УДК 004.931.2

ПОСТРОЕНИЕ ДИКТОРСКИХ ЭМБЕДДИНГОВ НА ОСНОВЕ МОДЕЛЕЙ С МЕХАНИЗМОМ ВНИМАНИЯ.

Хмелев Н.А. (Университет ИТМО)

Научный руководитель – к.т.н., доцент Волохов В.А. (Университет ИТМО)

Введение. Вопрос защиты конфиденциальной информации в крупных ИТ-компаниях на сегодняшний день является одним их самых важных. Стойкость привычных паролей стоит под вопросом, так как такой тип информации можно очень легко забыть, потерять или украсть. Перспективной отраслью развития защиты информации являются биометрические характеристики человека, такие как голос, лицо, отпечаток пальца. Теоретически такие системы защиты должны быть более устойчивые и удобные для использования, однако такие системы очень сложны, требуют высокой компетентности, большого количества вычислительных ресурсов и имеют свою погрешность. Хотя уже имеется большое количество качественных биометрических систем, ни одна из них не может работать без ошибок, поэтому в данной работе будут рассмотрены современные системы и методы повышения качества таких систем.

Основная часть. Цель моей работы — построение моделей извлечения низкоразмерных дикторских представлений для быстрой и качественной верификации и идентификации дикторов. В данной работе описываются и изучаются модели глубокого обучения, обученные методом тонкой настройки существующих решений, обучаемых на больших объемах данных без учителя, под открытую задачу классификации дикторов. В работе рассмотрены SOTA модели, такие как Wav2Vec2.0[1], Whisper[2] и SpeechT5[3], Обученные модели тестировались на задаче верификации, идентификации, кластеризации и диаризации дикторов. В качестве тестовых данных, для сравнения моделей, использовались базы данных VoxCeleb1[4], которая является наиболее популярной в данной задаче, Voices[5], в которой представлены записи в сложных условиях, и база данных AMI[6], популярная база для оценки качества диаризации дикторов. Все указанные базы данных записаны в микрофонных условиях. Также полученные модели сравнивались с существующими решениями, предложенными за последний год. Помимо модели построения дикторских эмбеддингов, предлагается алгоритм валидации полученных эмбеддингов, для детектирования эмбеддингов, построенных по неречевым сегментам, с целью повышения надежности системы голосовой биометрии.

Выводы. Результатами моей работы являются сравнение предлагаемых биометрических моделей с существующими, по критериям скорости и качества работы биометрических систем.

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. Radford A. et al. Robust speech recognition via large-scale weak supervision //arXiv preprint arXiv:2212.04356. – 2022.
3. Ao J. et al. Speecht5: Unified-modal encoder-decoder pre-training for spoken language processing //arXiv preprint arXiv:2110.07205. – 2021.
4. Nagrani A., Chung J. S., Zisserman A. Voxceleb: a large-scale speaker identification dataset //arXiv preprint arXiv:1706.08612. – 2017.
5. Richey C. et al. Voices obscured in complex environmental settings (voices) corpus //arXiv preprint arXiv:1804.05053. – 2018.
6. Kraaij W. et al. The AMI meeting corpus. – 2005.