EasyChair Preprint

№ 4819

The Lottery Ticket Hypothesis

Devansh Punj, Mehar Mutreja and Kunal Khandelwal

December 29, 2020

The Lottery Ticket Hypothesis

Devansh Punj, Mehar Mutreja and Kunal Khandelwal

AbstractMachine learning is a branch of artificial intelligence (AI) which helps the systems to gain abilities in such a way that they can automatically learn and improve from the experience without being precisely programmed. Moreover, it focuses on the development of computer programs so that they can access data and use it to learn by adapting to the environment themselves.

Deep learning is a subset of machine learning, which uses neural networks with many layers. Here, we have used Neural Network pruning to reduce the size of neural networks by compression. We find that these methods can reduce parameter counts of pre-trained subnetworks by 90%. Such techniques help in reducing the size, improving the computational performance and at the same time accuracy remains the same if not better.

A standard pruning technique naturally exhibits subnetworks whose initializations make the sub-networks capable of training efficiently. On the basis of the results above we vocalize the Lottery Ticket Hypothesis. These networks contain sub-networks which when trained in isolation are able to reach the test accuracy equivalent to the original network in a similar number of iterations. We present an algorithm to find the winning tickets which is supported by a few experiments in favour of the lottery ticket hypothesis.

Keywords: - Neural Network | Pruning | Artificial Intelligence | Deep Learning

1. Introduction

The idea of pruning originated in the 1990's and was re-popularized in 2015 and to explain the process of pruning in simple terms, it typically follows a relatively standard process. The steps of the process are:

1) Start with training the network wherein each connection in a neural network has been set with a random weight.
2) Remove the superfluous structure or unwanted parts. Networks can be pruned in several ways such as Weights, Neurons, Filters and Channels etc. This paper will majorly focus on pruning on the basis of weights which is also known as Sparse Pruning.

3) Thereafter, Fine-tune the network. Once the network undergoes pruning, it actually gets damaged a little. Thus, Fine-tuning is done in order to clean up the network.
4) Finally, you can choose to repeat step 2-3 iteratively and gradually remove more and more networksafter each round.

So, we ask ourselves why didn't we train a pruned network from the start? It would reduce the costs of network and training. Let us say we do that and follow the following steps:

1) Randomly initialise the pruned network
2) Train it to convergence

Unfortunately, at extreme levels of sparsity, training a pruned network from scratch is actually less accurate than training the full network by the process mentioned earlier and then pruning it.

The lottery ticket hypothesis is "A randomly-initialized; dense neural network contains a subnetwork that is initialized such that — when trained in isolation — it can match the test accuracy of the original network after training for at most the same number of iterations."

So, the new method that they discovered was that by preserving the original weight from the pruned network instead of randomly initiliazing it, one can actually train a network with topology of pruned networks with almost same if not better accuracy within the same number of steps.

2. The Analogy:

The reason behind naming the paper "The Lottery Ticket Hypothesis" was that your goal is to find a winning lottery ticket. So, let us suppose a person bought one ticket to a lottery, his odds of winning the lottery are very low. But if you buy a million tickets, there is a chance that a lot of tickets will get you some money.

Basically, it is an analogy from the gambling world where the training of an AI is compared to that of winning money through a lottery ticket. When we train machine learning models, they produce a large number of neural network structures which are considered as a bag of lottery tickets and when pruning takes place it removes the unwanted parts within the network so as to reduce the size of the model. A similar logic can be applied in the gambling world, where you find the winning tickets and get rid of the rest.

In the context of what is explained in the paper, every small subnetwork that exists within a network is considered to be a lottery ticket. If the idea behind

the lottery ticket hypothesis is correct, then we need to move to the next step, i.e., finding a way to identify the winning tickets.

The paper has two pruning methods to find the subnetworks that are referred to as lottery tickets:

One Shot Pruning:

1) Randomly initialize a neural network
2) Train the network
3) Set p% of weights with the lowest magnitude from each layer to 0
4) Reset the pruned network weights to their original random initializations

In this method, the third step is where the process of pruning really takes place.

Iterative Pruning:

Iterative pruning just iteratively applies the steps of one-shot pruning. The authors found that iterative pruning yielded smaller pruned subnetworks than one-shot pruning.

3. Conclusion:

The pruned networks that were produced were almost 10-20% smaller in size than the original network and the accuracy was the same or higher than the original network. They also had better generalization as they were smaller difference between train and test accuracies.

Even though the paper showed good results, but the iterative pruning methods shown do not provide significant practical benefits as the computational expense for iterative pruning is very high because it involves training a network 15 times per trial.

The authors of the paper found if harder to review large datasets and suggested that in the future, a non-magnitude based pruning strategy could find smaller pruned subnetworks.