



Ученые AIRI нашли способ оптимизировать большие ИИ-модели на 10-15% без потери качества

Ученые из Института искусственного интеллекта AIRI при поддержке коллег из SberAI и Сколтеха обнаружили принципиально новое свойство больших языковых моделей и научились контролировать его. Это позволит оптимизировать модели на 10-15% без потери в качестве, снижая количество вычислительных мощностей, необходимое для их использования.

Трансформер – тип архитектуры моделей, который привел к революции в развитии искусственного интеллекта. Именно благодаря ему стало возможным развитие популярных диалоговых ботов. Любая архитектура состоит из большого количества слоев, через которые от «входа», например, запроса «нарисуй котенка», информация доходит до «выхода» и преобразовывается в картинку. Принято считать, что линейность слоев — свойство самых слабых моделей, а нелинейность – сильных, например, тех самых трансформеров. Линейность позволяет делать модели более простыми и эффективными в вычислениях, но не позволяет модели решать сложные задачи, например, выучивать необычные закономерности в данных.

Научные сотрудники лаборатории FusionBrain Института AIRI, SberAI и Сколтеха изучили устройство 20 известных open-source языковых моделей типа декодер и выяснили, что между эмбедингами (числовыми представлениями данных) есть высокая линейная зависимость. Это значит, что при переходе от слоя к слою информация не претерпевает нелинейных преобразований, и сложную архитектуру трансформера на самом деле можно заменить намного более легкими слоями нейросети.

Чтобы избежать проявления негативных свойств линейности во время предобучения, коллектив разработал специальный «регуляризатор». Он позволяет контролировать проявления линейности и улучшать метрики качества.

Благодаря возможности контролировать проявления линейности, ученые смогли заменить сложные блоки слоев модели на более простые. В ходе экспериментов выяснилось, что облегчать без потери качества можно 10-15% слоев. Далее модель начинает терять полезные навыки.

«Нам удалось изучить модели под микроскопом. Простыми средствами описать сложные процессы внутри трансформеров и сразу предложить эффективный регуляризатор. Уже проверили всё на маленьких моделях, проверки на больших моделях и обучении — впереди. Обнаруженный эффект кажется очень контринтуитивным, он противоречит многим представлениям о глубоком обучении. В то же время, именно он позволяет тратить меньше вычислительных ресурсов на развертку и инференс больших языковых моделей. На днях мы выложили препринт статьи, а она уже обогнала публикации от Google, Microsoft, MIT, и Adobe в списке статей дня на HuggingFace. Понимая важность работы для научного сообщества, мы поделились регуляризатором с коллегами и опубликовали его в открытом доступе»

Иван Оселедец, д.ф.-м.н., генеральный директор Института AIRI, профессор Сколтеха

«Одним из вызовов развития AI-технологий, в особенности больших языковых моделей (представителей ветки GenAI), остаётся потребность в вычислительных ресурсах для обучения следующего поколения SOTA-моделей. В основе большинства архитектур GenAI лежат блоки трансформеров, и в опубликованной работе выявлена линейность в некоторых представлениях данных внутри этих блоков. Как следствие, это позволяет существенно оптимизировать архитектуры с точки зрения вычислительных мощностей, снизить нагрузку, получить результат за меньшее время за счёт адаптивной регуляризации. Потенциал сокращения вычислительных ресурсов на обучении оценивается в порядке до 10%. Мы, в Сбере, планируем провести тестирование рассмотренной идеи и в случае успеха — тиражировать её на флагманские модели GenAI. Поиск таких смекалок в AI-архитектурах позволяет частично компенсировать вычислительный голод, поэтому продолжим поддержку таких исследований в направлении обучения больших моделей».

Андрей Белевцев, старший вице-президент, руководитель блока «Технологическое развитие» Сбербанка

Статья принята на одну из наиболее престижных конференций в сфере ИИ – ACL 2024 (Main Track, Core A*). Получить доступ к регуляризатору можно по [ссылке](#).

.....

Вопросы: pr@airi.net

Научно-исследовательский Институт искусственного интеллекта [AIRI](#) — автономная некоммерческая организация, занимающаяся фундаментальными и прикладными исследованиями в области искусственного интеллекта. На сегодняшний день более 150 научных сотрудников AIRI задействовано в исследовательских проектах Института для работы совместно с глобальным сообществом разработчиков, академическими и индустриальными партнерами.