Surrogate-Assisted Diversity Estimation in Neural Ensemble Search

Udeneev Alexandr Vladimirovich

Scientific Advisor: Ph.D. in Physics and Mathematics, Oleg Yurievich Bahteev

Moscow Institute of Physics and Technology
My First Research Paper

April 17, 2025

Goals

Research Goal:

Reduce the time spent on building an effective ensemble.

Task:

Describe the space of neural network architectures. Describe the surrogate function and its application in ensemble construction.

Introduction

The neural network architecture space $\mathcal{A}$ is exceedingly large (on the order of 10^{24} architectures in our case), raising the question of developing an efficient method for searching for the optimal ensemble $S \subset \mathcal{A}$, i.e., the ensemble that achieves the highest accuracy.

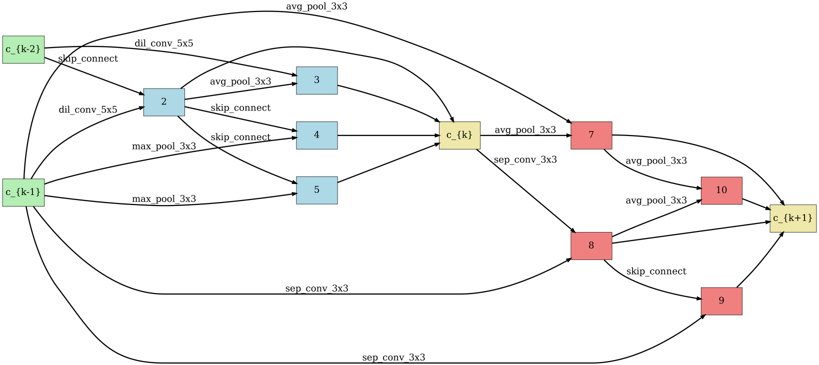


Рис.: Example of one architecture in NAS-Bench-201 format

Literature

Our work is based on the following articles:

- Similarity surrogate-assisted evolutionary neural architecture search with dual encoding strategy (Xue, Zhang, Neri, 2024)
- Neural Predictor for Neural Architecture Search (Wen et al., 2019)
- Few-shot Neural Architecture Search (Zhao et al., 2020)

Problem statement

The primary objective of NES is to find an optimal ensemble of neural networks whose architectures lie within the NAS search space.

Let denote $\alpha \in \mathcal{A}$ as a network architecture and $\omega(\alpha)$ as its corresponding weights. The action of this network on an input x is denoted by $f_\alpha(x, \omega(\alpha))$. Let $S \subset \mathcal{A}$ be a subset of architectures. Then, the NES problem can be formally described as follows:

$$\min_{\mathcal{S}} \mathcal{L}_{val} \left(\frac{1}{|S|} \sum_{\alpha \in S} f_\alpha(x, \omega^*(\alpha)) \right)$$

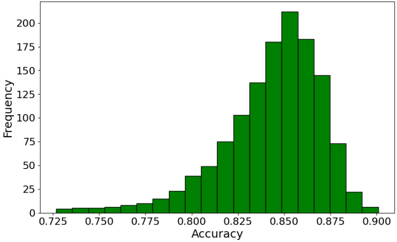
$$\text{s.t.} \quad \forall \alpha \in \mathcal{S} : \ \omega^*(\alpha) = \arg\min_{\omega(\alpha)} \mathcal{L}_{train}(f_\alpha(x, \omega(\alpha)))$$

Thus, in addition to searching over a vast number of architectures, we now also need to find the optimal ensemble composition.

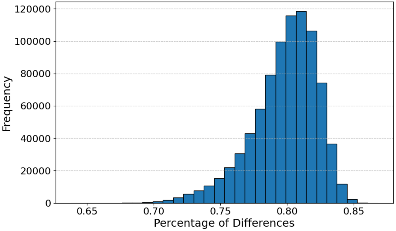
Problem solution

Surrogate Function

We propose a function to estimate model similarity (e.g., using the Jensen-Shannon distance on model predictions from the test dataset) to guide the selection of an optimal ensemble, where "optimal"refers to maximizing the ensemble's overall predictive accuracy.



Distribution of model accuracies



Distribution of model diversity

Problem solution

The algorithm for training the surrogate function is as follows:

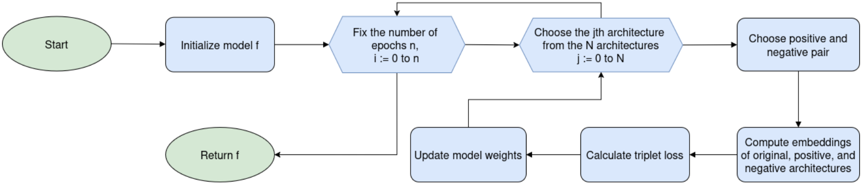


Рис.: Training the surrogate function

Problem solution

Our main algorithm is as follows:

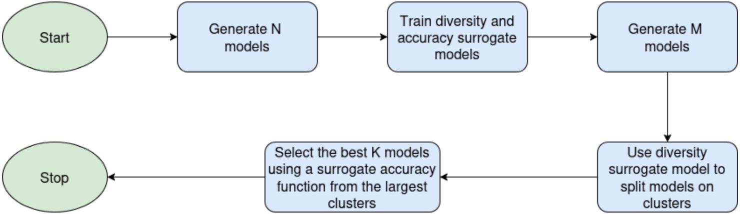


Рис.: The main algorithm of our model.