

«Анализ роли контекста в прогнозировании с помощью больших языковых моделей»

Дата: 2025-01-11 00:00:00

Ссылка на исследование: <https://arxiv.org/pdf/2501.06496>

Рейтинг: 75

Адаптивность: 85

Ключевые выводы:

Исследование оценивает эффективность языковых моделей (LLM) в прогнозировании бинарных вопросов. Основная цель - изучить, как контекстная информация влияет на точность прогнозов LLM. Результаты показывают, что включение новостных статей значительно улучшает производительность моделей, в то время как использование примеров few-shot снижает точность. Более крупные модели стабильно превосходят меньшие модели.

Объяснение метода:

Исследование предоставляет практически применимые стратегии улучшения прогнозов через обогащение контекста. Результаты показывают, какие типы контекста наиболее полезны (фоновая информация + новости), а какие избыточны (few-shot примеры). Понимание склонностей моделей к определенным типам ответов и влияния контекста критически важно для эффективного использования LLM в прогностических задачах.

Ключевые аспекты исследования: 1. **Анализ влияния контекста на прогнозирование с помощью LLM** - исследование изучает, как разные уровни контекста (только вопрос, фоновая информация, новостные статьи, критерии разрешения, примеры few-shot) влияют на точность прогнозирования бинарных событий.

Создание нового набора данных - авторы собрали 614 бинарных прогнозных вопросов с платформы Metaculus, дополнив их новостными статьями и их краткими резюме для обеспечения релевантного контекста.

Сравнение эффективности разных моделей - исследование сравнивает прогностические способности трех моделей разного размера и даты обучения: GPT-3.5-turbo, Alpaca-7B и Llama2-13B-chat.

Выявление оптимального уровня контекста - результаты показывают, что наилучшая точность достигается при предоставлении моделям фоновой

информации и новостных статей, а добавление примеров few-shot ухудшает производительность.

Обнаружение паттернов в прогнозировании - исследование выявило, что при минимальном контексте модели склонны давать преимущественно отрицательные прогнозы, а с увеличением контекста это соотношение меняется.

Дополнение:

Применимость методов без дообучения или API

Методы исследования полностью применимы в стандартном чате без необходимости дообучения или специального API. Исследователи использовали API только для удобства проведения масштабного эксперимента, но выявленные принципы работают в любом диалоговом интерфейсе с LLM.

Концепции и подходы для стандартного чата

Структурирование запросов с оптимальным контекстом Добавление фоновой информации и 2-3 релевантных новостных статей значительно повышает точность прогнозов Избегание слишком длинных запросов с избыточными примерами

Двухэтапный подход к прогнозированию

Сначала использовать LLM для поиска и резюмирования релевантной информации
Затем использовать эту информацию как контекст для прогностического вопроса

Критическая оценка ответов

Учитывать склонность моделей к отрицательным прогнозам при недостатке контекста Запрашивать обоснование прогноза для оценки его надежности

Итеративное уточнение контекста

При неуверенном ответе можно запросить, какой дополнительной информации не хватает Добавить эту информацию и повторить запрос #### Ожидаемые результаты применения

- Повышение точности прогнозов на 5-10% по сравнению с запросами без контекста
- Более сбалансированное соотношение положительных и отрицательных прогнозов
- Получение более обоснованных и аргументированных прогнозов
- Лучшее понимание факторов, влияющих на вероятный исход событий

Анализ практической применимости: 1. **Анализ влияния контекста на прогнозирование** - Прямая применимость: Пользователи могут значительно

улучшить точность прогнозов, добавляя в запросы к LLM релевантные новостные статьи и фоновую информацию. Это практическая техника, которую можно сразу применять. - Концептуальная ценность: Понимание, что контекст критически важен для точности прогнозов, помогает пользователям осознать необходимость предоставления достаточной информации. - Потенциал для адаптации: Подход с добавлением контекста можно применять для любых прогностических задач, не ограничиваясь форматом исследования.

Создание нового набора данных Прямая применимость: Ограниченная для обычных пользователей, так как сам набор данных больше полезен исследователям. Концептуальная ценность: Понимание типов вопросов и контекста, которые помогают в прогнозировании, может помочь пользователям формулировать свои запросы. Потенциал для адаптации: Методология сбора релевантных новостей и создания резюме может быть адаптирована пользователями для своих прогностических задач.

Сравнение эффективности разных моделей

Прямая применимость: Пользователи получают понимание, что более крупные модели дают более точные прогнозы, что может влиять на выбор модели. Концептуальная ценность: Осознание разницы в возможностях разных моделей помогает устанавливать реалистичные ожидания от прогнозов. Потенциал для адаптации: Ограниченный, так как выбор модели обычно предопределен платформой.

Выявление оптимального уровня контекста

Прямая применимость: Пользователи могут следовать рекомендациям по оптимальной структуре запроса: включать фоновую информацию и релевантные новости, но избегать избыточных примеров few-shot. Концептуальная ценность: Понимание, что слишком большие запросы могут ухудшать производительность, помогает оптимизировать взаимодействие с LLM. Потенциал для адаптации: Высокий, принцип "достаточного, но не избыточного контекста" применим к любым задачам с LLM.

Обнаружение паттернов в прогнозировании

Прямая применимость: Пользователи должны осознавать склонность моделей к отрицательным прогнозам при недостатке контекста. Концептуальная ценность: Понимание возможных предвзятостей моделей помогает критически оценивать получаемые прогнозы. Потенциал для адаптации: Знание о склонностях моделей может быть применено для корректировки интерпретации ответов в различных контекстах.

Prompt:

Как использовать знания из исследования в промптах для GPT ## Ключевые выводы исследования

Исследование показало, что: - Включение релевантного контекста (новостных статей, фоновой информации) значительно улучшает точность прогнозов - Избыточная информация (особенно примеры few-shot) может ухудшать результаты - Критерии разрешения вопроса помогают модели лучше понимать задачу - Краткие резюме новостей работают лучше, чем полные тексты

Пример эффективного промпта для прогнозирования

[=====] # Задача прогнозирования

Вопрос Превысит ли цена Bitcoin \$50,000 к концу 2023 года?

Фоновая информация Bitcoin - ведущая криптовалюта, торгуемая на глобальных рынках. Текущий курс составляет \$35,000.

Релевантные новости (краткие резюме) 1. ФРС США объявила о приостановке повышения процентных ставок на последнем заседании. 2. Крупные институциональные инвесторы, включая BlackRock, подали заявки на запуск биткоин-ETF. 3. Технические аналитики отмечают формирование бычьего паттерна на графике BTC.

Критерии разрешения Вопрос будет считаться положительно разрешенным, если цена Bitcoin на бирже Coinbase превысит \$50,000 хотя бы на 1 минуту до 23:59:59 31 декабря 2023 года по UTC.

Основываясь на предоставленной информации, спрогнозируйте, произойдет ли это событие, и объясните свой прогноз. [=====]

Почему этот промпт работает эффективно

Структурированный формат делает информацию легко воспринимаемой для модели **Фоновая информация** предоставляет базовый контекст **Краткие резюме новостей** содержат релевантные факты без перегрузки модели **Четкие критерии разрешения** помогают модели точно понять, что именно прогнозируется **Отсутствие примеров few-shot** убирает потенциально вредные элементы Такой подход к составлению промптов, согласно исследованию, может повысить точность прогнозов до 68% (для GPT-3.5-turbo), что значительно выше базовой точности при использовании только вопроса без контекста.