Julian Singkham
CS 2300/11
Dr. Berisha

Paper Review 4: Horovod

1. Why did you pick the paper compared to the others?

I chose this paper over the others as I find parallelization a fascinating topic. This paper discusses parallelization between multiple GPU's and how TensorFlow had poor support for this. Distributed computing allows for both cheap scaling as one doesn't have to continuously buy the next best GPU and better performance as the computational load can be split between multiple physical GPUS.

2. Write a few sentences that describe the main algorithm or concept presented.

This paper details how Uber created the Horovod package to easily add distributed processing to TensorFlow models. The issue at hand is that currently with TensorFlow there is a huge amount of overhead when utilizing multiple GPUs, to the point where efficiency is cut in half when 128 GPUs are used. Additionally programmers needed to be experts in parallelism and custom design their code for the amount of GPUs used.

3. Why is this idea/concept new/novel?

This concept is new in the sense that distributed computing is a brand-new field in the computational science realm and currently there exists huge inefficiencies with the standard method. By creating an easy to use library that requires minimal integration to existing code, the cost of computing drops as users can now utilize en-mass cheaper GPUs and dramatically increases processing speed.

4. What are the limitations/weaknesses of the approach?

The only issue with this approach is that it only achieves ~80% efficiency at 128 GPUs. Other then that I can't find any issues with using Horovod outside of its age as it hasn't been fully tested with very large datasets. Horovod seems easy to implement into existing code bases and the developers seem passionate about continuous support and utilizing the open-source community.

5. Was the paper well-written? Identify the most difficult part to grasp/understand.

The paper was very easy to understand and read as the authors simplified a lot of the technical aspects. Compared to the other papers I've read this seems to be a summary rather than a full fledged technical document that discusses every aspect. To me this is a great example of how a paper should be written as it doesn't contain unnecessary jargon and only contains relevant information.

6. What is your main takeaway from this paper?

The main takeaway from this paper is how complicated distributed computing is. Without the help form Nvidia libraries the Horovod package would've been many times more complicated to develop as it would require intricate knowledge of how the Nvidia GPUs worked and parallelization. Additionally, utilizing a vendor developed library allows for more efficient codebase as the vendor has complete knowledge of their system base on top of proprietary knowledge. NCCL very much expediated the development of Horovod.