THE LOTTERY TICKET HYPOTHESIS: FINDING SPARSE, TRAINABLE NEURAL NETWORKS

Abraham Jose
ID :5068109, CAP6614
abraham@knights.ucf.edu

1 Summary

The hypothesis says that a dense, randomly-initialized, feed-forward networks contain subnetworks (winning tickets) that—when trained in isolation— reach test accuracy comparable to the original network in a similar number of iterations. And if the network size is above the winning size, the network trains faster and reaches higher test accuracy at the earliest.

1.1 Strengths of the proposal

1. Empirically proven that we may need only 10-20% params for best accuracy(test,train) in very less number of iterations or 3.6%params for the original network's performance.

2. Iterative pruning method with initialization of parameters of the original network. The author, through his experiments,conjecture that we can obtain consistent results.

3. Through the experiments, author made a conjecture that the subset of the pruned network, in the original neural network space is dependent on the initial weights. This implies, for any given random initialization of weights, there exists a sub-network with fewer(20%) parameters compared to original network.

1.2 Weaknesses of the proposal

1. Computationally expensive method. No proofs given why it is required to initialize the model with same weights as the original model.

2. The method has been tested only in small image datasets.

3. It is not very clearly mentioned how to use the network knowledge to develop the training schemes while searching for pruned networks.

1.3 Results

The author was able to experimentally prove that it is possible to train a iteratively pruned network using the initial weights of the original network in less number of iterations to reach higher accuracy than the original network.

2 Discussion

2.1 Future Work

A study that proves that the initialization is very important while finding the subsets of the network and subset network is biased by the initialization of the weights. The effect of initialization is connected to the optimization algorithm, dataset used, and model or the structure of the neural network. It is important to define the relationship between them and the sub-network(pruned network).

2.2 Q & A

- Every network is over parameterized by a factor of 10! Why it is not possible to build an original network with 1/5th or even 1/10th of the actual parameter size.
- In one another ICLR 2019 paper *Rethinking the Value of Network Pruning* from Zhuang Liu and Mingjie Sun, "winning ticket" initialization as used in the proposal does not bring improvement over random initialization.
- The effect of pruning in the network's ability to propagate error during back-propagation.

3 Reference

[1] Zhuang Liu & Mingjie Sun (2019) Rethinking the Value of Network Pruning.*ICLR Conference 2019*,