Low-rank self-play fine-tuning for small LLMs

Павел Юрьевич Мун

Московский физико-технический институт

Курс: Инновационный практикум (научный трек)

Эксперт: А. В. Грабовой

Консультант: Н. В. Охотников

2025

Задача и цель исследования

Задача

Дообучение небольших языковых моделей в условиях ограниченных ресурсов

Цель

Исследовать оправданность применимости предложенного метода

Продвижения

- Написана большая часть статьи
- Обосновано использование метода LoRA к методу SPIN
- Запущен вычислительный эксперимент

Трудности

- Формальный язык статьи
- Недостаток знаний в области NLP
- Вычислительная сложность эксперимента
- Большое количество новых библиотек

Вычислительный эксперимент

Гипотеза

Значения лосса у модели, обученной предложенным методом будет ниже, чем у модели на этапе SFT

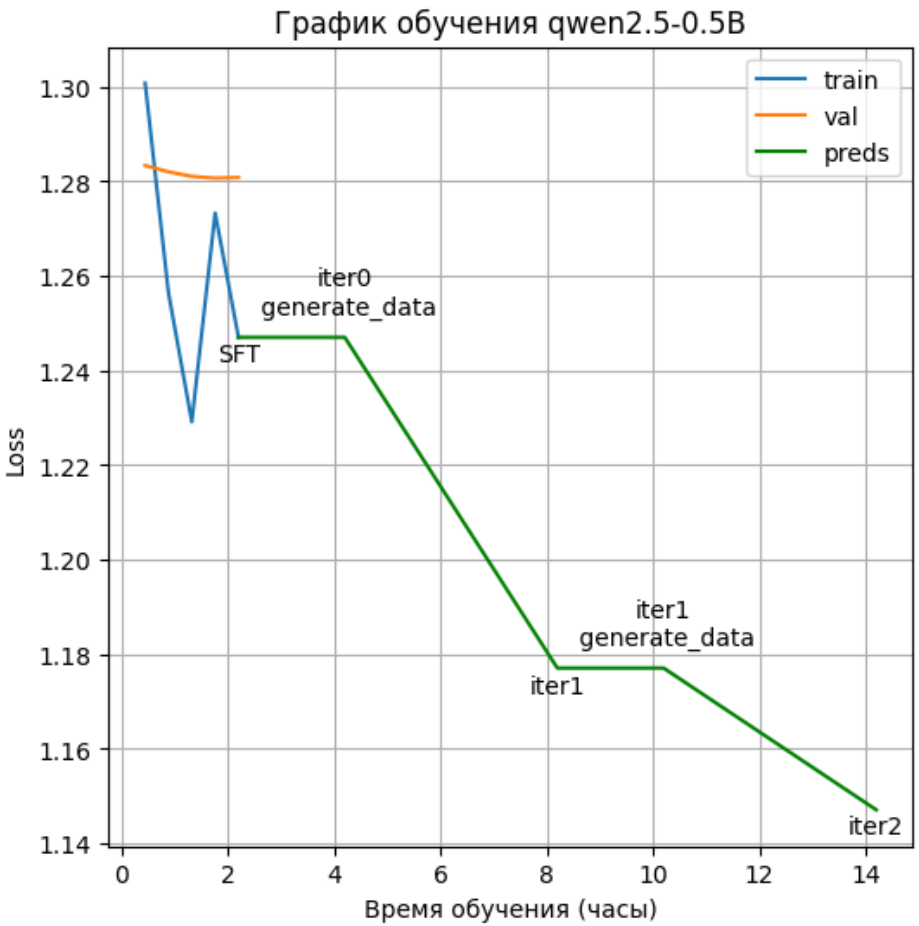
Цель

Обучение моделей предложенным методом и просто с адаптерами LoRA и сравнение по метрикам, лоссу и времени обучения

Данные

Рассматривается датасет ultrachat_200k, модели обучаются на 1% данных, примерно 2000 объектов

Сравнение итераций SPIN



- SFT - момент окончание этапа SFT
- iter(k)geneerate_data - окончание генерации данных противником для k + 1 итерации SPIN
- iter(k) - окончание k-ой итерации SPIN

Дальнейшие планы

- Обучение иных моделей
- Использование других датасетов
- Разобраться с квантизацией и, по возможности, встроить ее

Литература

1. Zixiang Chen и др. . «Self-Play Fine-Tuning Converts Weak Language Models to Strong Language Models». Openreview,2024 URL: https://openreview.net/forum?id=O4cHTxW9BS..
2. Edward J Hu и др. LoRA: Low-Rank Adaptation of Large Language Models Openreview, 2022 URL: https://openreview.net/forum?id=nZeVKeeFYf9.
3. Rafael Rafailov и др. Direct Preference Optimization: Your Language Model is Secretly a Reward Model// NeurIps, 2023