

Parameter Prediction for Unseen Deep Architectures

Практик-исследователь: Никита Морозов

Конференция

- NeurIPS 2021, Poster
- Оценки: 7, 6, 7, 8
- На другие конференции статья не подавалась
- На архиве статья была опубликована только после публикации на NeurIPS
- Про статью [написали](#) в научно-популярном издании Quanta Magazine. Броский заголовок: “Researchers Build AI That Builds AI”

Авторы

- Boris Knyazev, PhD student in University of Guelph and Vector Institute. 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. Различные работы по DL в CV.
- Michal Drozdal, Facebook AI Research. Работы по нейросетевым архитектурам в CV и приложениям DL в медицине. Руководитель стажировки у Boris Knyazev.
- Adriana Romero-Soriano, Facebook AI Research. Работы по графовым сетям, приложениям DL в медицине. Много совместных работ с Michal Drozdal. Руководитель стажировки у Boris Knyazev.

Опорные работы

- Hanxiao Liu et al, “DARTS: Differentiable Architecture Search”. Непрерывная релаксация задачи поиска оптимальной архитектуры, позволяющая подсчет градиентов.
- Chriz Zhang et al, “Graph HyperNetworks for Neural Architecture Search”. Отсюда авторы берут графовую сеть GHN, с которой идет основная работа в статье.
- При постановке оптимизационной задачи для обучения гиперсети авторы во многом вдохновлялись парадигмой meta-learning.

Кто цитирует

- Or Litany et al, “Federated Learning with Heterogeneous Architectures using Graph HyperNetworks”. Используют 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.

Дальнейшие исследования

- Эксперимент: что будет, если доучить предсказанные гиперсетью веса на том же датасете с помощью SGD/Adam?
- Для обучения гиперсети авторы не используют конкретные веса ранее обученных сетей, что в каком-то смысле противоречит описанной в начале статьи мотивации. Можно ли видоизменить оптимизационную задачу, 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.