaviors and their propensity for moral hazard? What if they could better predict how policyholders perceive value and out sell the rest of the industry? What if they were perceived more transparent and trustworthy to policyholders because of their brand recognition? Suddenly, the insurance industry could be in the same position as Nokia.

The next part of Dr. Wang's speech was about why Nokia was blinded by their big data model. All of Nokia's data was collected in the past. The questionnaires, surveys and other market research was based upon existing business models which greatly biased their insights to well-established historical trends. It is important to realize that predictive models work well in closed systems, such as delivery logistics, genetic code, electric power grids, death and disease. Big data fails in dynamic systems, especially when modeling human behavior, because once a pattern is established a new dynamic comes in to destroy it. Plus, if the modelers are not forward-looking, then how can their models be forward-looking? The important point of Dr. Wang's speech was to point out that it is not good enough to look at the behavior the model is predicting today. It is important to deeply understand the reinforced biases in the data and try to supplement with other sources to validate the accuracy of the model's predictions.

Dr. Wang coined the phrase "thick data" in her TED talk. It is the data that is small in quantity, gathered from various unorthodox sources and very difficult to quantify. "It needs stories, emotions and human interactions. What gives thick data its meatiness is its ability to explain the human narrative. Thick data grounds the business questions in human questions."

What happens if regulations suddenly changed to allow Facebook and Google to sell life insurance or [P&C] insurance?

It was this thick data she was using to validate the results of the Nokia predictive models. It was her ability and education to look outside of the traditional data collection and see the emerging trends. She stated, "It is the mixture of thick data and big data that gives companies their insights. Relying on big data alone increases the chances we will miss something, while giving us the illusion we know everything."

As actuaries we are bombarded by models every day. We are either using results coming from someone else's model or we are producing results that someone else will use in their model. As we look at the behavior of our model, regardless of whether it is a predictive, valuation or hedging model, we need to be using sources outside of the model to validate its correctness. We can't be looking at the model's results as an Oracle without looking at the thick data to make sure the model is capturing the emerging trends. We need to ask, why are these the results? We need to step back and look at the big picture to see the dynamics of the system as a whole. As you build economic scenario generators to value the business based on some probability distribution, do you ever stand back and ask what events would lead to the worst case scenarios? Or do you just take them as gospel and move on with your life? This is the difference between producing model inputs and collecting thick data to ask the important question of why.

This leads to the next important topic of quantification bias, which is the unconscious belief of valuing the measurable over the immeasurable. As a profession it is really easy to fall into this trap because all the ASA exams focus on weeding out candidates based on their ability to crank out precise values to existing actuarial models. It is only in the fellowship courses that there is any importance placed on practical qualitative models. This approach to giving exams can lead our profession to have an overreliance on our models and place much less importance on qualitative measures. As actuaries we are trained to be technical, detailed oriented and Excel loving calculating machines, which runs counter to looking outside our models for answers. Dr. Wang explained that she sees a lot of companies throw away data because it doesn't fit nicely into an existing model or insights weren't produced by a quantitative model. The more we rely on the models, their complexity grows, and they become more automated, the more we are removed from their details so we get comfortable with them,

disconnect ourselves from them and accept their results without question. The most important point of all is that more data doesn't mean better output or more predictive power.

The variable annuity business is a perfect example. A large driver of the value of that business is wrapped up in the policyholder's propensity to lapse and the utilization pattern of their benefits which are based on the perceived value of the annuity, the surrounding market conditions and the competition among variable annuity writers. If we look at the short history of variable annuities then we can see that pre-great recession there was a huge arms race to write variable annuities and their benefits became riskier to win customers. Was there thick data available to tell us that the market was going to tank the way it did? Back in 2003 to 2008 my dad was a real estate agent in Lincoln, Nebraska. When he would come to visit he would talk about the housing market. He would say things like, "We sold a $500,000 house to a couple that made $60,000 per year. They had less than 3 percent down. The banks had no issues accepting them. I don't know how this is sustainable, except that the banks are selling the loans to the market." With hindsight, he was predicting the major cause of the recession. I ask myself if I would have been responsible for managing a book of variable annuities at the time would I have been wise enough to research my father's insight to hedge potential losses. This

is exactly what Dr. Wang suggests that we do. This is the true nature of thick data.

In conclusion, the TED talk by Dr. Wang is an important reminder to actuaries. It is important to not get oversold on the huge hype of predictive analytics and big data. In a dynamic system like insurance, looking at past data has very little use unless you are using thick data to supplement it. It is important to validate that your predictive model is relevant to capture future behavior and understand its inherent biases. Thick data is nothing more than using quantitative data along with human questions to gain further insights into the results of the big data model. More importantly, regardless of the type of model, it is important to always ask if the model is still relevant and why is that. Don't ignore data or analysis just because it doesn't fit into the current model. The more dynamic the system being modeled, the more important it is to constantly question the model and its results. ∎

Bryon Robidoux, FSA, is director and actuary at AIG in Chesterfield, Mo. He can be reached at *bryon. robidoux@aig.com.*