3Qs with Adam Stotz

As TROVE's Chief Technology Officer, Dr. Adam Stotz leads the research, design, and commercialization of the company's Predictive Data Science software, advancing state-of-the-art possibilities in prediction to help clients make their data more useful. Prior to TROVE, he led Data Fusion and Predictive Analytics projects for the DoD and Intelligence Community, demonstrating the efficacy of technologies now being commercialized by TROVE. Stotz's primary research interests are in applied Operations Research and Machine Learning.

1

What predictive data science technologies or techniques are you finding most exciting or useful now?

Deep learning is an area that fascinates me. Let me compare it to machine learning to demonstrate why.

If you look at machine learning, it's very dependent on how well you structure the inputs into the process, what's called **feature engineering**, and this greatly dictates the quality of the outputs. Deep learning on the other hand is a lot less sensitive to feature engineering. If you think of machine learning as discovering relationships between variables that you provide it, with deep learning you simply hand it data, and it not only finds the relationships between the variables, it also finds the variables!

We use deep learning right now in some really advanced image-processing work we are doing for a large utility. With deep learning, all we have to show the model are examples of transmission poles. We don't have to tell it to look for a "brown cylindrical object sticking out of the ground." You don't *tell* it what to look for, you just *show* it examples. And in the future, what we see as feature engineering today could become "data set engineering" tomorrow, where data scientists are spending most of their time making sure deep learning networks are given good examples.

Deep learning can be extended easily into an area like load forecasting, where we currently do a lot of work at the meter-level. Today, we instruct our models to look at data such as hourly load magnitude, including how often to look and for what historical duration, along with a number of other parameters.

If we put these inputs through a deep-learning process, we can potentially have a model that becomes a lot more accurate because it will use features of those input data sets that we didn't think to model. For example, if we tell the current model to look at hourly load magnitude for the last 45 days, a deep-learning model might look at the last 40 or 35 days instead based on what it sees. This takes a lot of repeat testing out of the picture.

So, instead of throwing 100 darts and hoping to hit the bullseye with the right combination of variables and values, deep learning will throw the best bullseye it can with one dart.

→

TROVE™

2

What is the difference between AI and predictive data science?

I think there is a lot of imprecision and confusion around much of the data-science lexicon. Let me see if I can clarify things a bit.

First, we need to separate "prediction" from "data science." Data Science is an approach, a rigorous methodological approach, for developing models that deliver value. Those models may be predictive or not, be machine learning or deep learning or neither.

To me, AI is broader than just prediction. So, if you want to compare AI to predictive data science, I would say that predictive data science is a scientific methodology we would go through to produce a particular type of AI model, i.e., a predictive one.

Then I think it is useful to compare "AI" with "machine learning." Back in the 1950s, the fathers of the field, Minsky and McCarthy, described artificial intelligence as any task performed by a program or a machine that, if a human carried out the same activity, we would say the human had to apply intelligence to accomplish the task. This is a little broad, but point being that AI can *use* machine learning to develop its intelligence, but other forms of AI models exist such as agent-based models which develop their intelligence differently.

So let me put the two together: Machine learning is something that AI uses to learn about a data set or situation to develop the intelligence it needs to make informed decisions about a matter at hand.

3

What do you think the enterprise is getting right or wrong in terms of their expectations of both AI and predictive data science?

I think there is a misconception that both are still some kind of magic wand, i.e., that you just have to run data through a machine-learning model and you're going to get good results. The rigor of the data science process continues to show us that we haven't moved beyond "garbage in, garbage out."

Another misconception is that you are not always going to find predictability in data. Actually, when you embark on a project and are trying to predict something from the data you have, a valid result might be that there is no predictability in the data whatsoever. In that case, you learn that the nature of the problem itself is unpredictable given your observations of it, which is a valuable result.

Sometimes people don't like to hear that, but think about how valuable it is to know that looking at a certain problem in a certain way isn't going to produce the result you want. It's better to know, so you can move on quickly to a more fruitful opportunity.

The good news is that AI is game-ready when it comes to *narrow intelligence* – the kind of intelligence that can be used to solve specific use cases within an enterprise. When it comes to *general intelligence* – the kind that people refer to about Skynet conquering the human race – well, I think we're a long way off from that.

TROVE™

For more information, please contact Sales@trovedata.com.