thesis

The power of machine learning

Last year researchers released a mechanical bird into the air in the mountains of California. The bird, with wings stretching two metres across, was equipped with a computer running machine-learning algorithms processing real-time feedback on local vertical winds and the roll-wise torques. Real soaring birds face a tough task, their efforts to ride rising thermals hampered by turbulent eddies that make winds fluctuate. In mountainous regions, convection-driven air currents shift continuously as thermals form, disintegrate or get swept away by horizontal winds.

Nevertheless, the neural networks in this mechanical bird's control system learned from experience over five days, managing to adaptively alter the bird's bank angle and pitch behaviour to gain altitude successfully (G. Reddy et al. *Nature* **562**, 236–239; 2018).

As this glider illustrates, the power of machine learning is rapidly transforming a lot of modern science. In a quick informal search, I found recent applications in predicting the progression of kidney disease, improving the LIGO gravitational-wave detectors, tracking cyclones more effectively or developing more efficient simulations of fluid flows. Others include discovering new chemical compounds and quantum computing. In a recent review in Reviews of Modern Physics (in the press, preprint at https://arxiv.org/abs/1903.10563; 2019), Giuseppe Carleo and colleagues note that machine learning is coming to hold a crucial position in fields of physical science ranging from physical chemistry to particle physics and the engineering of quantum states of matter. Paradoxically, however, we're still not completely clear on why it works so well.

Machine learning hit the public awareness after spectacular advances in language translation and image recognition. These are typically problems of classification - does a photo show a poodle, a Chihuahua or perhaps just a blueberry muffin? Surprisingly, the latter two look quite similar (E. Togootogtokh and A. Amartuvshin, preprint at https://arxiv.org/abs/1801.09573; 2018). Less widely known is that machine learning for classification has an even longer history in the physical sciences. Recent improvements coming from so-called 'deep learning' algorithms and other neural networks have served to make such applications more powerful.

In experimental particle physics, machine classification has found two major uses — particle identification and event selection.

The first task is to answer — which particle or particles were involved in what we just recorded? The second problem is more general — was that an interesting event worth closer investigation, or not? In particle physics, data are plentiful, precisely the setting where machine learning is effective.



Machine learning is coming to hold a crucial position in fields from physical chemistry to particle physics and the engineering of quantum states of matter.

Algorithms for particle identification play a key role at the Large Hadron Collider (LHC), where collisions generate a proliferation of high-energy quarks and gluons. These particles radiate further quarks and gluons that finally combine into colour-neutral particles, mesons or baryons, which hit the detector as a jet. A key challenge is to characterize the structure of such jets - referred to as 'jet tagging' and to test the predictions of quantum chromodynamics. Of great recent interest are methods for 'top tagging', or identifying jets linked to the production of top quarks. These often produce multiple overlapping jets, making signal extraction difficult. Yet studies have found that machinelearning algorithms working just with crude calorimeter data - recorded energy deposited into every detector pixel - can outperform the previous best statistical analyses (G. Kasieczka et al., preprint at https://arxiv.org/abs/1902.09914; 2019).

Event selection is equally important. Collider experiments generate huge amounts of data, yet interesting events are exceedingly rare. Hence, experimenters use a real-time data reduction system — called a trigger — to focus only on the few collisions of interest, keeping as few as one in 100,000. As Carleo and colleagues note, for example, the LHCb experiment, which aims to measure the parameters of CP violation in the interactions of b-hadrons, now relies on machine-learning techniques for triggering, as it identifies interesting events more accurately than previous methods.

Another area where machine learning is leading the way is in exploring the vast space of possible material structures, and predicting, for example, structures more likely to be observed experimentally, or to have interesting properties. For a few molecules or materials, this can be done using computationally intensive methods from quantum chemistry, but not for a wide range of molecules. The route forward with machine learning is to use the precise methods to generate training data in the form of many pairs, X,Y, of molecular representations X and corresponding energies or properties Y. Once trained on this data, the neural network can make useful predictions for structures outside the training set. In one study (J. S. Smith et al. Chem. Sci. 8, 3192-3203; 2017), researchers showed how a deep neural network trained on quantum-mechanical calculations could learn and accurately predict total energies for organic molecules of up to eight atoms and containing four atom types: H, C, N and O. The network was then accurate in comparison with reference quantummechanical calculations even on much larger molecular systems, going up to 54 atoms, far beyond those included in the training dataset.

Carleo and colleagues go on to consider many other areas of application, including cosmology, quantum networks and quantum many-body simulations. Yet it remains somewhat mysterious why these methods work so well. One idea flowing back from physics into computer science is the notion of an information bottleneck (N. Tishby and N. Zaslavsky, in Proc. 2015 IEEE Information Theory Workshop, 2015; preprint at https:// arxiv.org/abs/1503.02406; 2015). This idea suggests that layers in a neural network, during training, navigate a delicate trade-off between keeping enough information about the input so as to predict the outputs, while also forgetting as much irrelevant information about the inputs as possible, thereby keeping the learned network representation simple.

Indeed, there appear to be deep links between important concepts in machine learning and statistical physics. This computational technique is transforming science, but physics may yet hold the key to explaining why.

Mark Buchanan

Published online: 2 December 2019 https://doi.org/10.1038/s41567-019-0737-8