

Автоматическая разметка с помощью открытых LLM, используя интеграцию динамической схемы меток

Дата: 2025-01-21 00:00:00

Ссылка на исследование: <https://arxiv.org/pdf/2501.12332>

Рейтинг: 78

Адаптивность: 85

Ключевые выводы:

Исследование направлено на изучение эффективности использования открытых LLM для автоматической маркировки данных с учетом ограничений приватности и ресурсов. Основным результатом - разработка метода Retrieval Augmented Classification (RAC), который динамически интегрирует схему меток и позволяет балансировать между качеством маркировки и охватом данных.

Объяснение метода:

Исследование предлагает метод RAC, который может быть адаптирован для стандартных чатов LLM. Ключевые преимущества: последовательная проверка категорий от наиболее вероятных, использование подробных описаний категорий, компромисс между точностью и охватом. Эти концепции применимы в повседневных задачах классификации без специальных инструментов и значительно улучшают точность взаимодействия с LLM.

Ключевые аспекты исследования: 1. **Retrieval Augmented Classification (RAC)** - метод, который превращает сложную многоклассовую классификацию в последовательность бинарных классификаций, начиная с наиболее релевантных категорий.

Интеграция схемы меток - исследование показывает, что включение не только названий категорий, но и их описаний значительно улучшает точность классификации.

Динамический компромисс между качеством и охватом - метод Truncated RAC позволяет находить баланс между точностью классификации и процентом размеченных данных.

Эффективное использование малых открытых LLM - исследование демонстрирует, как можно использовать небольшие модели (7B параметров) для

автоматической разметки данных с приемлемым качеством.

Дистилляция знаний - подход к обучению моделей на автоматически размеченных данных, что позволяет создавать специализированные классификаторы.

Дополнение: Для работы методов из этого исследования не обязательно требуется дообучение или API. Хотя авторы использовали специализированные модели (Mistral-7B, Llama-7B) с локальным хостингом, ключевые концепции могут быть применены в стандартном чате LLM.

Концепции и подходы, которые можно адаптировать для стандартного чата:

Последовательная бинарная классификация: Вместо одновременной проверки всех категорий, пользователь может проверять принадлежность текста к категориям по очереди. Например: "Относится ли этот текст к категории 'Финансовые новости'? Вот описание этой категории: ..."

Интеграция схемы меток: Включение подробных описаний категорий в запрос значительно улучшает точность классификации. Это легко реализуется в любом чате.

Приоритизация категорий: Пользователь может сначала проверять наиболее вероятные категории, что экономит время и повышает точность.

Chain of Thought (CoT): Использование подхода "цепочки рассуждений", когда модель объясняет свой ход мыслей, повышает точность классификации для большинства задач.

Компромисс между качеством и охватом: Пользователь может остановиться после проверки нескольких наиболее вероятных категорий, если получен удовлетворительный ответ.

Результаты от применения этих концепций: - Повышение точности классификации (в исследовании до 20% прироста F1-меры) - Более эффективное использование контекстного окна модели - Возможность работы со сложными задачами классификации (большое количество категорий) - Более понятные и обоснованные результаты классификации

Важно отметить, что хотя компонент ранжирования категорий по релевантности сложнее реализовать в стандартном чате, пользователь может сам определить наиболее вероятные категории на основе своих знаний о тексте и задаче.

Анализ практической применимости: 1. **Retrieval Augmented Classification (RAC)** - Прямая применимость: Высокая. Пользователи могут адаптировать этот метод для последовательной проверки принадлежности текста к категориям, начиная с наиболее вероятных. - Концептуальная ценность: Высокая. Демонстрирует, как разбивать сложные задачи на более простые для LLM. - Потенциал для адаптации: Очень высокий. Метод может быть адаптирован для любой задачи классификации с

использованием стандартных чат-интерфейсов.

Интеграция схемы меток Прямая применимость: Высокая. Пользователи могут немедленно начать включать подробные описания категорий в свои запросы. Концептуальная ценность: Высокая. Демонстрирует важность контекста для правильной интерпретации LLM. Потенциал для адаптации: Высокий. Применимо к любым задачам, где требуется точная классификация.

Динамический компромисс между качеством и охватом

Прямая применимость: Средняя. Требуется некоторой адаптации для использования в стандартном чате. Концептуальная ценность: Высокая. Понимание этого компромисса помогает пользователям настраивать взаимодействие с LLM. Потенциал для адаптации: Высокий. Можно реализовать через многоэтапные запросы.

Эффективное использование малых открытых LLM

Прямая применимость: Средняя. Большинство пользователей используют готовые сервисы, а не запускают модели локально. Концептуальная ценность: Высокая. Показывает, что даже небольшие модели могут быть эффективны при правильном подходе. Потенциал для адаптации: Средний. Применимо для пользователей с техническими навыками.

Дистилляция знаний

Прямая применимость: Низкая. Требуется специальных навыков в машинном обучении. Концептуальная ценность: Средняя. Показывает путь развития от ручной разметки к автоматизированным системам. Потенциал для адаптации: Средний. Применимо для технически подготовленных пользователей.

Prompt:

Применение знаний из исследования RAC в промптах для GPT ## Основные идеи исследования для промптов

Исследование о Retrieval Augmented Classification (RAC) предлагает несколько ценных подходов для улучшения классификации с помощью LLM:

Включение описаний меток значительно улучшает точность (до +20%)
Последовательная обработка меток вместо одновременной **Метод самосоогласованности** для повышения точности **Truncated RAC** для баланса между скоростью и качеством ## Пример промпта, использующего принципы RAC

[=====] # Задача классификации текста

Контекст Я предоставляю вам текст, который нужно классифицировать по одной из следующих категорий:

Запрос на перевод средств: Клиент хочет перевести деньги между счетами или другому лицу. Включает указания суммы, получателя или счета. **Проверка баланса:** Клиент интересуется текущим состоянием своего счета, доступными средствами или последними транзакциями. **Проблема с картой:** Клиент сообщает о потере, краже, блокировке карты или проблемах с транзакциями. **## Инструкции** 1. Прочитайте текст клиента внимательно. 2. Рассмотрите по очереди каждую категорию, начиная с наиболее вероятной. 3. Объясните ваш ход рассуждений для каждой категории (Chain-of-Thought). 4. Если текст не подходит ни к одной категории или вы не уверены, укажите "Требуется уточнение". 5. Дайте окончательный ответ в формате: "Категория: [название]".

Текст для классификации: [ТЕКСТ КЛИЕНТА] [=====]

Почему этот промпт работает лучше

Включение описаний меток - для каждой категории дано подробное описание **Последовательный анализ** - инструкция рассматривать категории по очереди **Chain-of-Thought** - запрос на объяснение рассуждений **Опция неопределенности** - возможность не классифицировать неясные случаи Этот подход позволяет достичь баланса между точностью классификации и охватом данных, как показано в исследовании RAC. Для более сложных задач с большим количеством категорий можно модифицировать промпт, разбивая классификацию на несколько этапов, начиная с наиболее вероятных категорий.