OFFICIAL ABSTRACT and CERTIFICATION

Mitigating Communication Bottlenecks in MPI-Based Distributed Learning

Abdullah B. Nauman
Ward Melville H.S, East-Setauket NY, USA

Current commercial and scientific facilities generate and maintain vast amounts of complex data. While machine learning (ML) techniques can provide crucial insight, developing these models is often impractical on a single process. Distributed learning techniques mitigate this problem; however, current models contain significant performance bottlenecks. Here, we conduct a detailed performance analysis of MPI_Learn, a widespread distributed ML framework for high-energy physics (HEP) applications, on the Summit supercomputer, by training a network to classify simulated collision events from high-energy particle detectors at the CERN Large Hadron Collider (LHC).

We conclude that these bottlenecks occur as a result of increasing communication time between the different processes, and to mitigate the bottlenecks we propose the implementation of a new distributed algorithm for stochastic gradient descent (SGD). We provide a proof of concept by demonstrating better scalability with results on 250 GPUs, and with hyperparameter optimization, show a ten-fold decrease in training time.

Category
Pick one only — mark an "X" in box at right

- Animal Sciences
- Behavioral & Social Sciences
- Biochemistry
- Biomedical & Health Sciences
- Biomedical Engineering
- Cellular & Molecular Biology
- Chemistry
- Computational Biology & Bioinformatics
- Earth & Environmental Sciences
- Embedded Systems
- Energy: Sustainable Materials and Design
- Engineering Mechanics
- Environmental Engineering
- Materials Science
- Mathematics
- Microbiology
- Physics & Astronomy
- Plant Sciences
- Robotics & Intelligent Machines
- Systems Software ■
- Translational Medical Sciences

1. As a part of this research project, the student directly handled, manipulated, or interacted with (check ALL that apply):

   ☐ human participants ☐ potentially hazardous biological agents

   ☐ vertebrate animals ☐ microorganisms ☐ rDNA ☐ tissue

2. I/we worked or used equipment in a regulated research institution or industrial setting: ■ Yes ☐ No

3. This project is a continuation of previous research. ☐ Yes ■ No

4. My display board includes non-published photographs/visual depictions of humans (other than myself): ☐ Yes ■ No

5. This abstract describes only procedures performed by me/us, reflects my/our own independent research, and represents one year's work only ■ Yes ☐ No

6. I/we hereby certify that the abstract and responses to the above statements are correct and properly reflect my/our own work. ■ Yes ☐ No

This stamp or embossed seal attests that this project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.