A Principled Investigation of the Lottery Ticket Hypothesis for Deep Neural Networks

Adrian Vladu
adrian.vladu@irif.fr
IRIF, Université de Paris

Context

Sparsifying deep neural networks is a fundamental challenge in deep learning. Deploying massive trained models into the wild faces significant obstacles, since even for basic tasks like image classification, the size of the model can easily become prohibitively large. As the model size directly affects the time required to perform inference, methods for reducing network size by removing connections and neurons have become highly important.

The Lottery Ticket Hypothesis (LTH) defined a suprising phenomenon that was noticed when attempting to obtain such sparse models. In a recent paper, Frankle and Carbin [3] advanced the hypothesis that dense, randomly initialized neural networks contain small subnetworks which, when trained in isolation, reach training accuracy comparable to the original network in the same number of passes. Hence finding such a small subnetwork (i.e. *winning lottery ticket*) would enable one to increase efficiency for both training and inference.

While this highly intriguing result posits the existence of good sparse subnetworks, it is unclear how to find them. A great body of work [3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15, 16] has been dedicated to this task. However, current results are mostly empirical and present understanding of *why* is sparsification even possible is far from being complete.

Goals

We propose a systematic investigation of this phenomenon in order to (1) understand what is the underlying structure of dense neural networks which allow the existence of *"winning lottery tickets"* and (2) derive efficient methods to find these sparse subnetworks.

The intern will be involved with both experiments and theory. The early phase of the project will consist of implementing a range of established methods for network sparsification [3, 8, 11, 16], in order to put in place a solid testbed for our projects. In parallel, the student will get familiarized with the existing literature on this topic, together with relevant established methods from classical optimization [13, 1, 2]. Once the preliminary set of experiments is in place, we will use them to guide our choices towards designing and analyzing new algorithms. The basic approach will start from understanding sparsification in the case of convex functions, which will then move towards the specific nonconvex functions that arise from canonical neural network architectures.

The final goal is to develop a series of principled methods that can recover the sparse subnetworks efficiently. A successful internship would result in a technical writeup matching the standards of flagship conferences, such as ICML/NeurIPS/ICLR.

Additional Information

We seek motivated candidates with a solid mathematical background. They should be familiar with the basics of continuous optimization, and be willing to learn and develop new theory. Prior exposure to programming and PyTorch/Tensorflow is recommended. In case of a successful project, the candidate will be encouraged to transition to a PhD. Students interested in this topic are encouraged to send an email to adrian.vladu@irif.fr.

References

[1] Aharon Ben-Tal and Arkadi Nemirovski. *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*. SIAM, 2001.

[2] Sébastien Bubeck. Convex optimization: Algorithms and complexity. *arXiv preprint arXiv:1405.4980*, 2014.

[3] Jonathan Frankle and Michael Carbin. The lottery ticket hypothesis: Finding sparse, trainable neural networks. *arXiv preprint arXiv:1803.03635*, 2018.

[4] Jonathan Frankle, Gintare Karolina Dziugaite, Daniel M Roy, and Michael Carbin. Linear mode connectivity and the lottery ticket hypothesis. *arXiv preprint arXiv:1912.05671*, 2019.

[5] Jonathan Frankle, Gintare Karolina Dziugaite, Daniel M Roy, and Michael Carbin. Stabilizing the lottery ticket hypothesis. *arXiv preprint arXiv:1903.01611*, 2019.

[6] Jonathan Frankle, David J Schwab, and Ari S Morcos. The early phase of neural network training. *arXiv preprint arXiv:2002.10365*, 2020.

[7] Trevor Gale, Erich Elsen, and Sara Hooker. The state of sparsity in deep neural networks. *arXiv preprint arXiv:1902.09574*, 2019.

[8] Tao Lin, Sebastian U Stich, Luis Barba, Daniil Dmitriev, and Martin Jaggi. Dynamic model pruning with feedback. *arXiv preprint arXiv:2006.07253*, 2020.

[9] Zhuang Liu, Mingjie Sun, Tinghui Zhou, Gao Huang, and Trevor Darrell. Rethinking the value of network pruning. *arXiv preprint arXiv:1810.05270*, 2018.

[10] Ari Morcos, Haonan Yu, Michela Paganini, and Yuandong Tian. One ticket to win them all: generalizing lottery ticket initializations across datasets and optimizers. In *Advances in Neural Information Processing Systems*, pages 4932–4942, 2019.

[11] Alex Renda, Jonathan Frankle, and Michael Carbin. Comparing rewinding and fine-tuning in neural network pruning. *arXiv preprint arXiv:2003.02389*, 2020.

[12] Pedro Savarese, Hugo Silva, and Michael Maire. Winning the lottery with continuous sparsification. *Advances in Neural Information Processing Systems*, 33, 2020.

[13] Shai Shalev-Shwartz, Nathan Srebro, and Tong Zhang. Trading accuracy for sparsity in optimization problems with sparsity constraints. *SIAM Journal on Optimization*, 20(6):2807–2832, 2010.

[14] Hidenori Tanaka, Daniel Kunin, Daniel LK Yamins, and Surya Ganguli. Pruning neural networks without any data by iteratively conserving synaptic flow. *arXiv preprint arXiv:2006.05467*, 2020.

[15] Haoran You, Chaojian Li, Pengfei Xu, Yonggan Fu, Yue Wang, Xiaohan Chen, Richard G Baraniuk, Zhangyang Wang, and Yingyan Lin. Drawing early-bird tickets: Towards more efficient training of deep networks. *arXiv preprint arXiv:1909.11957*, 2019.

[16] Hattie Zhou, Janice Lan, Rosanne Liu, and Jason Yosinski. Deconstructing lottery tickets: Zeros, signs, and the supermask. In *Advances in Neural Information Processing Systems*, pages 3597–3607, 2019.