

УДК 004.934.2

РАСПОЗНАВАНИЕ ВОВЛЕЧЕННОСТИ И ЭМОЦИЙ СОБЕСЕДНИКОВ ВИРТУАЛЬНОЙ КОММУНИКАЦИИ

Двойникова А.А. (Санкт-Петербургский Федеральный исследовательский центр РАН,
Университет ИТМО)

Научный руководитель – д.т.н., профессор Карпов А.А.

(Санкт-Петербургский Федеральный исследовательский центр РАН, Университет ИТМО)

Введение. В настоящее время тема автоматического распознавания вовлеченности собеседников в полилог или студентов в онлайн обучение является востребованной и актуальной. В последние годы люди стали чаще коммуницировать друг с другом с помощью систем телеконференций. В таком онлайн общении понимание вовлеченности и эмоций оппонента становится затруднительным процессом. Поэтому разработка автоматической системы для распознавания вовлеченности участников виртуальной коммуникации может помочь решить описанные выше проблемы. В области психологии вовлеченность человека в коммуникацию или образовательный процесс имеет высокую корреляцию с чувствами и эмоциями данного человека [1]. Исходя из этого для наиболее эффективного распознавания вовлеченности необходимо анализировать проявления эмоций человека. Взаимокоррелирующие понятия обуславливают применение многозадачного подхода для анализа поведения человека во время коммуникации. При этом многомодальный подход к распознаванию вовлеченности и эмоций человека позволяет разработать робастную систему анализа невербальных проявлений коммуниканта [2]. В статье предлагается многозадачный и многомодальный подход к анализу вовлеченности и эмоций человека в коммуникацию.

Основная часть. Для экспериментальных исследований и апробации предложенного подхода был использован многомодальный корпус RECOLA [3]. Он содержит в себе данные попарной коммуникации участников на французском языке. Диалог участников происходил при помощи системы телеконференции Skype. Объем корпуса равен 2 ч (5-минутные записи каждого из 23 участников). Каждая запись была целиком размечена метками уровней вовлеченности, а также полярностью и интенсивностью эмоций каждые 4 мс записи. Для экспериментальных исследований были выделены по 3 уровня вовлеченности, полярности и интенсивности эмоций.

В статье предлагается новый многозадачный и многомодальный подход для распознавания вовлеченности и эмоций собеседников. Предложенный подход представляет собой анализ видео- и аудиоданных. Для извлечения признаков из видеоданных была использована предобученная нейронная сеть ResNet50, в качестве акустических признаков использовались спектрограммы. Затем видео и аудио признаки подавались на сверточную нейронную сеть, для видеоданных нейронная сеть состояла из одного сверточного блока, для аудиоданных – из 5 блоков. Далее полносвязные слои, состоящие из 10 нейронов, каждой сверточной нейронной сети были конкатенированы, и распараллелены на 3 полносвязных слоя, выходы которых выдают вероятностное предсказание 3 уровней вовлеченности, полярности и интенсивности эмоций, соответственно.

Выводы. В ходе проведенных экспериментальных исследований с предложенным подходом для распознавания вовлеченности и эмоций коммуникантов на данных многомодального корпуса RECOLA удалось достичь точности 72,53%, 41,72% и 64,90% по показателю невзвешенной средней полноты для 3 уровней вовлеченности, интенсивности и полярности эмоций, соответственно.

Работа выполнена при поддержке Совета по грантам Президента РФ (грант № НШ-17.2022.1.6).

Список использованных источников:

1. Avey J. B. et al. Meta-analysis of the impact of positive psychological capital on employee attitudes, behaviors, and performance // Human resource development quarterly. – 2011. – Т. 22. – №. 2. – С. 127-152.
2. Двойникова А. А. и др. Анализ информационного и математического обеспечения для распознавания аффективных состояний человека // Информатика и автоматизация. – 2022. – Т. 21. – №. 6. – С. 1097-1144.
3. Ringeval F. et al. Introducing the RECOLA multimodal corpus of remote collaborative and affective interactions // 2013 10th IEEE international conference and workshops on automatic face and gesture recognition (FG). – IEEE, 2013. – С. 1-8.