

Дифференцированный поиск ансамблей нейронных сетей с контролем их разнообразия

Бабкин Пётр, Яковлев Константин, Петрушина Ксения

Конференция МФТИ

2024

План работы

Цель исследования:

Предложить метод ансамблирования нейросетей на основе методов поиска нейросетевых архитектур (NAS)

Проблема:

Методы NAS являются вычислительно затратными для каждой отдельной архитектуры. При обобщении методов на ансамбль сложность поиска архитектур растет линейно

Предлагаемый метод:

Предлагается рассматривать структуру модели как сложную функцию (гиперсеть), которая зависит от параметра разнообразия моделей, входящих в ансамбль

Постановка задачи

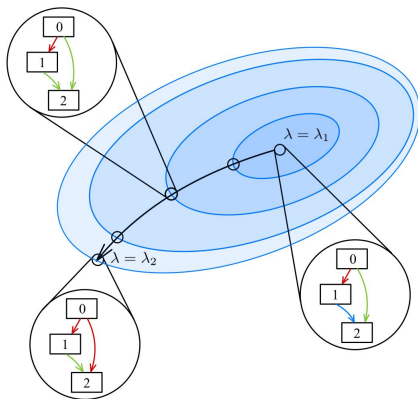
Общая задача по поиску ансамбля нейронных сетей

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

Проблемы:

- ▶ Экспоненциально большое пространство структур
- ▶ Экспоненциально большое количество возможных ансамблей

Модель и гипотезы



Пространство структур

- ▶ Пространство структур непрерывно
- ▶ Структуры в пространстве отличаются в смысле общих ребер
- ▶ Чем дальше структура от оптимальной, тем хуже ее предсказательные способности
- ▶ Разнообразие структур и их предсказательные способности важны для построения ансамбля

Мы получаем структуры с помощью гиперсети,
параметрического отображения

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

где \mathbb{R}^u — пространство параметров гиперсети, и \mathbb{R}^s —
пространство структур, Λ — максимальное возможное число
общих ребер

Наша постановка задачи

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

где λ это параметр, показывающий схожесть сетей в терминах
ребер, $\boldsymbol{\alpha}^*$ — оптимальная структура, а GS — эффективный
параметрический метод сглаживания структуры

Algorithm EdgeNES

Initialize: $n \in \mathbb{N}$, $\mathcal{S}' = \emptyset$, $N \in \mathbb{N}$

$\alpha^* \leftarrow$ result of NAS

Train hypernetwork α using problem above and α^*

for $i = 1, \dots, N$ **do**

 Sample $\lambda \sim p(\Lambda)$

$\mathcal{S} \leftarrow \mathcal{S} \cup \{\alpha(\lambda, \alpha^*)\}$

end for

$\mathcal{S} \leftarrow$ best performing ensemble of size n from \mathcal{S}'

Return: \mathcal{S} as a resulting ensemble

Цели вычислительного эксперимента

Провести эксперименты по предложенному методу на CIFAR100 с двумя основными целями

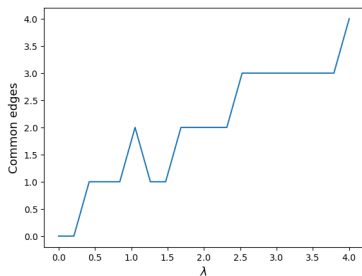
Исследовать сэмплируемые структуры

- ▶ Исследовать предсказательную способность структур
- ▶ Исследовать разнообразие структур

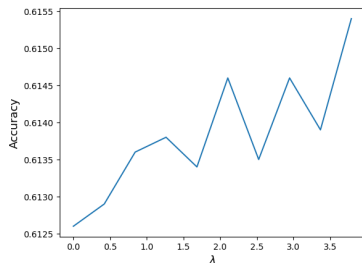
Проверить эффективность метода

- ▶ Проверить эффективность ансамблирования
- ▶ Сравниться с другими методами по времени работы и по качеству построенного ансамбля

Полученные структуры



Зависимость общих ребер от степени разнообразия



Зависимость точности предсказаний от степени разнообразия

Результирующий ансамбль

Метод	GPU-часы	Точность
Ансамблирование DARTS	~ 4.2	61.60 ± 0.43
Случайные ребра	~ 1.2	55.60 ± 6.03
Наш метод	~ 2.8	61.46 ± 0.62

Результаты экспериментов на CIFAR100.

Достигнутые результаты

- ▶ Был предложен алгоритм по сэмплированию ансамблей моделей глубокого обучения
- ▶ Показана возможность контроля разнообразия архитектур в ансамбле
- ▶ Показано, что данный метод работает быстрее базового алгоритма без сильного ущерба качеству

- ▶ 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](#)