Neural Layer Bypassing Network: A Novel Neural Network Architecture To Increase the Speed of Forward Propagation Without Sacrificing Accuracy, Network Structure, or CPU Resources

Palasamudram, Amogh (School: The International School of Bangalore (TISB))

Neural networks have improved deep learning, but at times, they can be inefficient. Current networks perform extensive computations on all input data to improve accuracy. However, not all inputs require such computations to be predicted accurately. This wastes time and CPU resources, especially for datasets with a range of input classification difficulty. To minimize this trade-off between speed and accuracy, this research proposes the Neural Layer Bypassing Network (NLBN), a new architecture that optimizes forward propagation based on inputs to reduce inefficiency and unnecessary calculations. The NLBN utilizes one additional 'fully connected' neural network layer after every layer in the main network (called the rejection layer). This new layer predicts the outputs of the semi-processed inputs to determine whether finishing the rest of forward propagation is useful to get accurate predictions. To test the NLBN's effectiveness, 5 image classification models were programmed using 5 different datasets. After training one CNN and NLBN per dataset (both of equivalent architectures), the accuracy and time period for predictions were measured. With the NLBN, the speed of forward propagation increased by 6% - 50% while the accuracy decreased by 0% - 4%; the results vary based on the dataset, model structure, and hyperparameters of the rejection layers; however, the increase in speed was always at least twice the decrease in accuracy. Also, due to the NLBN's complexity, it takes more RAM and 40% longer to train. The architecture can be more efficient if integrated into TensorFlow libraries. Overall, by autonomously skipping network layers, the NLBN can potentially teach itself to become more efficient by making faster, accurate, and less computationally intensive predictions.