

CRUNCH Seminars at Brown, Division of Applied Mathematics

Friday – February 28, 2020

Identifying Critical Neurons in ANN Architectures using Mixed Integer Programming

Zongren Zou

We introduce a novel approach to optimize the architecture of deep neural networks by identifying critical neurons and removing non-critical ones. The proposed approach utilizes a mixed integer programming (MIP) formulation of neural models which includes a continuous importance score computed for each neuron in the network. The optimization in MIP solver minimizes the number of critical neurons (i.e., with high importance score) that need to be kept for maintaining the overall accuracy of the model. Further, the proposed formulation generalizes the recently considered lottery ticket optimization by identifying multiple “lucky” sub-networks resulting in optimized architecture that not only perform well on a single dataset, but also generalize across multiple ones upon retraining of network weights.

Finally, the proposed framework provides significant improvement in scalability of automatic sparsification of deep network architectures compared to previous attempts. We validate the performance and generalizability of our approach on MNIST, Fashion-MNIST, and CIFAR-10 datasets, using three different neural networks: LeNet 5 and two ReLU fully connected models.