Differential neural ensemble search with diversity control

Petr Babkin

Moscow Institute of Physics and Technology

B05-003

Expert: O. Bakhteev

2023

Goals of research

- Propose a novel method of sampling deep learning models with diversity control
- Investigate sampled models in terms of diversity and performance
- Test different ensembles of the sampled models
- Compare performance with other state-of-the-art methods

Problem statement

Classic problem of searching for NN ensembles

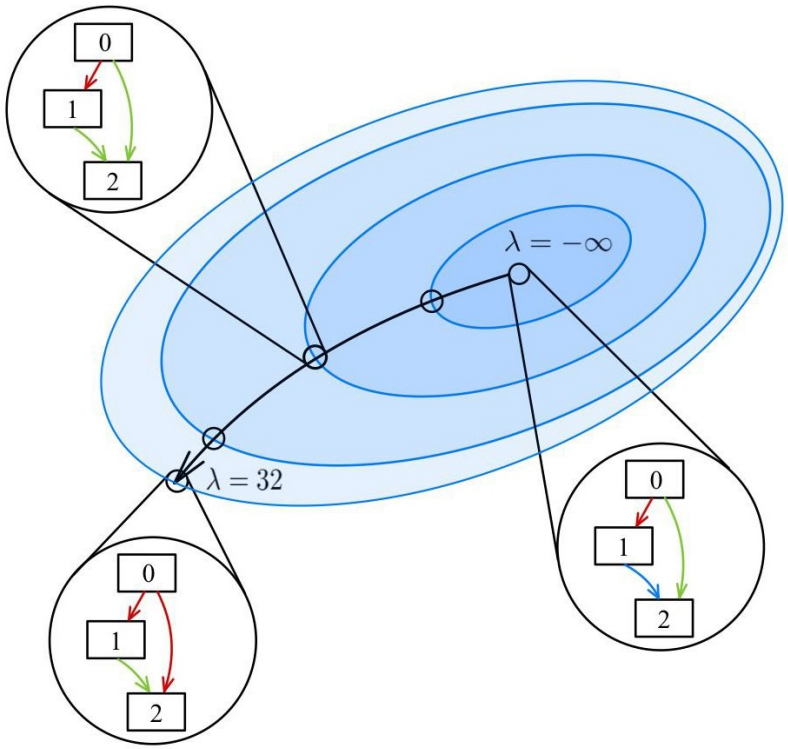
$$\min_{S} \mathcal{L}_{val}\left(\frac{1}{|S|}\sum_{\boldsymbol{\alpha}\in S} f(\boldsymbol{w}^*_{\boldsymbol{\alpha}}, \boldsymbol{\alpha})\right)$$
$$s.t.\ \forall \boldsymbol{\alpha} \in S\ \ \boldsymbol{w}^*_{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{w}} \mathcal{L}_{train}(f(\boldsymbol{w}^*_{\boldsymbol{\alpha}}, \boldsymbol{\alpha}))$$

Rearranged problem of searching for NN ensembles

$$\min_{\alpha} \mathbb{E}_{\lambda\sim U(0,\Lambda)}[\mathcal{L}_{val}(\mathbf{w}^*, \boldsymbol{\alpha}) - C(\lambda - \langle \boldsymbol{\alpha}^*, GS(\boldsymbol{\alpha})\rangle)^2]$$
$$s.t.\ w^* = \arg\min_{\mathbf{w}} \mathbb{E}_{\lambda\sim U(0,\Lambda)}[\mathcal{L}_{train}(\mathbf{w}, \boldsymbol{\alpha})]$$

where λ is an amount of common edges

Hypotheses and model



Architectural space

- Architectural space is continuous
- Architectures differ in terms of edges
- The further architecture locates the worse accuracy it performs
- Diversity and performance are both important for ensembling

Solution

We sample architectures using hypernetwork, a parametric mapping

$$h : [0, \Lambda] \times \mathbb{R}^u \to \mathbb{R}^s,$$

where $\mathbb{R}^u$ is hypernetwork parametric space and $\mathbb{R}^s$ is architectural space

