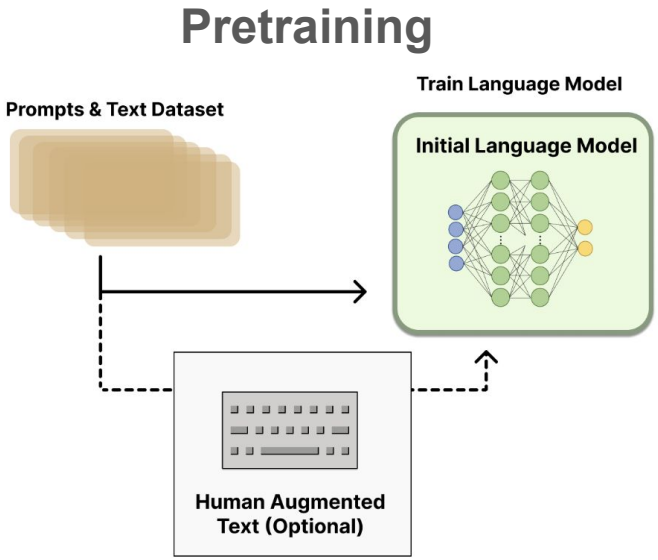
Тренировка диффузионных моделей при помощи DPO на нескольких изображениях

Куликов Антон

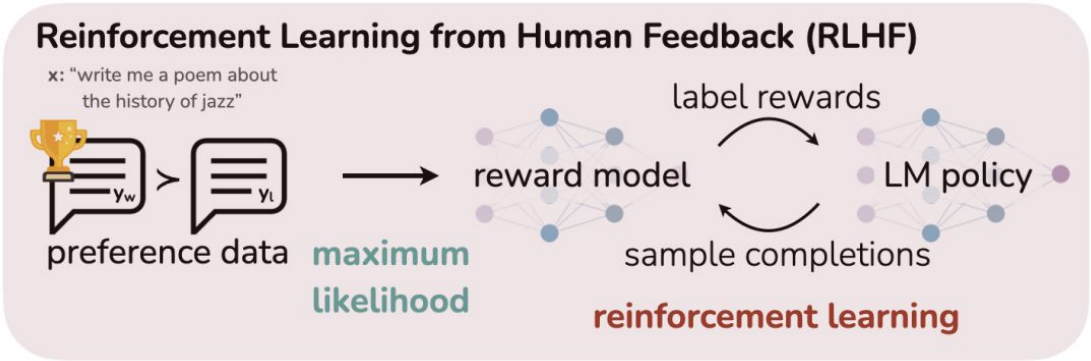
Предпосылки

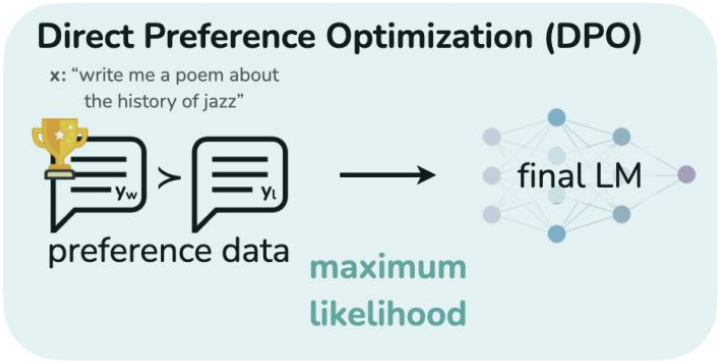
Для создания LLM, помимо предтренировки моделей как обычных LM и дотренировки на инструкциях, необходимо дообучать модель на ответах пользователей, иначе модели галлюцинируют.



Метод

Недавно ученые разработали новый подход - DPO, он производит такие же результаты, но проще в применении. Также он нашел свое применение в диффузионных моделях.





Метод

Помимо DPO ученые разработали еще один более эффективный подход - CPO.

1. Оригинальная статья
2. DPO в диффузиях
3. CPO

Метод

Основная идея метода заключается в применении DPO/CPO к предтренированной модели при помощи множественных отранжированных изображений.

Также одной из идей является добавление дискриминатора в лосс по образцу GAN.

$$\mathcal{L}_{\text{prefer}} = -\,\mathbb{E}_{(x,y_w,y_l)\sim\mathcal{D}}\Big[\log\sigma\Big(\beta\log\pi_\theta(y_w|x) - \beta\log\pi_\theta(y_l|x)\Big)\Big].$$

Дорожная карта проекта

1. Найти датасет с промптами и изображениями - picapick_v2.
2. Сгенерировать n изображений по каждому промпту - сделано при помощи sdxl_turbo с разным кол-вом шагов, но возможно использование sdxl-lightning.
3. Тренировка модели скоринга эстетичности изображений (дискриминатора) - сделано. Подробности далее.
4. Тренировка CPO/DPO на отскоренных изображениях - еще не сделано.

Тренировка дискриминатора

Задача дискриминатора - отличать менее эстетичные изображения от более эстетичных с учетом описания.

Данные:

Положительный класс - самые эстетичные изображения из coyo (score>=6.5).

Отрицательный класс - генерации sdxl_turbo.

Тренировка дискриминатора

Отрицательный класс

Отрицательный класс - генерация SDXL_turbo.

Генерация получена при помощи промпта, полученного при помощи мультимодальной модели LLAVA при инференсе на изображениях положительного класса.

Тренировка дискриминатора

Получение отрицательного класса

Реальное фото/картинка

llava

The image captures a serene scene of a small, rusted watchtower standing guard over a tranquil lake. The tower…

sdxl_t

Тренировка дискриминатора

Результат работы дискриминатора.

Изображения отсортированы по реалистичности слева направо

sdxl_t_1_coyo

sdxl_t_4_coyo_llava

sdxl_t_4_coyo_llava_prep

real

Тренировка дискриминатора

Результат работы дискриминатора.

Изображения отсортированы по реалистичности слева направо

Модель выбирает реальные изображения в 66% случаев

Спасибо за внимание!