Power Saving Mode For Neural Networks

Introduction

Ever since the bginning humans have tried to create intelligent machines. Only recently have be begunn to see some sucess. The best mehods uses some form of artificial neural networks (ANN) drawing some inspiration from the mammalian brain. The brain is an extraordinary biologiacal machine capable of great tasks and it is very energy effecient. The main principle of the brain that most ANNs uses is the perceptron—a simple model of a neuorn, connecting multiple of these thogheter you get a multilayer perceptron (MLP). If the goal is to mimmic the brain in order to crete intellignet machines, this technique may seem like a gross oversimplification, however it works quite well in pratcie leading to gpt, alphafold etc.

You may think that this i all we need to create intellignet machines and that this is essentualy a solved problem, we just need more compute and perhaps fine tune the architechtures a bit.

This does not scale that well we cannot throw infinite money and power at the problem

A 100 token promt uses ?MW [citation needed]

If we revisit the human brain is is apparent that it has some tricks that cannot be modeld by a simple MLP the brain uses 30W on what the current ANNs can't even do

We can try to draw more inspiraton from the brain.

It is diffucult to distill the fundamental properties that allows for effecient computation. Some models try to perfectly fit the intrcate molecular dynamics of the neurons with ion channels and what not This can be a great tool to understand the brain, but this might be a means to an end and not a key component of an intelligent system.

MLPs do matrix mul. can be inneficent if most elements do not contrube much mlps are relativly easy to train, use backprop this cannot be done using spiking NNs

Properties of the Brain

The spikes in a brain are delay to intensity ecoders[citaion needed]

The brain is on the edge of chaos, or critical system[citation needed]

visual cortex

Bibliography