

How machine learning helps scientists hunt for particles, wrangle floppy proteins and speed discovery

1 October 2020, by Erika K. Carlson



Physicist Daniel Ratner leads SLAC's machine learning initiative. Credit: Dawn Harmer/SLAC National Accelerator Laboratory

Machine learning is ubiquitous in science and technology these days. It outperforms traditional computational methods in many areas, for instance by vastly speeding up tedious processes and handling huge batches of data. At the Department of Energy's SLAC National Accelerator Laboratory, machine learning is already opening new avenues to advance the lab's unique scientific facilities and research.

For example, SLAC scientists have already used machine learning techniques to operate accelerators more efficiently, to speed up the discovery of new materials, and to uncover distortions in space-time caused by astronomical objects up to 10 million times faster than traditional methods.

The term "machine learning" broadly refers to techniques that let computers "learn by example" by inferring their own conclusions from large sets

of data, as opposed to following a predetermined set of steps or rules. To take advantage of these techniques, SLAC launched a machine learning initiative in 2019 that involves researchers across virtually all of the lab's disciplines.

An accelerator physicist by training, Daniel Ratner has worked to apply machine learning approaches to accelerators at SLAC for many years and now heads up the initiative. In this Q&A, he discusses what machine learning can do and how SLAC is uniquely equipped to advance the use of machine learning in fundamental science research.

What is machine learning?

Machine learning programs solve tasks by looking for patterns in examples. This is similar to the way that humans learn. So machine learning tends to be effective at tasks that humans are good at, but it's hard to explain why.

For instance, you can teach your teenager how to drive a car by example. But it's hard to write down a set of rules for how to drive a car in every possible situation you might encounter when you're driving. That's the kind of case where machine learning has been successful. Just by watching someone drive a car for long enough, a machine learning model can begin to learn the rules of driving.

It usually boils down to learning how to do something by watching enough data.

How is machine learning different from traditional methods?

It's a different conceptual approach to problems. Rather than writing a sophisticated computer program for an entire complex data analysis process by hand, machine learning shifts the

emphasis to developing a data set of examples and a way to evaluate solutions. At that point, I can hand over my raw data to a [machine learning model](#) and train it to predict solutions for new data it hasn't seen before. You can think of it as a way of avoiding that onerous, expensive programming by extracting rules hidden in a data set.

Let's say we're doing some high-energy physics, analyzing particle tracks in a detector, which is how particle physicists learn about nature's fundamental components. Rather than writing algorithms by hand that manage each part of the analysis—removing noise, finding tracks and identifying what all the particles are that created them—and building the analysis process up bit by bit, we can just take a big chunk of simulated data and learn how to do that entire analysis pipeline with a single big neural network, a machine learning technique inspired by human neural systems. And in practice, the machine learning methods often out-perform their human-written counterparts.

Why is SLAC interested in using machine learning?

In science, the area where machine learning has gotten the most press is in data analysis. A typical task would be: "I have a big data set and I want to extract some science from it." And certainly we do a lot of that as well at SLAC. But there's actually a lot more that we can do.

Because we run all these big scientific facilities, we think about how machine learning applies not just to data analysis, but to how scientific experiments at these facilities work.

We can use machine learning to address questions like "How do I design a new accelerator?", "Once I've built it, how can I run it better?" and "How can I identify or even predict faults?"

For example, we're building the next-generation X-ray laser LCLS-II, which will generate terabytes of data per second. A new project led by SLAC will develop machine learning models on the facility's detectors to analyze this enormous amount of data in real time. This model can be flexible and adapt to

the individual needs of every future user of LCLS-II.

Every level of a scientific experiment—from design to operations to experimental procedure to data analysis—can be changed with machine learning. I think that's a particular emphasis for a place like SLAC, where our bread and butter is running big facilities.

What are other examples of machine learning at SLAC?

One example is in improving our ability to analyze how a protein molecule changes over time on the atomic level. A protein is a floppy, flexible thing, and that motion is essential to the protein's function. Rather than trying to learn the average structure of a protein by taking a blurred picture of a moving object, we would like to make a movie and actually watch how that molecule is moving. There's been some very interesting research on using machine learning models to make these protein movies.

As an example of the particle tracking idea mentioned earlier, we have a group at SLAC applying machine learning to neutrino detectors. The task here is to look for very small tracks when particles fly through huge three-dimensional detectors. The scientists have been doing that using something called sparse models, which speed up the learning process by not allocating computational resources to empty space without any tracks. These sparse models are both faster to train and more accurate compared to the standard neural networks developed for the analysis of everyday images.

And it turns out that we can actually use that same concept in very different areas. For example, in materials science, you might want to be able to identify a single atom and ignore the vast area around that atom. So even with different scientific goals in mind, we can use the same boundary-pushing machine learning models. Having the machine learning initiative allows all these different people to talk to each other, share experiences and ideas, and make progress faster.

What motivates you personally to work on

machine learning?

I've always been intrigued by the possibility of extracting valuable information from seemingly random data. I think this search for structure in noisy data is what draws me to science in general. There's also a lot of overlap with the non-machine learning science projects I've chosen the last 15 years.

Science is a place where you often have large datasets and ask concrete questions, and that is exactly the setup that makes machine learning successful. And machine learning allows us to ask entirely different types of questions that we couldn't before. That's going to lead to very exciting science.

There are many people at SLAC who have been doing machine learning for a long time. Now we're codifying our lab-wide approach to machine learning and providing more structure and support for everyone who wants to apply these new tools in their research. My goal for our initiative is to provide a central locus for people to discuss, collaborate and come up with new ideas and educate themselves. This lets us scale up machine learning efforts across the lab and make everyone more effective. We have this big community of people who are actively using machine learning every day in their science research, and that number is only going to grow.

What do you hope to accomplish in the next few years?

One of the things we emphasize is that the goal of [machine learning](#) is not to do what we're doing today and do it 10% better. We want to do completely new science. We want to do things 10 times better, 100 times better, a million times better. And we want to start seeing examples of that in the next couple of years and enable [science](#) at SLAC that wasn't possible before.

Provided by SLAC National Accelerator
Laboratory

APA citation: How machine learning helps scientists hunt for particles, wrangle floppy proteins and speed discovery (2020, October 1) retrieved 27 November 2020 from

<https://techxplore.com/news/2020-10-machine-scientists-particles-wrangle-floppy.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.