

УДК 004.912

РАСПОЗНАВАНИЕ ЭМОЦИЙ ЧЕЛОВЕКА В ПОЛИЛОГЕ С ИСПОЛЬЗОВАНИЕМ БОЛЬШОЙ ЯЗЫКОВОЙ МОДЕЛИ

Козлов М.А. (ИТМО)

Научный руководитель – кандидат технических наук, доцент Махныткина О.В. (ИТМО)

Введение. Эмоция – психический процесс, отражающий субъективное оценочное суждение человека. Подобное выражение желаний и потребностей является неотъемлемой частью общения и взаимодействия между людьми. Автоматическое распознавание эмоций используется для совершенствования виртуальных собеседников и цифровых помощников, для контроля за психическим состоянием людей, для анализа реакции пользователей торговых и развлекательных платформ.

Основная часть. В рамках работы был проведен анализ исследований по распознаванию эмоций человека с использованием рекуррентных и графовых нейронных сетей, трансформеров и больших языковых моделей (LLM). В связи со значительным превосходством модели, основанной на большой языковой модели, было решено в дальнейшем исследовании использовать модель на основе InstructERC [1].

Для улучшения результатов распознавания эмоций человека решались следующие четыре задачи:

- 1) задача уменьшения требований к видеопамяти;
- 2) задача оптимизации запроса для LLM;
- 3) задача сериализации датасета с сохранением имён участников полилога;
- 4) задача добавления информации о типе личности говорящего.

Использование 4-битной квантизации [2] и градиентного чекпоинтинга [3] позволило сократить требования к видеопамяти до 6 ГБ, благодаря чему модель стало возможным запускать на потребительской видеокарте Nvidia GeForce RTX 3070 Ti с видеопамтью 8GB.

Запрос оптимизировался с помощью использования эмоционального стимула [4], изменения формата запроса и изменения наименования говорящих. Рассматривалось четыре варианта изменения формата запроса: перевод в верхний регистр только текста самого запроса без реплик, перевод в верхний регистр только важных слов, добавление в вопрос формулировки «говорящий обладает типом личности <тип_личности>» и перевод всего текста запроса в верхний регистр. Рассматривалось три новых варианта наименования говорящих: имена, тип личности и тип личности с приставкой «Speaker ».

В процессе сериализации датасета MELD [5] были обнаружены ошибки, влияющие на точность обучения моделей: разные персонажи могут иметь идентичное имя, персонажи могут иметь несколько имён в диалоге, диалоги могут быть разделены на диалоги из одной реплики. После исправления обнаруженных недочётов количество диалогов было уменьшено на 23.

Для получения типов личности использовалась открытая база данных, основанная на системе голосования. Из 13708 высказываний для 12524 были определены типы личности из базы данных, то есть более 91% данных датасета.

В рамках работы было выявлено, что точность распознавания увеличилась больше всего при изменении формата запроса на текст полностью в верхнем регистре, при использовании типов личности в наименовании говорящих и при использовании эмоционального стимула. Точность распознавания оказалась меньше при частичном переводе текста в верхний регистр или добавлении типа личности в вопрос в конце запроса.

Выводы. Использование LLM для распознавания эмоций человека в полилоге позволило увеличить значение F1-Weighted для датасета MELD [5] с 69.15% до 70.36%.

Список использованных источников:

1. Shanglin Lei, Guanting Dong, XiaoPing Wang, Keheng Wang, Sirui Wang. InstructERC: Reforming Emotion Recognition in Conversation with a Retrieval Multi-task LLMs Framework. – 2023. Сентябрь. – Текст : электронный. – URL: <https://arxiv.org/pdf/2309.11911v2.pdf> (дата обращения: 28.01.2024). – С. 11.
2. Cheng Li, Jindong Wang, Kaijie Zhu, Janice Zhang, Wenxin Hou, Jianxun Lian, Xing Xie. EmotionPrompt: Leveraging Psychology for Large Language Models Enhancement via Emotional Stimulus. // Symposium on Advances and Open Problems in Large Language Models (LLM@IJCAI'23) – 2023. Август. – Текст : электронный. – URL: <https://arxiv.org/pdf/2307.11760.pdf> (дата обращения: 28.01.2024). – С. 32.
3. Shih-yang Liu, Zechun Liu, Xijie Huang, Pingcheng Dong, Kwang-Ting Cheng. LLM-FP4: 4-Bit Floating-Point Quantized Transformers. // Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing – 2023. Декабрь. – С. 592–605 – Текст : электронный. – URL: <https://arxiv.org/pdf/2310.16836.pdf> (дата обращения: 28.01.2024). – С. 14.
4. Nimit S. Sohoni, Christopher R. Aberger, Megan Leszczynski, Jian Zhang, Christopher Ré. Low-Memory Neural Network Training: A Technical Report. – 2022. Апрель. – Текст : электронный. – URL: <https://arxiv.org/pdf/1904.10631v2.pdf> (дата обращения: 28.01.2024). – С. 38.
5. Soujanya Poria, Devamanyu Hazarika, Navonil Majumder, Gautam Naik, Erik Cambria, Rada Mihalcea. MELD: A Multimodal Multi-Party Dataset for Emotion Recognition in Conversations . // Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics – 2019. – С. 527-536. – Текст : электронный. – URL: <https://arxiv.org/pdf/1810.02508.pdf> (дата обращения: 28.01.2024). – С. 10.