

## НЕЙРОННЫЕ СЕТИ С ПАМЯТЬЮ ДЛЯ УЧЁТА КОНТЕКСТА В ДИАЛОГОВЫХ СИСТЕМАХ С ПОИСКОМ ОТВЕТА

Маслюхин С.М. (Университет ИТМО, ООО «ЦРТ-инновации»)

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».

Недавние исследования, посвящённые диалоговым системам с поиском ответа, показали важность использования информации из истории диалога (контекста диалога). Исследованию методов учёта контекста в диалоговых системах с поиском ответа, позволяющих эффективно извлекать ключевую информацию из истории диалога, посвящена данная работа.

Диалоги, обрабатываемые диалоговыми системами, могут содержать большое количество высказываний и иметь сложную структуру. Сложная структура подразумевает наличие смысловых связей между высказываниями, отделёнными друг от друга последовательностью других высказываний (поддиалогом). Для обработки таких сложных зависимостей между высказываниями необходимо учитывать весь контекст диалога. В условиях ограниченного размера входных данных в нейросетевых моделях, данная задача является критически важной, так как модель не может обработать весь диалог целиком. Существует несколько основных способов учёта истории диалога, применяемых в существующих диалоговых системах с поиском ответа: конкатенация и трекинг. Конкатенация является наиболее распространённым приёмом на практике. При использовании конкатенации вся история диалога объединяется в единую текстовую последовательность и обрезается слева с учётом максимальной длины входной последовательности нейросетевой модели. Преимуществами такого подхода являются простота и эффективность, поскольку в данном случае обращение к нейросетевой модели выполняется всего один раз. Главным недостатком подхода является то, что информация из обрезанной части полностью теряется, несмотря на то, что она могла быть ключевой. Трекинг предполагает векторизацию всех высказываний диалога и последующее вычисление их взвешенной суммы. Таким образом используется информация обо всей истории диалога, однако при выборе ответа не учитывается значимость отдельных высказываний. При этом подходе наиболее значимыми оказываются высказывания более близкие к концу диалога, что не всегда является верным. При поиске подходящего ответа важно не просто учитывать всю историю диалога, но и уметь извлекать ключевую информацию для обработки текущего запроса. Решению этих двух задач посвящена данная работа.

В данной работе предлагается использовать нейронную сеть с памятью. Применение нейронных сетей с памятью показало хорошие результаты на таких текстовых задачах как: моделирование языка и поиск ответа на вопрос в открытом домене. Поиск ответа в диалоговых системах является схожей задачей и может решаться с помощью подобных подходов. При этом структура нейронной сети с памятью определяется характеристиками компонента памяти и операциями чтения и записи. Предлагаемый в данной работе компонент памяти представляет собой матрицу размерности, количество высказываний в диалоге на размер векторного представления высказывания, и поддерживает операции чтения и записи. Операция записи в память предусматривает последовательную запись векторных представлений высказываний диалога в соответствующие слоты памяти. Рассматриваются различные способы векторизации текста такие как: объединение векторных словарных представлений `word2vec` и использование контекстных векторных представлений, генерируемых предобученной языковой моделью BERT. Операция чтения выполняется путём применения механизма внимания. На вход механизма внимания подаётся векторное представление текущего запроса и вектора, хранящиеся в памяти, на выходе получается

взвешенное векторное представление истории диалога относительно текущего запроса. Затем полученное векторное представление контекста диалога подаётся в нейросетевую модель вместе с векторным представлением текущего запроса. Таким образом модель обрабатывает запрос и историю диалога с учётом значимости каждого высказывания в контексте данного запроса.

Исследование различных способов учёта контекста диалога показало эффективность предложенной техники относительно существующих решений. Применение техники учёта контекста с памятью на практике позволяет повысить качество работы диалоговой системы с поиском ответа на диалогах, содержащих сложные зависимости.

Маслюхин С.М. (автор)

Подпись