УДК 004.931

Разработка модели для мультимодального распознавания валентности высказываний и доминантности дикторов в полилогах

Васюк М. 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. Эмоционально-ориентированные системы позволяют лучше понимать поведение человека и взаимодействие людей между собой. Применение подобных систем обширно: от чат-ботов до специализированных систем анализирования психоэмоционального состояния людей. Одной из сложностей решения задачи является разнообразие форматов данных, недостаточное количество данных для обучения моделей. Для корректного распознавания эмоций и их характеристик возникает дополнительная задача – извлечение высказываний отдельной персоны из полилога.

Основная часть. Целью моей работы является построение модели для автоматической оценки валентности высказываний и доминантности дикторов, используя визуальную и акустическую модальности. В качестве глубокой нейронной сети для оценки валентности/доминантности на основе визуальной модальности используется ResNet50 [1]. В качестве глубокой нейронной сети для оценки валентности/доминантности на основе акустической модальности используется MatchBoxNet [2]. Для обучения, валидации и тестирования используются следующие базы данных: AffectNet[3], AMI [4], MELD [5], IEMOCAP[6].

Выводы. Результатами моей работы являются реализованная и обученная модель автоматической оценки валентности высказываний и доминантности дикторов в полилогах, а также метрики работы модели на тестовых данных.

Список использованных источников

1. He, K. Deep residual learning for image recognition. / K. He, X. Zhang, S. Ren, J. Sun // Открытые СУБД. – URL: https://openaccess.thecvf.com/content_cvpr_2016/papers/He_Deep_Residual_Learning_CVPR_2016_paper.pdf.
2. Majumdar, S. MatchboxNet: 1D Time-Channel Separable Convolutional Neural Network Architecture for Speech Commands Recognition. / S. Majumdar, B. Grinsburg // Открытые СУБД. – URL: https://arxiv.org/abs/2004.08531.
3. Mollahosseini, A. Affectnet: Adatabase for facial expression, valence, and arousal computing in the wild / A. Mollahosseini, B. Hasani, M. H. Mahoor // Открытые СУБД. – URL: https://arxiv.org/pdf/1708.03985.
4. Kraaij W. et al. The AMI meeting corpus. – 2005.
5. Poria, Soujanya. MELD: A Multimodal Multi-Party Dataset for Emotion Recognition in Conversations. / Hazarika, Devamanyu & Majumder, Navonil & Naik, Gautam & Cambria, Erik & Mihalcea, Rada. - 2018.
6. Busso, C. IEMOCAP: Interactive emotional dyadic motion capture database. Language Resources and Evaluation. 42. 335-359. / Bulut, Murtaza & Lee, Chi-Chun & Kazemzadeh, Abe & Mower Provost, Emily & Kim, Samuel & Chang, Jeannette & Lee, Sungbok & Narayanan, Shrikanth. - doi:10.1007/s10579-008-9076-6. - 2008.