Cyclical Learning Rates (CLR'S) for Improving Training Accuracies and Lowering Computational Cost

Pawar Shrikant

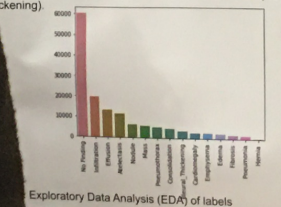
Department of Computer Science and Biology, Claflin University, Orangeburg, SC, 29115, USA

Introduction

- Prediction of different lung pathologies using chest X-ray images is a challenging task requiring robust training and testing accuracies.
- Cyclical learning rates (CLR's) allow the learning rates to vary between a range of boundary values. Selecting learning rate manually is a time consuming and computationally costly task.
- Optimal learning rate is important as the model can converge slowly if the learning rate is too slow or the model can diverge from the minima of the cost function if the learning rate is too high.
- To overcome this issue, instead of using constant single learning rate, a learning rate decay policy can be used to obtain better results. However, the learning rate decay also has several drawbacks. CLR's can be an effective technique to make the model converge faster in minimal number of epochs and to decrease the efforts of finding optimal learning rates.

Dataset

- In this project, one-class classifier (OCC) and binary classification algorithms have been tested to classify 14 different diseases (atelectasis, cardiomegaly, consolidation, effusion, edema, emphysema, fibrosis, hernia, infiltration, mass, nodule, pneumonia, pneumothorax and pleural-thickening).



Exploratory Data Analysis (EDA) of labels

Method

- We have utilized 3 different neural network architectures' (MobileNetV1, Alexnet, and DenseNet-121) with four different optimizers (SGD, Adam, and RMSProp) for comparing best possible accuracies.

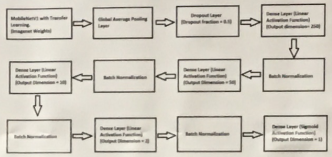


Figure 2: Network architecture

Result

- Cyclical learning rate (CLR), a tuning hyperparameters technique was found to have a faster convergence of the cost towards the minima of cost function.

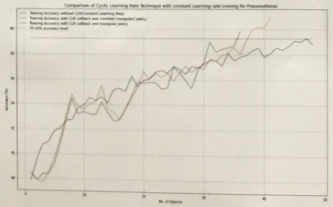


Figure 3: Accuracy Comparison of CLR

Result

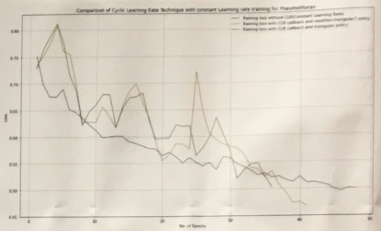


Figure 4: Loss comparison of CLR

Discussion & Conclusion

- Depthwise separable convolutions like MobileNets have been gradually pruned for improving the speed of dense network.
- MobileNetV1 Imagenet weights with SGD optimizer is found to outperform other optimizers and architectures in terms of training time taken and accuracy attained.
- Achieving a high test accuracy is directly depended on learning rate hyper-parameter for training neural networks
- Here, we present a unique approach of utilizing previously trained binary classification models with a learning rate decay technique for retraining models using CLR's.
- Doing so, we found significant improvement in training accuracies for each of the selected conditions.
- Thus, utilizing CLR's in callback functions seems a promising strategy for image classification problems.

Acknowledgements

- This work was primarily supported by the National Science Foundation EPSCoR Program under NSF Award # OIA-2242812.