УДК 004.932.2

МЕТОД АНАЛИЗА КОММУНИКАТИВНЫХ ЖЕСТОВ УЧАСТНИКОВ ВИРТУАЛЬНОЙ КОММУНИКАЦИИ

Двойникова А.А.[1, 2]
Научный руководитель – д.т.н., профессор Карпов А.А.[1]
[1]Санкт-Петербургский Федеральный исследовательский центр РАН, [2]Университет ИТМО
aadvoinikova@itmo.ru

Введение

В настоящее время люди часто взаимодействуют друг с другом с использованием программных средств телеконференций. При групповой виртуальной коммуникации понимание невербальных аспектов всех участников в разговор становится затруднительным процессом. Анализ вовлеченности участников в разговор, проявлений эмоциональных состояний дикторов помогает отследить групповую динамику разговора. Изучение групповой динамики активно используется в педагогике, фасилитации, психологии. Одним из проявлений уровней вовлеченности и эмоционального состояния является наличие коммуникативных жестов у диктора, например зевота, скрещивание рук, нахождение рук у лица, кивки головой. Автоматическое распознавание коммуникативных жестов является актуальной задачей и может быть применима в системах автоматического распознавания вовлеченности и эмоций участников виртуальной коммуникации. На сегодняшний день не существует работ по автоматическому анализу коммуникативных жестов. Некоторые ученые [1] проводили исследования только в области анализа жестового языка. Автоматические методы анализа жестового языка могут быть успешно применимы в задаче распознавания коммуникативных жестов собеседников. В настоящем исследовании представлен метод анализа коммуникативных жестов.

Основная часть

Метод анализа коммуникативных жестов состоит из трех этапов: извлечение различных информационных признаков из видеоданных; построение моделей анализа различных информационных признаков и объединение моделей для итоговой классификации коммуникативных жестов. На первом этапе из видеоданных извлекаются информационные признаки: на основе точек лица и рук (англ. landmarks), производные геометрические признаки (англ. derived) и визуальные нейросетевые признаки (англ. video). Landmarks отражают пространственную структуру жеста, они могут моделировать динамику движения и имеют небольшую размерность. Landmarks хорошо описывают такие жесты как наклоны головы, положение рук относительно лица и др. Derived признаки представляют собой межточечные расстояния между ключевыми точками, а также углы между сегментами. Такие признаки позволяют явно моделировать контакт рук и лица, являются более устойчивыми к шуму координат. Derived признаки хорошо описывают такие жесты как «рука на подбородке», «рука закрывает рот», «рука подпирает голову» и др. Video признаки извлекаются из локальных областей видеокадров с помощью предобученной сверточной нейронной сети R(2+1)D [2]. Video признаки содержат в себе информацию о текстуре, степени окклюзии, мимических изменениях.

Для всех информационных признаков используются различные модели классификации жестов. Так, для landmarks и video признаков используются временной Transformers (англ. Temporal Transformers), для derived признаков – временные сверточные сети (англ. Temporal Convolution Network, TCN). Временные модели применяются к признакам, описывающим динамические характеристики жеста, поскольку коммуникативные жесты являются процессами, развёрнутыми во времени. Затем выходы

всех моделей конкатенируются и подаются на полносвязную нейронную сеть (англ. Fully Connected Neural Network, FCNN). Выходом полносвязной нейронной сети является вероятность принадлежности объекта к одному из классов коммуникативных жестов.

Для экспериментальных исследований был использован корпус данных ENERGI (ENgagement and Emotion Russian Gathering Interlocutors) [3]. Он содержит аудиовизуальные данные участников групповой коммуникации с использованием систем телеконференций. Корпус ENERGI имеет аннотацию по трем уровням вовлеченности, валентности и активации эмоций, а также по 13 классам коммуникативных жестов.

Выводы

Результаты экспериментальных исследований доказывают эффективность объединения нескольких различных признаков для задачи распознавания коммуникативных жестов.

Литература

1. Рюмин Д. Метод автоматического видеоанализа движений рук и распознавания жестов в человеко-машинных интерфейсах // Научно-технический вестник информационных технологий, механики и оптики. – 2020. – Т. 20. – №. 4. – С. 525-531.
2. Tran D. et al. A closer look at spatiotemporal convolutions for action recognition // Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. – 2018. – С. 6450-6459.
3. Двойникова А. А., Величко А. Н., Карпов А. А. Многомодальный корпус данных взаимодействия участников виртуальной коммуникации ENERGI // Известия высших учебных заведений. Приборостроение. – 2025. – Т. 68. – №. 12. – С. 1011-1019.