Brett Duncan
CSc 8530 Parallel Algorithms
Annotated Bibliography

1. **Parallel Training of An Improved Neural Network for Text Categorization**
   This paper proposes to parallelize the improved neural network. The improved neural network is an improvement over the traditional back-propagation learning algorithm to make learning faster, to avoid local minima, and to improve generalization ability. There are generally two approaches for parallelizing neural networks: pattern parallel and network parallel. Pattern parallel is used in this paper.
2. **A Parallel Computing Platform for Training Large Scale Neural Networks**
   This paper implements cNeural, a parallel computing platform for training large scale neural network. Training is divided into two phases: training data loading and training process executing. Large scale training datasets are stored in Hbase to reduce the time cost of data loading. The platform could be deployed on Amazon EC2 or general PCs interconnected in a network.
3. **Parallel Neural Network Training with OpenCL**
   OpenCL is used to implement two parallel neural network training algorithms: parallel backpropagation, and parallel particle swarm optimization (PSO). Programming elements in OpenCL are based on work items or kernels, and are processed in parallel by a compute unit called a workgroup. Maximum efficiency is achieved on a GPU, but it could also be executed on a general purpose CPU. Training is more efficient with larger networks.
4. **Parallel Deep Neural Network Training for Big Data on Blue Gene/Q**
   Training in deep neural networks is slow, so the Blue Gene/Q computer system is used to parallelize it. The deep neural network is trained using the Hessian-free 2^{nd} order optimization algorithm. In what would take a month to train a deep neural network with 100 million+ parameters, it took 6 hours with two racks of Blue Gene.
5. **Parallel Batch Pattern Training Algorithm for Deep Neural Network**
   Each layer is treated as an unsupervised Restricted Boltzmann Machine (RBM) and pre-trained one at a time. Afterwards, the standard supervised backpropagation is used to train the layers. The batch pattern training algorithm updates neurons' weights and thresholds after all training patterns are processed (an epoch), instead of after each training pattern.
6. **Parallel Implementation of Neural Networks Training on Graphic Processing Unit**
   Backpropagation is used as a fine-tune technique of a deep belief network, and pre-training is not considered. In the training algorithm, sum and max are reduced on the GPU. The CPU and GPU implement different instructions in parallel, with the CPU handling logic instructions and the GPU handling complexity computing.
7. **Parallel Batch Pattern Training of Neural Networks on Computational Clusters**
   A batch pattern training algorithm is used instead of implementing a parallel MLP of the standard sequential training algorithm, because the parallelization using the standard sequential training algorithm has high synchronization and communication overhead. Batch pattern training updates neurons' weights and thresholds at the end of an epoch. Experiments using the ccNuma architecture showed performance improvements.
8. **Sequence Training of Multiple Deep Neural Networks for Better Performance and Faster Training Speed**
   A multiple deep neural network (mDNN) is used for acoustic modeling. Sequence training on mDNN is used, based on the maximum mutual information (MMI) criterion. Sequence training has 3 steps: i) DNN forward past, in which posterior probabilities of all HMM states are computed for all feature frames. ii) Word graph processing: performs forward-backward

algorithm in each word graph. iii) DNN back propagation.

9. **The Improved BP Algorithm Based on MapReduce and Genetic Algorithm**
   This paper combines MapReduce and a genetic algorithm to improve upon previous algorithms based on MapReduce. In in the MapReduce BP algorithm, batch training is used, in which weights are updated after all records in training have been processed. In the genetic algorithm, all the weights of the BP neural network are encoded as one chromosome. The reduce function is where the selection, crossover, and mutation of the genetic algorithm occur.
10. **Efficient Parallelization of Batch Pattern Training Algorithm on Many-core and Cluster Architectures**
    Batch training is also used in this paper, in which the weights and thresholds of each neuron is updated after all training patterns are processed (after an epoch). A recirculation neural network is used, which compresses the input pattern space of X to an output vector $\overline{X}$ containing compressed data. Open MPI, Mvapich, and Intel message passing libraries are used in the experiments and results compared.