View Reviews

Paper ID
3549

Paper Title
Power-law in Sparsified Deep Neural Networks

Reviewer #1

Questions

1. Please provide an "overall score" for this submission.
6: Marginally above the acceptance threshold. I tend to vote for accepting this submission, but rejecting it would not be that bad.

2. Please provide a "confidence score" for your assessment of this submission.
2: You are willing to defend your assessment, but it is quite likely that you did not understand central parts of the submission or that you are unfamiliar with some pieces of related work. Math/other details were not carefully checked.

3. Please provide detailed comments that explain your "overall score" and "confidence score" for this submission. You should summarize the main ideas of the submission and relate these ideas to previous work at NIPS and in other archival conferences and journals. You should then summarize the strengths and weaknesses of the submission, focusing on each of the following four criteria: quality, clarity, originality, and significance.
This paper proposes the meaningful connectivity of nodes in an artificial neural network follows a power law property w.r.t to the degree of connectivity between adjacent layers. The authors demonstrate their hypothesis by using standard pruning techniques to reveal the important connectivity in a trained neural network follows this distribution (Truncated Power Law). They further go to show that using this distribution to grow neural networks outperform the baseline of assigning random connectivity.

This work shows interesting avenues of research in optimizing the connectivity of neural networks in the context of continual learning. This appears to be a promising departure from the standard framework of neural networks which assumes a fixed connectivity topology through training and evaluation.

Comments:

Given multiple definitions of sparsity in the representation learning community, it would be helpful to define sparsity earlier in the text and mention that the sparsity described in this work is w.r.t parameters (as opposed to activations).

The authors should justify the definition of a node in a ConvNet as an entire feature map rather than as an x,y position within a feature map. It is unclear why this definition is preferred over others. This likely depends on the form of pruning being applied to the convolution kernel. These definitions are particularly important as the paper makes connections to biological neural networks.

As the increased prevalence of skip connections (residual) layers and its variants (DenseNet, etc) are becoming common in the neural network community, the paper should comment about how this particular type of power-law analysis extends to these well-known architectures.

Minor issues:

Can the performance differences between AlexNet on ImageNet before and after pruning be specified to be consistent with the other sections (and as a necessary sanity check)?

4. How confident are you that this submission could be reproduced by others, assuming equal access to data and resources?
2: Somewhat confident

Reviewer #2

Questions

1. Please provide an "overall score" for this submission.
7: A good submission; an accept. I vote for accepting this submission, although I would not be upset if it were rejected.

2. Please provide a "confidence score" for your assessment of this submission.
2: You are willing to defend your assessment, but it is quite likely that you did not understand central parts of the submission or that you are unfamiliar with some pieces of related work. Math/other details were not carefully checked.

3. Please provide detailed comments that explain your "overall score" and "confidence score" for this submission. You should summarize the main ideas of the submission and relate these ideas to previous work at NIPS and in other archival conferences and journals. You should then summarize the strengths and weaknesses of the submission, focusing on each of the following four criteria: quality, clarity, originality, and significance.
This interesting paper studies power laws in deep neural networks in several ways. First the authors study whether pruning (sparsification) procedures on neural networks lead to results in accord with the power law and concludes that they do, at least in the datasets they looked at. Second, inspired by processes known to give rise to power law distributions, they proposed a preferential attachment model that automatically grows a neural network in order to learn new tasks in a continual learning setting.

I wish the authors had tried a bit harder to convey an intuition for why power laws make since in neural networks (biological or artificial). They could at least have set out a few hypotheses to help the reader gain some intuition for the domain. They looked mostly at continual learning problems. Are there other kinds of problems where this connectivity makes sense? Are there any problems where it is expected to be counterproductive? Some additional comments along these lines could improve the paper.

4. How confident are you that this submission could be reproduced by others, assuming equal access to data and resources?
2: Somewhat confident

Reviewer #3

Questions

1. Please provide an "overall score" for this submission.
7: A good submission; an accept. I vote for accepting this submission, although I would not be upset if it were rejected.

2. Please provide a "confidence score" for your assessment of this submission.
3: You are fairly confident in your assessment. It is possible that you did not understand some parts of the submission or that you are unfamiliar with some pieces of related work. Math/other details were not carefully

checked.

3. Please provide detailed comments that explain your "overall score" and "confidence score" for this submission. You should summarize the main ideas of the submission and relate these ideas to previous work at NIPS and in other archival conferences and journals. You should then summarize the strengths and weaknesses of the submission, focusing on each of the following four criteria: quality, clarity, originality, and significance.

This paper tests the degree distribution in both DNN and CNN for a number of different scenarios. Previously there are some works regarding pruning DNN, and some works in continual learning.

The paper is really clear and quite enjoyable to read because there are tons of experiments and figures, and the figures are also well presented. The results are also very promising because (1) it fits power law very well. (2) good performance on continual learning

The paper is also original since I haven't read similar work before. Especially for the preferential attachment part in continual learning, the strategy is quite interesting.

Q1: Line 28, "biological neural networks are often scale-free", should it be biological network? Since the degree distribution follows power law, then it is a discrete distribution, so should the power-law has a probability mass function rather than PDF? Please correct me if I was wrong.

Q2 As for experiments in section 3, is it possible to add one experiment (either in paper or in supplementary material) that shows what will happen if we just remove edges randomly (with the same s) other than prune them by weights?

Q3 I still don't quite get the intuition about why the resulted network will follow power law because each node only have chance to connect to nodes in previous and next layer. What will happen if you train a ResNet with 100 layers on CIFAR 10 (100) or Imagenet?

4. How confident are you that this submission could be reproduced by others, assuming equal access to data and resources?

3: Very confident