

ADO: Автоматическая оптимизация данных для ввода в подсказках LLM

Дата: 2025-02-16 00:00:00

Ссылка на исследование: <https://arxiv.org/pdf/2502.11436>

Рейтинг: 75

Адаптивность: 85

Ключевые выводы:

Исследование направлено на улучшение производительности больших языковых моделей (LLM) путем оптимизации входных данных в промптах. В отличие от предыдущих исследований, фокусирующихся на улучшении инструкций или добавлении контекстных примеров, данная работа предлагает новый подход - оптимизацию самих входных данных. Результаты показывают, что предложенный метод ADO (Automatic Data Optimization) значительно улучшает производительность LLM в различных задачах.

Объяснение метода:

Исследование представляет ценную концепцию оптимизации входных данных для LLM. Двухуровневая стратегия (оптимизация содержания и структуры) может быть адаптирована пользователями любого уровня. Хотя полная реализация фреймворка требует технических навыков, основные принципы применимы вручную. Эффективность доказана на разных задачах и моделях, а возможность комбинирования с другими техниками повышает практическую ценность.

Ключевые аспекты исследования: 1. **Автоматическая оптимизация данных (ADO)**: Исследование представляет фреймворк для оптимизации входных данных в промптах для LLM путём автоматического улучшения как содержания данных, так и их структурного представления.

Двухуровневая стратегия оптимизации: Фреймворк ADO включает две ключевые стратегии - инженерию содержания (восполнение пропущенных значений, удаление нерелевантной информации, обогащение профилей) и структурное переформулирование (оптимизация формата представления данных).

Алгоритм DPS (Diverse Prompt Search): Разработан алгоритм поиска разнообразных промптов, который генерирует множество кандидатов с контролируемым разнообразием для более эффективного исследования пространства оптимизации.

Интеграция с существующими методами: ADO может быть объединен с другими техниками промпт-инжиниринга (CoT, ICL, PE2), что приводит к значительному повышению производительности LLM.

Эмпирическая валидация: Исследование демонстрирует улучшение производительности на 9 различных наборах данных с использованием 3 различных моделей LLM.

Дополнение:

Применение методов исследования в стандартном чате

Действительно, хотя в исследовании используется сложный фреймворк с несколькими LLM и API, основные концепции и подходы могут быть адаптированы для использования в стандартном чате без какого-либо дообучения или специального API.

Концепции для применения в стандартном чате:

Инженерия содержания: Восполнение пропущенных значений: Пользователь может явно указать в запросе известную информацию и попросить LLM сначала дополнить недостающие данные, а затем решить задачу. **Удаление нерелевантной информации:** Перед отправкой запроса пользователь может самостоятельно очистить текст от нерелевантных деталей или попросить LLM выделить ключевую информацию. **Обогащение профилей:** Пользователь может попросить LLM вывести дополнительную информацию из имеющихся данных перед решением основной задачи.

Структурное переформулирование:

Табличное представление: Преобразование текстовых данных в табличный формат. **XML/JSON структуры:** Структурирование данных в формате XML или JSON. **Визуальное форматирование:** Использование маркеров, отступов, группировки для улучшения читаемости.

Комбинирование с другими техниками:

Применение CoT ("давай подумаем шаг за шагом") после оптимизации данных. Включение примеров (ICL) с оптимизированными данными. ##### Пример применения в стандартном чате:

Исходный запрос: "У меня есть данные о пациенте: Иван, мужчина, имел боли в груди, курит, давление 150/90, уровень холестерина неизвестен, диабет отсутствует. Оцени риск сердечно-сосудистых заболеваний."

Оптимизированный запрос (двухэтапный):

"Проанализируй следующие данные о пациенте и восполни пропущенную информацию, если возможно: `<xml><пациент><имя>Иван</имя><пол>мужчина</пол><симптомы>боли в груди</симптомы><вредные_привычки>курение</вредные_привычки><давление>150/90</давление><холестерин>неизвестен</холестерин><диабет>отсутствует</диабет><возраст>неизвестен</возраст></пациент>` Какая дополнительная информация может быть выведена из этих данных?"

"Теперь, используя дополненные данные, оцени риск сердечно-сосудистых заболеваний. Давай подумаем шаг за шагом." Такой подход позволит получить более качественный результат без использования специального API или дообучения модели.

Анализ практической применимости: Автоматическая оптимизация данных (ADO): - Прямая применимость: Средняя. Обычным пользователям сложно напрямую использовать весь фреймворк ADO, так как он требует доступа к API нескольких LLM и написания кода. - Концептуальная ценность: Высокая. Идея о том, что оптимизация входных данных так же важна, как и инструкции, является ценной для всех пользователей LLM. - Потенциал для адаптации: Высокий. Пользователи могут вручную применять принципы оптимизации данных к своим запросам.

