Machine learning versus human understanding

1. **Self-Learning Monte Carlo Method: Continuous-Time Algorithm**
   Authors: Yuki Nagai, Huitao Shen, Yang Qi, Junwei Liu, Liang Fu
   arXiv:1705.06724

2. **Can Boltzmann Machines Discover Cluster Updates?**
   Authors: Lei Wang
   arXiv:1702.08586

Recommended with a Commentary by Anton Akhmerov, Kavli Institute of Nanoscience, Delft

Just a couple of days after AlphaGo beat the best human player and ended its Go career due to the lack of worthy opponents, it is very easy to predict that machine learning will be successful in all domains of human activity. A natural extension of this thought is the expectation of a deep neural network as a literal deus ex machina outperforming the best physicists in analysing physical models. Consequently there are now a handful of papers applying neural networks of varying complexity to the study of phase transitions in condensed matter and other conventional problems in condensed matter physics. I would like to explain what limitations I expect from complex machine learning algorithms, and describe two applications of 'simple' machine learning that I consider promising examples of an alternative approach.

Condensed matter physics, just like most of physics is extremely reductionist. Teaching someone to distinguish between ferro- and paramagnetic phases of Ising model takes a couple of sentences, while describing a reasonable Go strategy can hardly be confined to a single book. This is not a coincidence but rather the way all of physics was developed: the community spent centuries removing unimportant degrees of freedom, and bundling similar problems together into universality classes. Where this approach fails and alternative methods of study are required, other sciences appear: compare physics with biology, chemistry, or economics. This, I believe, is the first problem of applying advanced machine learning algorithms in physics: finding problems complex enough to warrant the use of these algorithms.

The other important limitation of using neural networks to predict the behavior of physical systems is the limited amount of new understanding that the researchers can extract from the trained neural network. These are highly nonlinear multidimensional functions with a lot of parameters that are therefore very hard to inspect or meaningfully modify. The consequence of the inability to inspect a neural network is the difficulty of finding the limitations of its performance: in order to answer whether a neural network trained with open

boundary conditions will apply to periodic boundary conditions, there is no easier strategy than to apply the network and check if it works.

The preprint by Nagai *et al.* is the latest in a series of works by some of the same authors about the "Self-learning Monte Carlo algorithm". The defining idea of the method is constructing an efficient Monte-Carlo model by introducing an effective unobservable parameter and providing it with effective theory of its own that should capture as much of the original model as possible. In the simpler case [1], such parameter may be the Hubbard-Stratonovich auxiliary field, and its effective theory is a local mean-field Hamiltonian. Alternatively, in the latest work, such hidden model is the imaginary time analog of the same idea, namely the correlation function of the auxiliary field spins. Describing the auxiliary field dynamics alone only requires a few parameters. Indeed, the authors are able to plot the complete fitting result as a single function: the effective mean-field Hamiltonian either in real space or imaginary time, while a similarly descriptive visualization is impossible for a neural network. Despite the simplicity of the auxiliary field dynamics, this method differs from approximate solutions of interacting quantum problems (e.g. DMFT) because it preserves the exact dynamics of the microscopic degrees of freedom rather than integrating them out completely.

The work by Wang sheds some extra light on the reason why such simple approaches are effective. He relates the problem of finding good updates in Monte Carlo simulations to the properties of *restricted Boltzmann machines*. These machines generate likely samples of physical degrees of freedom by coupling them to the hidden ones using an Ising coupling. In principle, given sufficiently many hidden degrees of freedom and a lot of fitting data a Boltzmann machine can represent any probability distribution. Boltzmann machines in general are a canonical example of artificial neural networks and come with all the corresponding problems: in the previous work [2], the authors only gained a modest advantage over the regular local Monte-Carlo updates by using a rather expensive training data set for the Boltzmann machine. On the other hand, in the follow-up Lei Wang shows that introducing an auxiliary hidden spin on every link of the Ising lattice, and assuming only local couplings with equal strength reproduces the conventional cluster update algorithm. Here the dynamics of the hidden spins captures the selection of clusters of correlated spins. While this algorithm is also a Boltzmann machine, it only has two free parameters, and therefore is much easier to train and evaluate. Likewise slight modifications of this Bolzmann machine reproduce other algorithms for the global updates.

It may seem these two preprints do not hold as much promise as a more advanced techniques: the most advanced numerical part in both is a single linear fit! They also return us to the known tools: mean-field Hamiltonians and cluster update algorithms. The new ingredient, however, is the straightforward ability to automatically increase the model complexity whenever it is necessary. Further, both works show how to "encode" our accumulated understanding of condensed matter systems into algorithms that have provable performance and predictive power. Looking ahead I hope that we will be able to produce simple and efficient descriptions of complex systems, and extract new knowledge from the machine learning algorithms.

References

[1] Junwei Liu, Huitao Shen, Yang Qi, Zi Yang Meng, Liang Fu, arXiv:1611.09364 (2016).

[2] Li Huang, Lei Wang, Phys. Rev. B **95**, 035105 (2017)