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 introduced in the paper.

1. CIFAR-10

2. MNIST-Variation

Data preprocessing

Authors of the paper say, that in the MNIST-Variation the handwritten digits are rotated to arbitrary angles and have noise in the background, which makes the prediction task more challenging.

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

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, for each class, $5,000$ out of $6,000$ 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:

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

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.

We use two-layer network with 312-128 neurons with ReLU activation. When it comes to the Convolutional networks, we use LeNet, with two convolutional layers and three fully-connected layers.

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, so we came up with our own.

Randomization control

Authors of the paper do not state how they set seed values. On the other hand, we use constant seed value.

Software and Hardware Environment

All models and algorithms in the DEN paper were implemented using the Tensorflow library. Package version is not specified.
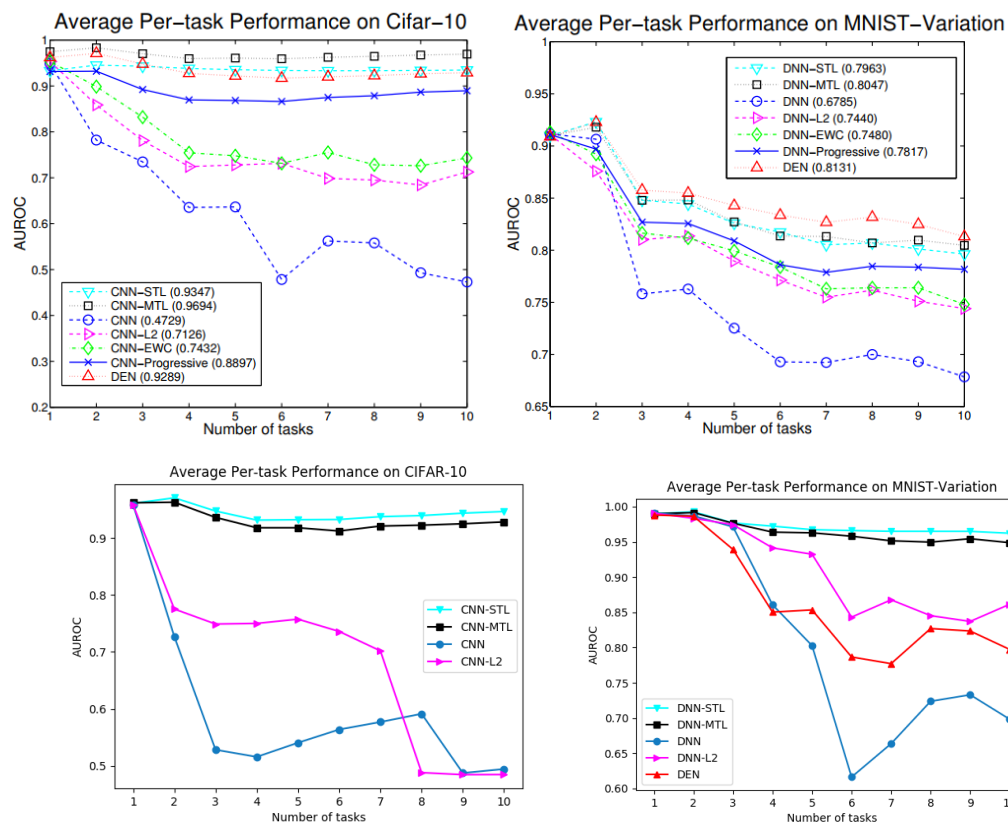
In order to implement our version of the network, we used PyTorch (version 0.3.0.post4).

Our hardware setup (GCloud virtual machine):

- 4 x vCPU

- 16GB RAM

- NVIDIA Tesla K80

No hardware information is given about the machine used in the DEN paper.

Results

Conclusion

Conclusion.

Acknowledgments

References

[1] Jaehong Yoon, Eunho Yang, Sung Ju Hwang, *Lifelong Learning with Dynamically Expandable Networks*, arXiv:1708.01547v1.

[2] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, *Gradient-based Learning Applied to Document Recognition*, Proceedings of the IEEE, 86(11):2278–2324, 1998.

[3] Babatunde K. Olorisade, Pearl Brereton, Peter Andras, *Reproducibility in Machine Learning-Based Studies: An Example of Text Mining*, 16 Jun 2017, ICML 2017 RML Submission.