

УДК 004.89

**ПРИМЕНЕНИЕ ГЕНЕРАТИВНОЙ ЯЗЫКОВОЙ МОДЕЛИ GPT-3 ДЛЯ
ГЕНЕРАЦИИ ВОПРОСОВ К
ТЕКСТУ**

Свищев А.Н. (Университет ИТМО),

Научный руководитель – к. ф.-м. н. Рыбин С.В.

(Университет ИТМО)

Решается задача генерации вопросов к тексту в диалоге (CoQG - conversational/contextual question generation). Для ее решения уже предлагались специализированные нейросетевые архитектуры, обученные на больших, специально собранных наборах данных. Однако, общее качество таких решений остается неудовлетворительным и требующем доменной адаптации под конкретные сценарии использования. В данной работе предлагается другой подход, не требующий сбора больших специализированных наборов данных, разработки и настройки специализированных нейросетевых архитектур, сложной доменной адаптации, и тонкой настройки параметров моделей. При этом предлагаемый метод генерации вопросов в контексте диалога демонстрирует удовлетворительное качество и разнообразие вопросов.

Исследования выполнены за счет финансирования университета ИТМО в рамках НИР №620183 «Разработка виртуального диалогового помощника для поддержки проведения дистанционного экзамена на основе моделей-трансформаторов и понимания естественного и математического языка».

Генерация вопроса требует глубокого понимания контекста беседы, информационного контекста (рационали), наличия стратегии опроса (определенная последовательность вопросов в беседе, стратегия переноса внимания в информационном контексте - источнике рационалей). Достаточно длинная рациональ может быть источником генерации большого числа вопросов (к различным ключевым словам и понятиям, различных типов вопросов). Все это создает неопределенность в работе модели, требует ввода дополнительной информации, которая существенно изменяет архитектуру решения и может требовать изменения набора обучающих данных под различные сценарии использования модели. Вопросы, генерируемые системой могут быть общими, специальными, закрытыми, открытыми. Для многих релевантных теме вопросов недостаточно информации только из текста. Требуется некоторое априорное знание модели о мире.

В новом подходе предлагается использовать универсальную языковую модель высокой емкости для русского языка ruGPT-3 и так называемое “затравочное программирование” (“prompt engineering”) или генерация на основе шаблона (zero-shot learning). Это позволяет избавиться от необходимости подготавливать обширный специализированный набор данных, а также, гибко изменяя шаблон “затравки”, получать решения для различных сценариев решения задачи и различных доменов знаний.

Обучение без примеров или zero-shot (“затравочное программирование”) - это когда генеративной модели предъявляется шаблон данных в некотором формате и ожидается, что модель продолжит генерацию текстом, связанным по смыслу с текстом из шаблона, в заданном формате.

В работе продемонстрирована способность модели ruGPT-3 поддерживать заданный формат шаблона. Исследована возможность применения затравочного программирования в контексте решения задачи генерации вопросов в контексте диалога. Предложены различные варианты шаблонов для генерации вопросов к заданной рационали. Затронута тема управления генераций через ключевые слова и предъявляемые ответы. В работе также поднимается тема процедуры быстрой адаптации модели к заданному формату шаблонов на малом наборе данных.

Основной вывод в работе - предлагаемый метод генерации вопросов в контексте диалога позволяет получать качество решение задачи сравнимое или выше других, значительно более трудоемких подходов. Гибкость предлагаемого метода позволяет значительно удешевить и ускорить разработку решений в рамках проводимого исследования.