Contribution ID: **131**   Type: **Poster**

A New Method for Enforcing Training Time Sparsity with Constrained Optimization

Enforcing sparsity, the number of zeros in a neural network's weight matrices, has a variety of uses in machine learning such as improving computational efficiency and controlling encoding efficiency. Often achieving a specified sparsity requires trial and error or multiple retrainings of the same model until the criteria are met, which can be labor intensive and prone to error. Using the modified differential multiplier method (Platt and Barr, NIPS, 1987), we suggest a new method for enforcing the sparsity of layers in a neural network at training time that makes sparsity a hyperparameter of the loss function being optimized and only requires one training pass. In this contribution we will discuss our methodology and demonstrate its efficacy in image processing and autoencoders.

Focus areas

Authors: DAS, Arghya Ranjan;   GRAY, Lindsey (Fermi National Accelerator Lab. (US))

Co-author: LIU, Miaoyuan (Purdue University (US))

Presenter: DAS, Arghya Ranjan