

The Need for Collaboration on a Theory of Brain Computation

DJ Passey

*Scientific Data Division
Computing Sciences Research
Lawrence Berkeley National Laboratory*

August 3, 2022

Introduction

For problems where solutions can be augmented by machine learning approaches, those of us who opt not to use deep learning feel a need to justify ourselves. And if we choose criticism of deep learning as our justification of choice, there are a lot of places where we might raise issues. You’ve probably heard the frustrated remarks about how deep learning can’t capture common sense reasoning, etc.. But despite the criticisms, deep learning marches on, producing one incredible demo after another, followed closely behind by business solutions.

To be clear, there are few outright attacks on those those who choose a different path than deep learning. But the difference in interest and attention for other methods is palpable. Not being able to keep the attention of others is enough. And so, justify we must.

But this attention deficit is not baseless. It isn’t rooted in peer pressure or a follow the crowd mentality. It is rooted in fact. Non deep learning methods get less attention, because *they can’t compete*.

Those of us who study “alternate” approaches have a shortlist for why our approach *does* compete. Maybe it is more efficient, more robust to noise or able to do one shot learning. We have our justifications. But outside of these justifications, when we compare our accuracy to a deep learning algorithm with optimized hyper-parameters, we lose. That’s it. And it has even been my unfortunate experience that for the majority of applications, my justifications didn’t really matter anyway.

Why do something besides deep learning?

Deep learning’s massive success has put us in a difficult position. If we want maximum prediction accuracy on our dataset, the smartest choice is to do the

time consuming work of architecture and hyper-parameter optimization. Once this is done, you can be confident that you have a model that is about as accurate as possible with today's technology.

It is likely that any alternate approach will result in accuracy loss. *So why try?*

Deep learning is incredible. It has created a massive societal change in a short period of time. It feels only a short leap away from an algorithm that can reason and problem-solve with the capacity of a human researcher. It's a future that has the best of us starry-eyed.

And whether deep learning can *achieve* this milestone is a hotly debated topic (see [2, 4] for two high profile examples). But if an algorithm gains the problem solving capacity of a human researcher, do we want it to achieve this by maximizing an objective function, or by employing the same memory and learning techniques used by our brain? Re-enforcement learning agents are notorious for exploiting rules and behaving in unexpected ways. And maybe they don't behave the way we expect them to because we are quick to believe that they learn like we learn. But they are not like us. Here's to instead, making an algorithm that thinks more like we do.

Fundamentally, deep learning does not work the way the brain works. This much is clear. While it is possible that the brain does back propagation, our brain is so much more than a pattern recognition machine. Here is a simple example. Our brain can read two different research papers and come up with a way to combine the two approaches in a useful way. Deep learning is amazing at pattern recognition, but no amount of pattern recognition alone can do this. This is because combining research approaches meaningfully is different in every situation, there is no pattern for the algorithm to find.

In theory, you could make a dataset where the inputs were text from two research approaches and the targets were an approach that combined the corresponding first two. But it would be ridiculous to try to train a deep neural net to take in the text of two research approaches and map them to a third approach that combines the two. You could use a massive model pretrained on every textbook from the field but it would produce useless gibberish after training. To solve this problem, you need to extract *meaning*, and deep learning does not learn meaning, it learns patterns.

The State of Brain Computation

There are three main fields interested in brain computation, cognitive science, computer science and mathematics (considered together) and neuroscience. We briefly describe the state of these three fields in relation to brain computation here.

Cognitive Science

Cognitive science is tightly focused on the problem central to this paper: How does the brain compute? By what mechanisms do we reason, infer and solve problems?

Cognitive science can be experimental with qualitative descriptive theories of the mind, but it can also be computational with theories of brain computation expressed in equations.

Many of the most important artificial intelligence algorithms of our time were originally published in cognitive science journals. This is the field where scientists are most actively working on developing the theory of how the brain solves problems.

However, cognitive science is not a technical field and so students of cognitive science do not receive a high level of training in math and programming. It is possible that this impedes the speed that mathematical theories of brain computation can spread through the field and limits the ability of cognitive scientists to critique new mathematical theories.

Computer Science and Mathematics

Computer scientists have been concerned with brain computation for decades. Where cognitive scientists develop new theories of how the brain works, computer scientists mainly contribute to advancing the theory of brain computation but taking cognitive models and mobilizing them for computational problems. This is was done with amazing results for the neural network and sparse distributed representations. But with several notable counter examples, this is done without regard for biological mechanisms and therefore makes little progress at advancing our understanding of brain computation.

