

ИССЛЕДОВАНИЕ МЕТОДОВ ПЕРСОНОФИКАЦИИ РАНЖИРУЮЩИХ ДИАЛОГОВЫХ СИСТЕМ

Посохов П. А.

(Университет ИТМО)

Научный руководитель – к. т. н. Махныткина О. В.

(Университет ИТМО)

Исследования выполнены за счет финансирования университета ИТМО в рамках НИР №621296 «Разработка технологий для персонификации разговорного искусственного интеллекта».

По прогнозам, объем рынка чат-ботов вырастет с 2,6 млрд. долларов в 2021 году до 9,4 млрд. долларов к 2024 году при среднегодовом темпе роста (CAGR) 29,7%. На сегодняшний день диалоговые системы принято разделять на целеориентированной и нецелеориентированной (с открытым доменом), вторые в свою очередь представляют наибольший исследовательский интерес, поскольку способны к поддержанию беседы в различных предметных областях. Вместе с тем модели с открытым доменом обладают рядом проблем, в числе которых: непоследовательность и противоречивость ответов, отсутствие явной долговременной памяти, склонность давать размытые и бессмысленные ответы. Персонофикация диалоговых систем частично моделирует долговременную память агента за счет использования последовательной персоны, делает его ответы более конкретными и логичными. На фоне этого исследования методов персонофикации являются одной из наиболее актуальных задач для развития диалоговых систем.

По принципу работы диалоговые модели можно классифицировать на ранжирующие модели поиска и генеративные. Первые на данный момент показывают наилучшие результаты и обладают более информативными метриками качества (MRR, R1), ввиду чего в данной работе рассматривается этот тип моделей. В частности проводится тестирование различных методов представления текста (BERT, GPT, tf-idf BoW и др.), сравнительный анализ Vi-Encoder и Poly-Encoder архитектур, сравнение функций подобия для ранжирования кандидатов, в том числе комплексных (CoBERT) и стратегий токенизации и различных типов агрегации выходного вектора контекста и кандидата. Так же проводится исследование и анализ эффективности персонофикации путем конкатенации вектора персоны и контекста и нахождением среднего этих векторов.

Для исследования был выбран свободно распространяемый набор данных «Toloka.ru PersonaChat», содержащий диалоги людей с размеченными персонами участников. Набор данных содержит 10 000 диалогов и 1505 описаний персон на русском языке, где диалоги представлены перепиской в мессенджере участников, каждый из которых играет роль человека, описанного набором характеристик (например, «у меня есть кот»). История разговора представлена в формате html, а персоны формируются 5 фактами.

По результатам работы делаются выводы об эффективности применяемых архитектур, возможных методах их персонофикации и предлагаются направления дальнейших исследований.