

УДК 004.931.2

ПОСТРОЕНИЕ МОДЕЛИ АВТОМАТИЧЕСКОГО АНАЛИЗА ДОМИНАНТНОСТИ ДИКТОРОВ В МАЛЫХ ГРУППАХ НА ОСНОВЕ АКУСТИЧЕСКИХ ПРИЗНАКОВ

Хмелев Н.А. (Университет ИТМО)

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

(Университет ИТМО)

Исследования выполнены за счет финансирования университета ИТМО в рамках НИР №622281 «Разработка методов и алгоритмов для мультимодального распознавания валентности высказываний и доминантности дикторов в полилогах».

Введение. В современных крупных компаниях, все чаще требуются инструменты для анализа различного типа разговоров (например, бизнес переговоры с конкурентами, диалог клиента и оператора). Данный анализ необходим для эффективной оценки качества, поиска информации о договоренностях, а также для повышения качества человеко-машинного взаимодействия. В данной работе будет рассмотрена задача автоматического анализа доминантности дикторов в полилогах от 3 до 4 человек.

Основная часть. Цель работы — построение модели для автоматической оценки доминантности на основе аудио и текстовых данных. В работе были исследованы современные предобученные модели Wav2Vec2.0[1] для аудио модальности и T5[2] для текстовой модальности. Проведены эксперименты по решению задачи ранжирования дикторов по уровню доминантности в полилоге, а также для выявления наиболее доминантного диктора. В качестве баз обучения и тестирования в настоящем исследовании использовались базы DOME[3] и ELEA[4], каждая из которых содержит до 10 часов записи полилогов, с участием от 3 до 4 персон, записанных на дальний микрофон. Экспериментальное исследование проводилось методом кросс-валидации.

Выводы. Результатами данной работы являются реализация и исследование моделей для оценки доминантности каждого участника в малых группах, а также проведено сравнение полученных моделей с существующими решениями.

Список использованных источников:

1. Baevski A. et al. wav2vec 2.0: a framework for self-supervised learning of speech representations //Advances in neural information processing systems. – 2020. – Т. 33. – С. 12449–12460.
2. Raffel C. et al. Exploring the limits of transfer learning with a unified text-to-text transformer //The Journal of Machine Learning Research. – 2020. – Т. 21. – №. 1. – С. 5485-5551.
3. Aran O., Hung H., Gatica-Perez D. A multimodal corpus for studying dominance in small group conversations //Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality 18 May 2010. – 2010. – Т. 22.
4. Sanchez-Cortes D. et al. A nonverbal behavior approach to identify emergent leaders in small groups //IEEE transactions on multimedia. – 2011. – Т. 14. – №. 3. – С. 816-832.