

Parameter Prediction for Unseen Deep Architectures

Практик-исследователь: Никита Морозов

Март 2022

Конференция

[Статья](#) была опубликована на конференции NeurIPS 2021 и представлена в формате постера. Она получила относительно высокие оценки от рецензентов: 7, 6, 7, 8. На другие конференции статью не подавали, а также на архиве она была опубликована только после публикации на NeurIPS.

Про статью [написали](#) в научно-популярном издании Quanta Magazine, где она была представлена под броским заголовком "Researchers Build AI That Builds AI".

Авторы

- [Boris Knyazev](#), PhD student in University of Guelph and Vector Institute. Имеет работы по графовым сетям и face recognition. Сама статья была написана в рамках его стажировки в Facebook AI Research.
- [Graham W. Taylor](#), University of Guelph and Vector Institute. Имеет различные работы по DL в CV.
- [Michal Drozdal](#), Facebook AI Research. Имеет работы по нейросетевым архитектурам в CV и приложениям DL в медицине. Был руководителем стажировки у Boris Knyazev.

- [Adriana Romero-Soriano](#), Facebook AI Research. Имеет аботы по графовым сетям, приложениям DL в медицине, в том числе много совместных работ с Michal Drozdal. Была руководителем стажировки у Boris Knyazev.

Опорные работы

В качестве ключевых работ, на которые опирались авторы, можно выделить две статьи.

- [Hanxiao Liu et al, “DARTS: Differentiable Architecture Search”](#). В ней предложена позволяющая подсчет градиентов непрерывная релаксация задачи поиска оптимальной архитектуры, которую авторы используют в своей работе.
- [Chriz Zhang et al, “Graph HyperNetworks for Neural Architecture Search”](#). Отсюда авторы берут графовую сеть GHN, которая используется в предложенном методе.

Можно отметить, что в обоих случаях авторы расширили и дополнили эти подходы для своих нужд.

Также при постановке оптимизационной задачи для обучения гиперсети авторы во многом вдохновлялись парадигмой meta-learning, но какую-то отдельную работу здесь выделить сложно.

Цитирования статьи

На данный момент у статьи имеется два цитирования:

- [Or Litany et al, “Federated Learning with Heterogeneous Architectures using Graph HyperNetworks”](#). Используется прогресс авторов в GHN для решения некоторой задачи распределенного обучения.
- [Xinran Liu et al, “Teaching Networks to Solve Optimization Problems”](#). Замена оптимизации на прямое предсказание параметров в классических задачах: линейная регрессия, РСА, задачи квадратичного программирования, и т. д.

Предложения по дальнейшим исследованиям

По моему мнению, будет интересно поставить следующий эксперимент: взять предсказанные гиперсетью веса для некоторой архитектуры сети, а затем доучить их с помощью SGD/Adam на том же самом наборе данных, и посмотреть, как итоговое качество соотносится с качеством у обученной с нуля модели.

Также для обучения гиперсети авторы не используют конкретные веса ранее обученных сетей, что в каком-то смысле противоречит описанной в начале статьи мотивации:

“When optimizing the parameters for a new architecture a, typical optimizers disregard past experience gained by optimizing other nets. However, leveraging this past experience can be the key to reduce the reliance on iterative optimization and, hence the high computational demands”.

Тогда возникает вопрос, можно ли видоизменить оптимизационную задачу, используемую для обучения гиперсети, чтобы включить туда эти веса?

Экологическое значение

State-of-the-art модели используют бешеные вычислительные мощности и требуют огромное количество времени для обучения, что приводит к увеличению environmental footprint. Как отмечают авторы, их работа делает шаг в сторону решения этой проблемы.