

УДК 004.81

Валиева Н. Л., Буянов И. О., Гончарова А. Б.

Перевод на русский язык датасета для оценки эмпатии

1. Введение. Эмпатия играет важную роль в коммуникации между людьми, и в частности, в психологической помощи. Чат-боты для оказания такой помощи получили широкое распространение. Для них способность эмпатично отвечать становится критически важным навыком, иначе сеанс может показаться безэмоциональным, что может навредить пользователю, находящемуся в уязвимом психологическом состоянии. Поскольку для русского языка нет датасетов для оценки эмпатии, в рамках данной работы мы предлагаем метод перевода англоязычного датасета EPITOME [1] на русский язык с помощью больших языковых моделей (LLM). Этот датасет включает тексты пользователей Reddit, размеченные по трем типам эмпатии и уровнями выраженности: слабый и сильный. Кроме самих типов датасет содержит аннотированные подстроки — носители эмпатии. Они указывают, какие именно части текста отражают эмпатичный отклик.

2. Задачи. Работа состоит из нескольких задач: подбор LLM, которая лучше всего справится с переводом, разработка набора инструкций для перевода, реализация полной процедуры перевода датасета и обучение модели для классификации эмпатии на русском языке, используя оригинальную модель из [1].

3. Подбор LLM и разработка затравки. Для перевода были протестированы несколько больших языковых моделей (LLM), включая GPT-4o, семейство Qwen-2.5, Mistral-Small-24B-Instruct-2501, YandexGPT Pro. Тестирование производилось на 20 текстах,

Валиева Нафиса Ленисовна – студент, Санкт-Петербургский государственный университет; e-mail: st117021@student.spbu.ru, тел.: +7(900)321-12-07

Буянов Игорь Олегович – аспирант, Федеральный Исследовательский центр «Информатика и управление» Российской Академии Наук; e-mail: buyanov.igor.o@yandex.ru, тел.: +7(909)944-49-70

Гончарова Анастасия Борисовна – доцент, Санкт-Петербургский государственный университет; e-mail: a.goncharova@spbu.ru, тел.: +7(911)736-48-92

включающих особенности текстов с Reddit. Параллельно разрабатывался набор инструкций, в который итеративно включались идеи из анализа ошибок моделей. Было выявлено три модели, показывающие лучший результат по убыванию: GPT-4o, Qwen-2.5-72b-instruct и YandexGPT Pro.

4. Реализация процедуры перевода. Перевод датасета осуществлялся в два этапа: общий перевод и перевод носителей эмпатии. Для двух этапов в качестве основной модели для перевода использовалась YandexGPT Pro, поскольку она демонстрировала лучшее соотношение по качеству и цене для большинства примеров. Для проблемных случаев, в которых YandexGPT ошибалась, применялась Qwen-2.5-72b-instruct. Если проблема сохранялась, то применялась GPT-4o. Такой ступенчатый процесс позволил достичь наилучшего баланса между стоимостью и качеством. На втором этапе специальный скрипт проверял, чтобы каждая переведенная подстрока носителя эмпатии входила в состав основного перевода без изменений.

5. Обучение модели. Для оценки применимости полученного датасета была обучена оригинальная модель классификации и выделения носителей эмпатии. В качестве архитектур были протестированы две базовые модели на основе архитектуры Трансформер: rubert-base-cased и xlm-roberta-base.

6. Заключение и обсуждение. Предложена воспроизводимая процедура перевода и адаптации англоязычного датасета эмпатии на русский язык. Путем эмпирического сравнения качества моделей на английском и русском языках показана адекватность приведенного метода. Основная проблема заключается в стиле общения на Reddit, как в социальных сетях вообще, которая включает множество особенностей, такие как сокращения, удлинённые междометия, нестандартная пунктуация и т.д.

Литература

1. Sharma A., Miner A., et al. A computational approach to understanding empathy expressed in text-based mental health support // Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP) / ed. by B. Webber, et al. Online, Pa: Association for Computational Linguistics. 2020. P. 5263–5276.