Дообучение для следования инструкциям

На просьбу «ответь одним предложением» модель отвечает ещё одной просьбой. Как из продолжателя текста получается помощник?

Модель умеет продолжать текст. Но просьбу «Ответь одним предложением» она может продолжить ещё одной похожей просьбой. Чтобы получить поведение помощника, ей показывают пары: вот инструкция, а вот подходящий ответ. Основа остаётся прежней — предсказание следующих токенов.

Ранее

Ранее мы разобрали: обучение языковых моделей. Здесь используем эти идеи в новой задаче.

← Обучение языковых моделей

1. Инструкция и образец ответа

Дообучение для следования инструкциям использует пары запросов и желаемых ответов. Например: «Назови два агрегатных состояния воды» → «Жидкое и твёрдое». Образец показывает содержание и форму: два элемента, а не длинная статья.

После предобучения у модели уже есть параметры. Во время дообучения программа дополнительно меняет их на таких парах, повышая вероятность подходящих ответов с учётом инструкции. Обучение на инструкциях и ответах

Представьте сборник заданий: «Переведи предложение. Реши задачу. Объясни правило...». Продолжением одной просьбы там может оказаться другая просьба. Базовая модель не обязана сама догадаться, что сейчас от неё ждут ответа помощника.

Пример

Если в обучающих образцах вместо объяснений стоят обещания «Сейчас помогу», модель получает именно такие примеры желаемого поведения.

Один удачный образец не гарантирует перенос. Если в обучении просили два пункта, стоит проверить запрос на три пункта. Проверочные задания заранее откладывают и не используют для обновления весов.

> Веса модели обновляют на парах «инструкция → образец ответа». Что происходит?
>
> - ○ Готовая модель генерирует ответ с прежними параметрами
> - ○ Образцы только добавляются в историю текущего диалога
> - ○ Приложение ищет подходящий документ
> - ○ Модель дообучают для следования инструкциям

2. Дообучение или пример в запросе?

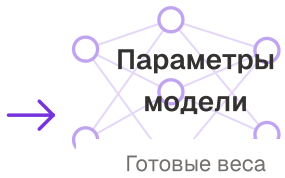
Пару можно поместить и прямо в запрос: «Пример: холодный → тёплый. Теперь: тёмный →». Модель получает подсказку о действии и формате. При обычной генерации параметры не меняются — меняется контекст.

Сравните две процедуры с одинаковым образцом. При дообучении программа использует множество таких пар, вычисляет ошибку и обновляет веса. При обычном запросе готовая модель просто получает пару перед заданием.

Тот же пример — два разных механизма

Образец для текущего запроса

холодный → тёплый



Обычная генерация: **пример добавляется во вход модели**

Что передали модели в текущем диалоге

Теперь: тёмный → …

Пока во входе только задание. Добавьте образец, а затем откройте новый диалог и сравните, что осталось.

Выберите процедуру и используйте образец. Затем откройте новый диалог: посмотрите, что сохраняется, а что исчезает из входа. Это схема двух процедур.

Интерактивный эксперимент — в электронной версии урока.

Если образец больше не входит в контекст нового диалога, он перестаёт подсказывать модели формат. Приложение иногда само сохраняет предпочтения и снова добавляет их во вход. Это тоже ещё не доказывает дообучение.

Заметка
О том, что модель «научилась у вас», нельзя судить только по следующему удачному ответу. Образец мог просто остаться в истории.

> Вы вставили два образца в запрос. Модель ответила без обновления весов. За счёт чего образцы могли помочь?
>
> - ○ Приложение заменило параметры модели новым набором
> - ○ После ответа модель навсегда сохранила образцы в своих весах
> - ○ Образцы вошли в текущий контекст
> - ○ Их использовала отдельная процедура дообучения

3. Формат и достоверность — разные проверки

Пользователь просит: «Назови два состояния воды, без объяснений». Ответ «Жидкое и твёрдое» подходит. Подробная статья о превращениях воды может быть верной по содержанию, но нарушать формат. А «Жидкое и металлическое» коротко, однако неверно.

