Lifelong Learning With Dynamically Expandable Networks – Reproducibility Report

Błażej Sowa, Łukasz Siudek

Motivation

Lifelong learning (Thrun, 1995), the problem of continual learning where tasks arrive in sequence, is an important topic in transfer learning. The primary goal of lifelong learning is to leverage knowledge from earlier tasks for obtaining better performance, or faster convergence/training speed on models for later tasks. While there exist many different approaches to tackle this problem, we consider lifelong learning under deep learning to exploit the power of deep neural networks. Fortunately, for deep learning, storing and transferring knowledge can be done in a straightforward manner through the learned network weights. The learned weights can serve as the knowledge for the existing tasks, and the new task can leverage this by simply sharing these weights.

Introduction

In this report we are going to show our results of attempted reproduction of a recent conference paper from ICLR, called "Lifelong Learning With Dynamically Expandable Networks" which proposes a novel deep neural network for lifelong learning, called "Dynamically Expandable Network". It performs partial retraining of the network trained on old tasks by exploiting task relatedness, while increasing its capacity when necessary to account for new knowledge required to account for new tasks, to find the optimal capacity for itself, while also effectively preventing semantic drift.

Reproducibility

Available Information Overview

In this section, we list important reproducibility metrics.

Dataset

We use the same datasets as those used in the paper.

1. CIFAR-10
2. MNIST-Variation

Data preprocessing

We use a few preprocessing methods on the MNIST dataset. At first, we aplly random rotation to the images. Then, we add Gaussian noise with parameters $\mu = 0$ and $\sigma = 0.2$.

No detailed preprocessing information was given, when it comes to the preprocessing used in the DEN paper.

Dataset Partitions

For the MNIST dataset, the authors of DEN paper use 1,000/200/5,000 images for train/val/test split for each class. They form each task to be one-versus-rest binary classification.

For the CIFAR-10 dataset, $5,000$ out of $60,000$ images are used for training and the remainder is used for test.

Model training

Out of introduced in the paper several models, we have implemented the following Feedforward networks:

1. DNN-STL. Base deep neural network, each task has its own network and exacltly one output.

2. DNN-MTL. Base DNN trained for all tasks at once. One network with many outputs.

3. DNN-L2. Base DNN, where at each task t, W^t is initialized as W^{t-1} and continously trained with SGD, with l_2-regularization between W^t and W^{t-1}. For this purpose, we use the equation:

$$\underset{W_l^{\mathcal{N}}}{\text{minimize}}\ \mathcal{L}(W^t; \mathcal{D}_t) + \lambda \left\| W^t - W^{t-1} \right\|_2^2,$$

   where $\lambda = 0.005$.

4. DNN. Same as previous, but no l_2-regularization is used.

5. DEN. Dynamically Expandable Network, but only with Selective Training algorithm used.

Also, we have implemented following Convolutional networks:

1. CNN-STL.
2. CNN-MTL.
3. CNN.
4. CNN-L2.

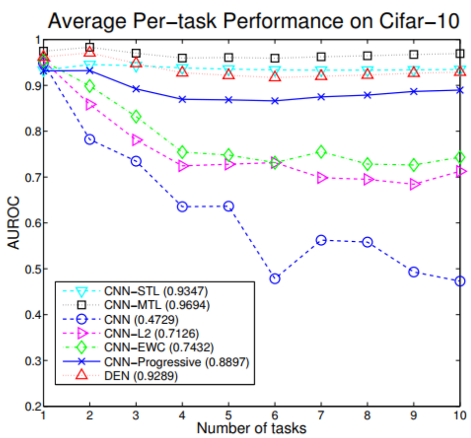
In model training, we used SGD with momentum 0.9 and weight decay parameter equal to 10^{-4}. These parameters were not stated in the original paper.

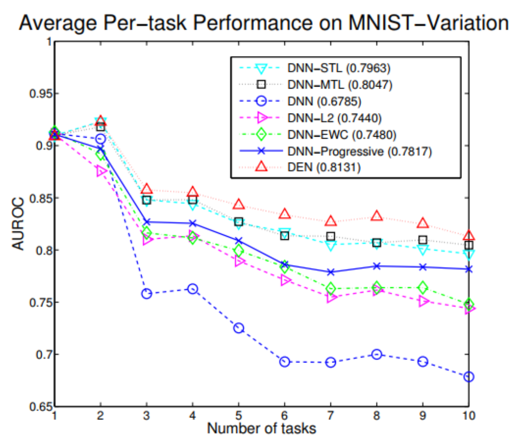
Model Assessment

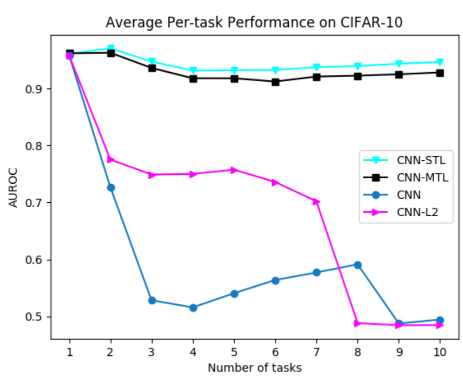
Randomization control

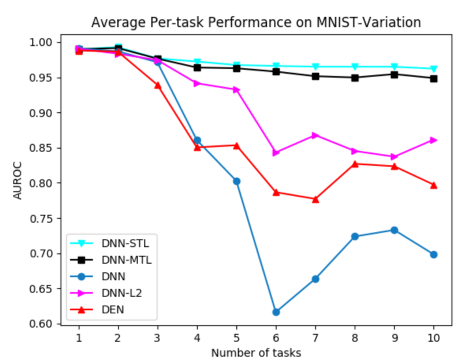
Software and Hardware Environment

Results









Conclusion

Conclusion.

References

[1] Leslie Lamport, *LaTeX: a document preparation system*, Addison Wesley,

Massachusetts, 2nd edition, 1994.