

УДК 004.852

ПРОРЕЖИВАНИЕ НЕЙРОННЫХ СЕТЕЙ С ИСПОЛЬЗОВАНИЕМ РАДАМАХЕРОВСКОЙ СЛОЖНОСТИ

Рогачев К.О. (Университет ИТМО)

Научный руководитель – кандидат технических наук, доцент Забашта А. С.

(Университет ИТМО)

Аннотация. Глубокие нейронные сети позволяют решать многочисленные задачи, благодаря большому числу настраиваемых параметров и наличию больших данных, но это требует огромного количества вычислений как во время обучения, так и во время применения к данным. В данной работе продемонстрирована структурная регуляризация нейронной сети архитектуры ResNet на примере задачи классификации, с использованием радамахеровской сложности промежуточных тензоров данных в нейронной сети для выбора порогов прореживания. Этот подход позволяет уменьшить число параметров нейронной сети после обучения с учетом , тем самым уменьшив необходимые для ее применения вычисления.

Введение. Задача прореживания нейронной сети крайне актуально в силу ресурсоемкости обучения и использования нейронных сетей, а также явления переобучения. В рамках данной задачи существует этап выбора весов на удаление с помощью разных критериев и подходов. Наиболее распространенный - удаление с помощью порогового значения. Но для каждого слоя такое число должно быть разнообразно, из-за различий распределений значений весов. В данной работе рассматривается зависимость весов от задачи, за счет оценки сложности задачи с помощью радамахеровской сложности.

Основная часть. Количество параметров глубоких нейронных сетей в рамках обучения закладывается больше, чем необходимо, из-за отсутствия знания о необходимом количестве. Эти параметры ведут к увеличению вероятности переобучения и дополнительным накладным затратам на применение нейронной сети к данным. В рамках большинства архитектур можно выделить функциональные единицы, например тензор свертки или колонка в рамках архитектур-трансформеров. Последние исследования показывают возможность оценки вклада слоя или группы весов в решаемую задачу с помощью оценки разнообразия данных получаемых после применения весов и функции активации. В рамках данной работы рассматривается метод удаления таких групп параметров на примере нейронной сети архитектуры ResNet, с анализом влияния разных групп на решаемую задачу с помощью применения радамахеровской сложности к промежуточным тензорам данных внутри нейронной сети.

Выводы. Реализованное программное обеспечение способно уменьшить размер нейронной сети, без серьезной потери качества при сохранении общих принципов построения ее архитектуры.

Список использованных источников:

1. Tristan Deleu, Yoshua Bengio. Structured Sparsity Inducing Adaptive Optimizers for Deep Learning // arXiv:2102.03869v2 [cs.LG] 5 Jan 2023
2. Graf F. et al. On measuring excess capacity in neural networks //Advances in Neural Information Processing Systems. – 2022. – Т. 35. – С. 10164-10178.

3. Tartaglione E. et al. Serene: Sensitivity-based regularization of neurons for structured sparsity in neural networks //IEEE Transactions on Neural Networks and Learning Systems. – 2021. – Т. 33. – №. 12. – С. 7237-7250.

Рогачев К.О. (автор)

Подпись

Забашта А.С. (научный руководитель)

Подпись