Osnabrück University
Institute of Cognitive Science

UNIVERSITÄT OSNABRÜCK

Exposé for the Bachelor Thesis

A Comparison Of Sparse Neural Networks

An analysis of networks created with lottery tickets and rare gems

Submitted by

Tjark Darius
Marienstr. 4
49074 Osnabrück
Email: tdarius@uos.de

Matriculation Nr. 983821
BA Cognitive Science
6th Semester

1. Supervisor: Ulf Krumnack, Ph.D.
2. Supervisor:

Date: 22.10.2022

Artificial Neural Networks (ANNs) are computationally imitating the workings of a human brain, in which many connections between neurones combined with certain electrical and chemical impulses are able to perform complex tasks. Instead of biological mechanisms these computational networks use mathematical formulas to make predictions for their given task, and in comparison to the human brain, these networks are constructionally quite simple, but nonetheless are able to outperform human capacities in special trained for purposes like Google's deep neural network AlphaGo winning in the game Go against the European Go Champion (Silver, 2016). Apart from games these ANNs also have many real life applications from predicting the stock market to recommending products to customers based on their online behaviour.
These models are getting increasingly bigger to master the increasingly more complex tasks they are dealing with and thus they require better computational resources and longer training periods than their earlier smaller predecessors. However the Lottery Ticket Hypothesis (LTH) (Frankle et al., 2019) states that in these big neural networks there exists a small sub-network which performs similar in terms of accuracy to the whole network, thus introducing the idea of pruning these unnecessary parts of the network to keep the ANN smaller overall.
Since the publication of the Lottery Ticket Hypothesis there have been proposed many methods (Lee et al. 2018; Sreenivasan et al. 2022; Tanaka et al. 2020; Wang et al. 2020) of finding the so called „winning tickets" in a neural network architecture. A winning ticket refers to a sparse sub-network that achieves comparable accuracy to the complete network, which would make training this sub-network faster and more efficient compared to the complete network.
These methods of determining winning tickets strive to achieve comparable or better accuracy than the Iterative Magnitude Pruning (IMP) algorithm introduced by the authors of the LTH. The IMP finds winning tickets reliably, but needs multiple rounds of training and pruning to retrieve such a winning ticket, which results in an overall more complex procedure than simply training the complete network without any pruning, since the accuracy on average does not improve for the generated sparse sub-network. So finding winning tickets without loosing accuracy or investing too much time becomes a worthwhile endeavour. One technique for faster pruning is to prune at the initialisation of the network in a one-shot manner, hence avoiding the computationally ineffective iteratively pruning like the IMP does.

One of the first methods introduced to find winning tickets at initialisation is called *Single-shot network pruning* (SNIP) (Lee et al., 2018) which eliminates weights that do not influence the loss, resulting in a sparse network that can be normally trained afterwards. The successor called *Gradient Signal Preservation* (GraSP) (Wang et al. 2020) is pruning using a slightly different criterion, mainly GraSP is trying to preserve the gradient flow of the neural network, so taking connections into account, while SNIP simply eliminates weights in isolation. Unfortunately both these pruning algorithms suffer from layer-collapse, which renders a neural network incapable of training by eliminating all weights in a single layer, so the *Iterative Synaptic Flow Pruning* (SynFlow) (Tanaka et al. 2020) is avoiding this problem of layer-collapse by introducing another criterion for pruning that also looks at the gradient flow through multiple layers. Another important point to a more abstract picture of winning tickets is that SynFlow is data-agnostic, meaning that it calculates the scores for removing weights by using random noise instead of real data, which may lead to more generalised instead of task-specific winning tickets worth investigating.
Next GEM-MINER (Sreenivasan et al. 2022) is build on the premise that all the algorithms mentioned above simply find good sparse sub-networks but they do not achieve a better accuracy than randomly selected sub-networks with a similar sparsity ratio, so GEM-MINER is finding winning tickets at the initialisation of the network that achieve comparable accuracy to the originally proposed but much more resource and time intensive IMP algorithm by iteratively sampling training data and performing back propagation to determine a supermask for pruning.
This results in so called rare gems (winning tickets) that achieve significant accuracy pre-training while also being up to 19x faster than IMP.

