Abstract to general public

Artificial intelligence (AI) has quietly revolutionized our daily lives through neural networks (NN) – from advanced specialized pattern recognition engines in medicine and science up to automating tasks and improving decision-making with personal assistants. NN, partially inspired by how our brains process information are digital systems with thousands or even trillions of learnable parameters nodes called artificial neurons that learn by adjusting their connection strengths based on provided examples (learning).

Despite massive fascination with the recent advances of AI, what the public rarely sees are the astronomical resources required to train these systems. This has become a huge bottleneck as there are limited numbers of organizations able to perform such training, not only due to data ownership, but also because of computational requirements. This slows scientific progress and limits our ability to develop AI systems overall. Furthermore, enhancing efficiency of NN training would also bring substantial impact on reducing negative environmental impacts. Thus, it is essential not only to develop architectures and approaches that increasingly improve the model capacities but also focus on efficient training which may bring some partial answers to the main problems such as generalization gap or convergence challenges.

The learning process, also called optimization or model fitting, involves continuously adjusting these billions of parameters to minimize errors, enforcing the underlying network to yield desired results. Often times, current learning paradigms treat this process as uniform throughout training, applying the same optimization techniques from start to finish. Here lies the crux of the problem: most optimization paradigm treats neural network training as a uniform, steady process or introduces sensitive scheduling which demands separate finetuning. Arguably, this approach works, however it requires an appropriate set of parameters controlling optimization trajectory, missing critical insights about how learning actually unfolds. Learning rate - perhaps the most critical parameter - is typically chosen through exhaustive and computationally expensive grid searches without principled understanding of how these choices affect the learning process.

We propose a fundamental shift: we aim at correctly understanding what happens in the first stage of optimization trajectories. It is an enormously difficult task as there is lots of stochasticity, huge number of interconnected subprocesses in massive parameter space. However, we claim that even small enhancements may push networks towards regions of higher interest. The early stages of training are crucial as they set the foundation for the model's learning trajectory. Analyzing these stages can provide insights into how neural networks begin to form representations of data and how initial conditions influence final performance. In the proposed work, we strive for better analysis in the early stages of optimization which induces neural networks to intrinsically regularize themselves and control the speed and strength of gradient updates.

This approach promises to address fundamental challenges plaguing AI development: reducing training times, minimizing energy consumption, improving final performance, and making AI development more accessible and sustainable. Most importantly, understanding early training dynamics represents a fundamental advance in our comprehension of how and what artificial intelligence systems actually learn. There is a need to shift the paradigm and investigate innovative techniques that address the core challenges of neural network training.

Finally, we would like to emphasis that suggested work aligns well in the modern line of research. More and more work focuses less on more parameters but rather on re-using them better. The future of artificial intelligence depends not just on building bigger models, but on building them smarter, and we hope that our struggles will be a piece of the puzzle.