FID3018 Opposition Report 1

For Tianze Wang session on Device Placement

By Abubakrelsedik Karali
karali@kt.se

The topics of the session had a unified theme of optimizing the device placement for neural networks training. Which is a vital research problem that deals with finding the optimal approach to train neural network on the available hardware resources. The topic is very relevant to the course and for me as part of the audience. In the session three main papers was discussed while the two of them are from the same research group:

- A hierarchical model for device placement
- Spotlight: Optimizing Device Placement for Training Deep Neural Networks
- Post: Device Placement with Cross-Entropy Minimization and Proximal Policy Optimization

The first paper discusses an approach which learns to optimize device placement for training neural networks that have tens of thousands of operations with no need for manual grouping. The method consists of a two-level hierarchical network, in which the first model groups the operations of the graph (the Grouper) and the second model places those groups onto devices (the Placer). The Grouper is a feed forward network which reads in information about each operation and its context within the graph, to predict the group to which that operation should be assigned. The Placer is a sequence-to-sequence model that reads in the embedding of the group and predicts the device placement for that group. In the second paper is based on the claim that using reinforcement learning to learn placement skills by repeatedly performing Monte-Carlo experiments is efficient but due to its equal treatment of placement samples, They proposed new joint learning algorithm, called Post, that integrates cross-entropy minimization and proximal policy optimization to achieve theoretically guaranteed optimal efficiency. So that the training process can complete within the shortest amount of time. The third paper proposed Spotlight, a new reinforcement learning algorithm based on proximal policy optimization, designed for finding an optimal device placement. The design relies upon a new model of the device placement problem: by modeling it as a Markov decision process with multiple stages, we will be able to achieves a theoretical guarantee on performance improvements.

The composition of the presentation in terms of highlighting the interesting aspects of the papers was very good. However, some figures were in low resolution which wasn't very appealing to the eye and not very clear. The presenter gave a very good exposure if the technical details of individual papers and showed solid understanding and could easily go a bit more. In the QA session, the presenter managed to address the questions put forward by the audience successfully. One question I still have for the presenter and in general for this specific field of research is. How are they going to cope with the advances in hardware resources which goes faster and faster and putting the whole research field in danger.

Mirhoseini, Azalia, Anna Goldie, Hieu Pham, Benoit Steiner, Quoc V. Le, and Jeff Dean. "A hierarchical model for device placement." In *International Conference on Learning Representations*. 2018.

Yuanxiang Gao, Li Chen, Baochun Li: Spotlight: Optimizing Device Placement for Training Deep Neural Networks. *ICML 2018: 1662-1670*

Gao, Y., Li Chen and B. Li. "Post: Device Placement with Cross-Entropy Minimization and Proximal Policy Optimization." *NeurIPS* (2018).