

Обобщение против запоминания: прослеживание возможностей языковых моделей до данных предварительной тренировки

Дата: 2025-03-01 00:00:00

Ссылка на исследование: <https://arxiv.org/pdf/2407.14985>

Рейтинг: 60

Адаптивность: 75

Ключевые выводы:

Исследование направлено на изучение баланса между способностью больших языковых моделей (LLM) к обобщению и запоминанию предобучающих данных. Основной вывод: разные способности LLM имеют разную природу - задачи, требующие фактических знаний, больше зависят от запоминания, а задачи рассуждения и перевода - от обобщения.

Объяснение метода:

Исследование имеет высокую концептуальную ценность, объясняя разницу между меморизацией и генерализацией в LLM для разных типов задач. Практическая ценность включает методы оптимизации промптов и понимание, что фактические вопросы требуют меморизации, а рассуждения — генерализации. Однако многие технические аспекты недоступны широкой аудитории без специальных знаний.

Ключевые аспекты исследования: 1. **Дистрибутивная меморизация и генерализация** - Исследование вводит новую концепцию "дистрибутивной меморизации", измеряющую корреляцию между вероятностями выходных данных LLM и частотой данных в предобучающем корпусе. Генерализация определяется как расхождение между этими распределениями.

Task-gram языковая модель - Авторы предлагают новый метод для моделирования распределений языка путем подсчета семантически связанных пар n-грамм из входных и выходных данных задачи, что позволяет эффективно анализировать большие предобучающие корпуса.

Различные типы задач имеют разные шаблоны меморизации/генерализации - Исследование показывает, что задачи, основанные на знаниях (например, фактические вопросы-ответы), больше зависят от меморизации, в то время как задачи рассуждения и перевода больше опираются на генерализацию.

Влияние размера модели - С увеличением размера модели баланс между меморизацией и генерализацией меняется в зависимости от типа задачи, с тенденцией к большей генерализации в более сложных задачах.

Оптимизация промптов - Исследование демонстрирует, что понимание того, требует ли задача меморизации или генерализации, может быть использовано для оптимизации промптов и улучшения производительности модели.

Дополнение:

Применение методов в стандартном чате

Для работы методов этого исследования не требуется дообучение или API в полной мере. Хотя авторы использовали расширенные техники (поиск по предобучающему корпусу, подсчет n-грамм) для научного анализа, основные концепции можно применить в стандартном чате:

Выбор типа формулировки в зависимости от задачи: Для фактических вопросов: использовать прямые, конкретные формулировки, близкие к учебным текстам Для задач рассуждения: использовать формулировки, поощряющие новизну и креативность

Адаптация промптов:

Метод "максимизации меморизации": использование более формальных, учебных формулировок для фактических вопросов Метод "максимизации генерализации": использование необычных, нестандартных формулировок для задач рассуждения

Практические результаты:

Улучшение точности фактических ответов при использовании промптов, способствующих меморизации Получение более креативных и необычных решений для задач рассуждения при использовании промптов, способствующих генерализации Понимание, что для сложных задач рассуждения более крупные модели могут давать более качественные результаты не из-за лучшей меморизации, а из-за лучшей генерализации Эти подходы можно применять в стандартном чате без необходимости доступа к предобучающим данным или API для анализа n-грамм.

Анализ практической применимости: 1. **Дистрибутивная меморизация и генерализация** - Прямая применимость: Низкая. Концепция требует доступа к предобучающим данным и специализированных инструментов для анализа. - Концептуальная ценность: Высокая. Помогает пользователям понять, что LLM могут по-разному обрабатывать разные типы задач (знания vs. рассуждения). - Потенциал для адаптации: Средний. Пользователи могут адаптировать свои запросы, зная, что фактические вопросы больше опираются на меморизацию, а сложные рассуждения — на генерализацию.

Task-gram языковая модель Прямая применимость: Очень низкая. Требуется технических навыков и доступа к предобучающим данным. Концептуальная ценность: Средняя. Помогает понять, как LLM связывают входные и выходные данные на основе предобучения. Потенциал для адаптации: Низкий. Метод сложен для адаптации обычными пользователями.

Различные типы задач имеют разные шаблоны меморизации/генерализации

Прямая применимость: Средняя. Пользователи могут формулировать запросы с учетом того, требует ли задача фактических знаний или рассуждений. Концептуальная ценность: Высокая. Помогает понять сильные и слабые стороны LLM в разных типах задач. Потенциал для адаптации: Высокий. Пользователи могут структурировать свои запросы в соответствии с типом задачи.

Влияние размера модели

Прямая применимость: Средняя. Пользователи могут выбирать модели разного размера в зависимости от типа задачи. Концептуальная ценность: Высокая. Помогает понять, что более крупные модели не всегда лучше для всех задач. Потенциал для адаптации: Средний. Пользователи могут адаптировать свои ожидания от моделей разного размера.

Оптимизация промптов

Прямая применимость: Высокая. Пользователи могут применять принципы оптимизации промптов для улучшения ответов LLM. Концептуальная ценность: Высокая. Демонстрирует, как формулировка запроса влияет на баланс между меморизацией и генерализацией. Потенциал для адаптации: Высокий. Представленные методы можно упростить для использования обычными пользователями.

Prompt:

Использование исследования о запоминании и обобщении в промптах для GPT ##
Ключевое понимание из исследования

Исследование показывает, что языковые модели по-разному обрабатывают различные типы задач: - **Задачи с фактическими знаниями** (например, TriviaQA) больше опираются на **запоминание** - **Задачи рассуждения и перевода** (например, MMLU, GSM-8K) больше опираются на **обобщение**

Пример промпта, учитывающего эти знания

[=====] # Запрос на решение математической задачи

Я хочу, чтобы ты решил следующую математическую задачу.

Поскольку исследования показывают, что языковые модели лучше справляются с задачами рассуждения при использовании обобщения, а не запоминания, я прошу тебя:

Не пытайся вспомнить похожую задачу из твоих тренировочных данных Вместо этого разбей задачу на логические шаги Используй общие математические принципы Объясняй свое рассуждение на каждом шаге Вот задача: [математическая задача] [=====]

Объяснение эффективности

Этот промпт работает, потому что:

Направляет модель на использование обобщения вместо запоминания, что согласно исследованию более эффективно для задач рассуждения **Структурирует процесс мышления** модели, запрашивая пошаговый подход **Явно указывает не полагаться на запоминание** конкретных примеров из тренировочных данных ## Другие применения исследования в промптах

- Для фактических вопросов: запрашивайте информацию в форматах, близких к обучающим данным
- Для творческих задач: явно запрашивайте новизну и минимизацию повторения шаблонов
- Для гибридных задач: разделяйте запрос на части, требующие запоминания и обобщения

Понимание того, как работает баланс запоминания и обобщения, позволяет более целенаправленно формулировать запросы к языковым моделям для получения оптимальных результатов.