Algorithm

Initialize: $N \in \mathbb{N}$, $\mathcal{S} = \varnothing$, hypernetwork
S $\leftarrow \{ \boldsymbol{\alpha}^* \}$ ▷ Result of NAS
for $i = 1, \ldots, N$ **do**
    Sample $\lambda \sim U(0, \Lambda)$
    S $\leftarrow$ S $\cup \{\boldsymbol{\alpha}(\lambda, \boldsymbol{\alpha}^*)\}$ ▷ α gained from hypernetwork
end for
Return: S as a resulting ensemble

Goals of computational experiment

To conduct experiments on CIFAR100 dataset following two main problems

Comparison of architectures

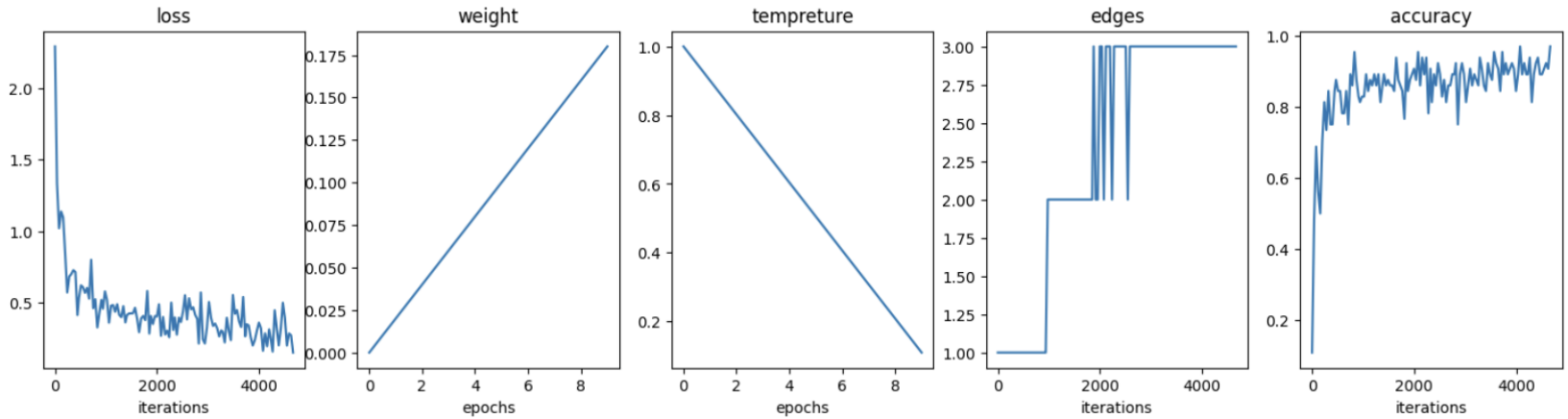
- Investigate performance of architectures
- Investigate diversity of architectures

Ensembling effectiveness investigation

- Try different ensembles
- Compare results with DARTS ensembling

Learning and warming

Weight increases while temperature decreases with iterations, regularizer become strict



Learning process

Experiments show that this trick significantly increases performance

Scores

Performance in terms of accuracy			
λ	Constant weight	Weight warming	Temperature warming
1	0.306	0.400	0.403
2	0.283	0.387	0.310
3	0.265	0.354	0.314

Conclusion

Achieved results

- A novel method of sampling deep learning models architectures with diversity control
- Base experiments directed to beat baseline
- Diversity of sampled models was controlled, ensemble shows compatible performance

Further investigations

- Hypernetwork implementation, comparison with SOTA results
- MIPT conference
- International conference

Literature

- Yao Shu1, Yizhou Chen, Zhongxiang Dai, Bryan Kian, Hsiang Low: Neural Ensemble Search via Bayesian Sampling
- Hanxiao Liu, Karen Simonyan, Yiming Yang: DARTS: Differentiable Architecture Search
- Konstantin Yakovlev, Olga Grebenkova, Oleg Bakhteev, Vadim Strijov: Neural Architecture Search with Structure Complexity Control
- Ashwin Raaghav Narayanan, Arber Zela, Tonmoy Saikia, Thomas Brox, Frank Hutter: Multi-headed Neural Ensemble Search