Xiaowen Chu et al., Adv Robot Autom 2018, Volume:7
DOI: 10.4172/2168-9695-C2-017

Xiaowen Chu
Hong Kong Baptist University, Hong Kong, ROC

Co-Authors
Shaohui Shi, Qiang Wang
Hong Kong Baptist University, Hong Kong, ROC

Performance modeling and evaluation of distributed deep learning frameworks on GPUs

Deep learning frameworks have been widely deployed on GPU servers for deep learning applications in both academia and industry. In training of deep neural networks (DNNs), there are many standard processes or algorithms, such as convolution and stochastic gradient descent (SGD), but the running performance of different frameworks might be different even running the same deep model on the same GPU hardware. In this paper, we evaluate the running performance of four state-of-the-art distributed deep learning frameworks (i.e. Caffe-MPI, CNTK, MXNet and TensorFlow) over single-GPU, multi-GPU and multi-node environments. We first build performance models of standard processes in training DNNs with SGD and then we benchmark the running performance of these frameworks with three popular convolutional neural networks (i.e., AlexNet, GoogleNet and ResNet-50), after that, we analyze what factors that result in the performance gap among these four frameworks. Through both analytical and experimental analysis, we identify bottlenecks and overheads which could be further optimized. The main contribution is that the proposed performance models and the analysis provide further optimization directions in both algorithmic design and system configuration.

Biography

Xiaowen Chu is an Associate Professor in the Department of Computer Science, Hong Kong Baptist University. He is also the Director of High Performance Cluster Computing Centre of HKBU. He has received his Bachelor's degree in Computer Science from Tsinghua University, China, in 1999 and PhD degree in Computer Science from the Hong Kong University of Science and Technology in 2003, respectively. His current research interests include distributed and parallel computing, cloud computing and also computer networks and wireless networks. He has published more than 150 research papers in a variety of international journals and conferences. He is serving as an Associate Editor of IEEE Internet of Things Journal and IEEE Access.

chxw@comp.hkbu.edu.hk

Notes: