Hewlett Packard Enterprise

Hewlett Packard Labs

What I Thought About Data and the Physical Sciences and Why I Was Wrong

Author: Ray Beausoleil

I was certain, five years ago, that machine learning techniques and data-driven techniques would never be relevant to doing the kind of experiments you see as the face of the physical sciences.

When you’ve got enormous amounts of data and need to analyze it, sure, there’s a place for data techniques. But at the front end of science, where you're asking questions, trying to find answers, and then from those answers developing new questions, I didn't think data science would have an effect on how the physical sciences were done.

Here’s why I thought I was right and how I turned out to be wrong.

The reason for my skepticism was that the linchpin of science is explainability. When you’re using neural networks to identify images, for instance, it can be very, very difficult to understand how it’s actually doing the identification. So I thought there was no way that you were going to be able to apply neural networks, in particular, or machine learning techniques in general, to understand physical systems.

Explainability, sparsity, and top-down examination

It is, in fact, possible to use data-driven techniques in such a way as to preserve explainability. Here’s an example. One of the ways to achieve explainability is through a technique called sparsity. You don’t need to understand all of the features of a system to understand the system as a whole as long as you can understand which features of that physical system are the most important in determining its dynamic behavior.

Over the last 20 years, there has been a pretty significant mathematical advance in system identification. It's kind of trivial when you're doing regression with the L1 norm instead of the L2. It turns out to be the thing that helps you to really drill down and enforce sparsity in a system to reduce the number of characteristics of that system that you really have to pay attention to in order to understand why it's doing what it's doing, or how it's doing what it's doing.

This is a really big deal because the fewer the elements that you have to hold in your puny human mind to understand what's going on in a physical system, the better off you're going to be. This is why for centuries, scientists have tried to reduce systems down to the minimum number of microscopic constituent parts because doing anything more than that is extremely complicated.

Imagine trying to understand how the brain works and how the mind might arise from the brain by starting out with a description that is a physically and chemically accurate picture of how neurons fire and then trying to deduce from the behavior of those neurons what it is that the brain is doing on a very high level.

That's how scientists have been thinking for a very, very long time. But now we have the ability to look at a complex system from the top down and to begin appreciating patterns of dynamic evolution in these physical systems by pulling out a couple of key features and realizing that those are, in fact, the most important things that you have to be able to understand.

This realization took me completely by surprise, but now that it's out in front of us, I'm absolutely delighted and excited about it. I think it's phenomenal.

A Taylor series

One of the techniques that we have used in science for a very long time has been perturbation theory. And so you start from a point that you understand quite well, usually in a linear system, and then you say, "Okay, well, suppose I move in parameter space, just a tiny distance away." You invoke something called a Taylor series to do that.

Perturbation theory has been used in a huge number of physical theories for a long time. In fact, there was a paper published several years ago in which Max Tegmark started from the standpoint of, "Hey, all there is in the universe that is interesting can be described by Taylor series."

That's why neural networks work because Taylor series allow a massive reduction in the amount of data that you have to understand in order to describe a physical system. In retrospect, I think that paper got it absolutely backward and upside down. If what you understand is how to use perturbation theory and Taylor series, then the sorts of phenomena that you try to investigate are going to be those that you would use perturbation theory and Taylor series to describe.

Using machine learning to examine the physical world

We're people, and our brains are small, so we look for ways to reduce problems to something that we can understand. This is one of the reasons why we embrace the process of looking at big, complicated physical systems from the top down, and with the fewest possible number of elements, we need to understand what we're looking at.

For the longest time, we thought we'd have to attack these problems by starting at the bottom and working our way up. The thing that is surprising and wonderful is that we can now look at them from the top down, using machine learning techniques to reduce the problem to something that we can fit inside of our heads.

Think of this exciting question, one which we couldn't even ask before. Is it going to be possible at some point in the future to look at the brain from the top, at the results of MRIs and other brain scans, and deduce patterns that are, in fact, quite simple but hidden and which help us understand what the brain might be doing without having to worry about the underlying physical mechanism of neurons and any other microscopic connections in the brain?

I say yes.

Now I am not saying that AI will by itself come to conclusions about the physical sciences. My colleague Paolo Faraboschi put it best. He said scientists who know how to use computers have pretty much replaced those who don't. Now it's the same thing for AI. Today's science is like the rest of life. You are so drowned in data that if you don't know how to use AI to manage it, to assist in understanding its patterns and implications, you're pretty much out of luck.

The state of things is such that the physical sciences, which for a long time looked resistant to data-driven techniques and machine learning, are beginning to embrace them to great effect. This union of physical and digital investigation is destined to produce more and better exciting insights.

The more we investigate, the more we understand that silos are only good for grain. For science and for data, for understanding and for revelation, collaborative interaction between disciplines is (to use a metaphor derived from the new LiDAR-powered archaeology) the only method that will allow us to see beneath the trees.

Learn more at

hpe.com/us/en/hewlett-packard-labs

a00128628enw

Hewlett Packard
Enterprise