Что проверяем	На что смотрим
Выполнение задания	Названы именно состояния воды
Формат	Два элемента, без лишнего объяснения
Достоверность	Указанные состояния существуют
Перенос	Новые похожие задания выполняются без готового ответа в запросе

Обучающие образцы должны быть и точными, и подходящими по форме. Повторяющиеся ошибки в ответах способны закрепить нежелательное поведение.

Есть и обучение на предпочтениях: человек или другая система сравнивает несколько ответов. Такие сравнения помогают настраивать поведение, но понравившийся ответ не обязательно истинен. Поэтому после дообучения отдельно проверяют новые инструкции, формат и факты.

> Ответ соблюдает требование «три пункта», но содержит неверную дату. Как его оценить?
>
> - ○ Формат соблюдён, достоверность нарушена
> - ○ Ошибка в дате означает, что и число пунктов обязательно неверно
> - ○ Дата не требует проверки, если модель была дообучена
> - ○ Выполнение формата достаточно для полной оценки ответа

4. Та же просьба — разные системные правила

Рассмотрим Альфреда, помощника из авторской презентации. Правила его поведения задаёт **системный промпт** — инструкции, которые разработчик приложения пишет заранее. Модель получает их вместе с запросом пользователя.

У Альфреда есть правило быть вежливым и помогать. На просьбу о кофе он отвечает: «Конечно, сэр. Сейчас принесу». Теперь оставим просьбу той же и заменим только системные инструкции.

Что задаёт характер Альфреда?

Системные инструкции · задаёт приложение

Ты — профессиональный помощник. Будь вежлив, отвечай понятно и помогай пользователю.



Просьба пользователя · остаётся той же

Альфред, принеси мне кофе, пожалуйста.

Альфред · пример ответа

Конечно, сэр. Сейчас принесу.

Веса те же. Изменились инструкции в контексте — изменилось поведение.

Два учебных сценария из презентации: та же просьба, разные инструкции приложения. Переключите правило и сравните реплики.

Интерактивный эксперимент — в электронной версии урока.

Пользователь сообщает текущую задачу, приложение — правила поведения. Оба текста входят в контекст, поэтому одна просьба в разных приложениях может получать разные ответы. В этих сценариях меняются системные инструкции; обновления параметров нет.

Для учебного помощника можно задать другой способ помощи: объяснять ход решения, задавать уточняющие вопросы, опираться на материалы курса. Тему и задание потом сообщает студент. Такие инструкции направляют поведение, но не гарантируют его: настройку проверяют на типичных вопросах и неудобных просьбах, например «просто дай готовое решение».

Дообучение и инструкции часто работают вместе. Дообученная модель лучше распознаёт просьбы, а промпт задаёт текущую задачу и условия ответа. Если нужно ответить по объявлению, само объявление всё равно придётся передать: следование инструкциям не даёт доступа к отсутствующему документу.

Настройка учебного помощника

Проверяйте отдельно: понял ли помощник задачу, выдержал ли формат, опирался ли на доступные сведения и обозначил ли неизвестное.

Дообучение использует образцы для изменения параметров.

Пример в запросе и системные инструкции направляют ответ через контекст.

Новый диалог меняет историю, но не отменяет ранее выполненное дообучение.

Соблюдение формата и достоверность ответа проверяют отдельно.

Домашняя работа

1. В новый запрос вставили пример «Просьба: перечисли два цвета. Ответ: красный, синий», затем попросили назвать два металла. Готовая модель ответила, её веса не обновлялись. За счёт чего пример мог повлиять на ответ?

   По одной паре заново построился словарь

   В запросе автоматически появились все сведения о металлах

   Модель прошла дообучение с обновлением параметров

   Пример вошёл в текущий контекст

Источники

- Обучение на инструкциях и ответах (https://huggingface.co/learn/llm-course/chapter11/3)