SOCIETY OF ACTUARIES

Article from

Forecasting and Futurism

Month Year July 2015
Issue Number 11

Stepping Out

By Bryon Robidoux

An overview of the Predictive Analytics World (PAW) conference.

At the end of March, I attended the PAW conference in San Francisco. I first discuss the general structure of the conference so you can relate it to the SOA conferences. Then I will discuss the R class that I attended. Lastly, I will explain the many important overtones that rang through the different presentations at the conference.

The conference was setup in three parts: pre-conference, conference, and post-conference. The pre- and post-conference were day-long classes. They had classes for beginners all the way to experts. The conference itself was only two days long. It was divided into two tracks. The first track was for anyone. The second track was for practitioners and experts. I mainly stayed with the first track. For each day of the conference, there was a breakfast keynote address and a lunch keynote address and then after each keynote we would break off into our track of interest. I really enjoyed having classes before and after the conference. Given that I am only allowed one paid conference per year, this allowed me to attend a couple of training courses in addition to going to the conference. If the Society is not already doing this, I think this would a great idea to try.

The *R for Predictive Modeling: A Hands-On Introduction* class was by far my favorite part of the conference. I easily could have attended five days of this class and skipped the rest of the conference. It was supposed to be a hands-on class, but there was so much information that there was no way to listen to him and have time to run the code myself. Max Kuhn coauthored a book *Applied Predictive Modeling*, which I highly recommend. This class was really just a cliff note version of his book. In this class, he focused a lot on preparing data features for the models, such as centering, scaling, and removing skew using the Box-Cox transformation. He discussed which models require the transformations, such as linear regression or principal component analysis, versus models that do not need the data transformation, such as decision trees. He focused on preparing testing and training data. It is obvious that the less data the actuary has the more important it is to "spend" the data wisely. He explained the importance of stratified sampling to make sure the training and test sets have the same distribution of event outcomes. Given the correct features are in the model, he discussed the difference between bootstrapping and K-fold cross-validation of the model. Bootstrapping is sampling with replacement and the sample is the same size as the original dataset. K-fold cross-validation is sampling a subset of the test data and using the remaining portion to test the prediction. This is done K times, where K is usually five or 10. He talked at great length about comparing different model measures, such as R squared, root mean square error. His preference was to use mean square error over R squared because you can get a high R squared by just having a very volatile distribution. He talked at great length about comparing different classification models using Lift Curves and the Receiver Operating Curves (ROC). Overall I would give this class an A. If you can't make it to the Advanced Business Analytics Seminar in Philadelphia this year, then this would be a great alternative. As an added bonus, it is given four or five times a year. I hope to attend one of Max's classes next year.

In terms of the conference, it was very interesting because I was able to see all the different uses for predictive modeling outside of an insurance context. I really enjoyed seeing how other disciplines frame their problems and derive solutions. This usually gives a fresh insight into problems that I am trying to solve. There definitely were some interesting themes throughout the conference. In the keynote address, "The Revolution in Retail Customer Intelligence," the major theme was to look past the demographic information and collect or transform the data in a form that has a better "signal" of intent and behavior of the individual. His example was trying to explain why customers behave as they do when shopping on the Internet and how can you turn this knowledge of behavior into more sales. For an actuarial example of a GMWB policy, the age of an annuitant really gives little information about the potential behavior of an annuitant, but knowing the number of years until retirement or number of alternative income sources gives far more color on utilization of the benefits. If the data is encoded in such a way that the modeler can determine intent and behavior, the predictive model will be better at prediction.

The second keynote, “New Challenges: From Predictive to Prescriptive to Automated Analytics” was trying to address the difference between data versus information. Data can seem infinite and it is currently growing at an exponential rate, but the information extracted from the data is finite. The issue is how are actuaries, statisticians, and data scientists going to be able to process all of the data to extract the information. The short answer is that they are not. Dell’s answer to the solution is to automate model calibration, modeling, and responses to the output. In the future, the models will self-calibrate and have little or no interface. In the future, the data scientist will be mostly involved in making sure the process is behaving as expected. Personally, I would not want to be a data scientist at that point because the creativity and art would be sucked out of the science.

The third presentation I want to mention is “Making Impacts Through Analytics,” by Bin Mu from MetLife. In many presentations including this one, presenters discussed the importance of communication. If the modeler does not have the ability to communicate the importance of the model and its potential return on investment, then the model is of little use. I especially enjoyed this presentation because Bin took this concept a step further and described a framework for modeling: define business objectives, analyze information and draw insights, identify actions from insights, and measure business impact. At the end of the day, if you are able to describe your projects with these four steps, then you will always be able to justify your work to the C-Suite. The point he drove home over and over again was to design experiments so that the results and outcomes are easily measurable. One of his major frustrations was with internal clients asking him to do studies, but the business owner really had no plans for using the information outside of a standard report that will more than likely be ignored. He turned the question around and asked the client how to make a request actionable and how can the request improve the business.

I REALLY ENJOYED SEEING HOW OTHER DISCIPLINES FRAME THEIR PROBLEMS AND DERIVE SOLUTIONS. THIS USUALLY GIVES A FREASH INSIGHT INTO PROBLEMS THAT I AM TRYING TO SOLVE.

The last item that I would like to mention is about the expert panel on “Education and Training Options for Predictive Analytics.” In this panel, they debated the lack of data scientists in the future and the education required in getting the industry up to speed. It was explained that there will be a deficit of 140,000 to 190,000 jobs in data science in the next three years. Schools are scrambling to train people for these positions. They discussed at length the steps of evolving into the field:

1. Read high-level books;
2. Watch instructions on YouTube, which I thought was interesting;
3. Massively Open Online Classes (MOOC), for example Coursera.org;
4. Certification from a university to supplement your professional experience; and
5. Get a Master’s Degree.

The last topic discussed was the qualities of a good data scientist, which is very similar to the qualities of an actuary. Besides being good at math, a person needs to be creative, innovative, and most importantly, have good communication skills.

I do think it would be an interesting debate to determine when someone becomes a data scientist versus a statistician versus an actuary. How and where do you draw the line between these disciplines? The bigger question that went through my mind as I was enjoying the conference was, as

CONTINUED ON **PAGE 42**

STEPPING OUT | FROM PAGE 41


Bryon Robidoux

actuaries, how do we fit into the big data/predictive modeling revolution? Are we on the sidelines or are we active participants? Does the combination of course P and course C qualify us as a data scientist? It was interesting, while talking to a gentleman in my R class, he explained that he wanted to use predictive modeling to determine when Pepsi-Cola machines were most likely to fail. I have 25,000 good friends that can help him solve that problem! The disturbing part of the conference for me was that I think I was the only actuary that attended. (They provided a mobile app, which allowed me to see the occupation of other participants at the conference.) Notice that the room full of statisticians and data scientists did not put on their list of possibilities of becoming a data scientist to first become an actuary. I would love to see how actuaries were perceived by this audience and the role we should play.

In conclusion, I was glad that I attended this conference. I definitely think it as worth my time to go, especially because of the day-long training classes before and after the conference. If possible, I would like to go again next year and take more advanced classes. (It doesn't hurt that I got to enjoy fresh crab on the Fisherman's Wharf while gazing at the Golden Gate Bridge in the background.) But the question that churns in mind is how do actuaries fit into the predictive analytics revolution?

Bryon Robidoux, FSA, CERA, MAAA, is director of Hedging, Global Financial Solutions, at RGA Reinsurance Company in Chesterfield, Mo. He can be reached at brobidoux@rgare.com.