

УДК 004.934.2

МУЛЬТИМОДАЛЬНЫЙ РУССКОЯЗЫЧНЫЙ КОРПУС ДАННЫХ ДЛЯ РАСПОЗНАВАНИЯ ЭМОЦИОНАЛЬНОГО СОСТОЯНИЯ ЧЕЛОВЕКА

Матвеева А.А. (ИТМО)

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

Введение. В данной работе представлен анализ и оценка качества [1,2] оригинального корпуса данных, включающего записи полилогов, сегментированных на отдельные реплики. Каждая реплика сопровождается аудиофрагментом и многомерной разметкой, основанной на психолингвистической модели VAD (валентность, включенность, доминантность), а также категориальными эмоциональными метками.

Основная часть. Корпус данных включает записи полилогов на 11 тематик с участием 14 носителей русского языка (9 мужчин и 5 женщин). Материал сегментирован на 7288 реплик общей продолжительностью 11 часов (662 минуты). Каждая реплика представлена аудиофрагментом (средняя длительность — 7 секунд, максимум — 15 секунд), транскрибированным текстом и многоуровневой эмоциональной разметкой. Транскрибация выполнена с использованием автоматической системы распознавания речи с последующей ручной коррекцией для устранения ошибок. Синхронизация аудио и текста обеспечена файлами в формате TextGrid, содержащими временные метки начала и конца реплик, а также идентификаторы спикеров.

Эмоциональная разметка реализована по четырём меткам: валентность (позитивная, негативная, нейтральная), включенность (возбужденная, нейтральная, заторможенная), доминантность (высокая, средняя, низкая) и категориальные эмоции (гнев, отвращение, страх, радость, нейтральность, грусть, удовлетворенность, расслабленность). Доминантность рассчитывалась на основе адаптированного словаря NRC-VAD-Lexicon: каждой лексической единице присваивалось значение от 0 (низкий уровень влияния) до 1 (высокий уровень), а итоговый показатель для реплики определялся как среднее арифметическое оценок входящих слов, категориальная метка доминантности выводилась на основе численного значения доминантности. Категориальные эмоции выводились через комбинацию меток VAD [3]. Аннотация валентности и включенности выполнена тремя независимыми экспертами (носители русского языка, специалисты по машинному обучению), а финальные метки формировались по принципу консенсуса с привлечением четвертого эксперта в случае отсутствия согласия.

Выводы. В данной работе представлено описание собранного мультимодального корпуса данных. Согласованность разметки оценена через коэффициент Каппа Коэна. Помимо этого, для валидации корпуса было проведено обучение доработанной модели RoBERTa-base, адаптированной для предсказания эмоций и метрик VAD. Корпус может быть использован для решения задач распознавания эмоционального состояния человека по аудио и текстовым данным.

Список использованных источников:

1. Kim T., Vossen P. EmoBERTa: Speaker-Aware Emotion Recognition in Conversation with RoBERTa // arXiv preprint arXiv:2108.12009. – 2021.
2. Bălan, O., Moise, G., Petrescu, L. Emotion Classification Based on Biophysical Signals and Machine Learning Techniques. // Symmetry. – 2020, P. 12–21.
3. Mohammad S. Obtaining Reliable Human Ratings of Valence, Arousal, and Dominance for 20,000 English Words // In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, P. 174–184, 2018.