Summary 16

Shaun Pritchard

Florida Atlantic University

CAP 6778

November 16, 2021

M. Khoshgoftaar

An Examination of Neural Networks on Cluster Computers

The purpose of this study is to evaluate the effects of computational resources on deep learning models. A comparison was performed between convolutional neural networks (CNNs) with image data that does not have imbalances and multi-layer perceptron models (MLPs), trained on highly class-imbalanced Medicare Part-B fraud big data sets. The overall purpose of this study was to evaluate and compare deep learning models across distributed cluster computers using data parallelism and hardware acceleration to prove overall performance by shorter training times and fewer resources. Random undersampling techniques were applied to both deep learning models using both image data and high-class imbalanced Medicare Part B big data. Moreover, models performance was measured using AUC-when trained across multiple workers with the MLP.

Deep learning algorithms, such as CNN, can learn weights and biases that are applied to input images. The large size of image data and the high computational complexity of this type of experiment make it ideal for a study like this.

MLP is a class of feedforward artificial neural networks (ANN). MLP employs an algorithm known as backpropagation for training. MLPs differ from linear perceptrons in that they have multiple layers and non-linear activation. They are capable of distinguishing data that is not linearly separable. An algorithm for supervised learning of artificial neural networks using gradient descent. Backpropagation implements the backward propagation of errors.

An Artificial Neural Network (ANN) consists of n layers and neurons. A neuron is a mathematical function that represents the functioning of a biological neuron. It is best to train neural networks using massive big data sets and to include a larger number of neurons and layers. However, a significant amount of computational power and time is required for this. CPUs and GPUs are typically needed for evaluating deep learning models. The purpose of this study was to examine the effects of parallelization on neural network training time using a variety of processors and a number of workers managed by Slurm workload manager cluster management tools.

Based on the results of this study, GPU acceleration can reduce training time depending on the model type and dataset type, but is not guaranteed across all types of NN architecture. According to the results, distributed training has a greater effect on training time and model performance when training on high-class imbalanced data, compared to training on data that is not class imbalanced. There was a significant drop in model performance when measured with AUC, with the largest drop occurring around 8 worker nodes with MLP parallelism not scaling as well as CNN on image data, and A smaller MLP model on Medicare Part B data does not scale as well as a CNN.

According to the researchers, more research is needed on this topic. I would find it fascinating to see the results of a new study that compares parallel processing used in this study with batch processing and scheduling techniques commonly used in high-performance computing systems using Kubernetes.