A conversation with Professor Konrad Kording, September 11, 2019

Participants

- Professor Konrad Kording - Penn Integrated Knowledge Professor, University of Pennsylvania
- Joseph Carlsmith - Research Analyst, Open Philanthropy

Note: These notes were compiled by Open Philanthropy and give an overview of the major points made by Prof. Kording.

Summary

Open Philanthropy spoke with Prof. Konrad Kording of the University of Pennsylvania as part of its investigation of what we can learn from the brain about the computational power ("compute") sufficient to match human-level task performance. The conversation focused on different ways of generating brain-based compute estimates.

Approaches to generating estimates

We can distinguish between the goal of replicating the inputs and outputs relevant to a particular task, and replicating the internal dynamics involved in a particular human brain performing that task. The former may be substantially easier. We know that for a lot of tasks (e.g., chess), there are specialized solutions that can be employed using limited hardware.

In general, though, we don't know how hard being human is, as a computational problem. Here are some different approaches to generating compute estimates.

Low-end estimate based on vision

The human brain dedicates roughly half of its hardware to processing vision (this can be seen by looking at diagrams created by David Van Essen). And we can solve a lot of the vision problem (e.g., detecting objects, segmenting scenes, storing information) using very modest compute. Indeed, computers have long been much better than humans at memory storage.

There are other tasks we don't know how to do (e.g., high-level decision-making), and vision might be comparatively easy. Still, there is a history of changing our conception of how hard the problems the brain solves are. For example, we used to think that multiplying large numbers and playing chess were strong signs of intelligence, but now we know that these tasks can be done fairly easily with the right hardware. If half of the brain is vision, maybe the other half isn't particularly difficult either, and the brain's full task-performance can be replicated with today's hardware.

It's also possible, though, that learning is the main challenge, and that most of the compute burdens come from the brain's need to update itself every second, or every couple of milliseconds, rather than from the processing involved in performing already-learned tasks. That said, artificial systems might need to do less self-updating than the brain does.

Estimate based on spikes through synapses

You have roughly 10^{11} neurons in your brain, each with roughly 10^{3} synapses, and each firing at roughly a 10 Hz rate. If we assume that each neuron is performing a very trivial computation -- e.g., multiply each input by some number, and then add them up -- then this suggests roughly 10^{15} multiplications per second. This is a manageable compute burden.

High-end estimate

Examination of neurons reveals that they are actually very non-linear, and the computations involved in plasticity probably include a large number of factors distributed across the cell. In this sense, a neuron might be equivalent to a three-layer neural network, internally trained using backpropagation. In that case, you'd need to add another factor of roughly 10^{5} to your compute estimate, for a total of 10^{20} multiplications per second. This would be much less manageable.

Overall uncertainty

The difference between the estimates generated by these different approaches is very large -- something like ten orders of magnitude. It's unclear where the brain is on that spectrum.

Specific sources of compute

Learning

Prof. Kording thinks that learning in the brain requires the same amount of compute as processing. If you have a compute graph, going forwards and backwards comes at roughly the same cost.

Glia

Glial cells would imply a factor of two in required compute, but we are likely to be so many orders of magnitude wrong already that incorporating glia will not make the difference.

Estimates based on V1

There is a traditional view in systems neuroscience that each brain area does something pre-assigned and simple. E.g., V1 detects edges, V4 pulls out colors and curvature, etc. But this view is dying at the moment.

It was always suspicious on theoretical grounds. The fact that you know so much, about so many types of things, is in conflict with the view that each specific brain area is simple, as this view does not explain where all of the information available to you comes from.

But it's also empirically wrong. If you look at the literature, when you take a type of signal that matters to animals and looks for it in the brain, you find it everywhere. For example, you can find movement signals and expectations in the primary visual cortex, and rewards explain more of the variance in the primary motor cortex (the "movement area") than movement. Basically, it's all a complete mess.

It's true that simple models of V1 can describe 30 percent of the variance in V1's activity. But you can describe half of the variance in the activity of your transistors just by realizing that your computer is turned off at night. It could be that some V1 neurons are silent until certain image features cause them to perform a certain type of analysis, but in our ignorance, we assume that they just detect local edges.

Of course, there's some specialization. Sound explains more of the variance in auditory cortex than in visual cortex. But the specialization isn't simple. It's just easier to publish papers saying e.g. "X is the brain area for romantic love," than e.g. "here are another ten variables X region is tuned to."

Defining vision

“What things are” isn’t the only question at stake in vision. You want answers to questions like “can I grasp this water bottle? Can I hold it there?”. Indeed, there are a vast number of questions that we want to be able to ask and answer with vision systems, and the “solution” to vision will depend on the exact thing that other parts of the brain need from the visual system. It’s not an easily definable space, and the only way to figure it out is to build a system that fully learns all of the relevant pieces.

Upper bounds

If you want upper bounds on required compute, you can look at the parts list of the computing elements in the brain, the noisiness of which will put physical limits on the amount of computation they can do. This might result in very high estimates. For example, it might say that every ion channel does a bit roughly every ten milliseconds.

This approach doesn’t necessarily rule out molecules and proteins as possible avenues of computation. However, some molecules may equilibrate so fast that you can replace them with a variable that describes their average state (e.g., mean field theory is applicable).

You can’t do this across a neuron: there are NMDA spikes and other complexities. So the question is: what is the compartment size where local averaging is possible? People disagree. Some think the brain has organized as itself to be mean-field modelable, but they have never shown much evidence for that. Still, at some length-scale (say, ten micrometers) and some time-scale (much faster than electrophysiology), everything will equilibrate.

Best guesses

Prof. Kording’s hunch is that in order to replicate firing decisions in neurons, you’d need to break the neuron into pieces of something like ten microns (this would hundreds, maybe thousands of compartments per neuron). This hunch is grounded in a belief that neurons are very non-linear.

Here is one non-standard argument for this degree of non-linearity in neurons. Adjusting synapses in helpful ways requires computing how that synapse should adjust based on its contribution to whether the neuron fires. But this computation applies in basically the same way to individual ion channels in the cell: e.g., if the brain can signal to the synapse how to adjust in order to improve neuron firing, it can do the same for ion channels, at no additional cost. This makes Prof. Kording thinks that the brain is optimizing both.

However, current techniques are very bad at measuring ion channel plasticity. Neuroscientists don't tend to focus on it for this reason.

There are considerably more ion channels than synapses, and ion channels change how synapses linearly and nonlinearly interact with one another. This suggests an uglier computational space.

Opinions in the field

Some neuroscientists think that the brain is a deep learning system. They generally believe that neurons are trivial. Others, such as some experimentalists, think that neurons are extremely complicated.

In general, people are often willing to take a philosophical position, without much evidence, if it makes their research more important.

All Open Philanthropy conversations are available at
http://www.openphilanthropy.org/research/conversations