Двухуровневая стратегия оптимизации: - Прямая применимость: Высокая. Пользователи могут сразу начать применять принципы инженерии содержания и структурного переформулирования к своим запросам. - Концептуальная ценность: Высокая. Разделение оптимизации на содержание и структуру помогает пользователям системно подходить к улучшению запросов. - Потенциал для адаптации: Высокий. Эти стратегии можно адаптировать для любого типа данных и задач.

Алгоритм DPS: - Прямая применимость: Низкая. Сложный для реализации обычным пользователем. - Концептуальная ценность: Средняя. Концепция исследования разнообразных вариантов полезна, но технические детали сложны для широкой аудитории. - Потенциал для адаптации: Средний. Идею о проверке разных форматов представления данных можно применять вручную.

Интеграция с существующими методами: - Прямая применимость: Высокая. Пользователи могут комбинировать методы оптимизации данных с другими техниками промпт-инжиниринга. - Концептуальная ценность: Высокая. Понимание комплементарности различных подходов к промпт-инжинирингу важно для эффективного использования LLM. - Потенциал для адаптации: Высокий. Этот подход может быть адаптирован для любых задач.

Эмпирическая валидация: - Прямая применимость: Средняя. Результаты подтверждают эффективность подхода, но не дают конкретных рекомендаций для всех типов задач. - Концептуальная ценность: Высокая. Доказательство эффективности оптимизации данных мотивирует пользователей уделять внимание этому аспекту. - Потенциал для адаптации: Высокий. Методы, показавшие

эффективность на разных наборах данных, вероятно, будут работать и в других контекстах.

Prompt:

Применение ADO для оптимизации промптов Исследование ADO (Automatic Data Optimization) предлагает методы улучшения эффективности LLM путем оптимизации самих входных данных в промптах. Вот как можно применить эти знания на практике.

Ключевые принципы ADO - Инженерия контента: заполнение пропусков, удаление нерелевантного, обогащение информацией - **Структурное переформулирование:** оптимизация представления данных - **Разнообразие кандидатов:** поиск оптимальных вариантов с разными подходами

Пример промпта, использующего принципы ADO

[=====] Проанализируй следующие данные о транзакции и определи, является ли она мошеннической:

ОПТИМИЗИРОВАННЫЕ ДАННЫЕ: { "transaction_id": "TX78923", "amount": 1450.00, "time": "02:34 AM", "location": "Online", "device": "Unknown mobile device", "ip_address": "185.76.xx.xx", "country": "Нидерланды", "user_location": "Россия", "account_age": "6 дней", "previous_transactions": 2, "average_transaction_amount": 50.00, "similar_transactions_last_24h": 3 }

СТРУКТУРИРОВАННЫЙ АНАЛИЗ: 1. Сравни сумму транзакции (1450.00) с обычной активностью пользователя (средняя сумма: 50.00) 2. Оцени географическое несоответствие (транзакция из Нидерландов, пользователь из России) 3. Учи возраст аккаунта (6 дней) относительно активности 4. Проанализируй частоту похожих транзакций за последние 24 часа (3 транзакции)

На основе этого анализа, является ли данная транзакция мошеннической? Объясни свое решение шаг за шагом. [=====]

Как работают принципы ADO в этом промпте

Инженерия контента: Добавлены важные атрибуты (возраст аккаунта, средняя сумма транзакций) Выделены ключевые сравнения (несоответствие локаций, необычная сумма) Удалены потенциально отвлекающие детали

Структурное переформулирование:

Данные представлены в JSON-формате для лучшей обработки моделью Добавлена секция "СТРУКТУРИРОВАННЫЙ АНАЛИЗ" для направления рассуждений

Применение разнообразия:

Промпт комбинирует фактические данные и инструкции по анализу Включает элемент пошагового рассуждения (подобно Chain-of-Thought) ## Практический совет

При создании промптов с использованием принципов ADO, попробуйте: 1. Обогащать исходные данные дополнительной релевантной информацией 2. Структурировать данные в формате, который легче обрабатывается LLM (JSON, XML, таблицы) 3. Добавить явные инструкции по анализу ключевых элементов данных 4. Комбинировать ADO с другими техниками (CoT, ICL) для максимального эффекта

Такой подход позволяет значительно повысить качество ответов LLM в различных задачах, особенно связанных с классификацией, рекомендациями и логическим анализом.