
Differentiable algorithm for searching ensembles of deep learning models with diversity control

A Preprint

K. Yakovlev, O. Bakhteev, K. Petrushina, P. Babkin

Abstract

This paper is developed to introduce a new method of creating ensembles of deep learning models. Many modern researches were focused on creating effective and efficient algorithms of differentiable architecture search, missing opportunity to create ensembles of deep learning models. This approach gives impressive results as it was shown in few modern papers. In our research we investigate an algorithm of sampling deep learning models using hypernetwork, which controls diversity of the models. This method allows us to sample deep learning models in one-shot, without any additional calculational losses, this independence means that algorithm is able to sample as many models as we want. To evaluate the performance of the proposed algorithm, we conducted experiments on the Fashion-MNIST and CIFAR-10 datasets and compare the resulting ensembles with ones sampled by other searching algorithms.

Keywords differential search · neural ensembles · hypernetwork · diversity control

1 Introduction

Nowadays algorithms of neural architecture search (NAS) are well-explored and proved to be an effective way of creating more effective and efficient neural networks. Such algorithms use different ways to make problem differentiable so it can be solved by wide range of methods for smooth optimization problems. On the other hand, neural ensemble search (NES) is modern and not as well investigated problem as NAS, although it is known that ensembles of deep learning models show better results in different applied problems.

Our paper investigates an algorithm of sampling deep learning models in a new way. Despite the fact that similar algorithms were described and investigated, our scheme has unique set of techniques that gives compatible results and has its own sphere of implementations.

First of all, we use DARTS as our basic NAS algorithm. It is proved to be an effective architecture search algorithm. Some modern investigations have shown that the algorithm can be upgraded, but the modernisations do not make tangible difference and are mainly made for some specific cases. So in our paper we are focusing on ensemble sampling, so slightly different base model does not change anything in resulting ensemble.

Second of all, we use hypernetwork in our sampling algorithm. Hypernetwork is small network so it does not consume too much computational capacity. This network contains information about another network, which is called target network. This approach is not new: hypernetworks and supernet networks were intended to control different characteristics such as complexity of architecture or parameters of the target model in several modern investigations. In our algorithm it controls diversity of the target models, so every sampled model differs from previously sampled ones.

The hypernetwork uses Jensen-Shannon divergence (JSd) to measure difference between two architectures which is symmetric and finite in contrast to more popular Kullback–Leibler divergence. Our main idea of sampling different model is to use a regularizer, based on JSd as a source of diversity.

This way we are able to sample deep learning models in one-shot, without and additional computational losses. To sum up the scheme of our method:

1. find a base architecture using DARTS
2. sample architectures in one-shot via differentiable algorithm
3. inferencing answer is ensemble of the sampled deep learning models