Distributed Neural Networks: A Parallel Approach

Tasneem Alowaisheq and Devendra Singh Dhami

ABSTRACT

In this work we are implementing a parallel mini-batch back propagation algorithm using Hadoop framework. Our results show that the parallel algorithm converges faster than the sequential algorithm, remaining comparable in the accuracy achieved.

Problem Definition

Our main goal: design a distributed neural network to get a classification performance with less training time comparable to the sequential neural network.

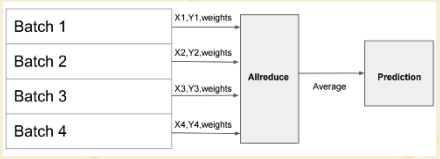
Motivation:

1. The data is large and cannot be handled by one process.
2. For complex structures the size of a layer becomes too large for one process to handle.

Proposed Method

Our method is based on the parallelization using model averaging manner. [1]
The method used for training the neural network is mini-batch gradient descent.
1. Each process will train on its local mini-batch data (X : feature vector and Y: corresponding targets).
2. Update local weights, across all the processes, then combined using **Allreduce** and averaged.
3. These weights are the final model used for prediction.

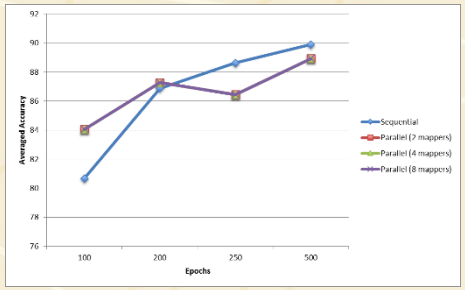
Proposed Method



The Parallel NN Algorithm

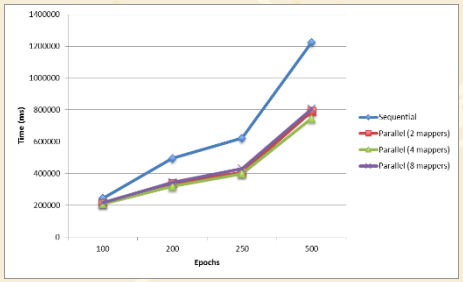
Data: The input data and corresponding labels
Result: List of accuracies across N splits
for *the number of splits, N* **do**
  Divide the data randomly into training and testing;
  Initialize the weights;
  for *(number of iterations / 5)* **do**
    Randomly sample the mini batch from training;
    Train for 5 epochs to update the weights locally;
    Allreduce;
    Average the models;
  end
  if *master* **then**
    Predict on test and report accuracy;
  end
end

Performance Comparison between Sequential NN and Parallel NN

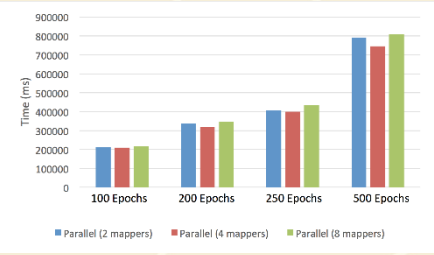


Faster Convergence

The parallel NN converges faster than the sequential NN algorithm



Time Comparison Between Different Number of Mappers



CONCLUSIONS

In this paper we presented a parallel implementation for the mini-batch neural network. We demonstrated our method on a digit classification task. Our results show that the parallel version is comparable to the sequential one in terms of accuracy. The parallel method converged significantly faster than the sequential method.

REFERENCE

[1] H. Su and H. Chen. Experiments on parallel training of deep neural network using model averaging. arXiv preprint arXiv:1507.01239, 2015.

INDIANA UNIVERSITY BLOOMINGTON
SCHOOL OF INFORMATICS AND COMPUTING