Critique 2 of 10

Written by: Victor Palffy

Submitted on: 05/25/2020

Publication venue: SOSP

Title: “PipeDream: Generalized Pipeline Parallelism for DNN Training“

In this article, the author opens by describing a common paradigm found in today’s approach to Deep Neural Networks (DNNs) parallelization training. In this approach, communication overhead cost are high. In fact, as DNNs scale, communication is believed to be the bottleneck. The author proposes a solution to this constraint, and, that solution is PipeDream.

PipeDream is a model that utilizes pipeline parallelism to accelerate a DNNs training by merging both the inter-batch parallelization with intra-batch parallelism. Because of this, PipeDream drastically reduces communication between input workers, which ultimately improves system performance. Although pipelining in general is not a new idea, PipeDream has been developed to use high level hardware more effectively and steady state stalls have been eliminated. Its secret is its scheduling algorithm know as one forward one backward or (1F1B).

The 1F1B scheduling algorithm was design to more efficiently manage the pipeline by adding weight updates with every stage backward pass. The updated weight is used and an additional copy is stored on the forward pass. The stored copy is later retrieved for use for the corresponding backward pass. Although, this is not a perfect system it has proved to be more effective than previous methods and has drastically improved training speeds. PipeDream’s pipeline parallelism training was improved further by incorporating round-robin.

Round-robin scheduling was added as an addition to PipeDreams 1F1B algorithm. The improved 1F1B-RR scheduling algorithm helped pipedream achieve its accuracy goal when utilizing multi-GPU machines. enhanced image classification, machine translation, language modeling and video capturing tasking’s. Pipedream has also outperformed model parallelism and hybrid parallelism.

What I liked most about this article was, how the author described the evolution of pipedream and its inner workings. I felt that the author did a great job by including statistical data on pipedreams improvement over other models. I spent a long time reading and rereading this article in hopes to build a better understanding of this type of technology. I felt more confident towards the completion of my summary. Overall, I learned a lot from reading this article on how parallelism is being utilized and deep neural networks.