

Theme: Artificial Intelligence

Subject: Fast Parallel and Distributed Framework for Deep Learning

Introduction

Since the mid of 2000s, deep learning (DL) has been one of the leading machine learning techniques. Many experts on artificial intelligence (AI) often say that DL is a big step towards realizing Searle's strong AI hypothesis. Current research shows that DL outperforms previous the state-of-the-art on many fields, such as image recognition, speech recognition, and natural language processing. However, many DL-related algorithms seem to work with the support of high-end resources. And, many algorithms are based on a single machine and parallel and distributed ones are beginning to get attention. However, the current versions are not well documented yet nor be applicable into a broad range of DL algorithms. The goal of this research project is to explore novel framework for deep neural network in parallel and distributed environment. We expect the new framework to enable model parallelism within a machine via multithreading and across machines via message passing.

Scope

Challenges that significantly advance the state-of-the-art of parallel and distributed method in deep learning which include:

- Method to partition a neural network in an optimal way for effective distribution.
- Method to train distributed network weights in an efficient manner.
- Method to find the best (or near-optimal) configuration (i.e., the connection between nodes in adjacent layers) of neural network for the given learning task.

Research questions

We are interested in the following research questions. These questions are not exhaustive but different research questions are open to discuss with research partners.

- What would be the most effective way to partition a neural network in the parallel and distributed environment for best performance?
- What would be the optimal configuration of neural network for the given the resources (i.e, number of process within a machine and clusters)?
- What would be an efficient distributed optimization method for learning weights?

Expected Deliverables

The following is open to discussion:

- Suggestion of new materials or new structure
- Detailed progress reports every 3 months summarizing accomplishments.
- Prototype samples
- Patents with Samsung SDI (if agreed)