

МУЛЬТИМОДАЛЬНОЕ РАСПОЗНАВАНИЕ ЭМОЦИОНАЛЬНОГО СОСТОЯНИЯ СОБЕСЕДНИКОВ В ДИАЛОГЕ

Зубаков А. Н.

(Университет ИТМО)

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

(Университет ИТМО)

Введение. В работе рассматривается метод определения эмоционального состояния собеседников в диалоге на русском языке по двум модальностям (аудио и текст). Исследование проводилось с использованием мультимодального корпуса данных MELD, переведенного на русский язык.

Основная часть. Распознавание эмоционального состояния человека используется в диалоговых системах для оценки удовлетворенности пользователя и адаптации диалога в соответствии с его состоянием, при проведении социологических и маркетинговых исследований, в диагностических системах для выявления состояний депрессии и прочих эмоциональных расстройств, в устройствах комплексной регистрации психофизиологических параметров, а также в тех случаях, когда важно своевременно получать достоверные и объективные данные об эмоциональном состоянии собеседника.

В настоящее время не существует доступных русскоязычных диалоговых датасетов, сравнимых по объему и качеству аннотирования с наиболее популярными англоязычными наборами данных MELD [1] и IEMOCAP [2]. Вместе с тем, английская речь имеет ряд особенностей, не позволяющих с одинаковой эффективностью применять для русского языка модели, обученные на англоязычных данных, в особенности, для определения эмоций. При этом, специфика корпуса данных MELD, а именно то, что он собран из фрагментов многосерийного телевизионного фильма «Друзья», транслировавшегося по отечественному телевидению, позволяют воссоздать набор данных MELD на русском языке практически без потери качества.

В рамках исследования для аудио модальности русскоязычного корпуса данных использован профессиональный перевод фильма «Друзья», а для текстовой модальности — автоматизированный перевод. Для выделения признаков из аудиоданных использована предобученная модель Wav2Vec, для выделения признаков из текстовых данных — предобученная на русском языке разновидность модели BERT. Непосредственно для классификации эмоций по дискретной модели Пола Экмана [3] использовалась конфигурация рекуррентной нейронной сети.

Выводы. По результатам работы делаются выводы о целесообразности подобного подхода как к формированию набора данных, так и к перспективности методов извлечения признаков и классификации, а также предлагаются направления для дальнейших исследований.

Список использованных источников:

1. Soujanya Poria, Devamanyu Hazarika, Navonil Majumder, Gautam Naik, Erik Cambria, Rada Mihalcea. MELD: A Multimodal Multi-Party Dataset for Emotion Recognition in Conversations. Conference: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Vol. -, 2019, pp. 1–10. DOI: 10.18653/v1/P19-1050
2. C. Busso, M. Bulut, C.C. Lee, A. Kazemzadeh, E. Mower, S. Kim, J.N. Chang, S. Lee, and S.S. Narayanan. IEMOCAP: Interactive emotional dyadic motion capture database. Journal of Language Resources and Evaluation. Vol. 42, no. 4, 2008, pp. 335–359. DOI: 10.1007/s10579-008-9076-6.
3. Ekman, Paul. Are there basic emotions?. Psychological Review. Vol. 99(3), 1992, pp. 550–553. DOI: 10.1037/0033-295X.99.3.550.