

1 Introduction

The human brain is truly remarkable. It ...

The study of its functionality is more in the domain of computer science: We try to model the brain with computer algorithms, which is inspired by the idea that the human brain consists of elementary computational units, neurons, which itself can be sufficiently modeled with computers, and whose interplay creates the emergent complex behavior we see.

We structure this work by logical statements like *axiom*, *premise* and *conclusion* etc. similar to mathematical writing. I believe the chain of argument is more evident this way.

Axiom 1.1. *In theory, the human brain can be sufficiently modelled by a computer program.*

The human brain consists of a very large number of neurons, around 86 billion of them. Every neuron is connected to others via *synapses*. On average, 1,000 synapses are emanating per neuron. If we were to store each synapse by one byte of computer memory, we would need 86 trillion bytes \approx 86 terabytes of storage. Sounds manageable, but how long would it take to run this program? Additionally, note that we stated that each neuron itself acts as an elementary computational unit, i.e. all neurons operate in parallel. Clearly, the complexity of modelling the brain adds up enormously.

We also must not forget that the human brain is a dynamic system which operates *in space and over time*. This is in strong contrast to basic neural networks, which operate in simple input \mapsto output fashion.

Example 1.1. Given the task to multiply 96459 and 29537. We humans can use the space dimension to grab ourselves a pen and a piece of paper, and use the time dimension to repeatedly perform simple multiplications and additions in order to solve the original task. A basic neural network is incapable of that.

Thus, we should think of the human brain as a device which can *create* and *perform* instructions. Hence, we may focus on the operations it can perform as a beginning. We can think of this as a human sitting in an empty room without any auxiliary items around performing a task like adding two numbers.

Premise 1.1. *Among other things, the human brain can perform operations like adding and multiplying numbers (for a restricted domain of \mathbb{Z} for example), or inflecting verbs.*

Thus, we should be able to reconstruct this behavior with our models as well. Of course, it might be the case that all the operations the human brain can perform is an emergent phenomenon of the entire brain working as a whole. However, it also could be the case that the brain has certain sub-circuitry for specific tasks like adding numbers. Assuming this is the case, we also might conclude that these circuits shouldn't be much more complex than necessary, as the brain is a product of the energy minimization of evolution.

Remark 1.1. However, this begs the question whether there is some efficient circuitry "preinstalled" due to evolution, or whether everything needs to be learned. But are the learned circuits efficient themselves, or are they rather big and all operations the brain is capable of are emergent phenomena from different brain regions working together?

Based on these observations, our main focus is on how models, specifically neural networks, can efficiently implement a certain subset of tasks humans are naturally good at.