

Review:

Written by: Леванов Валерий

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

Author: Павел Мун

Введение:

- 1) Большое количество опечаток (“прдлагается”, “человеского” и т.п.)
- 2) Возможно следует ввести расшифровки аббревиатур SFT, RLHF.
- 3) “В частности предлагается метод дообучения LLM, который значительно снижает потребление видеопамати, а также убирает необходимость в прямом человеческом участии”. Вряд ли получится полностью убрать человеческое влияние.

Постановка задачи:

- 1) Пропущена скобка после вектора из y_i
- 2) лосс-функция = функция потерь

Метод:

- 1) Пропущенные слова, опечатки.
- 2) Не указано как ищем $\arg\min/\arg\max$
- 3) Непонятно, что такое дивергенция Кульбака-Лейблера. Либо определить, либо ссылку где почитать

Результаты:

- 1) Цветовая палитра на графике не очень, желтый текст плохо видно.
- 2) Непонятно с чем сравнивать значение памяти с LoRA.
- 3) Указание лосса нам ничего не говорит, желательна другая более наглядная метрика.
- 4) Непонятно, почему мы берем такую модель. что такое leaderboards

Понятно что статья не до конца доведена, поэтому часть ошибок легко будет исправлена позже.

Мнение:

Мне нравится тема доклада, Проблема, которые пытается решить авторы, действительно важна. Нравится, что все расписано достаточно

формально. Из минусов: мне было сложно разобраться в формулах, и в целом сложно разобраться если ты не подкован в теме. Ну и еще, конечно, необходимо дописать статью.