Hessisches Kompetenzzentrum
für Hochleistungsrechnen

Combining Regenerative and Degenerative Methods in Growing Neural Networks

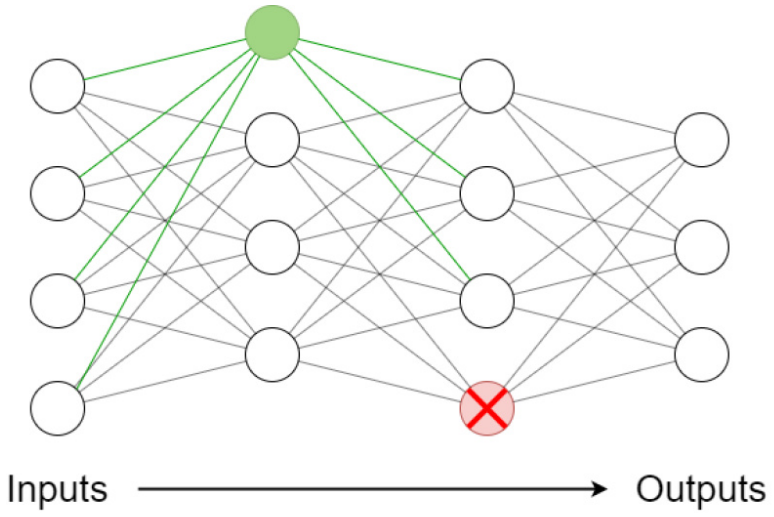


Figure 1: Neuron-level growing and pruning in a generic neural network.

Introduction

Neural networks are usually trained with a static architecture. However, the fields of growing and pruning, or artificial neurogenesis, or structure adaption, or dynamic neural networks, investigate settings were the architecture is adapted during training to accelerate the training process or better fit the underlying function to be learned. While pruning (the selective removal of parts of a neural network) has been investigated quite extensively, growing (the counterpart) is a rather novel research direction. The combination of the two may have a lot of potential and was investigated with real world machine learning applications.

Methods

Since it was found that no existing software framework sufficiently supported growing and pruning together, a new framework was implemented based on Tensorflow. The framework is capable of supporting unstructured growing and pruning, as well as structured neuron-level, layer-level and cell-level growing and pruning. Two pruning- and four growing-methods were investigated together in different scenarios on a CIFAR10 image recognition task. Namely the pruning methods are l1-regularization- and polarization regularization pruning. The growing methods are gradmax, gradmax-optim, firefly and random growing. To reduce complexity and for a more detailed comparison, the scheduling of the growing and pruning steps was fixed. The VGG11 network architecture was used as a basis

Project Manager
Roman Frels

Researchers
Sascha Hauck

Principal Investigator
Prof. Dr. Heinz Koeppl

Project Term
2022 - 2022

Clusters
Lichtenberg II Cluster Darmstadt

Software
TensorFlow

Institute
Self-Organizing Systems

University
Technische Universität Darmstadt

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN
TECHNISCHE UNIVERSITÄT DARMSTADT
UNIKASSEL VERSITÄT
Philipps Universität Marburg

for the experiments. In three settings the network size was either growing, shrinking, or held constant, while varying degrees of growing and pruning in different permutations were employed.

Results

The limited number of methods that were trialled, only revealed results that were en-par to the baseline in training performance. Furthermore, growing and pruning severely impacts the training and evaluation loss, what is to be expected, since it introduces a large departure from the previous network function. It was found that batch-normalization poses a major challenge to the growing methods and must be considered in their design, due to its wide usage. Especially gradmax and firefly were affected so severely, that they worsened the training error permanently. During the development of the software framework, non-sequential architectures were considered as well. The theory for altering the architecture in a permissible way was developed. Furthermore, algorithms were introduced that identify all permissible elements to be grown or pruned in a neural network.

Discussion

Due to limited time, only very few methods could be trialled and the scheduling had to be fixed. It would be desirable to continue the research by forming a unified theory of growing and pruning and identifying the most promising methods. Currently the formalities and terms are very fragmented and additionally many papers build on heuristics instead of a proper mathematical derivation. Furthermore, common benchmarks need to be defined. These would greatly reduce the entry barrier and help to compare methods on their performance. To achieve the full potential, the scheduling should be a part of the methods as well.

Last Update: 2023-10-06 11:28