Performance Evaluation of Distributed Deep Learning Frameworks in Cloud Environment

Authors : Shuen-Tai Wang, Fang-An Kuo, Chau-Yi Chou, Yu-Bin Fang
Abstract : 2016 has become the year of the Artificial Intelligence explosion. AI technologies are getting more and more matured that most world well-known tech giants are making large investment to increase the capabilities in AI. Machine learning is the science of getting computers to act without being explicitly programmed, and deep learning is a subset of machine learning that uses deep neural network to train a machine to learn features directly from data. Deep learning realizes many machine learning applications which expand the field of AI. At the present time, deep learning frameworks have been widely deployed on servers for deep learning applications in both academia and industry. In training deep neural networks, there are many standard processes or algorithms, but the performance of different frameworks might be different. In this paper we evaluate the running performance of two state-of-the-art distributed deep learning frameworks that are running training calculation in parallel over multi GPU and multi nodes in our cloud environment. We evaluate the training performance of the frameworks with ResNet-50 convolutional neural network, and we analyze what factors that result in the performance among both distributed frameworks as well. Through the experimental analysis, we identify the overheads which could be further optimized. The main contribution is that the evaluation results provide further optimization directions in both performance tuning and algorithmic design.