

Living up to the promise

Machine learning — in other words, computers running sophisticated algorithms to identify meaningful patterns in data — has spread from automated language translation and facial recognition to financial analysis. It's having a huge impact on business, with companies investing over US\$70 billion in research and development. But its biggest influence may ultimately come not in self-driving cars or super-smart robots, but in something far more fundamental — science itself.

From drug discovery to cancer genetics, machine learning is beginning to transform how science is done. Chemists are using it to discover new materials with exotic properties, and physicists to speed the race to practical fusion energy. Perhaps even more surprising: researchers recently showed how machine learning can speed up computer simulations used across all the sciences, making them run as much as a billion times faster.

It's certainly true that machine learning and artificial intelligence often get too much hype, the reality lagging behind the promises. But not always.

Researchers in nearly every field of science now rely on computer simulations. Some reflect the laws of fluid dynamics or thermodynamics and help predict the future weather or climate. Others steer researchers in useful directions towards new materials or drug molecules. Typically, these simulations make it possible to explore complex theoretical models in practical ways, producing approximate solutions and useful insight.

Of course, what's possible with simulations is limited by available computing power. Today, on a supercomputer, simulating about one year of the climate takes about 1,100 hours — a month and a half of continuous running. As computers get faster, of course, scientists can do more — make better predictions of climate or weather, or discover useful materials more quickly. Computers get faster and more powerful on a steady basis, and this will continue.

But some discontinuous leaps forward may also be possible, including the discovery of ways to make some simulations run far faster, even on the same computer. This possibility is the focus of a new study by researchers from Europe and the United States. The technique they report is conceptually simple and remarkably powerful, and — while there are some caveats — will likely have a huge impact on

future research (Kasim, M. F. et al. Preprint at <https://arxiv.org/abs/2001.08055>; 2020).

An old idea in computer science is emulation — using one algorithm or computing system to emulate the running of another. It's often employed as a technique to allow software designed for an old hardware system to run without problems on a newer system, by having the new hardware emulate the operation of the older. But another aim, in the context of software, is to emulate a complex and resource-demanding algorithm with a simpler one, gaining speed and freeing resources, while accepting the loss of some accuracy.



Researchers in nearly every field of science now rely on computer simulations. Machine learning is beginning to transform how science is done.

This is the idea proposed here as well — to replace time-consuming simulations with algorithms that run many times faster yet give results nearly as good. The recipe for doing this is as follows. First, run the real simulations a small number of times to generate data on which a machine learning algorithm can be trained. Next, train the algorithm on that data until it can reproduce results much like the simulation. Finally, generate any desired new results for further simulations by using the trained algorithm rather than the slow simulation. The slow, hard-to-use thing gets replaced by something much faster, yet still accurate.

In one remarkable demonstration, Muhammad Kasim of Oxford University and colleagues showed that a global climate model that normally takes 1.5 months to run can be run 110 million times faster, yet gives results of comparable accuracy. And that is possible by training the machine learning algorithm on only 39 actual simulations of the original model; not a lot of data. In another example, the researchers tested how well the emulator could reproduce realistic X-ray emissions from dense plasmas created in inertial confinement fusion experiments. In this case, many runs of the physical simulation code generate histograms for key

parameters describing the fusion plasma hot spot, such as mass and temperature. It takes 22 days to run the simulation 200,000 times. After training on 14,000 examples, the emulator could produce similar output in only a few seconds.

It's hard to overstate how much impact this could have. In the paper, the authors give further explicit examples of how the method works on simulations in astrophysics, climate science and seismology, in one case speeding up the simulation by two billion times. There ought to be similar opportunities in areas ranging from the behaviour of supply networks to finance. Give up a tiny bit of accuracy, and the algorithmic approach can speed up modelling by a huge amount.

I mentioned a caveat. I went to Oxford University and discussed this work with Kasim and Peter Hatfield, one of his collaborators. One limitation, they told me, is that the method doesn't always work for chaotic systems — systems in which what happens is extremely sensitive even to tiny changes in conditions or parameters. There are plenty of important systems of that sort. Even so, their approach did work pretty well for a global climate model. They suspect the method might work for chaotic systems if one is interested mostly in the lowest spatiotemporal modes, rather than higher-frequency details. For now, when it will or will not work remains a little unclear.

Where it will have most impact, they believe, is in using simulations in design. For example, in the quest to achieve nuclear fusion, researchers rely on detailed simulations of the physics to test out new reactor designs. This approach greatly speeds up the process, letting them test out hundreds of possibilities in less than the time it would otherwise take to test one. The same will likely be true in design areas ranging from the composition of steels to medicines.

There's plenty of room for concern over machine learning — for example, the way algorithms, being used to make decisions on anything from mortgages to sentencing, tend to propagate the biases in the data they analyse. But these algorithms have an undeniable positive side, and perhaps nowhere more than in helping science itself. □

Mark Buchanan

Published online: 7 July 2020
<https://doi.org/10.1038/s41567-020-0962-1>