УДК 004.522

ОБУЧЕНИЕ МОДЕЛИ ЯЗЫКОВОЙ ИДЕНТИФИКАЦИИ ДЛЯ РЕШЕНИЯ ПРОБЛЕМЫ МУЛЬТИЯЗЫЧНОСТИ В РЕЧЕВЫХ СИСТЕМАХ

Марчевский В.Д. (ИТМО)
Научный руководитель – кандидат технических наук Новосёлов С.А.
(ИТМО)

Введение. Речевые технологии — это бурно развивающаяся область исследований, дающая множество возможностей в человеко-машинных интерфейсах. Одноязычные системы на основе речевых технологий показали полезность в повседневной жизни, поэтому следующий очевидный шаг в эволюции подобных систем - это способность работать с более чем одним языком.

Основная часть. Общим ограничением большинства современных речевых систем, в частности систем автоматического распознавания речи, является возможность работать на нескольких языках независимо, но не в случае когда несколько языков присутствуют на одной записи. Поэтому, для использования подобных моделей в случаях, когда язык произнесения заранее неизвестен, либо когда таких языков несколько, необходима дополнительная модель языковой идентификации и диаризации, способная определять сегменты речи, в течение которых произношение ведётся на одном конкретном языке [1]. Альтернативный подход заключается в обучении так называемого “переключаемого” ASR, то есть модели, которая способна обрабатывать сразу несколько языков [2]. В качестве базового варианта для разработки модели языковой идентификации была выбрана архитектура свёрточной нейронной сети на основе ResNet [3]. Подобные архитектуры широко применяются в таких задачах как тексто-независимое распознавание диктора [4] и автоматическое распознавание речи [5]. В качестве базы данных для обучения выступает смесь из открытых датасетов микрофонного канала. Результат обучения показал сравнимое с лучшей на данный момент открытой моделью, при этом работая значительно быстрее. Это делает данный подход отличным базовым решением, на котором можно строить модуль языковой диаризации для полноценного решения проблемы мультиязычности в речевых системах.

Выводы. Проведен анализ проблемы мультиязычности в речевых системах и разработана модель языковой идентификации.

Список использованных источников:

1. Radford A. et al. Robust speech recognition via large-scale weak supervision // International Conference on Machine Learning. – PMLR, 2023. – С. 28492-28518.
2. Ghadekar P. et al. ASR for Indian regional language using Nvidia’s NeMo toolkit // AIP Conference Proceedings. – AIP Publishing, 2023. – Т. 2851. – №. 1.
3. He K. et al. Deep residual learning for image recognition // Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016. – С. 770-778.
4. Jakubec M., Lieskovska E., Jarina R. Speaker recognition with ResNet and VGG networks // 2021 31st International Conference Radioelektronika (RADIOELEKTRONIKA). – IEEE, 2021. – С. 1-5.
5. Synnaeve G. et al. End-to-end asr: from supervised to semi-supervised learning with modern architectures // arXiv preprint arXiv:1911.08460. – 2019.