Incidentally all the these algorithms that prune at initialisation start to resemble Early-Bird Tickets (You et al. 2019) which are winning tickets that are found during the early training stages of a network, since both methods need minimal „training" effort to determine good sparse sub-networks.
Lastly there is an approach that does not require any amount of training, because once a winning ticket is found, that ticket then can be transferred to different network structures by adapting the depths of the ticket carefully to suit the new network (Chen et al. 2021). Once adapted that ticket

can be used as a supermask to get the winning ticket from the new network, thus removing the efforts of creating good sparse sub-networks every time a neural network is trained.

Due to these diverse approaches of finding winning tickets, there arises the question whether these winning tickets determined by different methods share any structural similarities when created on the same data set or if they differ vastly, thus giving insight in the structure of winning tickets that possibly could inform a better build of artificial neural network architecture.

Research Question:

I. What similarities do sparse neural networks created by different methods share when found on the same dataset ?

Possible Methodology:

I. Using Pytorch to implement the different approaches of finding winning tickets
II. Either building upon distance measures from similar works or implementing them myself to compare the winning tickets
III. Analyse results, evaluation and interpretation

Possible Problems:

I. Coming up with an effective and novel way of comparing two sparse networks
II. Difficult to assume structural improvements inspired by sparse networks
III. Understanding the thousands of lines of code for the implementation of these algorithms

General Schedule:

Begin ~15.11.2022
End ~15.02.2023

References

(1) Silver, D., Huang, A., Maddison, C. et al. (2016, January 27). Mastering the game of Go with deep neural networks and tree search. Nature 529, 484–489 (2016). https://doi.org/10.1038/nature16961

(2) Frankle, J., & Carbin, M. (2018, March 9). The Lottery Ticket Hypothesis: Finding Sparse, Trainable Neural Networks. arXiv.Org. https://arxiv.org/abs/1803.03635

(3) Lee, N., Ajanthan, T., & Torr, P. (2018, October 4). SNIP: Single-shot Network Pruning based on Connection Sensitivity. arXiv.Org. https://arxiv.org/abs/1810.02340

(4) Wang, C., Zhang, G., & Grosse, R. (2020, February 18). Picking Winning Tickets Before Training by Preserving Gradient Flow. arXiv.Org. https://arxiv.org/abs/2002.07376

(5) Tanaka, H., Kunin, D., Yamins, D., & Ganguli, S. (2020, June 9). Pruning neural networks without any data by iteratively conserving. . . arXiv.Org. https://arxiv.org/abs/2006.05467

(6) Sreenivasan, K., Sohn, J., Yang, L., Grinde, M., Nagle, A., Wang, H., Xing, E., Lee, K., & Papailiopoulos, D. (2022, February 24). Rare Gems: Finding Lottery Tickets at Initialization. arXiv.Org. https://arxiv.org/abs/2202.12002v2

(7) Chen, X., Cheng, Y., Wang, S., Gan, Z., Liu, J., & Wang, Z. (2021, March 30). The Elastic Lottery Ticket Hypothesis. arXiv.Org. https://arxiv.org/abs/2103.16547

(8) You, J., Leskovec, J., He, K., & Xie, S. (2020, July 13). Graph Structure of Neural Networks. arXiv.Org. https://arxiv.org/abs/2007.06559v2

(9) You, H., Li, C., Xu, P., Fu, Y., Wang, Y., Chen, X., Baraniuk, R., Wang, Z., & Lin, Y. (2019, September 26). Drawing Early-Bird Tickets: Towards More Efficient Training of. . . arXiv.Org. https://arxiv.org/abs/1909.11957