Distributed Pytorch Performance Analysis

By Harshit Yadav & Shraddha Naik

Abstract

- Deep Learning and Data Mining Algorithms are Increasingly Important in Processing Large amount of data
- Many New Libraries and Framework built on top of legacy Drivers and codebase lack support for Multi Core Execution Support
- Parallelism helps in Decreasing the Processing time
- Deep Learning Computation is one of the Most Resource Intensive yet Most Efficient in Building any model in field of ML
- Given Enough Computation power Deep Learning can beat accuracy and efficiency of any existing solution

Introduction

- Deep Learning Algorithms Emulate the Brain Activity of Several layers of neurons and learn the weights for the Synapses by using Gradient Descent Method
- Open Source Efforts such as THeano , CuDNN and Caffe have gained wide acceptance among Researchers
- These Frameworks are useful for experimenting new Algorithms and Techniques but fail to utilize the Hardware capabilities for processing large amount of Data

Literature Review

- Documentation and Performance Analysis done for various datasets and frameworks on different Hardware was studied
- Identified the Knowledge gap and Analysis not available online for various popular and uprising and frameworks
- Narrowed down the results to frameworks that are readily available for deployment with Multi-Core
- Among the top 3 frameworks used in Industry Level widely were

Microsoft-DMTK   Google- TensorFlow   Facebook - Pytorch

Methodology- MPI (Message Passing Interface)

- MPI was Introduced in 1992
- MPI Interface can be accessed using many languages such as C++,C,
- Since many important Core Mathematical Libraries were built using these languages many of the modern ML frameworks use wrapers around these libraries
- Optimising and Parallelising these libraries code-base at the root level gives us a significant boost in performance to all the frameworks utilizing it

Methodology-Pytorch

- Facebook Developed Pytorch to answer their need of a Machine Learning Algorithm to analyse the Visual Data generated by the Users on their Social Media Platform
- Torch which is based on the programming language LUA which is highly efficient in computing machine learning
- Unfortunately , Pytorch framework can not be included with MPI implementation because of various proprietary and patented technologies being used in both of them individually owned by separate entities

Proposed Model (Pytorch Using MPI)

- Pytorch binaries have to be manually recompiled which is fairly simple and time consuming since the compiling part is done sequentially
- On execution the PyTorch the will look by itself for the available MPI backend implementation in the Binaries
- Proposal based on reading various recent article and studies which are being practically implemented in the industry at production level and deployed across datacenters in the Tech Giants like Facbook

Result and Analysis

-
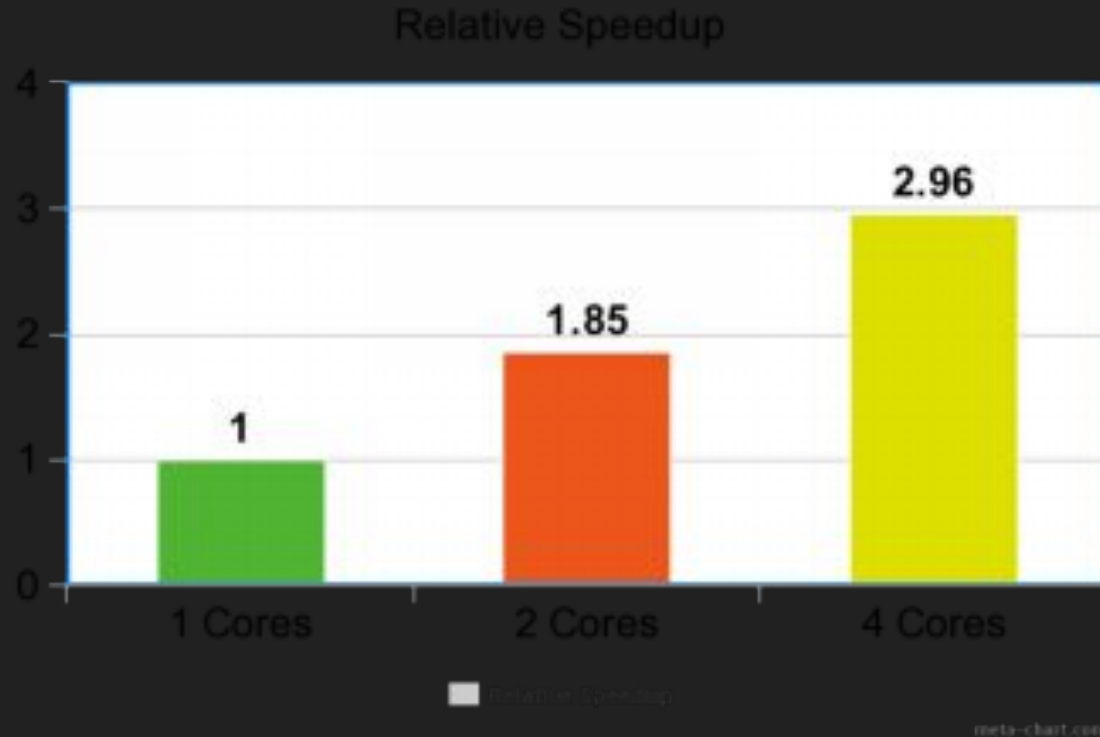
Conclusion

- Evaluating the MNIST dataset gives the best speedup when executed parallely on 4 core using Pytorch with MPI. we checked the speedup i.e. sequential execution with parallel execution using 4 cores and we got 69% efficiency.
- Further research can be carried out to implement Pytorch using NCCL and other standard datasets to analyze its efficiency .

Reference

- Abhinav Vishnu, Charles Siegel Jeff Daily, Distributed TensorFlow with MPI, Macrh 7, 2016.
- https://medium.com/@esaliya/pytorch-distributed-with-mpi-acb84b3ae5fd

- All code Open source and Posted Open source on Github repository

Thank You