

Дифференцируемый поиск ансамблей нейронных сетей с контролем их разнообразия

Бабкин Пётр

Научный руководитель: к.ф.-м.н. Бахтеев Олег

Кафедра интеллектуальных систем ФПМИ МФТИ

Специализация: Интеллектуальный анализ данных

Направление: 03.04.01 Прикладные математика и физика

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

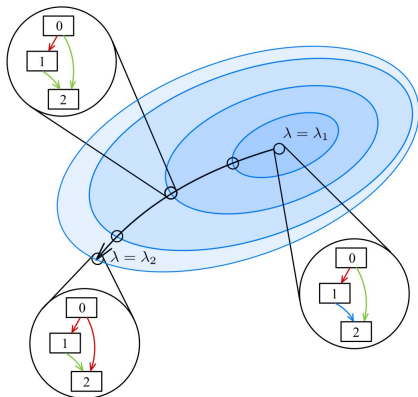
Общая задача по поиску ансамбля нейронных сетей

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

Проблемы:

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

Модель и интуиция метода



Пространство структур

1. Пространство структур непрерывно
2. Есть некоторая оптимальная архитектура
3. Структуры в пространстве отличаются в смысле общих ребер
4. Чем дальше структура от оптимальной, тем хуже ее предсказательные способности
5. Выберем архитектуры с высокой предсказательной способностью, контролируя разнообразие

Гиперсеть для получения архитектур

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

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

где λ — это параметр, показывающий схожесть сетей в
терминах ребер, α^* — оптимальная структура, а GS —
Gumbel-softmax, и \mathbf{a} — вектор параметров гиперсети

Предложен алгоритм EdgeNES

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

Теорема. (Бабкин, 2024) Рассмотрим задачу нахождения ансамбля размера N . Предположим, что максимальное количество ребер в финальных структурах обозначено как Λ . Пусть $p(\lambda) = U(\overline{1}, \Lambda)$. Тогда ожидаемое количество ребер, участвующих в обновлении параметров и включенных в финальный ансамбль, оценивается как

$$\mathbb{E}[m] = N \frac{(\Lambda + 1)^2}{4\Lambda}.$$

- ▶ Эффективность ансамбля растет с размером ансамбля
- ▶ Предложенный метод позволяет одновременно оптимизировать несколько структур моделей, имеющих общие ребра

Цели вычислительного эксперимента

Провести эксперименты по предложенному методу на CIFAR-100 с двумя основными целями

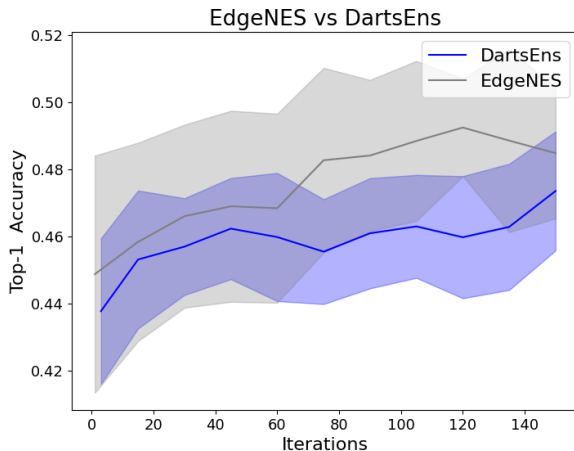
Оценка скорости работы

- ▶ Проверить, как быстро разные методы строят эффективный ансамбль

Оценка качества итогового ансамбля

- ▶ Проверить эффективность работы итогового ансамбля

Оценка скорости работы



AUC для DartsEns 2.90 ± 0.19 , AUC для EdgeNES 2.78 ± 0.13 , p -value для t -test 0.0329.

Оценка качества итогового ансамбля

Метод	Точность, %	GPU-часы
DARTS	46.2 ± 2.2	~ 1.2
Случайная архитектура	44.2 ± 3.3	0
Ансамблирование DARTS	47.5 ± 1.4	~ 3.6
Случайный ансамбль	45.1 ± 2.8	0
Предложенный метод	47.9 ± 1.8	~ 2.1

Достигнутые результаты:

- ▶ Предложен алгоритм для сэмплирования ансамблей моделей глубокого обучения
- ▶ Показана возможность контроля разнообразия архитектур в ансамбле
- ▶ Приведен теоретический анализ скорости работы метода
- ▶ В экспериментах проверено, что данный метод работает быстрее базового алгоритма без сильного ущерба качеству
- ▶ Сделано выступление на конференции МФТИ