When computer scientists do produce models of the brain, the cognitive scientists and neuroscientists capable of providing appropriate critiques usually do not provide them, and so the projects are either abandoned or grow in isolation. This can lead to a lot of self citations and sweeping uncontested claims (in the case of Adaptive Resonant Architectures) [3]. The work on adaptive resonant architectures is real, but those of us wanting to contribute to the study of brain computation have no way of assessing the merits and weaknesses of the different approaches developed by mathematicians and computer scientists.

Neuroscience

Neuroscientists have the most knowledge of the mechanics of the brain, but are generally more concerned with modeling the mechanics than they are with models that explain how useful computation occurs. In 2018 Richard Axel, a Nobel prize winning neuroscientist said, "We do not have a logic for the transformation of neural activity into thought and action. I view discerning [this] logic as the most important future direction of neuroscience" highlighting the lack of a model that connects neural activity to computation [1].

The tools of experimental neuroscientists are well poised to make discoveries about the nature of computation in the brain, but unfortunately experimentalists often have little training in the formal mathematics needed to create a theoretical model.

Computational neuroscientists do have the training necessary to create such models, but the computational neuroscience results that I am familiar with are highly mechanistic, placing accurate models of the mechanics of neurons in the forefront and avoiding the simplifications necessary to make a useful model of brain computation.

Brain computation, in summary

For those of us hungry for meaningful advances in this space, the different dynamics of each of the three fields make collaboration difficult. This is a major source of frustration, because we know that the skills and knowledge of all three fields is needed to make real progress.

And what does real progress look like? It looks like elegant and creative new models of brain computation. Something like the original neural network or like sparse distributed representations. Models like this could change the world.

How does this happen? It looks like a team of neuroscientists who doing experiments that aim to verify or disprove the predictions of cognitive models, a mathematician or computer scientist who can be consulted to help create new models that explain experimental phenomenon *in the context of brain computation* rather than neural mechanisms. And a cognitive scientist who can teach the neuroscientist and computer scientist what has already been tried and put in to context the unanswered questions of cognitive science in a manner that the neuroscientist can devise experiments that dig deeper and the computer scientist and mathematician can wrestle with the logic problems and offer new unique answers up for critique by the neuroscientist and cognitive scientist.

One might argue that a cognitive scientist is not needed and that a computer scientist and neuroscientist can do this work alone. However, cognitive scientists have the best grasp of the challenges in creating models of thought and the best knowledge of our best attempts at solving these problems and where the solutions fail. Collaboration with a cognitive scientist greatly accelerates progress in the field because they prevent the computer scientist and neuroscientist from working on solved problems or from providing solutions that have already been shown to fail.

We conclude this section by listing a final challenges for brain computation: It is likely that initial results will have very little application, either for medicine or for artificial intelligence. Creating simplified models of brain computation results in a tool that is not useful for medicine because it doesn't capture real brain mechanisms and isn't useful for artificial intelligence because it is too simple. This could explain the lack of progress in this area to date as funding is often allocated based on the immediate impact of proposed work. Advances in brain computation have little immediate impact, but can have massive long term

payoff as AI architectures are designed and optimized based on the simplified models.

History has shown that the return on investment can be massive, but takes many years of work to come to fruition. Therefore, the best source of funding for this is the government rather than a private company. Therefore, we argue that the NIH or NSF should set aside grant money for interdisciplinary teams with at least one active participant from cognitive science, neuroscience, and computer science/mathematics. This will incentivize much needed collaboration and drive meaningful advances in brain computation.

This author offers his unlimited services as a reviewer of applications but would require assistance from cognitive scientists and neuroscientists.

References

- [1] Richard axel. *Neuron*, 99(6):1110–1112, 2018.
- [2] Ragnar Fjelland. Why general artificial intelligence will not be realized. *Humanities and Social Sciences Communications*, 7(1):10, 2020.
- [3] Stephen Grossberg. Adaptive resonance theory: How a brain learns to consciously attend, learn, and recognize a changing world. neural. *Neural networks*, 37, 10 2012.
- [4] David Silver, Satinder Singh, Doina Precup, and Richard S. Sutton. Reward is enough. *Artificial Intelligence*, 299:103535, 2021.