

Дистилляция Большой Языковой Модели

Технический отчёт

Гуляков Никита

ВМК МГУ
204 группа

Ноябрь 2024

1 Введение

В данном отчёте представлены результаты работы по дистилляции большой языковой модели. Цель исследования заключалась в том, чтобы дообучить большую языковую модель с возможностью следования инструкциям (instruct) и дистиллировать её в меньшую модель, не обученную следовать инструкциям. В качестве исходных моделей были выбраны Qwen2.5-1.5b-instruct (модель-учитель) и Qwen2.5-0.5b (модель-ученик). Учитель был дообучен на научных статьях с ресурса elibrary.ru и выравнен для следования инструкциям, а ученик — на синтетических данных, сгенерированных на основе ответов модели-учителя. Эффективность моделей оценивалась по перплексии, и полученные результаты показали значительное улучшение на тестовом наборе данных.

2 Дообучение

Для модели-учителя была выбрана Qwen2.5-1.5b-instruct, так как она демонстрирует наилучшие результаты на бэнчмарках среди моделей с таким количеством параметров.

2.1 Набор данных

Основной набор данных состоял из статей на русском и английском языках с ресурса elibrary.ru. Из

текстов были удалены артефакты, которые могли помешать обучению модели. Итоговый набор данных включал 80 000 статей, из которых 20% были на русском языке и 80% на английском. Дополнительный набор данных для настройки модели на следование инструкциям включал смесь из следующих наборов:

1. Vikrmodels/GrandMaster-PRO-MAX (для инструкций на русском языке) — 6 000 примеров,
2. HuggingFaceH4/no_robots (для инструкций на английском языке) — 1 000 примеров,
3. dim/camel_ai_chemistry (для вопросов по химии) — 1 000 примеров,
4. lgaalves/camel-ai-physics (для вопросов по физике) — 1 000 примеров,
5. dim/camel_ai_biology (для вопросов по биологии) — 1 000 примеров,
6. lighteval/MATH (для вопросов по математике) — 200 примеров,
7. taesiri/axiv_qa (для вопросов по научным статьям с arXiv) — 100 примеров.

Таким образом, общий обучающий набор состоял из 90 300 примеров: 80 000 научных статей и 10 300 примеров для настройки следования инструкциям.

2.2 Методика обучения

Обучение проводилось в формате Bfloat16. Для дообучения использовалась методика LoRA (Low-Rank Adaptation) с рангом 4 и параметром alpha 8. Скорость обучения была установлена на $2e-4$ с линейным убыванием, обучение проводилось в одну эпоху на двух видеокартах GPU T4 и заняло около 30 часов.



Рис. 1: Первые 30 000 итераций обучения

2.3 Результаты

Оценка проводилась на тестовой выборке из 100 научных статей из исходного набора данных. Изначально, модель без дообучения показывала следующие результаты: perplexity = 9.4 и loss = 2.24. Лучший показатель перплексии (perplexity = 6.9, loss = 1.93) наблюдался у модели, обученной на первых 67 000 итерациях (только на научных статьях без примеров инструкций), однако такая модель с трудом следовала инструкциям и больше стремилась имитировать стиль научных статей. Финальная модель (perplexity = 7.84, loss = 2.06) показала способность следовать инструкциям и на наборе данных allenai/sciq (вопросы без вариантов ответов по различным наукам) корректно отвечала на большинство вопросов.

3 Дистилляция

В качестве модели-ученика была выбрана Qwen2.5-0.5b, так как она принадлежит тому

же семейству моделей и показывает хорошие результаты на своём уровне.

3.1 Синтетические данные

Синтетические данные для обучения были сформированы на основе вопросов из наборов:

1. HuggingFaceH4/no_robots — 11 000 примеров,
2. allenai/sciq — 9 000 примеров.

Для генерации ответов к каждому вопросу добавлялись системные запросы, повышающие качество ответов (в ходе обучения модели-ученика эти системные запросы были удалены). Итоговый набор данных включал 20 000 примеров вопросов с ответами, сгенерированными моделью-учителем.

3.2 Метод

Процесс обучения модели-ученика был идентичен дообучению модели-учителя, с той разницей, что скорость обучения оставалась фиксированной на протяжении всего обучения. В общей сложности процесс занял около 5 часов.

3.3 Результаты

Финальная модель-ученик успешно научилась следовать инструкциям, воспроизводя примеры учителя. На тестовом наборе вопросов allenai/sciq ученик корректно отвечал на большинство вопросов, демонстрируя стиль модели-учителя.

4 Заключение

По итогам эксперимента удалось передать способность следовать инструкциям от модели-учителя к модели-ученику, изначально не настроенной на следование инструкциям. Также модель-ученик унаследовала научные знания и стиль ответов модели-учителя, что подтвердилось при ручной оценке на тестовом наборе вопросов.