PAPER REVIEW
SNIP: SINGLE-SHOT NETWORK PRUNING BASED ON CONNECTION SENSITIVITY

Muhammad Umar Salman
21010241@mbzuai.ac.ae

Samar Fares
21010173@mbzuai.ac.ae

Yu Kang Wong
21010186@mbzuai.ac.ae

1. Summary

The goal of a pruned network is to learn a smaller network which mimics the performance of larger complex reference network. This paper takes a new approach on pruning where a network is pruned at initialization prior to training. For this purpose, the paper introduces a saliency criterion based on connection sensitivity which identifies structurally important connections in the network at a single shot. This eliminates the need for both expensive pretraining and the complex pruning schedule while making it robust to architectural variations. The main motivation behind this paper is to efficiently and effectively prune networks to make them less memory-consuming and less computationally expensive. The paper argues that its approach is more simple, versatile and interpretable in nature as compared to the modern iterative optimization procedure which requires prune-retrain cycles. The study evaluates its methods on MNIST, CIFAR-10 and Tiny-ImageNet classification datasets with widely varying architectures which obtains extremely sparse networks with virtually the same accuracy as the existing baselines. The paper then shows the approach is generally applicable to other complex models without any change in architecture. The paper finally goes on to discuss future work in the topics of neural network architectures, multi-task transfer learning, structural regularization and the generalization of sparse networks

2. Strengths

Firstly, a major strength the paper uses is a solid motivation to build upon its arguments. It highlights the that deep neural networks are highly overparameterized making them computationally expensive with excessive memory requirements which aren't suitable for mobile device applications. The authors then discuss how their pruning method eliminates the need for both pretraining and the complex pruning schedule. This motivation to find a solution for such a big problem captures the readers interest early on.

Secondly, the paper takes you through all the complex mathematics and logical flow step-by-step for the proposed pruning method assuming that the reader would only know the very basics and then builds upon those mathematical theories using proper mathematical notation and explaining why each change is done at each step, giving the reader an idea of the thought process taken by the authors.

Thirdly, another strength which differs from many other papers is that this paper has added a pseudo code for the proposed pruning algorithm and not only does it clearly show what's happening in each line of code but for further explanations in front of each line there is a reference to the section in which the idea of that algorithm has been discussed within the paper. This allows the reader to understand why each line of code has been added and where it has been added.

Fourthly, the experiments section is very well thought out. The paper first begins with the experiment setup along with providing the code. After that the authors make their proposed model's comparison with all the modern models that they have mentioned in their related work proving through well-constructed graphs and detailed tables how their model without expensive iterative prune cycles does just as good if not better. The paper further goes on to defend its argument about the model being robust to architectural variations by showing good performances to different type of networks.

Lastly, the paper is clear and precise in its language and choice of words. The sentences are neither too long nor too complex and are free of grammatical errors which makes the paper easy to read and follow leaving a good lasting impression for the readers.

3. Weaknesses

One of the few weaknesses that could be seen in the paper was the excessive redundant use of the authors mentioning how their model is simple, versatile and interpretable along with the fact that their approach prunes any network prior to training and its robustness to architectural variation which can get annoying for the reader.

Another weakness was the discussion regarding initialization of weights which in the paper was a very crucial part as the scale of weights would directly have an effect on the proposed model and even though throughout the paper the authors were very detailed in their methodology and reasoning when it came to weight initialization, they just referred the readers to an example without giving any proper insights into it.

One other weakness in the paper was that although multiple types of experiments were conducted by the authors to prove their model, the experiments were only done on mini-batches for the MNIST and CIFAR data as opposed to larger datasets. The large datasets are what bring about the issues of excessive memory requirements which the authors in their motivation said they wanted to solve.

Furthermore, even though the paper is very well written and easy to grasp for readers. The authors after the experiments leaves their readers without any conclusion section or conclusive remarks making the readers infer from the experimental results what they are trying to conclude.

4. Justification

The paper starts off with a meaningful and concise title and a very solid motivation which sets the base and tone of the paper. This strong start gives a huge positive weight to the overall paper in the mind of the reader. It is then followed by some well-reasoned arguments and a good logical flow which leaves a very positive lasting effect and keeps the reader interested. This is followed by step-by-step mathematics which is a huge plus for any reader trying to get a better understanding of the underlying work and this paper has done exactly that. Credit for this clear understanding also goes to the pseudo code algorithm which helps redirect confused readers trying to understand the code. Presenting experimental findings in such a tabulated manner is always a good sight for at a glance inference but what helps even more for digging deeper into experiments is the experiment setup and having access to the code which the authors have done boosting their rankings with the readers.

These strengths play a huge advantage in the authors favor as they create a superb idea backed by good formatting and grammatically flawless writing. As for the weaknesses, even though the redundancy doesn't create a huge negative impact it is still noticed by a careful reviewer. The negative parts which do majorly affect the paper is the missing of the conclusion section and working on small datasets. This not only gives the reviewer the idea that the authors seem a little inconclusive about their proposed model but questions the validity of the test results as well. This weakness further grows when the explanation an extremely crucial part of the algorithm is also seen missing in the paper. However, the paper all in all besides its few lows was insightful and enjoyable to read giving the reviewers confidence to rate this paper extremely well.