

Рецензия на работу "On the Periodic Behavior of DNNs training on the example of the Grokking effect" (Мельник Ю.М., Южаков Т.А., Ветров Д.П.)

Рецензент: Мелихов Дмитрий

1 декабря 2023 г.

В статье описывается эффект "гроккинга" для масштабно-инвариантных нейронных сетей. Данный эффект воспроизводится для трансформерной модели, которая решает задачу предсказания результата бинарной операции: Передаётся токены "a" И "b", нужно предсказать "c" такой, что $a \circ b = c$, где "o" - бинарная операция.

Получены интересные экспериментальные результаты по оптимизации масштабно-инвариантных нейронных сетей. Вводится эффективный темп обучения и норма эффективного градиента и объясняется почему во время обучения масштабно инвариантных моделей стоит обращать внимание именно на них.

Замечания:

- Использование нормализации для достижения периодического поведения: в статье периодический характер объясняется масштабной инвариантностью модели. Для подтверждения этой теории нужно провести эксперимент без LayerNorm, чтобы посмотреть на поведение модели не обладающей масштабной инвариантностью.

- Анализ статистик обучения:

"...При этом график эффективного темпа обучения имеет «пиловидный» характер, что можно объяснить резкими изменениями значения нормы весов модели..."

Можно добавить график нормы весов модели, чтобы это подтвердить.

- Вывод (и в некоторых других местах): "масштаб-инвариантных" \rightarrow "масштабно-инвариантных"